Ion Mobility-Mass Spectrometry and Collision Induced Unfolding of Multi-Protein Ligand Complexes

by

Shuai Niu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2015

Doctoral Committee

> Associate Professor Brandon T. Ruotolo, Chair
> Professor Kristina I. Håkansson
> Professor Robert T. Kennedy
> Assistant Professor Matthew B. Soellner

Acknowledgements

I feel truly fortunate for the decision to come to the U.S. for grad school five years ago, not only because of the professional advances I gained here, but also the many people who I met here and have impacted my life trajectory so much. In full gratitude I would like to acknowledge the following individuals who have inspired, encouraged and supported me along this long journey of earning my Ph.D.:

First and foremost, I would like to thank my advisor Prof. Brandon Ruotolo. I used to be trained under a pragmatic culture and biasedly considered the relationship between professors and students to be more like employers and employees. Brandon completely changed my mind set, and showed me educative, productive and harmonic manners to share knowledge, conduct research and organize a laboratory. You have always been extremely patient, supportive and insightful during our meetings and discussions, and as of now I'm really proud that the reason motivates me to do science is not that I have to, but rather I love to. You are not only a great mentor in science, but also an awesome friend in life. In addition, I'd also like to thank my dissertation committee members, Prof. Kristina Håkansson, Prof. Robert Kennedy, Prof. Mattew Soellner and Prof. Jason Gestwicki for your insightful comments about my work, and really helped me to expand my view beyond the analytical chemistry community. It has been a great pleasure to have you on my committee.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

3D	three-dimensional
AP	affinity purification
ATD	arrival time distribution
BLA	β-lactoglobulin A
CCS	collision cross-section
CD	circular dichroism
CE	crown ether
CID	collision induced dissociation
CIU	collision induced unfolding
CMC	critical micelle concentration
ConA	concanavalin A
CRM	charge residue model
CEM	chain ejection model
cryo-EM	cryo-electron microscopy
CXL	crosslinking
Cymal-5	5-cyclohexyl-1-pentyl-β-D-maltoside
DC	direct current
DDM	dodecylmaltoside
DM	decylmaltoside
DMA	differential mobility analyzer
DMSO	dimethyl sulfoxide
DSC	differential scanning calorimetry
DT	drift tube
E. Coli	*Escherichia coli*

ECD	electron capture dissociation
EDTA	ethylene diamine tetraacetic acid
EHSS	exact hard-sphere scattering
EM	electron microscopy
ESI	electrospray ionization
ETD	electron transfer dissociation
FDX	ferredoxin
FNR	ferredoxin-NADP+ reductase
FRET	Fluorescence resonance energy transfer
GDH	glutamate dehydrogenase
HDAC	human histone deacetylase
HDX	hydrogen-deuterium exchange
HTS	high-throughput screening
HSP	heat shock protein
IEM	ion evaporation model
IgG	immunoglobulin G
IM	ion mobility
IM-MS	ion mobility-mass spectrometry
IMoS	ion mobility spectrometry calculation package
ITC	isothermal titration calorimetry
K_D	dissociation constant
LC	liquid chromatography
MALDI	matrix-assisted laser desorption/ionization
MCP-1	Monocyte chemoattractant protein-1
MD	molecular dynamics
m-NBA	m-nitrobenzyl alcohol
MS	mass spectrometry

nESI	nanoelectrospray ionization
NG	nonyl glucoside
NM	nonylmaltoside
NMR	nuclear magnetic resonance
OFP	oxidative foot printing
OM	octylmaltoside
PA	projection approximation
PCBP1	poly-r(C)-binding protein 1
PDB	protein data bank
PG	phosphatidyl-glycerol lipid
PpIX	protoporphyrin IX
PSA	projection super-position approximation
RF	radio frequency
Rs	rhodobacter sphaeroides
SAP	serum amyloid P component
SAXS	small-angle X-ray scattering
SEC	size-exclusion chromatography
SID	surface-induced dissociation
SPR	surface plasmon resonance
TM	trajectory method
TMA	tetramethylammonium
ToF	time-of-flight
TRAP	tryptophan RNA-binding attenuator protein
Trap CV	trap collision voltage
TSPO	translocator protein
TTR	transthyretin
T-wave	traveling wave

UV-Vis	ultraviolet-visible
XRD	X-ray diffraction

Abstract

Mass spectrometry (MS) serves as an indispensable technology for modern pharmaceutical drug discovery and development processes, where it is used to assess ligand binding to target proteins and to search for biomarkers that can be used to gauge disease progression and drug action. However, MS is rarely treated as a screening technology for the structural consequences of drug binding. Instead, more time-consuming technologies capable of projecting atomic models of protein-drug interactions are utilized. In this thesis, ion mobility-mass spectrometry (IM-MS) methods are developed in order to fill these technology gaps. Principle among these is collision induced unfolding (CIU), which leverages the ability of IM to separate ions according to their size and charge, in order to fingerprint gas-phase unfolding pathways for non-covalent protein complexes.

Following a comprehensive introductory chapter, we demonstrate the consequences of sugar binding on the CIU of Concanavalin A (Con A) in Chapter 2. Our CIU assay reveals cooperative stabilization upon small molecule binding, and such effect cannot be easily detected by solution phase assays, or by MS alone. In Chapter 3, the underlying mechanism of multi-protein unfolding is systematically investigated by IM-MS and molecular modeling approaches. Our

results show a strong positive correlation between monomeric Coulombic unfolding and the tetrameric CIU process. This provides strong evidence that multi-protein unfolding events are initiated primarily by charge migration from the complex to a single monomer. In Chapter 4, the interactions between human histone deacetylase 8 (HDAC8) and poly-r(C)-binding protein 1 (PCBP1) are investigated by IM-MS. Our data suggest that these proteins interact with each other in a specific manner, a fact revealed by our optimized ESI-MS workflow for quantifying binding affinity (K_D) for weakly-associated hetero-protein complexes. In Chapter 5, the translocator protein (TSPO) dimer from *Rhodobacter sphaeroides,* as well as its disease-associated variant forms, is analyzed by IM-MS and CIU assays. By utilizing a combination of CIU and collision induced dissociation (CID) stability data, an unknown endogenous ligand bound to TSPO is detected and identified.

Chapter 1.Introduction

Mass spectrometry (MS) is a critical component of many drug development efforts worldwide, where it is used to assess small molecule binding to target substrates and the pharmacokinetics of drug action.[1, 2] However, MS is rarely used to screen for the structural consequences of drug binding, and more time-consuming technologies capable of projecting atomic models of protein-drug interactions are utilized instead.[3] As the intended phenotypes of drug interactions trend toward those that target critical conformation changes within the bound substrates, in many cases binding to proteins having no known enzymatic active site, developing rapid tools capable of screening subtle conformational changes within protein-ligand complexes is critical.[4]As such, the content of this thesis deals directly with these technology challenges and endeavors, in an effort to create MS-based novel technologies for protein-ligand screening.

Since the early 1990s, technological advances such as combinatorial chemistry[5] and high-throughput screening (HTS)[6] have been serves as indispensable components for modern pharmaceutical industry.[7]Well-established protein-ligand screening pipelines typically require some type of fluorescent- or radio-labeling to deduce the binding of a ligand molecule to its receptor. Most of these label-dependent screening platforms are based on measurements of fluorescence, such as fluorescence resonance energy transfer (FRET) and fluorescence polarization, or radioactivity, such as filter binding assays and scintillation proximity experiments.[8, 9] These techniques are widely applied in the pharmaceutical industry due to

their high throughput (up to 100,000 compounds per day)[10].The disadvantages of such methods are often linked to the labeling step. Such process can be costly and time-consuming, provide little information about the structural consequences upon binding, and may interfere with molecular interactions by disrupting the binding site, causing false negatives.

Our approach involves ion mobility separation coupled to MS (IM-MS). This technology detects not only the mass (m/z) of ions produced from protein-ligand mixtures, but can also detect their orientationally-averaged size, and we are developing methodologies that utilize such multi-dimensional measurements in order to screen protein-ligand complexes for drug discovery. Our main approach is termed collision induced unfolding (CIU), and it involves heating of protein-ligand complex ions following electrospray ionization (ESI) from solution so that the protein-ligand complex ions unfold in the absence of bulk solvent. There are still many unanswered questions and challenges for the development and applications of the CIU method, and the goal of this thesis is to develop tools and protocols capable of robustly assessing the binding consequences, including binding strength and stability, as well as subtle conformational change, within protein-small molecule and protein-protein complexes. In addition, the underlying mechanistic details of the CIU process itself, as well as its ability to be universally applied in ligand screening to a range of various protein targets, is still not fully understood.

1.1 Mass Spectrometry in Drug Discovery and Development

The process of introducing a new pharmaceutical compound as an approved therapeutics serves as a highly integrative endeavor, involving disciplines ranging from computer science to biology.[11, 12] Following the identification of viable biological targets, the majority of which

are proteins that fall into a few narrow functional classifications, the generation of small molecule drugs can be broken roughly into two equally important phases of work.[1] The first, termed 'discovery', involves the experiments needed to identify molecular scaffolds and structures that alter the function of the desired biological target and terminates with a candidate compound of promising potency. The second, generally referred to as 'development', refers to the experiments and clinical trials that seek to validate the toxicology and safety of the drug for use in patients. In both drug discovery and development, many analytical tools are utilized to assess the binding, structure, stability, and mechanism of action of potential drugs. For example, fluorescence and radiometric assays are routinely used during drug discovery to search vast compound libraries for any that bind to a target protein.[13, 14] Mass spectrometry (MS) of large bio-molecules is a versatile tool used within both drug discovery and development.[1]Technologies such as electrospray ionization (ESI)[15]enable MS to contribute to proteomic analyses[16], purity assessments of isolated targets, structural determinations of protein-ligand complexes during discovery[17, 18], and biomarker discovery that link candidate molecules to critical metabolic processes in pre-clinical evaluations.[2, 19]

Over 20 years ago, the first evidence that intact protein-ligand interactions could be retained during ESI-MS was reported.[20, 21] Following these key observations, many reports followed that illustrated the power of MS for assessing the stoichiometry and strength of small molecule binding to various protein targets.[22]Subsequent experiments expanded the role of such ESI-MS to include larger multi-protein targets,[23] complex mixtures of small molecule binders,[24, 25] accurate measurements of dissociation constants,[26, 27] and ultimately proteome-wide information on direct protein-ligand interactions.[28] In each of these cases, the power of the MS approach for drug discovery and development was clearly illustrated, and its complementarily

with other technologies within the pharmaceutical pipeline made clear. The ability of MS to gain a large amount of both qualitative and quantitative information from complex, dynamic biological mixtures is its chief advantage over other analytical tools. When labeling chemistries,[29, 30] chemical cross-linking,[31] hydrogen-deuterium exchange,[32-34] and other technologies are combined with MS, it becomes clear that the range of this information content can include significant structural data on the protein-ligand interaction that can rapidly inform the discovery of lead compounds.[14]

Ion mobility (IM) spectrometry, a gas-phase separation tool analogous to electrophoresis in solution,[35] is often combined with MS to generate tandem, multi-dimensional datasets for complex peptide mixtures,[36, 37] tissue imaging experiments,[38, 39]and protein-ligand complexes of pharmaceutical interest.[40, 41] IM separates gas-phase ions based on their differential transport through an environment of inert neutrals in the presence of a relatively weak electric field.[42, 43] Many versions of the technology exist, and in most cases precise measurements of ion transport properties can be used to generate ion size information, in the form of an orientationally-averaged ion-neutral collision cross section (CCS). While IM measurements of protein size take place in the absence of bulk water, and it is clear that surface groups likely rearrange during desolvation,[43, 44] many datasets now exist that point to a clear correlation between the CCS values recorded and X-ray or NMR datasets for the same proteins and complexes, thus enabling their utility for structural biology.[23, 45] These results, combined with the recent introduction of commercially-available technology make such IM-MS experiments a potentially-powerful addition to the discovery and development process for therapeutic small molecules.[46, 47]

1.2   Mass Spectrometry for Studying Protein Complexes

1.2.1 ESI

The desolvation and ionization of analytes serves as the first step in MS experiment. ESI is the most widely used soft ionization method to produce gas phase ions for labile bio-molecule analysis.[48]ESI generates gas-phase ions by applying a large electrostatic potential difference between the tip of a capillary emitter and a counter-electrode which typically comprises the entry inlet of the MS vacuum system (1-4 kV). For protein-containing solutions, a positive voltage is typically applied, making the liquid at the capillary tip enriched with positive ions. The high field causes the exposed liquid surface to form an extended structure, known as a Taylor cone, from which droplets are emitted. Aided by both parallel and then counter-directional flows of nebulizing and drying gases respectively, solvent evaporation from these droplets results in a reduction of their diameter. This reduction in droplet size continues until the Columbic repulsion between the increasingly crowded surface charges becomes strong enough to overcome the liquid surface tension. At this point, termed the "Rayleigh limit", droplet fission occurs. The upper limit of net droplet charge is given by the Rayleigh equation:

$$z_R = \frac{8\pi}{e}\sqrt{\varepsilon_0 \gamma R^3} \qquad\qquad (1)$$

Where z_R is the number of elementary charge e, R is the droplet radius, and ε_0 is the vacuum permittivity. The surface tension γ is 0.059 N m^{-1} for water at its boiling point.

There are two main mechanisms that govern ESI ion formation. Smaller analytes are thought to be primarily created through the Ion Evaporation Model (IEM)in which analyte ions are emitted directly from the droplet surface prior its complete evaporation.[15, 49]In the case of larger

analytes, like intact proteins, successive Rayleigh fission events result in a nano-scale droplet from which all solvent evaporates, leaving behind a charged analyte.[50] This process is typically referred to as the Charge Residue Model (CRM).[51]Recently, a combined CRM and IEM model was proposed by Hogan *et al*,[52] in which macromolecular ions are produced by CRM, but their charge state is determined by the field emission of charged ions from highly charged droplets. Alternatively, it's been proposed that for unfolded proteins, ion formation occurs via a chain ejection model (CEM), during which proteins are expelled from the droplet prior to complete solvent evaporation as disordered chains.[53, 54] Both computational and experimental investigations support this model, which is primarily driven by electrostatic repulsion and hydrophobicity-based arguments specific to unfolded proteins.[55]

Typically, ESI sources use pressure-driven flow, with rates in the μL /min range, utilizing emitter outlet sizes in the range of 50-100 μm. Ion desolvation typically requires the application of a drying gas and heating, and results in relatively low (<1%) ionization efficiencies.[56]To solve many of these problems, nano-electrospray ionization (nESI) was developed,[57] using emitter sizes of only 5-10 μm. Flow rates for these devices are in the low nL/min range, thus generating small droplet sizes with no application of additional heat or drying gas, and resulting in higher (~10%) ionization efficiencies with lower potential voltages required for Taylor cone formation. nESI also shows significantly improved sensitivity and higher tolerance to salts and other contaminants when compared to conventional ESI because of its ability to produce decreased initial droplets.[58, 59]

1.2.2Native Mass Spectrometry

Buffer conditions encountered by proteins prior to ESI plays a pivotal role in the resultant ions' structure in the gas phase and observed mass-to-charge ratios (m/z).[60] When proteins are prepared using conditions that favor denaturation, typically employing solvent compositions that include water/methanol or water/:acetonitrile mixtures and low pH, the gas-phase ions created tend to acquire a relatively large amount of ionic charges and unfolded gas-phase conformations[60].In contrast, if these same proteins are exposed to aqueous, buffered solutions at pH 7, compact sizes and native-like quaternary structures can be maintained in the gas phase.[43, 61] Ammonium acetate salts are often used to maintain ionic strength and establish a weak, volatile buffer for native MS experiments.[62-64]

In the early 1990s, Chait[21] and Henion[20] demonstrated that protein-ligand complexes can be directly observed by altering the conditions of the ESI experiment, specifically the pH.[65, 66] With the development of nESI, coupled with the use of volatile buffers, the field of native MS leapt forward to analyze larger, more complex non-covalent assemblies of biomolecules. For example, early work by Robinson[66] and Heck[67]demonstrated the capacity of MS to analyze megadalton-scale multi-protein complexes[68, 69]Such assemblies can now be routinely investigated by MS, leading to insights into the assembly pathways and ligand binding stoichiometries of macromolecular protein machines.[70, 71]

Several reports have demonstrated that ESI is capable of transporting intact protein complexes from solution into the gas phase, while still retaining native state stoichiometries for the gas-phase assemblies detected when low protein concentrations are used.[43, 61, 66]On the other hand, when high protein concentrations are used for native MS analysis, multiple biological units can be trapped in the same ESI droplet during the fission process, thus leading to the detection of false-positive complexes. This process produces complexes that are commonly referred to as

'non-specific' assemblies or 'ESI artifacts'. (See Figure 1-1 for detailed illustration) Such ion formation mechanisms can lead to highly-complicated MS spectra and decreased overall signal intensity, both of which can be detrimental to any quantitative analyses of the extracted MS signals.[26]To minimize the formation of ESI artifacts, several strategies can be employed, including using nESI to generate smaller droplets, applying mild activation to dissociate the artificial complexes, as well as lowering the concentration (<20 µM) of protein samples. For quantitative MS measurements that aim to determine binding affinities (K_D), a titration of protein concentration and/or protein : ligand molar ratio is typically applied, and the averaged binding affinity is reported from multiple measurements, taken over the concentration range probed.[26, 27, 72]
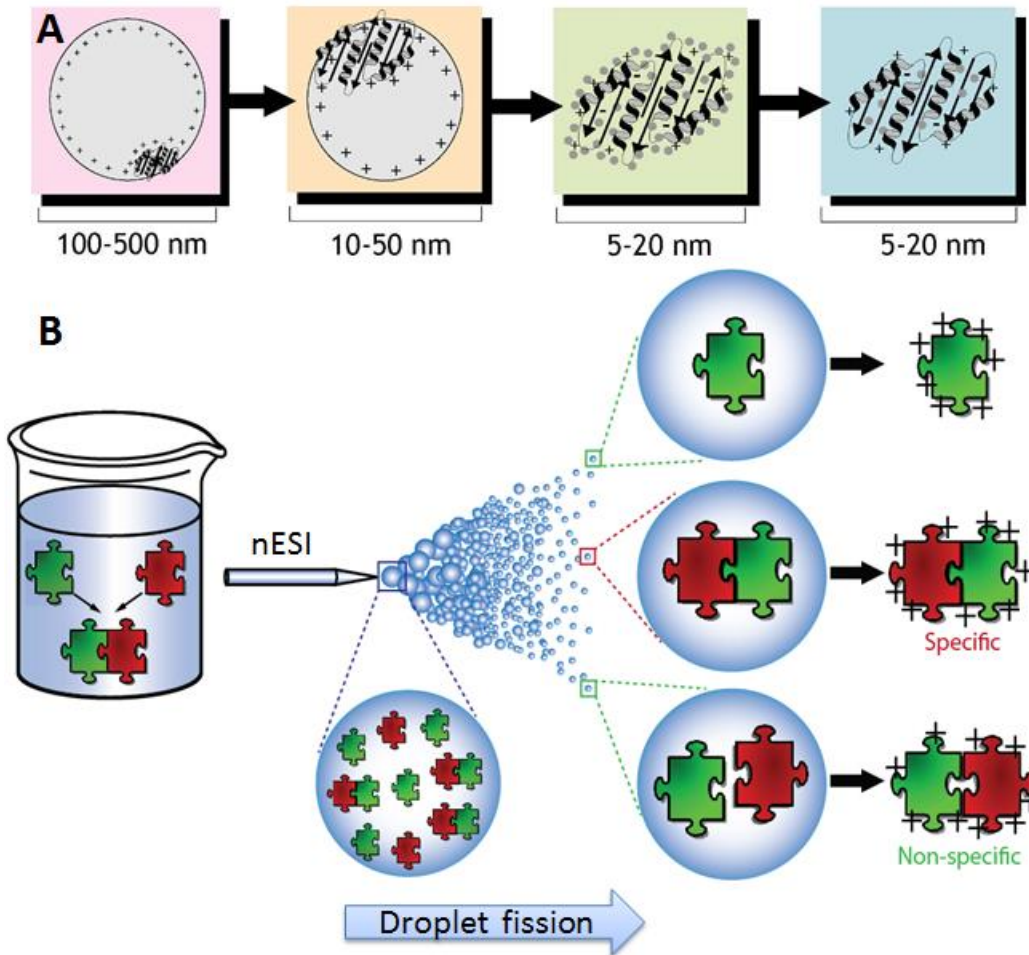
Figure 1-1. Mechanistic depiction of electrospray ionization **(A)** Schematic representation of the droplet desolvation process, depicting how non-covalent interactions of a protein is retained from solution phase to gas phase **(B)** A CRM (charge residue model) type schematic depiction of the nESI process where two distinct proteins (green and red)are present in a sample. Typically, signals for individual protein ions **(top right)**, specific interactions that exist in solution**(middle right)**, as well as non-specific ESI artifacts**(bottom right)** can all be detected by MS. Analytical workflows and software approaches are available to differentiate between the various channels shown (described in text).

1.2.3 Quadrupole Mass Filter

Quadrupole filters consist of two pairs of parallel conductive rods, positioned perpendicularly to each other. The two opposite electrodes have an applied potential of U+Vcos(ωt), and the other two have a potential of –(U+Vcos(ωt)), where U is a DC voltage and Vcos(ωt) is an RF potential. The applied voltage will affect the trajectory of ions traversing the flight path centered between the four rods. For given DC and RF voltages, only ions of certain m/z values can pass through the quadrupole, and all other ions deflected into the electrodes. The quadrupole serves as a narrow band mass filter in this mode, whereby ions of a desired m/z value can be selected by tuning the DC potential and RF field. Alternatively, quadrupole can also act as a broad-band mass filter when only RF field is applied. This mode is widely used since it has good tolerance to high pressures and capability to access a wide range of masses. To achieve enhanced transmission efficiency for large protein assembly ions, a modified quadrupole mass filter is housed in the Synapt G2 IM-MS system used for measurements described in this thesis, capable of selecting ions up to 32,000 m/z, in contrast to the typical m/z range for such devices (<4,000).Installing this capability within our MS instrument involves decreasing the operating RF frequency that drives the quadrupole rods.[47, 73] Selected ions can be subjected to collision induced dissociation (CID) or collision induced unfolding (CIU) following selection in the quadrupole.
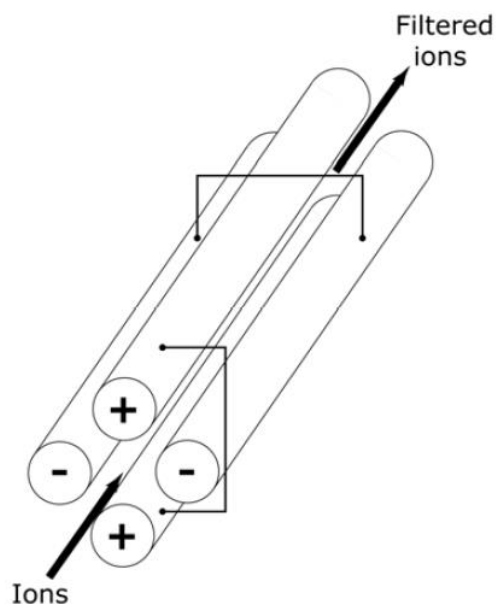
Figure 1-2. Schematic representation of a quadrupole mass analyzer. Ions are introduced in parallel with four electrodes arrange in a cylinder. Opposite rods are connected in pairs to the same DC voltage and RF field. Each pair has opposing DC polarity and RF field phase difference of π.

1.2.4 Time of Flight Mass Analyzer

In Time of flight (ToF) mass analyzers, ions are accelerated by an electric field of known strength. A classic instrumental setup consists of electrostatic rings to compact ion beams to produce packets of ions, a pusher where the ion packets are accelerated by a constant voltage (U) to the same kinetic energy, and a drift tube where ions are separated according to the time taken for ions of charge z to travel through a field-free drift region. Electrostatic lenses are often used to focus ion beams into the ToF analyzer, and then ion packets are injected into the drift region. A pulse may be delayed to correct for a spread of kinetic energy of ions. Optimizing the delay coefficient will average out the influence of kinetic energy distribution. In order to radially

confine ion beams, a space focus can be achieved in the orthogonal extraction source by carefully tuning of DC fields and the magnitude of the extraction pulse utilized at the pusher. All ions must traverse a same fixed distance in the flight tube to reach the detector. Since the ions of different mass possess different velocities, the time each ion takes to reach the detector is recorded and is directly related to the m/z by:

$$t = \frac{d}{\sqrt{2U}} \sqrt{\frac{m}{z}} \qquad\qquad (2)$$

Where t is the flight time, d is the fixed distance that ion travels through the drift tube and U is the electric potential applied in the pusher to give all like charged ions equivalent kinetic energy.

Most commercially-available ToF instruments are equipped with one or more reflectrons. When ions are ejected from the pusher into the flight tube, two ions with the same m/z theoretically have the same kinetic energy, and therefore the same velocity. However, due to differences in starting positions within the pusher, and differing initial kinetic energies in the direction of flight, this is not typically the case in real-world instruments. The purpose of the reflection is to compensate for the different kinetic energies for ions of identical m/z.[74]To achieve this, the reflectron acts as an ion mirror, where a series of equality spaced ring or grid electrodes creates an electric field to redirect the ions, in a fashion similar to light on a mirror, toward the detector or a successive reflectron. Ions of greater kinetic energy will penetrate the electric field more deeply and take a longer time to reach the detector, while ions with lower kinetic energy will penetrate less deeply and more quickly reach the detector. Properly configured, the difference is flight distance caused by reflection will normalize the total flight times of equal m/z ions for different starting kinetic energies, and therefore produce mass spectra with higher resolving power.[74]

1.3 Ion Mobility-Mass Spectrometry

Though the history of ion mobility spectrometry can be traced back to the early 20th century, IM-MS was only recently coupled to soft ionization techniques, for example ESI and MALDI, and released as a key commercially-available platform for biomolecular analysis. By measuring the changes in analyte CCS, important information, such as conformational dynamics[75, 76], folding and unfolding intermediates[77], ligand-induced stability/conformation changes[78, 79], aggregation intermediates and protein topologies[43] can be obtained. It should be noted that all the experimental measurements described in this thesis were all carried out on a Synapt G2 HDMS (Waters, Milford, MA) platform. Pictorial representation of the instrument in shown in Figure 1-3, which includes an nESI source, a quadrupole mass filter modified for high-mass (32k m/z range) operation, a T-wave ion mobility separator (TWIMS), and a time-of-flight (ToF) mass analyzer, all arranged in tandem.
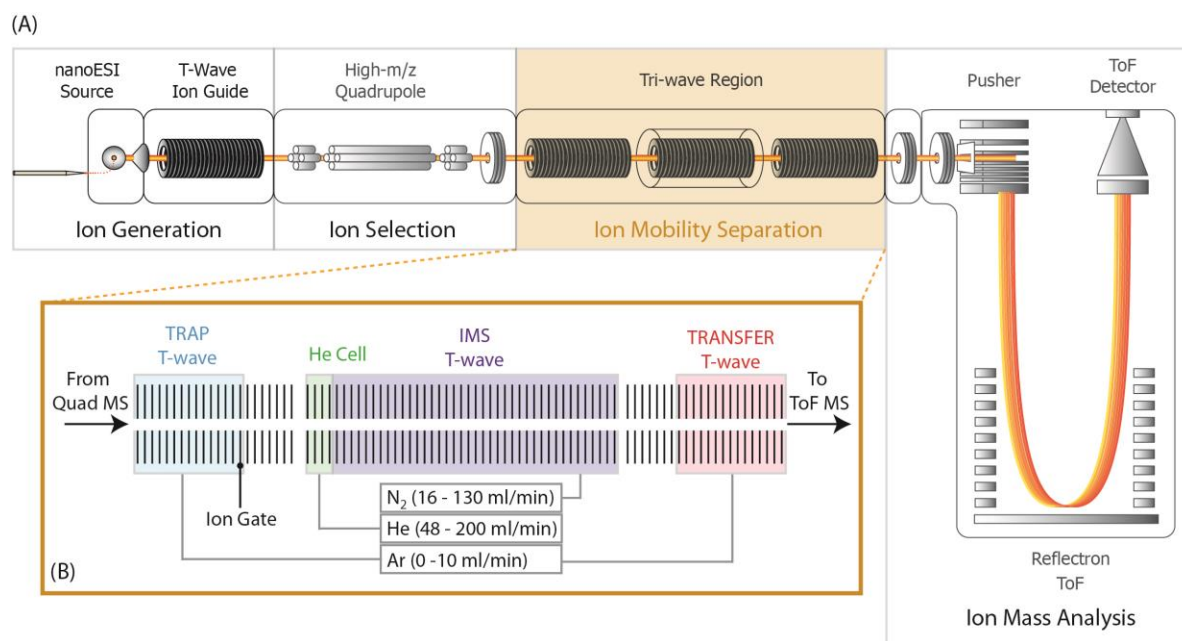
Figure 1-3.A Schematic diagram of the Synapt G2, quadrupole-ion mobility-time-of-flight mass spectrometry instrument used in this thesis work.**(A)** Overall schematic of the complete instrumentation indicating the four main regions of instrument operation: ion generation (using a nESI ion source), ion selection (using a modified quadrupole mass analyzer capable of selecting ions up to 32,000 m/z), ion mobility separation (carried out in the tri-wave region described in detail in B), and ion mass analysis (using a time-of-flight mass analyzer capable of ~40,000 mass resolving power). **(B)** Detail is shown for the ion mobility separation, or tri-wave region of the instrument. Three major T-wave ion guides dominate this section of the instrument, including the ion trap (blue), ion mobility (purple), and the transfer (red) regions. A fourth pressurized region is shown as the He cell (green), which facilitates the injections of ions across the pressure gradient that exists between the ion trap and the ion mobility regions with minimal ion activation. Typical mass flow controller values are also shown for the gas flow (in ml/min) into each enclosed region of the instrument.

1.3.1 IM Separation

IM separation, when coupled to MS, enables to collection of protein complex size information, and when this is combined with the connectivity information described above, coarse-grained or atomic models of the assemblies can be constructed.[80] IM separates protein ions based on their ability to traverse a chamber filled with inert neutrals under the influence of an electric field. Larger protein ions undergo a greater number of collisions with the inert neutrals filling the chamber, and therefore have a larger CCS values, than more compact protein ions of identical m/z (Figure 1-4).[42] While this description holds for most contemporary IM separations described currently in the literature, modern IM technology expands this basic principle into a variety of instrument platforms available for IM-MS experiments. Such instrumentation, as applied to multi-protein complexes, takes three basic forms: Drift Tube (DT)-type, Differential Mobility Analyzer (DMA)-type, and TWIMS instruments. All of these technologies have both strengths and weaknesses for the analysis of multi-protein assemblies.[35, 81]
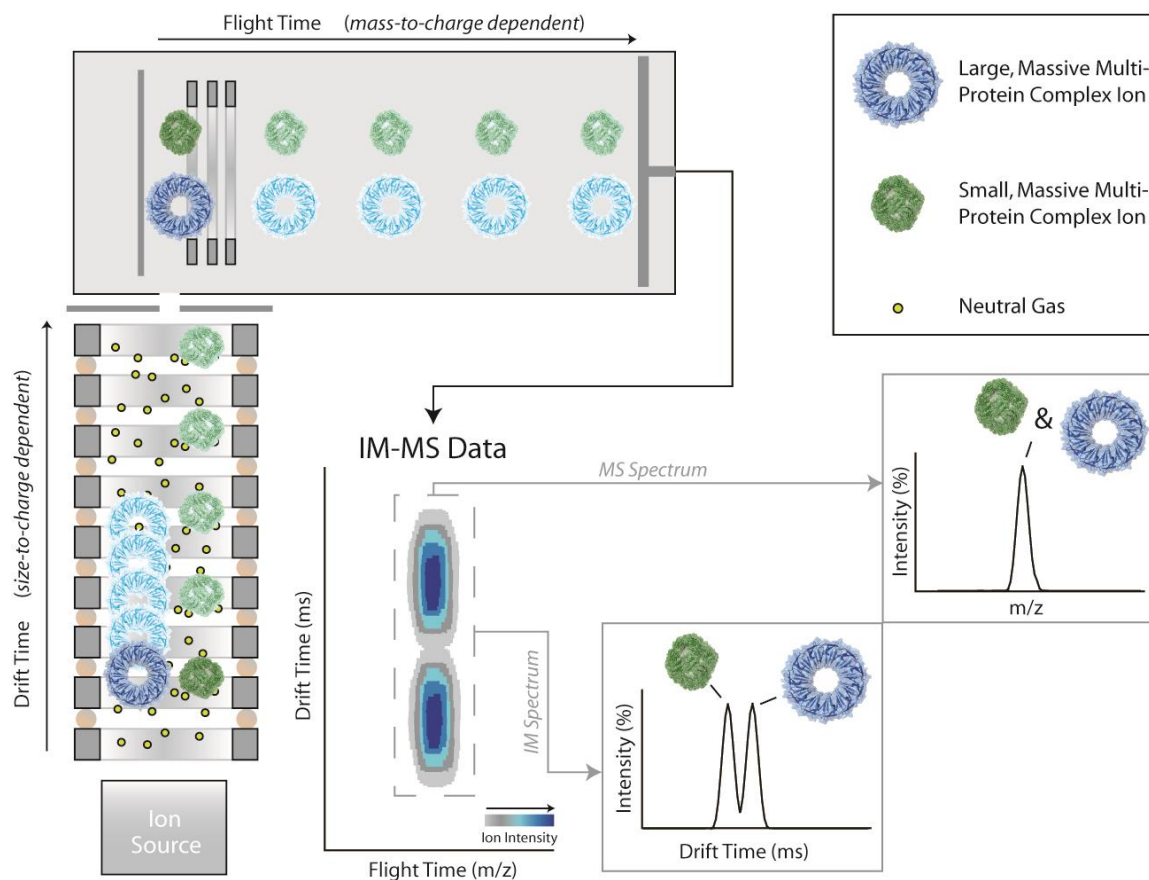
Figure 1-4. Fundamental separation principles and data acquisition schematic of ion mobility-mass spectrometry. Ions are generated at the ion source **(lower left)**, and are allowed to drift in an ion guide filled with neutral gas molecules under the influence of an electric field. The ions migrate through this region according to their size-to-charge ratio. They are then injected into a time-of-flight (ToF) mass analyzer under vacuum for mass-to-charge (m/z) analysis. The resulting data is 3-dimensional, containing ion intensity, size, and mass information. The various dimensions of the data can be shown as a contour plot **(middle, bottom)**, or 2D selections in drift time or m/z (lower right). Figure legend for the diagram is shown in upper right.

Early native IM-MS studies on intact proteins and small complexes were mostly conducted using DT type instruments.[35, 82] DT-IMS typically comprise a series of ring electrodes upon which a fixed axial gradient of electric potential is constructed. Measured IM drift time is then proportional to ion CCS, thus enabling direct access to ion size information. While early versions

15

of DT-type devices suffered from ion loss and poor sensitivity when used for biomolecule analysis, contemporary implementations of the technology use ion guides either during or after IM separation in order to refocus ions and preserve limits of detection.[83-86] Moreover, tandem DT analyzers have been used to assess micro-conformational states within the structural envelope of gas-phase protein ions, and have illustrated that such structural populations can be stable on the millisecond timescale, thus enabling detailed investigations of protein structure in the gas phase.[87, 88]

Since the introduction of the Synapt platform, the majority of IM-MS datasets for multi-protein complexes have been generated on IM-MS instruments using TWIMS. Such analyzers are similar in basic construction to DT-type IM devices, but differ significantly in their operation. Rather than a linear field gradient, ions are propelled through the analyzer using a series of low-voltage waves.[47, 89] Ions are carried by the waves relatively briefly before being subsumed by the wave front in a manner depending on the CCS of the ions being separated, generating a time-domain IM separation similar to DT-IM devices.[90] An important feature of this process is that, due to the nature of the separation mechanism employed, TWIMS drift times are most often calibrated using standard CCS values for protein complexes rather than calculated directly from drift time measurements.[91] Apart from being the first high-transmission efficiently IM analyzer currently incorporated into commercially available, TWIMS analyzers offer some modest advantages in terms of separation resolution when compared to other IM devices.[46, 92, 93]Often defined in terms of the centroid arrival time of the IM peak normalized to the IM peak width (t/Δt), drift time resolutions for DT analyzers range from 30-150 for research-grade instruments, with those at the high end of the range produced using instruments with very long

flight tubes (>1m) and high separation voltages (multiple kV).[94-96] Because of the physical principles involved in T-wave IM separation, drift time is correlated to CCS through an exponential relationship, which results in a TWIMS drift time axis that is effectively 'stretched' relative to those achieved on DT analyzers.[91]This relationship enables TWIMS separators to achieve 40-60 CCS resolution (CCS/ΔCCS) using comparatively shorter devices, operating at lower fields and pressures, than DT devices of equivalent dimensions.[46, 93, 97]

1.3.2 Collision Induced Unfolding (CIU)

ESI-MS plays a pivotal role in quantifying the bound states within complex multiprotein-ligand systems, and IM-MS can supplement such efforts by isolating protein-ligand complexes for stability measurements in the gas phase, following collisional activation. Such CIU assays can be used to investigate the stability and conformational difference of protein ions with/without ligand binding. Ions are collisionally activated in the ion trap T-wave region by collision with argon gas, the resulting ion conformations are further separated in the IMS drift cell and detected by ToF. By adjusting the 'trap' collision voltage, the kinetic energies of protein ions entering trap region can be controlled. Activation profiles are typically collected from step-wise increments. Such CIU experiments were first described for small monomeric proteins ions[98] but have rapidly expanded to include more detailed instrumentation and applications covering large multi-protein complexes.[76, 78, 79, 99]

Since the first observation of protein complex CID, protein unfolding events have been invoked to describe the details of the process,[100]. During protein complex CID, charges are partitioned asymmetrically on the eventual product ions observed, and thus an asymmetrically unfolded

transition state was invoked for such experiments early in the development of the field. For example, Jurchen and co-workers provided convincing evidence suggesting that the origin of such asymmetric charge partitioning within protein homodimers is linked directly to the foldness of the transition state, resulting in some monomeric units that carry away the majority of charge.[101, 102] In terms of the underlining mechanism of this collisional unfolding/dissociation phenomenon, it is proposed that the apparent gas phase basicity of the elongated proteins is greater than that of a folded one at the same charge state, rendering it energetically favorable to transfer protons from a folded to an unfolded protein.[101] Moreover, high transition-state entropies measured during blackbody infrared dissociation experiments on homo-pentameric protein complexes described by Klassen *et al* also indicate that unfolding events take place during the CID transition state.[100]

In modern experiments, CIU fingerprints, contour plots that tracks the size change of protein ion as a function of collision energy, have been developed to assess the subtle stability and conformational changes that occur within proteins in the presence and absence of ligand binding (detailed workflow in generating CIU fingerprints is illustrated in Figure 1-5), evaluate the binding potency of various sugar ligands to a lectin protein[4, 79], classify a broad set of kinase inhibitors[76], differentiate various lipid binders potencies upon binding a an ammonia channel membrane protein[99], infer the correlation between gas phase unfolding to protein domain structures in solution[103]and differentiate the monoclonal antibody structural iso-forms with various disulfide mapping patterns.
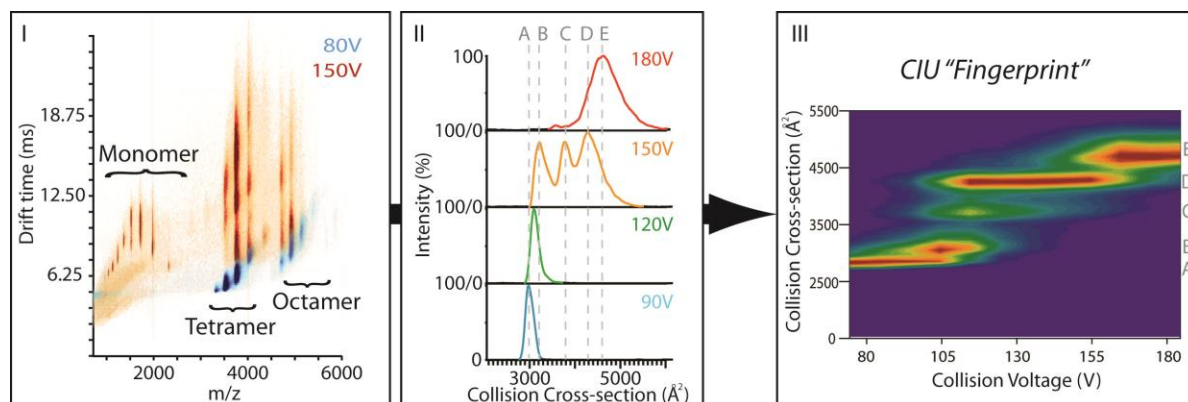
Figure 1-5. Schematic workflow for generating CIU fingerprints: **(I)** Protein ions are first generated by nano-electrospray ionization and different stoichiometries are identified by MS. Preliminary surveys are conducted which assess the change in IM drift time observed at different activation voltages. **(II)** A detailed collision voltage ramp is conducted with a single charge state selected in a quadrupole mass filter prior to the IM stage, collision cross-sections (calibrated to known ions) for all IM features are recorded, and all features having similar collision cross-sections are assigned to the same structural 'family' [shown as A-E]. **(III)** A three-dimensional contour plot is created that charts the change in intensity of each of the features observed in IM data as a function of collision voltage.

1.4 Technologies for Studying Protein-ligand Interactions

1.4.1 Non-Mass Spectrometry Technologies

The screening of bio-molecular interactions involves the determination of binding strength, activity, structural/conformational information and predicted *in vivo* availability of drug candidates, serves as a pivotal part of modern drug discovery process.[104] Well-established pharmaceutical screening pipelines currently require some type of fluorescent- or radio-labeling to deduce the binding of a ligand molecule to its receptor. Most of these label-dependent screening platforms are based on measurements of fluorescence, such as fluorescence resonance energy transfer (FRET) and fluorescence polarization, or radioactivity, such as filter binding assays and scintillation proximity experiments.[8, 9] These techniques are widely applied in the

pharmaceutical industry due to their high throughput (up to 100,000 compounds per day)[10].The disadvantages of such methods are often related to the labeling step, as well as the lack of structural and conformational information acquired during screens. The labeling process can be costly and time-consuming, and may interfere with molecular interactions by disrupting the binding site, causing false negatives. In contrast, label-free screening approaches provide enormous flexibility for protein-ligand screening and high throughput drug discovery efforts. Surface plasmon resonance (SPR) spectroscopy is a fast growing technique and serves as a key label-free technique for protein ligand binding assays.[105, 106] SPR involves spectroscopically interacting with a resonant oscillation of conducting electrons at a metal surface. A typical SPR assay involves tethering protein (bait) onto the surface, and a microflow cell is used to rapidly wash analytes (prey ligand) through, based on the change of SPR signal response, the 'on' and 'off' rates of the binding event can be deduced, and thus binding affinity (K_D) can be calculated. SPR is capable of measuring the real-time quantitative binding affinities and kinetics for protein-ligand complex with relatively low sample consumption. The interacting molecules may be proteins, peptides, lipids, viruses, nucleic acids, or small organic molecules such as fragments or drug candidates. The ease with which real time information can be obtained has changed many conventional approaches in both antibody and small molecule/fragment interaction analysis, from label based and affinity/IC50 (a measure of how effective a drug is. It indicates how much of a particular drug or other substance is needed to inhibit a given biological process by half.) based protocols towards a label free and kinetic based workflow.[107]

Isothermal titration calorimetry (ITC) has long been viewed as the 'gold standard' in characterizing the thermodynamics and stoichiometry of protein-ligand interactions. When titrant ligand is injected into the protein solution and binds to each other, heat is generated or absorbed,

which is proportional to the fraction of bound ligand. Sensitive thermocouple circuits detect the subtle heat change plots as a function of incoming ligand concentration. The complete thermodynamic profile(enthalpy, entropy and free energy), as well as the binding constants and reaction stoichiometries for protein-ligand complexes can be deduced with careful isothermal curve fitting.[108] ITC does not require target immobilization or modification of reactants with fluorescent tags. It has been routinely used to characterize various types of binding reactions including protein-small ligand, protein-protein, protein-membrane, as well as drug-receptor interactions. To achieve high quality ITC data, however, great care in concentration determination and sample preparation is required.[109]In addition, large amounts of sample (typically milligram level) having high purity is often required for successful ITC analyses. Moreover, in the context of multi-ligand binding proteins, sophisticated models are often needed to fully interpret data.[110]An application of ITC toward protein-sugar binding is discussed in detail in Chapter 2.

1.4.2 Mass Spectrometry Related Technologies

MS is wildly applied in probing the structure and dynamics of various protein-ligand complexes present at physiologically relevant concentrations over a wide range of solution conditions.[111] Numerous, sensitive strategies can be used for interrogating structural/conformational changes, folding and ligand binding via mass shift readouts, such as covalent labeling[112, 113] and HDX[114, 115]. The fundamental principle of these solution phase labeling methods is to alter the analyte mass in a conformation-dependent fashion. Protein structures that are disordered or unfolded experience more extensive labeling, whereas tightly folded structures, often achieved

upon ligand binding, undergo less modification from labeling. HDX-MS evaluates the solvent accessibility of a protein, or a protein-ligand complex. Such an information profile can be further used to annotate regions of protein structure/sequence according to its apparent flexibility and stability. Similarly, covalent cross-linking also serves as indispensable tool for uncovering the connectivity of multi-protein assemblies, and generating distance constraints that are highly important for constructing structural architectures.[116, 117]

1.4.3 Estimating K_D Values by Mass Spectrometry

Classically, K_D values for protein-ligand complexes can be quantitatively evaluated by a variety of techniques (details in section 1.4.1). MS combined with soft ionization techniques has been used for over a decade to perform similar assessments of protein-ligand K_D values, having the primary advantages of sensitivity, speed, and the direct detection of binding stoichiometries within mixtures. Klassen[26] and Zenobi[118] have both developed methodologies that directly evaluate protein-small molecule binding affinities using ESI-MS intensity values, and related these signal intensities directly to protein concentrations in bulk solution. By carefully analyzing the influence of factors including: solution pH, analyte absolute concentrations and relative molar ratios, and as in-source CID, binding constants for a host of protein/small molecule systems can be measured by nESI-MS in a manner that provides good agreement with ITC experiments.[26]The Klassen lab have further enhanced the throughput of MS based approaches specifically aimed at the analysis of protein-carbohydrate complexes through the development of catch and release ESI-MS methodologies [119],.

MS-based K_D values reported for protein-ligand binding are built upon several key assumptions, one of which that requires that the ionization and detection efficiencies for both the bound and apo states of the protein are similar, if not identical. Many studies have borne out the accuracy of this assumption, in general, for relatively large proteins bound to small molecule ligands, as such binding events do not alter the charge-accessible surface area or intact mass of the apo protein significantly.[26, 27] In contrast, very few reports of MS-based K_D values for protein-protein binding events have been reported, as it is challenging to predict the validity of the key assumption discussed above *a priori*. For example, Konermann and co-workers have studied the self-dissociation constants of both β-lactoglobulin and hemoglobin homo-dimer.[120]In addition, the Robinson group quantified an impressive number of protein-protein K_D values that define the Hsp90 chaperone complex interaction network using direct MS and kinetics assays.[121] Lingering questions still remain unanswered, however, regarding appropriate methods and interpretation frameworks surrounding MS-based K_D values recorded for weakly associated protein-protein complexes. A detailed discussion on the subject of such binding affinity measurements for hetero-protein complexes is provided in Chapter 4.

1.4.4 Binding Cooperativity in Protein Substrates

The activity of multi-protein systems is often subjected to allosteric regulation that is achieved by conformational changes induced or stabilized by ligand binding.[122]The allosteric control of protein activity is classically measured using sigmoidal plots of the initial ligand binding reaction velocity or through the fractional protein saturation observed as a function of ligand concentration. For example, ensemble measurements methods like ITC[123] can, in principle, be

analyzed to determine Hill coefficients which quantify the relative cooperativity of binding observed in experimental data, but such analyses are often difficult to execute, nor to extract any mechanistic insights, especially for large multi-protein systems, and highly dependent upon the absolute and relative ratio of ligand concentration.[124, 125] To assess the stability shifts and depict mechanistic insights behind the allosteric interactions for such systems, the separation of individual bound states of the bio-molecules is required, which is often not possible for conventional spectroscopic or chromatographic techniques. These difficulties have resulted in a general dearth of experimental evidence for cooperative stabilization effects in proteins upon ligand binding.

As a gas-phase analog of calorimetry experiments, CIU and MS measurements can be used to capture cooperative increases in protein stabilities as a function of ligand attachment within multi-protein complexes that are either challenging or impossible to record using solution-phase ensemble-based measurements.[76, 79, 126]MS provides a means to distinguish between various allosteric binding models, as individual bound states of the protein can be resolved by MS. As such, the binding constants for each individual bound state within large multimeric protein assemblies can be deduced, and any allosteric/cooperative mechanistic details can be inferred directly from such MS data.[126]Alternatively, stability shifts between adjacent protein bound states can be recorded through CIU measurements to reveal important insights to the allosteric binding behavior, especially for large proteins and complexes that bind multiple ligands. [79]

1.4.5 Strategies for Studying the Structure and Ligand Binding Properties of Membrane Proteins

Membrane proteins play a pivotal role in mediating the transport of solutes in and out of the cell, and translating the action of extra-cellular stimuli into function.[127, 128] These assemblies are of great pharmaceutical significance, and thus represent more than two thirds of all the druggable targets.[129, 130]However, their general insolubility in aqueous solvents, reliance upon fragile membrane-lipid interactions for structural stability, and their propensity to aggregate renders the structural and functional characterization of membrane proteins remain challenging.[131] While remarkable advances have been achieved by X-ray crystallography and NMR spectroscopy for individual membrane proteins and small complexes,[132, 133]as well as recent developments in Cryo-EM technique and technology that have yielded near atomic resolution for membrane protein complexes prepared in detergent micelles, comprehensive structural analysis of membrane proteins is well beyond our current structural tools.[134]

Within the past decade, MS has made profound contributions to the structural characterization of membrane proteins.[135-138] MS has already been applied to identify the sequence of the membrane proteome,[139]and by coupling such workflows to chemical labeling techniques, secondary and tertiary structure information for membrane-bound proteins can also be achieved.[140-142]Similarly, Native MS is capable of providing invaluable information on membrane protein structure, interactions and dynamics, and works primarily by releasing intact protein ions from gas phase from detergent micelles ions generated by nESI, formed originally in solution.[136, 138, 143] Such methods, when used in tandem with IM, can provide a direct measure of the binding consequences of small molecules on membrane protein structure and stability.[99, 144, 145]The Native MS membrane protein analysis workflow requires detergent molecules in order to stabilize and solubilize membrane protein in solution prior to MS, which then must be removed in the gas-phase through collisional activation. As shown in Figure 1-6,

the choice of detergent is critical as excessive stabilization leads to insufficient detergent release from the analyte protein ions, while insufficient stabilization results in protein unfolding in solution. . The optimal choice is highly dependent on the nature of the detergents and the structure of the proteomicelle formed,[138, 146] Hence, a systematic screen is often necessary to achieve optimal resolution for native MS membrane protein experiments.
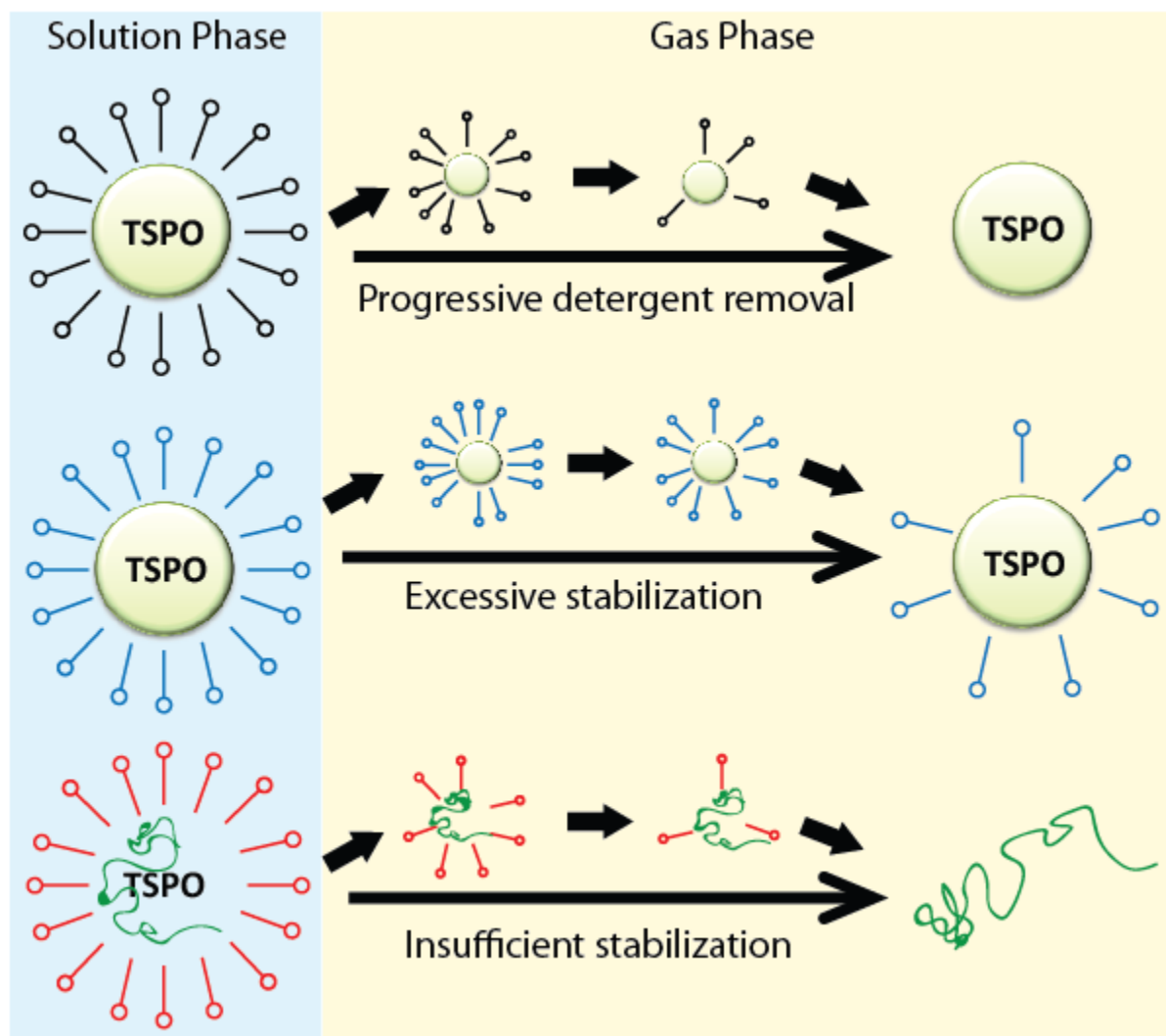
Figure 1-6. A schematic depiction of the transport of a membrane protein (Translocator Protein, TSPO used as an example), protected within a micelle, from solution (blue shaded region) to the gas phase (yellow shaded region) for subsequent detection by MS. Three scenarios are depicted: the detergent stabilizes the membrane protein optimally **(top)**, excessively **(middle)** and

insufficiently **(bottom)**. Note that ionization and other desolvation processes are not shown in this schematic.

1.5   Computational Strategies to Assist IM-MS analysis

1.5.1 Theoretical CCS measurements

Theoretical modeling is often used to provide structural explanations for IM-MS experimental measurements. A number of CCS calculation algorithms are broadly available, each developed for particular applications and molecular size ranges. The most accurate and complex of these is the trajectory method (TM)[147], which takes into account long-range interactions (Lennard-Jones potentials for example) and aims to approximate the momentum transferred upon each ion-neutral collision event. The TM typically provides the most accurate CCS estimates for a given model structure; however, the method is also computationally expensive and requires a precise understanding of the partial charges, charge placement, and charge mobility within the model to provide accurate results. The exact hard sphere scattering (EHSS) algorithm [148] ignores the ion-neutral long range interaction potential, and simply treats all atoms as hard spheres to compute the resultant momentum transfer collision integral. This method performs well for molecules with rigid structures like carbon clusters, but tends to over-estimate of the CCSs of most bio-molecular ions. The projection approximation (PA)[149, 150] method serves as the simplest and fastest CCS calculation method, equating the CCS to an orientationally-averaged projected area of the model. Because this algorithm ignores long range interactions and scattering effects, it is extremely fast but also tends to under-estimate the CCS. However, detailed comparisons between PA calculations performed on X-ray and NMR structures revealed

strong correlations to both experimental measurements and TM values computed for the same structures, enabling the generation of scaled PA values that can be used for the generation of solution-phase relevant multiprotein models and highly-accurate gas-phase structures respectively.[91, 111]

All of the above basic CCS calculation methods are available in the MOBCAL[147, 148] software package. Recent alternative software packages, such as IMoS[151], PSA[152] and IMPACT[153]are also now broadly available, and offer advantages over MOBCAL in the context of providing easy access to scaled PA values, or the ability to estimate CCS values in drift gases other than Helium. The theoretical CCS calculations presented in this thesis were all performed by either MOBCAL or IMoS, with specific details stated explicitly in chapter 3.

1.5.2 Molecular Dynamics: Simulated Annealing

Molecular dynamics (MD) simulations, specifically those employing simulated annealing-type MD, are often performed in order generate the model structures referenced in the sections above for comparison with IM-MS data.[154, 155]In a typical simulated annealing experiment aimed at generating models for protonated (positively charged) protein ions, an all-atom representation of the protein structure is subjected to an initial charge assignment phase, where ionic charge in the form of an ionizing proton is placed on the side chains of the most basic residues (Lys, Arg, His, and N-terminus) on the protein surface in a manner aided by an MD force field. After the resulting protein structure is subjected to an energy minimization calculation, in both NVT and NPT space, the resultant model structures are further subjected to a series of periodic gradual temperature ramps, where the elevated temperature enhances the energy of the system in order to

disorder the structure.[156] Each temperature increase is followed by a gradual decrease to anneal the protein structure from local minima on the potential energy landscape, eventually funneling the protein toward its lowest energy configuration in the gas phase.[157, 158] All the structures generated from the simulation can be evaluated in comparison with CCS measurements, using the CCS calculation algorithms described above. Models found in good agreement with experimental CCS measurements are considered as candidates for the final gas phase ensemble of structures.[159]

1.6 IM-MS for Protein-ligand Screening

1.6.1 Paradigms for Protein-ligand Screening by IM-MS

Early IM data for protein ions produced by ESI indicated a strong structural dependence on ion charge.[98, 100-102] Subsequent experiments have refined the ability of IM to separate subtly different conformational forms of proteins over a range of charge states produced during the ESI process.[45, 160] Computational methods are typically used in conjunction with IM data to generate atomic models of peptide and protein structure,[161] and have advanced significantly over the past few years in their ability to generate such models for ever larger systems.[162] Smaller protein-ligand systems can make use of such technology to deduce binding locations for small molecules within protein targets[76, 163-165] and, in some cases, atomic models of protein-ligand complexes.[166, 167] Larger protein-ligand complexes are currently beyond the scope of such detailed computational methods, and as such, screening in such cases first involves the observation of a key protein conformation shift as a function of a known binding event that can be linked directly to compound efficacy. Subsequent experiments can then be constructed to

search a broader library of compounds for similar conformation shifts upon binding the same target (Figure 1-7). This general mode of operation is currently the most-commonly employed approach for IM-MS in the context of protein-ligand analysis and screening, as reflected in the literature surveyed below.

In addition to the above-noted charge state dependence for gas-phase protein CCS, early studies noted other critical variables that affect the gas-phase structure of desolvated protein ions.[98, 168] Among these, altering the internal temperature of the ions produced had a dramatic influence on the size of the protein ion recorded by IM, primarily leading to a positive correlation between protein ion CCS and temperature with protein ions of high internal temperatures adopting large, string-like conformational states.[98] Subsequent data have extended these observations to include protein-protein[40, 43, 160, 169] and protein-ligand[78, 170] complexes, each of which displays similar yet distinct unfolding properties upon gas-phase activation. Most contemporary experiments utilize collisional activation to initiate unfolding,[78, 160, 170-173] however other activation methodologies have been shown to elicit conformational change,[98, 161, 174-176] although to a lesser extent. Collision induced unfolding (CIU) can be used in two basic modes in the context of protein-ligand screening experiments (Figure 1-7). In many experiments, the surviving population of the most-compact form of the protein, typically the form of the protein most-highly correlated to its solution structure, is tracked as a function of the voltage used accelerate ions and initiate unfolding. Differences recorded in protein ligand complex stabilities primarily relate to the stability of the gas-phase complex, and can be compared to both solution measurements and apo-protein CIU data to provide a workable screening methodology.[61, 78, 169, 171, 172] In addition to measuring the survival of a single conformational form of the protein-ligand complex upon activation, the unfolding pathway of the

protein can be followed in detail to generate additional points of comparison between either apo-states or alternate conformational families of the protein. Since many possible tertiary structures project identical ion CCS values, the detection of subtle conformational shifts in protein-ligand complexes is often challenging for IM-MS methods. CIU fingerprints can be a useful tool in circumventing such imitations, as the unfolding intermediates accessed by proteins during CIU can be uniquely related to specific protein-ligand binding modes.[78, 169, 171, 172, 177]

All of the above modes of operation can be combined into metrics that define the structural stability and conformation changes that occur upon ligand binding for an efficacious molecule, the properties for which are sought to be replicated in new molecular scaffolds (Figure 1-7). Alternatively, IM-MS results can be interpreted using other data, including NMR and X-ray structure information, or computational models in an attempt to link specific conformational shifts observed to desired ligand binding modes.[23, 40, 45, 76, 163, 165, 178] Once sufficiently-descriptive scoring algorithms are established, the developed screen can be applied to larger libraries to search for molecules that bear similar effects on target protein conformation and stability.
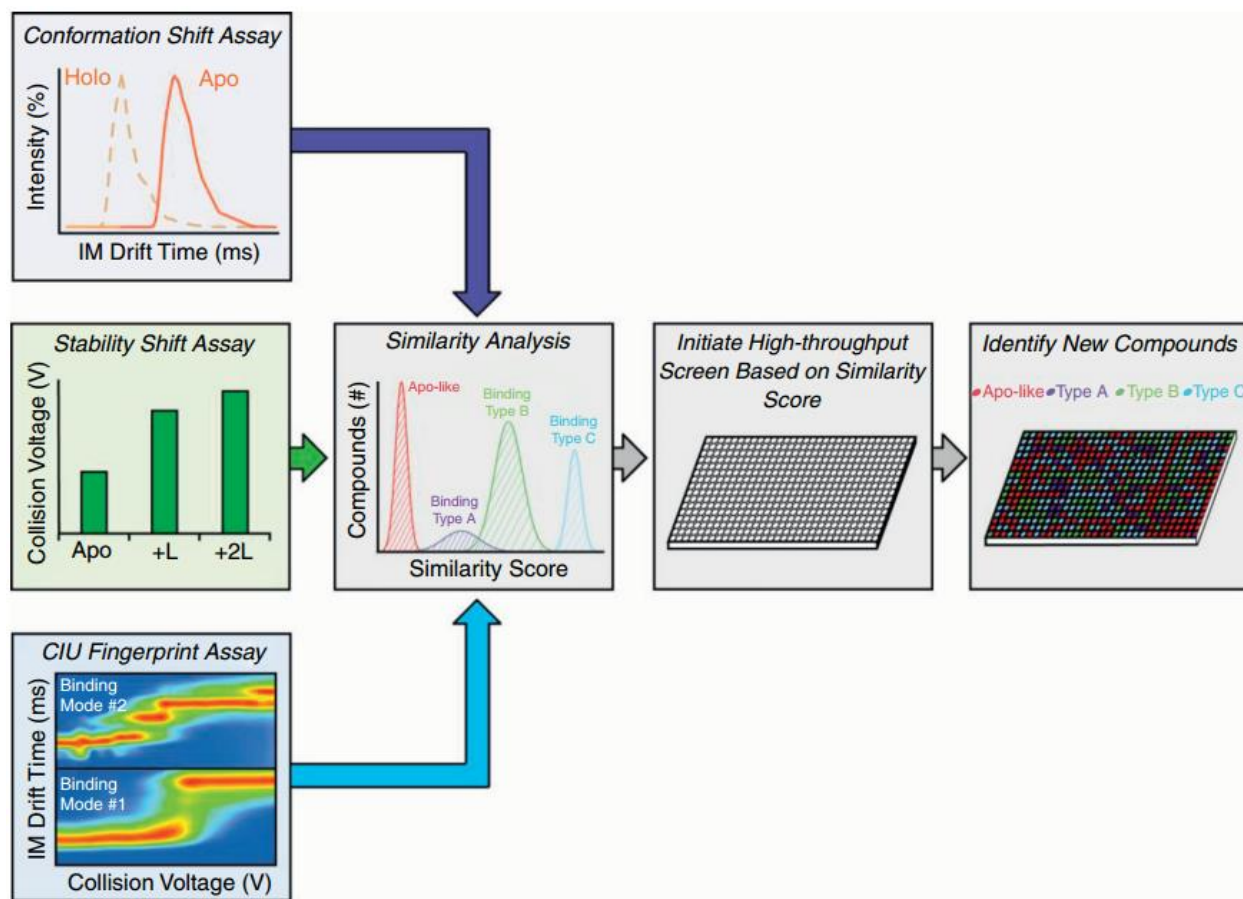
Figure 1-7. Basic screening strategies for intact protein–ligand complexes by IM-MS. Several different modes are available to assess the consequences of small molecule binding within intact proteins using IM-MS screens. Binding may result in a clear conformation shift by IM, and can thus be used as the basis for a conformational shift assay. Ligand binding may also alter the stability of the protein ion when compared to control data, enabling a stability shift assay. Finally, detailed protein unfolding data can be recorded by tracking the sizes occupied by protein–ligand complexes upon activation, and the differences observed between protein–ligand complexes of known binding modes or conformations can be used to construct CIU Fingerprint-based assays. Once known binders are analyzed and metrics assembled that allow for sufficiently accurate scoring of known data, resulting in clustered responses that differentiate a desired class of binder from other potential ligands, a library can be assembled from previously untested molecular scaffolds and used to search for new compounds that replicate the stability or conformation shifts observed in efficacious molecules.

1.6.2 Searching for Shifts in Protein-ligand Stability

Gas-phase protein-ligand stability measurements by ESI-MS have a long history.[61, 179, 180] While relative ion intensities can be used to generate binding constant information,[111, 161] collision induced dissociation (CID)[22, 181]and other tandem MS technologies have been used for many years to study the stability and dissociation of protein-protein and protein-ligand complexes.[101, 102] More recently, ESI-CID-MS has been applied to ever more complex protein-ligand complex systems of potential pharmaceutical interest. For example, a recent study investigated the dimeric monocytechemo attractant protein-1 (MCP-1), and found CID thresholds for the complex to be relatively low in the absence of Arixtra, a glycosaminoglycan analog binder. The results, which included IM-MS, indicated that the dimeric MCP-1 is significantly stabilized upon Arixitra binding, and that Arixtra interacts with both of the subunits within the MCP-1 complex.[182]

Such IM measurements have appeared with increasing frequency in conjunction with MS-based stability measurements of protein-ligand complexes. For instance, IM-MS was used to study the stability of ubiquitin-cis-[Pd(en)(H$_2$O)$_2$]$^{2+}$ complexes, and indicated that Pd-bound ubiquitin exhibits diminished gas-phase unfolding when compared to the apo protein. Furthermore, it was found that Pd^{2+} binding aided conformational stability to a greater extent than Pd(en)$^{2+}$.[183] IM measurements of protein ions bound to extensive anion and cation populations have been used to deduce a Hofmeister series analog for gas-phase protein structure.[171, 172] These studies identified several gas-phase specific mechanisms by which proteins can gain differential stability from bound ligands in the absence of solvent. Bound anions, for example, tend toward evaporation upon collisional activation of the complex, thus allowing the protein to dissipate excess internal energy and retain its shape over a broader array of IM-MS conditions. In contrast, cation adducts tend to stabilize complexes through remaining bound to the protein, serving to

tether regions of the protein through multi-dentate interactions.[177] More recently, IM-MS data for crown ether (CE) – protein complexes have suggested new modes of stabilizing protein structures in the gas-phase upon ligand binding.[44] The CE compounds studied non-covalently bind preferentially to primary amines, e.g. lysine side chains, and serve to solvate the ionic charge present. The IM-MS data collected showed that CE binding can compensate for rearrangements local to the charge site in a manner potentially similar to solvent molecules in the condensed phase, and thus suggests future routes for tuning and manipulating protein structures in the gas phase through ligand attachment.

Detailed CIU datasets have also been used to study protein-ligand stabilities, and serve as powerful tools to investigate the consequences of small molecule attachment in larger protein systems. In a recent study, ESI-IM-MS was used to evaluate the structural stability of natively-compact protein ions (FK-binding protein, hen egg-white lysozyme, and horse heart myoglobin) as a function of small molecule binding.[170]The results show clear shifts in the CIU stabilities of ligand bound complexes relative to apo protein, shifting the onset of CIU by up to 21 eV. CIU datasets were also used to assess the stability changes produced in familial amyloid polyneuropathy (FAP)-associated variant form of the tetrameric protein transthyretin (TTR) upon binding its natural ligand, thyroxine.[78] By combing CID and CIU datasets, it was found that thyroxine binding stabilizes the L55P disease-associated form of TTR to a greater extent than the wild type protein. Furthermore, CIU fingerprints were shown not to depend on the L55P point mutation, and that ligand binding primarily influenced the stability of the most compact tetramer conformations, rather than significantly unfolded intermediate unfolded forms of the complex. An example of this type of analysis is shown in Figure 1-8. Concanavalin A (Con A) is a 103 kDa lectin tetramer with well-known structure and sugar binding properties.[184]This data

shows CIU datasets recorded for three ConA-manosyl sugar complexes, and indicates strong

shifts in CIU stability for different ligand bound populations. Importantly, if selected areas of the

CIU plot are interrogated as shown, the relative stabilities recorded for the most compact form of

the protein track precisely with the relative binding strengths of the manosyl sugars used in our

experiments. While more data will be required to validate this result, it also illustrates the

potential utility of CIU based stability measurements for protein-ligand complexes.[79]
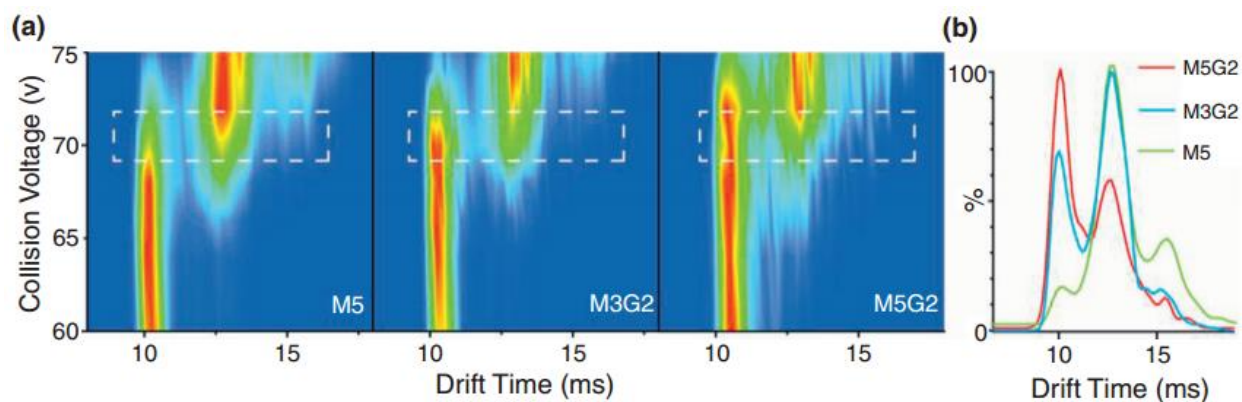
Figure 1-8. IM-MS stability measurements differentiate ligand binding in a tetrameric protein complex. **(a)** CIU fingerprints of the Concanavalin A (Con A) tetramer, bound in a 1:4 protein tetramer:ligand stoichiometry for three different manosyl sugars: 3α,6α-mannopentaose (M5, MW = 828.74), 3α,6α-mannotriose-di-(N-acetyl-d-glucosamine) (M3G2, MW = 910.84) and 3α,6α-mannopentaose-di-(N-acetyl-d-glucosamine) (M5G2, MW = 1235.10). Regions of the CIU fingerprints marked with a dashed box shown for each complex are selected for detailed analysis over a range of voltages. **(b)** Drift time data for the selected regions from (a) show differential stabilities for the ligand complexes that relate primarily to the strength of the protein–ligand complex isolated in the gas phase.

1.7 Summary

As we move forward into the future of drug discovery and development, it is clear that many

protein targets, and their associated disease mechanisms, will challenge existing paradigms.[185,

186] Many of the protein targets discussed above exist in multiple conformational states, linked

by rapid dynamic motions, present as polydisperse multi-protein ensembles, and lack any enzymatic active site for small molecule attachment. All of these factors conspire to create an environment where the rapid assessment of binding strength alone will likely be insufficient information to identify potential lead compounds. Instead, next-generation high-throughput screening technologies will need to detect small molecules that elicit an efficacious conformational change, shift in stability/flexibility, or oligomeric state alteration within a target protein system in a manner that may not correlate with overall binding strength.

Overall, the IM-MS protein–ligand screening methods discussed here are primarily limited by: the detection limits associated with ESI-MS and the software solutions currently available to rapidly extract scored responses and computational models of protein structure from IM-MS data. Significant development challenges also surround the throughput of IM-MS screening technology, which is currently limited to hundreds of samples-per-day, primarily through deficiencies in rapid sample introduction methods and post-analysis software tools. Despite these challenges, the advantages of IM-MS based screens, which are capable of detecting minor conformational changes in protein targets within mixtures at relatively low concentrations without the need for chemical labeling, are enabling a growing number of studies involving protein–ligand complexes of pharmaceutical interest. This trend is likely to continue in the future, and lead to both the support and the acceleration of continued efforts in the pharmaceutical sciences. In this thesis, IM-MS and CIU methods are developed in an effort to develop robust protein-ligand screening protocols focusing on the structural binding consequences, including binding strength and associated conformational changes.

In chapter 2, the binding and CIU profile of the Concanavalin A – manosyl sugar system, using a range of sugar ligands with known K_D values, binding modes and structures, were systematically investigated, in an effort to evaluate the impact of ligand binding strength and size on the CIU process. (**Published. Niu, S. and B.T. Ruotolo,** *Collisional unfolding of multiprotein complexes reveals cooperative stabilization upon ligand binding.***Protein Science, 2015.**)

In chapter 3, molecular modeling approaches were performed in conjunction with experimental measurements, to investigate the fundamental mechanism of collision induced unfolding of the multi-protein complexes.

In chapter 4, the interactions between Human histone deacetylase 8 and Poly-r(C)-binding protein 1 were systematically probed by IM-MS, and for the first time, a working protocol to measure the binding affinity between hetero-protein-protein complexes by MS was developed.

In chapter 5, the integral membrane protein Translocator Protein (TSPO) and its therapeutic ligand binding behavior was investigated by IM-MS, a native endogenous ligand was observed within the A139T mutant, and its identity was systematically screened via MS- and CIU-based approaches.

Part of the content in this introduction was published as a review on *Current opinion in chemical biology***, 2013. 17(5): p. 809-817.**

1.8 References

1. Geoghegan, K.F. and M.A. Kelly, *Biochemical applications of mass spectrometry in pharmaceutical drug discovery.* Mass spectrometry reviews, 2005. **24**(3): p. 347-366.
2. Rifai, N., M.A. Gillette, and S.A. Carr, *Protein biomarker discovery and validation: the long and uncertain path to clinical utility.* Nature biotechnology, 2006. **24**(8): p. 971-983.
3. Meyer, B. and T. Peters, *NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors.* Angewandte Chemie-International Edition, 2003. **42**(8): p. 864-890.
4. Niu, S., J.N. Rabuck, and B.T. Ruotolo, *Ion mobility-mass spectrometry of intact protein–ligand complexes for pharmaceutical drug discovery and development.* Current opinion in chemical biology, 2013. **17**(5): p. 809-817.
5. Terrett, N.K., et al., *COMBINATORIAL SYNTHESIS - THE DESIGN OF COMPOUND LIBRARIES AND THEIR APPLICATION TO DRUG DISCOVERY.* Tetrahedron, 1995. **51**(30): p. 8135-8173.
6. Silverman, L., R. Campbell, and J.R. Broach, *New assay technologies for high-throughput screening.* Current opinion in chemical biology, 1998. **2**(3): p. 397-403.
7. Hajduk, P.J. and J. Greer, *A decade of fragment-based drug design: strategic advances and lessons learned.* Nature reviews Drug discovery, 2007. **6**(3): p. 211-219.
8. Koehn, F.E. and G.T. Carter, *The evolving role of natural products in drug discovery.* Nature reviews Drug discovery, 2005. **4**(3): p. 206-220.
9. Solomon, B., et al., *Applications of positron emission tomography in the development of molecular targeted cancer therapeutics.* BioDrugs, 2003. **17**(5): p. 339-354.
10. Hann, M.M. and T.I. Oprea, *Pursuing the leadlikeness concept in pharmaceutical research.* Current opinion in chemical biology, 2004. **8**(3): p. 255-263.
11. Kitchen, D.B., et al., *Docking and scoring in virtual screening for drug discovery: methods and applications.* Nat Rev Drug Discov, 2004. **3**(11): p. 935-949.
12. Weston, A.D. and L. Hood, *Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine.* Journal of proteome research, 2004. **3**(2): p. 179-196.
13. Adrián, F.J., et al., *Allosteric inhibitors of Bcr-abl–dependent cell proliferation.* Nature chemical biology, 2006. **2**(2): p. 95-102.
14. Annis, D.A., et al., *A general technique to rank protein-ligand binding affinities and determine allosteric versus direct binding site competition in compound mixtures.* Journal of the American Chemical Society, 2004. **126**(47): p. 15495-15503.
15. Fenn, J.B., *Electrospray wings for molecular elephants (Nobel lecture).* Angewandte Chemie International Edition, 2003. **42**(33): p. 3871-3894.
16. Scholten, A., T.T. Aye, and A.J. Heck, *A multi‐angular mass spectrometric view at cyclic nucleotide dependent protein kinases: In vivo characterization and structure/function relationships.* Mass spectrometry reviews, 2008. **27**(4): p. 331-353.
17. Congreve, M., C.W. Murray, and T.L. Blundell, *Keynote review: Structural biology and drug discovery.* Drug discovery today, 2005. **10**(13): p. 895-907.
18. Wang, R., et al., *The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures.* Journal of medicinal chemistry, 2004. **47**(12): p. 2977-2980.

19.   Benesch, J.L., et al., *Protein complexes in the gas phase: technology for structural genomics and proteomics.* Chemical reviews, 2007. **107**(8): p. 3544-3567.

20.   Ganem, B., Y.T. Li, and J.D. Henion, *Detection of noncovalent receptor-ligand complexes by mass spectrometry.* Journal of the American Chemical Society, 1991. **113**(16): p. 6294-6296.

21.   Katta, V. and B.T. Chait, *Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry.* Journal of the American Chemical Society, 1991. **113**(22): p. 8534-8535.

22.   Afonso, C., Y. Hathout, and C. Fenselau, *Qualitative characterization of biomolecular zinc complexes by collisionally induced dissociation.* Journal of mass spectrometry, 2002. **37**(7): p. 755-759.

23.   Alverdi, V., et al., *cGMP-binding prepares PKG for substrate binding by disclosing the C-terminal domain.* Journal of molecular biology, 2008. **375**(5): p. 1380-1393.

24.   Annis, D.A., et al., *Affinity selection-mass spectrometry screening techniques for small molecule drug discovery.* Current opinion in chemical biology, 2007. **11**(5): p. 518-526.

25.   Hofstadler, S.A. and K.A. Sannes-Lowery, *Applications of ESI-MS in drug discovery: interrogation of noncovalent complexes.* Nature Reviews Drug Discovery, 2006. **5**(7): p. 585-595.

26.   Wang, W., E.N. Kitova, and J.S. Klassen, *Influence of solution and gas phase processes on protein-carbohydrate binding affinities determined by nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry.* Analytical chemistry, 2003. **75**(19): p. 4945-4955.

27.   El-Hawiet, A., et al., *Quantifying labile protein–ligand interactions using electrospray ionization mass spectrometry.* Journal of the American Society for Mass Spectrometry, 2010. **21**(11): p. 1893-1899.

28.   DeArmond, P.D., et al., *Thermodynamic analysis of protein–ligand interactions in complex biological mixtures using a shotgun proteomics approach.* Journal of proteome research, 2011. **10**(11): p. 4948-4958.

29.   Liu, F. and M.B. Goshe, *Combinatorial electrostatic collision-induced dissociative chemical cross-linking reagents for probing protein surface topology.* Analytical chemistry, 2010. **82**(14): p. 6215-6223.

30.   Leite, J.F. and M. Cascio, *Probing the topology of the glycine receptor by chemical modification coupled to mass spectrometry.* Biochemistry, 2002. **41**(19): p. 6140-6148.

31.   Sinz, A., *Chemical cross‐linking and mass spectrometry to map three‐dimensional protein structures and protein–protein interactions.* Mass spectrometry reviews, 2006. **25**(4): p. 663-682.

32.   Green, M.K. and C.B. Lebrilla, *Ion-molecule reactions as probes of gas-phase structures of peptides and proteins.* Mass spectrometry reviews, 1997. **16**(2): p. 53-71.

33.   Zhu, M.M., R. Chitta, and M.L. Gross, *PLIMSTEX: a novel mass spectrometric method for the quantification of protein–ligand interactions in solution.* International Journal of Mass Spectrometry, 2005. **240**(3): p. 213-220.

34.   Zhu, M.M., D.L. Rempel, and M.L. Gross, *Modeling data from titration, amide H/D exchange, and mass spectrometry to obtain protein-ligand binding constants.* Journal of the American Society for Mass Spectrometry, 2004. **15**(3): p. 388-397.

35.   Kanu, A.B., et al., *Ion mobility–mass spectrometry.* Journal of Mass Spectrometry, 2008. **43**(1): p. 1-22.

36.   McLean, J.A., et al., *Ion mobility–mass spectrometry: a new paradigm for proteomics.* International Journal of Mass Spectrometry, 2005. **240**(3): p. 301-315.

37.   Valentine, S.J., et al., *Developing liquid chromatography ion mobility mass spectometry techniques.* Expert review of proteomics, 2005. **2**(4): p. 553-565.

38.   Jackson, S.N., et al., *MALDI‑ion mobility‑TOFMS imaging of lipids in rat brain tissue.* Journal of Mass Spectrometry, 2007. **42**(8): p. 1093-1098.

39.   McLean, J.A., W.B. Ridenour, and R.M. Caprioli, *Profiling and imaging of tissues by imaging ion mobility‑mass spectrometry.* Journal of mass spectrometry, 2007. **42**(8): p. 1099-1105.

40.   Heck, A.J., *Native mass spectrometry: a bridge between interactomics and structural biology.* Nature methods, 2008. **5**(11): p. 927-933.

41.   Pacholarz, K.J., et al., *Mass spectrometry based tools to investigate protein–ligand interactions for drug discovery.* Chemical Society Reviews, 2012. **41**(11): p. 4335-4355.

42.   Mason, E.A. and E.W. McDaniel, *Transport properties of ions in gases.* NASA STI/Recon Technical Report A, 1988. **89**: p. 15174.

43.   Ruotolo, B.T., et al., *Evidence for macromolecular protein rings in the absence of bulk water.* Science, 2005. **310**(5754): p. 1658-1661.

44.   Warnke, S., G. von Helden, and K. Pagel, *Protein Structure in the Gas Phase: The Influence of Side-Chain Microsolvation.* Journal of the American Chemical Society, 2013. **135**(4): p. 1177-1180.

45.   Duijn, E.v., et al., *Chaperonin complexes monitored by ion mobility mass spectrometry.* Journal of the American Chemical Society, 2009. **131**(4): p. 1452-1459.

46.   Giles, K., J.P. Williams, and I. Campuzano, *Enhancements in travelling wave ion mobility resolution.* Rapid Communications in Mass Spectrometry, 2011. **25**(11): p. 1559-1566.

47.   Pringle, S.D., et al., *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): p. 1-12.

48.   Takats, Z., et al., *Electrosonic spray ionization. A gentle technique for generating folded proteins and protein complexes in the gas phase and for studying ion-molecule reactions at atmospheric pressure.* Analytical Chemistry, 2004. **76**(14): p. 4050-4058.

49.   Iribarne, J. and B. Thomson, *On the evaporation of small ions from charged droplets.* The Journal of Chemical Physics, 1976. **64**(6): p. 2287-2294.

50.   Kebarle, P. and U.H. Verkerk, *Electrospray: from ions in solution to ions in the gas phase, what we know now.* Mass Spectrometry Reviews, 2009. **28**(6): p. 898-917.

51.   Dole, M., et al., *Molecular beams of macroions.* The Journal of Chemical Physics, 1968. **49**(5): p. 2240-2249.

52.   Hogan Jr, C.J., et al., *Combined charged residue-field emission model of macromolecular electrospray ionization.* Analytical chemistry, 2008. **81**(1): p. 369-377.

53.   Konermann, L., et al., *Unraveling the mechanism of electrospray ionization.* Analytical chemistry, 2012. **85**(1): p. 2-9.

54.   Chung, J.K. and S. Consta, *Release mechanisms of poly (ethylene glycol) macroions from aqueous charged nanodroplets.* The Journal of Physical Chemistry B, 2012. **116**(19): p. 5777-5785.

55. Yue, X., S. Vahidi, and L. Konermann, *Insights into the mechanism of protein electrospray ionization from salt adduction measurements.* Journal of the American Society for Mass Spectrometry, 2014. **25**(8): p. 1322-1331.

56. Page, J.S., et al., *Ionization and transmission efficiency in an electrospray ionization–mass spectrometry interface.* Journal of the American Society for Mass Spectrometry, 2007. **18**(9): p. 1582-1590.

57. Wilm, M.S. and M. Mann, *Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last?* International Journal of Mass Spectrometry and Ion Processes, 1994. **136**(2): p. 167-180.

58. Bahr, U., et al., *High-sensitivity analysis of neutral underivatized oligosaccharides by nanoelectrospray mass spectrometry.* Analytical chemistry, 1997. **69**(22): p. 4530-4535.

59. Karas, M., U. Bahr, and T. Dülcks, *Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine.* Fresenius' journal of analytical chemistry, 2000. **366**(6-7): p. 669-676.

60. Heck, A.J. and R.H. van den Heuvel, *Investigation of intact protein complexes by mass spectrometry.* Mass spectrometry reviews, 2004. **23**(5): p. 368-389.

61. Ruotolo, B.T. and C.V. Robinson, *Aspects of native proteins are retained in vacuum.* Current opinion in chemical biology, 2006. **10**(5): p. 402-408.

62. Baca, M. and S.B. Kent, *Direct observation of a ternary complex between the dimeric enzyme HIV-1 protease and a substrate-based inhibitor.* Journal of the American Chemical Society, 1992. **114**(10): p. 3992-3993.

63. Huang, E.C., et al., *Application of electrospray mass spectrometry in probing protein-protein and protein-ligand noncovalent interactions.* Journal of the American Society for Mass Spectrometry, 1993. **4**(8): p. 624-630.

64. Light-Wahl, K., B. Winger, and R.D. Smith, *Observation of the multimeric forms of concanavalin A by electrospray ionization mass spectrometry.* Journal of the American Chemical Society, 1993. **115**(13): p. 5869-5870.

65. Chowdhury, S.K., V. Katta, and B.T. Chait, *Probing conformational changes in proteins by mass spectrometry.* Journal of the American Chemical Society, 1990. **112**(24): p. 9012-9013.

66. Robinson, C.V., et al., *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): p. 8646-8653.

67. Snijder, J., et al., *Studying 18 MDa virus assemblies with native mass spectrometry.* Angewandte Chemie International Edition, 2013. **52**(14): p. 4020-4023.

68. Rostom, A.A., et al., *Detection and selective dissociation of intact ribosomes in a mass spectrometer.* Proceedings of the National Academy of Sciences, 2000. **97**(10): p. 5185-5190.

69. Tito, M.A., et al., *Electrospray time-of-flight mass spectrometry of the intact MS2 virus capsid.* Journal of the American Chemical Society, 2000. **122**(14): p. 3550-3551.

70. Sobott, F., et al., *Subunit Exchange of Multimeric Protein Complexes Real-Time Monitoring of Subunit Exchange Between Small Heat Shock Proteins by Using Electrospray Mass Spectrometry.* Journal of Biological Chemistry, 2002. **277**(41): p. 38921-38929.

71. Hernández, H., et al., *Subunit architecture of multimeric complexes isolated directly from cells.* EMBO reports, 2006. **7**(6): p. 605-610.

72. Han, L., et al., *Protein–Glycolipid Interactions Studied in Vitro Using ESI-MS and Nanodiscs: Insights into the Mechanisms and Energetics of Binding.* Analytical chemistry, 2015. **87**(9): p. 4888-4896.

73. Sobott, F. and C.V. Robinson, *Characterising electrosprayed biomolecules using tandem-MS—the noncovalent GroEL chaperonin assembly.* International Journal of Mass Spectrometry, 2004. **236**(1): p. 25-32.

74. Mamyrin, B., et al., *The massreflect ron, a new non-magnetic time-of-flight mass spectrometer with high resolution.* Zh. Eksp. Teor. Fiz, 1973. **64**: p. 82-89.

75. Jenner, M., et al., *Detection of a Protein Conformational Equilibrium by Electrospray Ionisation‐Ion Mobility‐Mass Spectrometry.* Angewandte Chemie International Edition, 2011. **50**(36): p. 8291-8294.

76. Rabuck, J.N., et al., *Activation state-selective kinase inhibitor assay based on ion mobility-mass spectrometry.* Analytical chemistry, 2013. **85**(15): p. 6995-7002.

77. Shi, H., et al., *Conformation types of ubiquitin [M+ 8H] 8+ ions from water: methanol solutions: evidence for the N and A states in aqueous solution.* The Journal of Physical Chemistry B, 2012. **116**(10): p. 3344-3352.

78. Hyung, S.-J., C.V. Robinson, and B.T. Ruotolo, *Gas-phase unfolding and disassembly reveals stability differences in ligand-bound multiprotein complexes.* Chemistry & biology, 2009. **16**(4): p. 382-390.

79. Niu, S. and B.T. Ruotolo, *Collisional unfolding of multiprotein complexes reveals cooperative stabilization upon ligand binding.* Protein Science, 2015.

80. Zhong, Y., S.-J. Hyung, and B.T. Ruotolo, *Ion mobility–mass spectrometry for structural proteomics.* Expert review of proteomics, 2012. **9**(1): p. 47-58.

81. Harvey, S.R., C.E. MacPhee, and P.E. Barran, *Ion mobility mass spectrometry for peptide analysis.* Methods, 2011. **54**(4): p. 454-461.

82. Clemmer, D.E. and M.F. Jarrold, *Ion mobility measurements and their applications to clusters and biomolecules.* Journal of Mass Spectrometry, 1997. **32**(6): p. 577-592.

83. Gillig, K.J., et al., *Coupling high-pressure MALDI with ion mobility/orthogonal time-of-flight mass spectrometry.* Analytical chemistry, 2000. **72**(17): p. 3965-3971.

84. Gillig, K.J., et al., *An electrostatic focusing ion guide for ion mobility-mass spectrometry.* International Journal of Mass Spectrometry, 2004. **239**(1): p. 43-49.

85. Valentine, S.J., S.L. Koeniger, and D.E. Clemmer, *A split-field drift tube for separation and efficient fragmentation of biomolecular ions.* Analytical chemistry, 2003. **75**(22): p. 6202-6208.

86. Tang, K., et al., *High-sensitivity ion mobility spectrometry/mass spectrometry using electrodynamic ion funnel interfaces.* Analytical chemistry, 2005. **77**(10): p. 3330-3339.

87. Koeniger, S.L., et al., *An Ims-Ims Analogue of Ms-Ms.* Analytical chemistry, 2006. **78**(12): p. 4161-4174.

88. Koeniger, S.L., S.I. Merenbloom, and D.E. Clemmer, *Evidence for many resolvable structures within conformation types of electrosprayed ubiquitin ions.* The Journal of Physical Chemistry B, 2006. **110**(13): p. 7017-7021.

89. Giles, K., et al., *Applications of a travelling wave‐based radio‐frequency‐only stacked ring ion guide.* Rapid Communications in Mass Spectrometry, 2004. **18**(20): p. 2401-2414.

90. Shvartsburg, A.A. and R.D. Smith, *Fundamentals of traveling wave ion mobility spectrometry.* Analytical chemistry, 2008. **80**(24): p. 9689-9699.

91. Ruotolo, B.T., et al., *Ion mobility–mass spectrometry analysis of large protein complexes.* Nature Protocols, 2008. **3**(7): p. 1139-1152.

92. Giles, K., et al., *A method for direct measurement of ion mobilities using a travelling wave ion guide.* International Journal of Mass Spectrometry, 2010. **298**(1): p. 10-16.

93. Zhong, Y., S.-J. Hyung, and B.T. Ruotolo, *Characterizing the resolution and accuracy of a second-generation traveling-wave ion mobility separator for biomolecular ions.* Analyst, 2011. **136**(17): p. 3534-3541.

94. Merenbloom, S.I., et al., *High-resolution ion cyclotron mobility spectrometry.* Analytical chemistry, 2009. **81**(4): p. 1482-1487.

95. Dugourd, P., et al., *High-resolution ion mobility measurements.* Review of Scientific Instruments, 1997. **68**(2): p. 1122-1129.

96. Kemper, P.R., N.F. Dupuis, and M.T. Bowers, *A new, higher resolution, ion mobility mass spectrometer.* International Journal of Mass Spectrometry, 2009. **287**(1): p. 46-57.

97. Roscioli, K.M., et al., *Modular ion mobility spectrometer for explosives detection using corona ionization.* Analytical chemistry, 2011. **83**(15): p. 5965-5971.

98. Shelimov, K.B. and M.F. Jarrold, *Conformations, unfolding, and refolding of apomyoglobin in vacuum: An activation barrier for gas-phase protein folding.* Journal of the American Chemical Society, 1997. **119**(13): p. 2987-2994.

99. Laganowsky, A., et al., *Membrane proteins bind lipids selectively to modulate their structure and function.* Nature, 2014. **510**(7503): p. 172-175.

100. Felitsyn, N., E.N. Kitova, and J.S. Klassen, *Thermal decomposition of a gaseous multiprotein complex studied by blackbody infrared radiative dissociation. Investigating the origin of the asymmetric dissociation behavior.* Analytical chemistry, 2001. **73**(19): p. 4647-4661.

101. Jurchen, J.C. and E.R. Williams, *Origin of asymmetric charge partitioning in the dissociation of gas-phase protein homodimers.* Journal of the American Chemical Society, 2003. **125**(9): p. 2817-2826.

102. Jurchen, J.C., D.E. Garcia, and E.R. Williams, *Further studies on the origins of asymmetric charge partitioning in protein homodimers.* Journal of the American Society for Mass Spectrometry, 2004. **15**(10): p. 1408-1415.

103. Zhong, Y., L. Han, and B.T. Ruotolo, *Collisional and Coulombic Unfolding of Gas‐Phase Proteins: High Correlation to Their Domain Structures in Solution.* Angewandte Chemie, 2014. **126**(35): p. 9363-9366.

104. Cooper, M.A., *Label-free screening of bio-molecular interactions.* Analytical and bioanalytical chemistry, 2003. **377**(5): p. 834-842.

105. Rich, R.L. and D.G. Myszka, *Advances in surface plasmon resonance biosensor analysis.* Current Opinion in Biotechnology, 2000. **11**(1): p. 54-61.

106. Kerns, E.H. and L. Di, *Pharmaceutical profiling in drug discovery.* Drug Discovery Today, 2003. **8**(7): p. 316-323.

107. Englebienne, P., A. Van Hoonacker, and M. Verhas, *Surface plasmon resonance: principles, methods and applications in biomedical sciences.* Spectroscopy-an International Journal, 2003. **17**(2-3): p. 255-273.

108. Ladbury, J.E. and B.Z. Chowdhry, *Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions.* Chemistry & biology, 1996. **3**(10): p. 791-801.

109. Pierce, M.M., C. Raman, and B.T. Nall, *Isothermal titration calorimetry of protein–protein interactions.* Methods, 1999. **19**(2): p. 213-221.

110. Houtman, J.C., et al., *Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: application to adaptor protein complexes in cell signaling.* Protein Science, 2007. **16**(1): p. 30-42.

111. Benesch, J.L. and B.T. Ruotolo, *Mass spectrometry: come of age for structural and dynamical biology.* Current opinion in structural biology, 2011. **21**(5): p. 641-649.

112. Kiselar, J.G. and M.R. Chance, *Future directions of structural mass spectrometry using hydroxyl radical footprinting.* Journal of Mass Spectrometry, 2010. **45**(12): p. 1373-1382.

113. Konermann, L., et al., *Mass spectrometry combined with oxidative labeling for exploring protein structure and folding.* Mass spectrometry reviews, 2010. **29**(4): p. 651-667.

114. Konermann, L., J. Pan, and Y.-H. Liu, *Hydrogen exchange mass spectrometry for studying protein structure and dynamics.* Chemical Society Reviews, 2011. **40**(3): p. 1224-1234.

115. Engen, J.R., *Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS.* Analytical chemistry, 2009. **81**(19): p. 7870-7875.

116. Petrotchenko, E.V. and C.H. Borchers, *Crosslinking combined with mass spectrometry for structural proteomics.* Mass spectrometry reviews, 2010. **29**(6): p. 862-876.

117. Lee, Y.J., *Mass spectrometric analysis of cross-linking sites for the structure of proteins and protein complexes.* Molecular BioSystems, 2008. **4**(8): p. 816-823.

118. Daniel, J.M., et al., *Quantitative determination of noncovalent binding interactions using soft ionization mass spectrometry.* International Journal of Mass Spectrometry, 2002. **216**(1): p. 1-27.

119. Han, L., et al., *Identifying carbohydrate ligands of a norovirus P particle using a catch and release electrospray ionization mass spectrometry assay.* Journal of The American Society for Mass Spectrometry, 2014. **25**(1): p. 111-119.

120. Liu, J. and L. Konermann, *Protein–protein binding affinities in solution determined by electrospray mass spectrometry.* Journal of the American Society for Mass Spectrometry, 2011. **22**(3): p. 408-417.

121. Ebong, I.-o., et al., *Heterogeneity and dynamics in the assembly of the heat shock protein 90 chaperone complexes.* Proceedings of the National Academy of Sciences, 2011. **108**(44): p. 17939-17944.

122. Perutz, M.F., *Mechanisms of cooperativity and allosteric regulation in proteins.* Quarterly reviews of biophysics, 1989. **22**(02): p. 139-237.

123. Brown, A., *Analysis of cooperativity by isothermal titration calorimetry.* International journal of molecular sciences, 2009. **10**(8): p. 3457-3477.

124. Doyle, M.L., S.J. Gill, and M.A. Cusanovich, *Ligand-controlled dissociation of Chromatium vinosum cytochrome c'.* Biochemistry, 1986. **25**(9): p. 2509-2516.

125. Wyman, J., *Linked functions and reciprocal effects in hemoglobin: a second look.* Adv. Protein Chem, 1964. **19**(223): p. 91.

126. Dyachenko, A., et al., *Allosteric mechanisms can be distinguished using structural mass spectrometry.* Proceedings of the National Academy of Sciences, 2013. **110**(18): p. 7235-7239.

127. Lappano, R. and M. Maggiolini, *G protein-coupled receptors: novel targets for drug discovery in cancer.* Nature reviews Drug discovery, 2011. **10**(1): p. 47-60.

128. Almén, M.S., et al., *Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin.* BMC biology, 2009. **7**(1): p. 50.

129. Dorsam, R.T. and J.S. Gutkind, *G-protein-coupled receptors and cancer.* Nature reviews cancer, 2007. **7**(2): p. 79-94.

130. Overington, J.P., B. Al-Lazikani, and A.L. Hopkins, *How many drug targets are there?* Nature reviews Drug discovery, 2006. **5**(12): p. 993-996.

131. Calabrese, A.N., et al., *Amphipols outperform dodecylmaltoside micelles in stabilizing membrane protein structure in the gas phase.* Analytical chemistry, 2014. **87**(2): p. 1118-1126.

132. White, S.H., *Biophysical dissection of membrane proteins.* Nature, 2009. **459**(7245): p. 344-346.

133. Nygaard, R., et al., *The dynamic process of β 2-adrenergic receptor activation.* Cell, 2013. **152**(3): p. 532-542.

134. Kim, J., et al., *Subnanometre-resolution electron cryomicroscopy structure of a heterodimeric ABC exporter.* Nature, 2015. **517**(7534): p. 396-400.

135. Ong, S.-E. and M. Mann, *Mass spectrometry–based proteomics turns quantitative.* Nature chemical biology, 2005. **1**(5): p. 252-262.

136. Barrera, N.P., et al., *Micelles protect membrane complexes from solution to vacuum.* Science, 2008. **321**(5886): p. 243-246.

137. Barrera, N.P., et al., *Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions.* Nature Methods, 2009. **6**(8): p. 585-U49.

138. Laganowsky, A., et al., *Mass spectrometry of intact membrane protein complexes.* Nature protocols, 2013. **8**(4): p. 639-651.

139. Whitelegge, J.P., *Integral membrane proteins and bilayer proteomics.* Analytical chemistry, 2013. **85**(5): p. 2558-2568.

140. Chung, K.Y., et al., *Conformational changes in the G protein Gs induced by the [bgr] 2 adrenergic receptor.* Nature, 2011. **477**(7366): p. 611-615.

141. Padayatti, P.S., et al., *A hybrid structural approach to analyze ligand binding by the serotonin type 4 receptor (5-HT4).* Molecular & Cellular Proteomics, 2013. **12**(5): p. 1259-1271.

142. Efremov, R.G., et al., *Architecture and conformational switch mechanism of the ryanodine receptor.* Nature, 2015. **517**(7532): p. 39-43.

143. Barrera, N.P. and C.V. Robinson, *Advances in the Mass Spectrometry of Membrane Proteins: From Individual Proteins to Intact Complexes*, in *Annual Review of Biochemistry, Vol 80*, R.D. Kornberg, et al., Editors. 2011, Annual Reviews: Palo Alto. p. 247-271.

144. Wang, S.C., et al., *Ion Mobility Mass Spectrometry of Two Tetrameric Membrane Protein Complexes Reveals Compact Structures and Differences in Stability and Packing.* Journal of the American Chemical Society, 2010. **132**(44): p. 15468-15470.

145. Konijnenberg, A., et al., *Global structural changes of an ion channel during its gating are followed by ion mobility mass spectrometry.* Proceedings of the National Academy of Sciences, 2014. **111**(48): p. 17170-17175.

146. Bechara, C. and C.V. Robinson, *Different modes of lipid binding to membrane proteins probed by mass spectrometry.* Journal of the American Chemical Society, 2015. **137**(16): p. 5240-5247.

147.	Mesleh, M., et al., *Structural information from ion mobility measurements: effects of the long-range potential.* The Journal of Physical Chemistry, 1996. **100**(40): p. 16082-16086.

148.	Shvartsburg, A.A. and M.F. Jarrold, *An exact hard-spheres scattering model for the mobilities of polyatomic ions.* Chemical Physics Letters, 1996. **261**(1): p. 86-91.

149.	Mack Jr, E., *Average cross-sectional areas of molecules by gaseous diffusion methods.* Journal of the American Chemical Society, 1925. **47**(10): p. 2468-2482.

150.	von Helden, G., et al., *Carbon cluster cations with up to 84 atoms: structures, formation mechanism, and reactivity.* The Journal of Physical Chemistry, 1993. **97**(31): p. 8182-8192.

151.	Larriba, C. and C.J. Hogan Jr, *Ion mobilities in diatomic gases: measurement versus prediction with non-specular scattering models.* The Journal of Physical Chemistry A, 2013. **117**(19): p. 3887-3901.

152.	Bleiholder, C., T. Wyttenbach, and M.T. Bowers, *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): p. 1-10.

153.	Marklund, E.G., et al., *Collision Cross Sections for Structural Proteomics.* Structure, 2015. **23**(4): p. 791-799.

154.	Sakae, Y., et al., *Protein structure predictions by parallel simulated annealing molecular dynamics using genetic crossover.* Journal of computational chemistry, 2011. **32**(7): p. 1353-1360.

155.	Wyttenbach, T., G. von Helden, and M.T. Bowers, *Gas-phase conformation of biological molecules: Bradykinin.* Journal of the American Chemical Society, 1996. **118**(35): p. 8355-8364.

156.	Kirkpatrick, S., C.D. Gelatt, and M.P. Vecchi, *Optimization by simulated annealing.* science, 1983. **220**(4598): p. 671-680.

157.	Soper, M.T., et al., *Amyloid-β–neuropeptide interactions assessed by ion mobility-mass spectrometry.* Physical Chemistry Chemical Physics, 2013. **15**(23): p. 8952-8961.

158.	Bornschein, R.E., . *Ion Mobility-Mass Spectrometry Reveals Highly-Compact Intermediates in the Collision Induced Dissociation of Charge-reduced Protein Complexes.* Development of Charge Manipulation Nanoelectrospray Ion Mobility-Mass Spectrometry Techniques for Multiprotein Complex Analysis, 2015: p. 58.

159.	Hall, Z., A. Politis, and C.V. Robinson, *Structural modeling of heteromeric protein complexes from disassembly pathways and ion mobility-mass spectrometry.* Structure, 2012. **20**(9): p. 1596-1609.

160.	Benesch, J.L., *Collisional activation of protein complexes: picking up the pieces.* Journal of the American Society for Mass Spectrometry, 2009. **20**(3): p. 341-348.

161.	Mao, Y., et al., *Thermal unfolding of unsolvated cytochrome c: experiment and molecular dynamics simulations.* Journal of the American Chemical Society, 1999. **121**(12): p. 2712-2721.

162.	Alber, F., et al., *Determining the architectures of macromolecular assemblies.* Nature, 2007. **450**(7170): p. 683-694.

163.	Teplow, D.B., et al., *Elucidating amyloid β-protein folding and assembly: a multidisciplinary approach.* Accounts of chemical research, 2006. **39**(9): p. 635-645.

164.	Woods, L., S. Radford, and A. Ashcroft, *Advances in ion mobility spectrometry-mass spectrometry reveal key insights into amyloid assembly.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2012.

165. Woods, L.A., et al., *Ligand binding to distinct states diverts aggregation of an amyloid-forming protein.* Nature chemical biology, 2011.

166. Taverner, T., et al., *Subunit Architecture of Intact Protein Complexes from Mass Spectrometry and Homology Modeling†.* Accounts of chemical research, 2008. **41**(5): p. 617-627.

167. Politis, A., et al., *Integrating ion mobility mass spectrometry with molecular modelling to determine the architecture of multiprotein complexes.* 2010.

168. Suckau, D., et al., *Coexisting stable conformations of gaseous protein ions.* Proceedings of the National Academy of Sciences, 1993. **90**(3): p. 790-793.

169. Ruotolo, B.T., et al., *Ion Mobility–Mass Spectrometry Reveals Long‐Lived, Unfolded Intermediates in the Dissociation of Protein Complexes.* Angewandte Chemie International Edition, 2007. **46**(42): p. 8001-8004.

170. Hopper, J.T. and N.J. Oldham, *Collision induced unfolding of protein ions in the gas phase studied by ion mobility-mass spectrometry: the effect of ligand binding on conformational stability.* Journal of the American Society for Mass Spectrometry, 2009. **20**(10): p. 1851-1858.

171. Han, L., S.J. Hyung, and B.T. Ruotolo, *Bound cations significantly stabilize the structure of multiprotein complexes in the gas phase.* Angewandte Chemie, 2012. **124**(23): p. 5790-5793.

172. Han, L., et al., *Bound anions differentially stabilize multiprotein complexes in the absence of bulk solvent.* Journal of the American Chemical Society, 2011. **133**(29): p. 11358-11367.

173. Klassen, J.S. and P. Kebarle, *Collision-induced dissociation threshold energies of protonated glycine, glycinamide, and some related small peptides and peptide amino amides.* Journal of the American Chemical Society, 1997. **119**(28): p. 6552-6563.

174. Badman, E.R., S. Myung, and D.E. Clemmer, *Evidence for unfolding and refolding of gas-phase cytochrome c ions in a Paul trap.* Journal of the American Society for Mass Spectrometry, 2005. **16**(9): p. 1493-1497.

175. Zhou, M., S. Dagan, and V.H. Wysocki, *Protein subunits released by surface collisions of noncovalent complexes: nativelike compact structures revealed by ion mobility mass spectrometry.* Angewandte Chemie International Edition, 2012. **51**(18): p. 4336-4339.

176. Zhou, M., S. Dagan, and V.H. Wysocki, *Impact of charge state on gas-phase behaviors of noncovalent protein complexes in collision induced dissociation and surface induced dissociation.* Analyst, 2013. **138**(5): p. 1353-1362.

177. Han, L. and B.T. Ruotolo, *Traveling-wave ion mobility-mass spectrometry reveals additional mechanistic details in the stabilization of protein complex ions through tuned salt additives.* International Journal for Ion Mobility Spectrometry, 2013: p. 1-10.

178. Uetrecht, C., et al., *Interrogating viral capsid assembly with ion mobility–mass spectrometry.* Nature Chemistry, 2010. **3**(2): p. 126-132.

179. Uetrecht, C., et al., *Ion mobility mass spectrometry of proteins and protein assemblies.* Chemical Society Reviews, 2010. **39**(5): p. 1633-1655.

180. Loo, J.A., *Studying noncovalent protein complexes by electrospray ionization mass spectrometry.* Mass Spectrometry Reviews, 1997. **16**(1): p. 1-23.

181. Tabb, D.L., et al., *Influence of basic residue content on fragment ion peak intensities in low-energy collision-induced dissociation spectra of peptides.* Analytical chemistry, 2004. **76**(5): p. 1243-1248.

182. Schenauer, M.R. and J.A. Leary, *An ion mobility–mass spectrometry investigation of monocyte chemoattractant protein-1.* International journal of mass spectrometry, 2009. **287**(1): p. 70-76.

183. Giganti, V.G., et al., *Ion Mobility-Mass Spectrometry Study of Folded Ubiquitin Conformers Induced By Treatment with cis-[Pd (en)(H 2 O) 2] 2+.* Journal of the American Society for Mass Spectrometry, 2011. **22**(2): p. 300-309.

184. Goldstein, I., C. Hollerman, and E. Smith, *Protein-carbohydrate interaction. II. Inhibition studies on the interaction of concanavalin a with polysaccharides*.* Biochemistry, 1965. **4**(5): p. 876-883.

185. Verdine, G.L. and L.D. Walensky, *The challenge of drugging undruggable targets in cancer: lessons learned from targeting BCL-2 family members.* Clinical Cancer Research, 2007. **13**(24): p. 7264-7270.

186. Dandapani, S. and L.A. Marcaurelle, *Grand challenge commentary: Accessing new chemical space for'undruggable'targets.* Nature chemical biology, 2010. **6**(12): p. 861-863.

Chapter 2

Collisional Unfolding of Multiprotein Complexes Reveals Cooperative Stabilization Upon Ligand Binding

2.1 Introduction

Protein biochemistry is replete with binding and interaction mechanisms that rely upon cooperativity, which acts as a form of general control to drive protein-ligand selectivity and function in many higher-order complexes (1-3). Beyond well studied systems, such as the cooperative mechanism surrounding the binding of molecular oxygen and other ligands to hemoglobin (2; 4), many additional proteins and protein complexes have been identified that exhibit cooperative ligand binding mechanisms. For example, many protein-DNA complexes have well known cooperative binding mechanisms that functionally regulate DNA replication (5; 6). In addition, many protein-based motors and pumps rely upon cooperative binding of lipids

and other small molecules to allosterically control protein function (7). While many questions remain surrounding the details of cooperative protein-ligand interactions *in vitro* and *in vivo*(8), a combination of theoretical models of protein-ligand binding cooperativity (9; 10), in combination with detailed measurements of binding thermodynamics (11), have been used to describe the functional consequences of a broad range of protein-ligand complexes (12; 13).

In contrast to our understanding of protein-ligand binding cooperativity, detailed mechanisms that describe cooperative increases in protein stability as a function such ligand binding events remain relatively elusive. For many years, cooperative effects have been invoked to describe enhancements to protein stability upon folding (11). Computational chemistry approaches, for example, have been used to analyze the detailed cascade of non-covalent interactions, hydrogen bonds, and salt-bridges that give rise to folded structures, and have identified cooperative elements in many cases (14-18) . Similar examples centering on the protein stability acquired upon ligand binding are rare, but several have been reported (15; 19-22) . For example, density functional and *ab initio* methods in combination with molecular modeling have been used to quantify the hydrogen-bonding cooperativity in the context of biotin-avidin binding to be on the order of 4 kcal/mol(23) . Computational efforts dominate this area of research, as measurements of cooperative protein-ligand stabilization energies are tremendously challenging, beginning with the difficulties associated with recording evidence of cooperative binding patterns (10; 18; 24). Calorimetry data can, in principle, be analyzed to determine Hill coefficients which quantify the relative cooperativity of binding observed in experimental data, but such analyses are often difficult to execute, especially for large multiprotein systems, and dependent upon overall ligand concentration (4; 25; 26) . In order to assess stability shifts for such systems, the separation of

50

individual bound states of the biomolecules is required, which is often not possible using conventional spectroscopic or chromatographic techniques. These difficulties have resulted in a general dearth of experimental evidence for cooperative stabilization effects in proteins upon ligand binding.

Gas-phase structural biology methods, primarily based on nano-electrospray ionization (nESI) mass spectrometry (MS), possess the separation resolution and information content sufficient to address many of the challenges associated with the assessment of protein-ligand cooperativity and stability described above. MS methods can detect protein-ligand complexes (27-29), either intact or indirectly through mass shifts associated with chemical labeling (29-31), and have been used broadly to assess protein-ligand dissociation constants (K_D) and stability shifts in protein-ligand complexes (32-36). Recently, global methods, based on radical labeling and hydrogen deuterium exchange (HDX) have been developed, capable of the *in vivo* assessment of protein-ligand binding and stability shifts throughout an entire proteome (37). Similarly, MS of intact protein-ligand complexes has been used to resolve individual binding stoichiometries of small molecule ligands on large multiprotein targets, including the individual adenosine tri-phosphate (ATP) binding states of the 800 kDa GroEL chaperone assembly (22; 38; 39). In the most recent of these studies, MS was used to assess the cooperativity of ATP binding to GroEL, demonstrating a strong fit to the Monod–Wyman–Changeux model of cooperativity, which preserves the symmetry of the protein-ligand states created (40; 41).

In addition to quantifying the bound states within complex multiprotein-ligand systems, MS can also act to isolate protein complexes for stability measurements in the gas-phase, following collisional activation. Such collision induced unfolding (CIU) experiments were first described for small monomeric protein ions (42), but have rapidly expanded to include more detailed instrumentation (43) and applications covering large multiprotein complexes (35; 44; 45). In order to track gas-phase protein unfolding, MS must typically be coupled with ion mobility (IM), which acts to separate protein ions according to their orientationally-averaged size and charge (46). For example, CIU results have been used to record the gas-phase folding landscape of ubiqutin ions over a range of charge states using tandem IM instrumentation, with collisional activation regions between IM stages (47). Additionally, CIU of protein complexes has measured the stability of salt-adducted assemblies (48; 49), been used to assess stability enhancements in pathogenic mutants (35), and differentiate conformationally-selective kinase inhibitors (50). Most recently, IM-MS and CIU data have been used to ascertain the selectivity and stability of bound lipids within the mechanosensitive channel of large conductance from *Mycobacterium tuberculosis*, s well as E. *coli* aquaporin and ammonia channels (44).

Here we apply CIU and MS measurements to capture cooperative increases in protein stability as a function of ligand attachment in a multiprotein-ligand binding system. Our target system is Concanavalin A (Con A), a 103kDa lectin tetramer that has been well studied both in solution(51; 52) and in the gas-phase (53), due in part for its central role in lectin affinity chromatography (54). Beyond its well-understood structure, the affinities of the four carbohydrate binding sites on Con A (one per monomer) are well known for a variety of manosyl carbohydrate ligands (55; 56). The complex bound to many of these carbohydrates has been studied intact by MS (53), as has its structural transitions as a function of solvent composition

(57). Cooperative binding models for the assembly have been discussed in the literature (58; 59), but the extent of the cooperativity observed, and how that varies as a function of carbohydrate ligand, is currently relatively unknown. Our IM-MS and CIU data for Con A, which we acquired comprehensively over a range of carbohydrate ligands and binding stoichiometries, reveals evidence for differential cooperative stabilization that favors larger ligands. We discuss our methods and alternative explanations for our observations, as well as their potential implications for IM-MS, CIU and structural biology in general.

2.2 Experimental Methods

2.2.1 Sample Preparation

Con A was purchased from Sigma (St. Louis, MO), and associated manosyl carbohydrate ligand systems were purchased from V-LABS (Covington, LA). Con A is a lectin manosyl carbohydrate-binding protein tetramer, with well-studied sequence and structure.(52) Con A contains one carbohydrate binding site per protein subunit, with each monomer consisting of 237 amino acids (Mw = 25.7kDa), arranged into two anti-parallel β-sheets. While the biological unit of the complex is a tetramer, the assembly has an established pH-dependent equilibrium with a dimeric form, with the dimer dominating below pH 5.6 and at low temperatures(60). Con A has a high affinity to glucose / mannose carbohydrates, and exhibits the highest affinity for carbohydrates having a tri-mannoside, 3,6-di-O-(α-D-mannopyranosyl)-D-mannose core(55; 61). The Con A carbohydrate binding site is situated on a solvent exposed cap of each monomeric unit, proximal to two metal binding sites; a transition metal ion site (S1, typically Mn^{2+}) and a Ca^{2+} site (S2)(62).It has been reported that dimeric and tetrameric Con A bind similarly to a variety of carbohydrates, as reported by both calorimetry (58)and nESI-MS(53).We have chosen

five oligo-saccharide ligands with different binding affinities, having K_Ds ranging from 0.32 - 2.97μM and molecular weights ranging from 504-1235 Da (See Table 2-1 for details), to evaluate the CIU responses for Con A.

2.2.2 Ion Mobility-Mass Spectrometry

Protein-ligand samples (~10uL) were analyzed using our quadrupole ion mobility time-of-flight mass spectrometry (Q-IM-ToF MS) instrument (Synapt G2 HDMS, Waters, Milford, MA). Complex ions were generated using a nESI source and optimized to allow transmission of protein-ligand complexes. The capillary voltage of the nESI source was typically held around 1.6 kV, with the source operating in positive mode. The sampling cone was operated at ~90V. The traveling-wave IM separator was operated at an N_2 pressure of approximately 3.5 mbar, using a 40V wave amplitude traveling at 800-1000 ms^{-1}, in order to generate IM separation. Protein samples were prepared in 100 mM ammonium acetate at pH7 to a concentration of 10 μM following buffer exchange. Saccharide ligand samples were also prepared in aqueous solution at concentrations of 10 μM using 100 mM ammonium acetate buffer concentrations. All ligands were incubated with Con A for 30 minutes prior to nESI-IM-MS analysis.

2.2.3 Collision Induced Unfolding Measurements

Collisional activation in the ion trap traveling-wave ion guide prior to the IM separator was used for CIU of protein complexes in order to investigate the gas-phase stability of protein ions bound to different carbohydrate ligands. Experiments were initially performed in tandem-MS mode.

Ions were selected in the quadrupole mass filter at an m/z corresponding to the 19+ and 20+ charge state of Con A bound to different ligands. Results showed that carbohydrate ligands bind tightly to Con A under our experimental conditions, with no apparent ligand dissociation and charge stripping (up to 9% signal lost for 19+ and 13% for 20+, at trap collision energy of 100V). Data were then collected under a high-throughput native MS1 mode, where we transmitted all ligand bound states simultaneously into the ion trap, to undergo CIU simultaneously. CIU data was acquired by varying the trap collision voltage experienced by ions as they enter the ion trap region of the instrument in 2-5 volt increments and recording IM data for MS-isolated peaks at each discrete voltage value.

2.2.4 Data Analysis

All mass spectra were processed with Masslynx 4.1 software (Waters). The relative intensities of compact Con A tetramer / dimer ions for both apo and ligand bound species (I_f), the only features observed when no activation energy was applied, were calculated as a percentage of the total ion intensity observed at a selected m/z value corresponding to the intact corresponding tetramer or dimer ions using Eq.1. This value is used to chart the unfolding process of protein oligomers as a function of activation voltage used throughout our dataset. The typical standard deviation for the determination of $I_f(\%)$ was 2-4%.

$$I_f(\%) = \frac{I_{folded}}{\Sigma I_{conformers}} \times 100 \quad (1)$$

Figure 2-1 further demonstrates the typical experimental protocol we have developed for measuring the gas-phase stabilities of Con A-Carbohydrate complexes: Protein ions are first generated by nESI and different binding stoichiometries are identified by MS, in the positive ion mode. When low activation voltages are used (blue data, Figure 2-1A), all protein charge states and ligand bound states remain compact and monomodal with respect to their IM arrival time distributions. As the activation voltage is increased (shown in red), IM distributions increase in drift time and become multimodal, indicating protein unfolding in the gas phase. Prior to collecting complete datasets, we conducted preliminary surveys of the IM-MS data as a function of activation voltage in order to assess the course changes in IM drift time encountered by Con A over the activation voltage range available for CIU measurements. The IM drift time signals are further isolated according to the m/z values corresponding to each ligand bound species detected (Figure 2-1B), and the percentage of compact Con A complex ions observed is computed at each activation voltage used for CIU data collection (Figure 2-1C).The resultant compact ion intensities are then plotted against a range of activation voltages, typically ranging from 20-100V, for both the dimeric and tetrameric forms of the Con A – oligosaccharide complexes observed. When we compare protein CIU as a function of ligand binding, additional bound carbohydrates universally stabilize Con A(Figure 2-1D). To quantitatively evaluate the stabilizing influence of carbohydrate binding on Con A, histograms captured at the 50% intensity thresholds shown in Figure 2-1C are generated, which quantify the relative collisional energy required to unfold 50% of the selected protein complex ions (Figure 2-1E). CIU voltage differences captured at 50% relative ion intensity between two adjacent bound states are also calculated in order to assess potential cooperative increases in Con A stability (Figure 2-1E).

Figure 2-1. Experimental protocol for measuring the gas-phase stabilities of Con A – carbohydrate complexes. (A) Protein ions are first generated by nESI and different binding stoichiometries are identified by MS. Representative data at activation voltages of 50V (blue) and 75V (red) are shown, and exhibit different extents of collisional unfolding. (B) The IM drift time signals are further isolated according to the m/z values corresponding to each ligand bound species detected, and analyzed as a function of activation voltage (shown as a IM drift time stack plot). (C) The percentage of compact Con A complex ions observed is computed at each charge state, and each activation voltage, ranging from apo to all carbohydrate-bound species. (D) Representative IM drift time stack plots of protein CIU response as a function of ligand binding. The relative population shift from structural family *b* to *a* upon ligand binding is related to the stability conferred to the Con A complex upon ligand binding. (E) Histograms captured at 50% intensity thresholds shown in Figure 2-1C are generated, which quantify the relative collision voltage required to unfold 50% of the selected protein complex ions. Differential CIU stabilities extracted for proximal bound states are also calculated for our cooperative stabilization analysis.
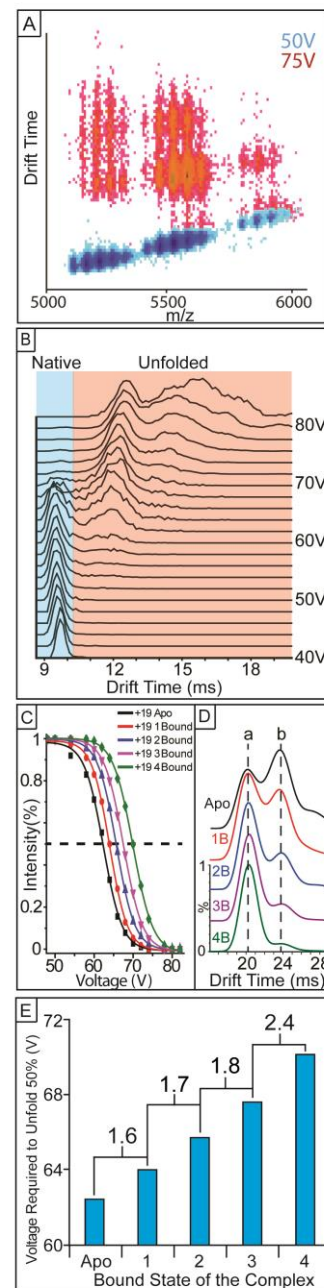
2.2.5 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) measurements were performed on a nanoITC 2G model system from TA Instruments (Waters LLC, New Castle, DE). Briefly, 20 μM Con A and 500~600 μM carbohydrate ligand, both in 100 mM ammonium acetate buffer at pH 7, were used to achieve optimal isotherm titration curves. Blank buffer runs were performed for baseline determination prior to protein ligand titration experiments. Con A solutions were degassed for 15 min, taking care to avoid air bubble formation in the sample, before loading them into the calorimeter cell. The titrations were performed at 25°C, a 300 rpm stir rate, with a >1800 sec equilibration time, and followed by 24 injections separate sample injections each with a 240 sec spacing. Raw ITC data were analyzed by NanoAnalyze software, and were exported to excel for further cooperativity analysis.

2.3 Results and Discussion

Figure 2-2A and 2B show nESI-MS results for dimeric and tetrameric Con A incubated with all five of the carbohydrate ligands shown in Table 2-1. For dimeric Con A, we observe largely apo, 1:1 and 1:2 Con A-ligand stoichiometries, with small amounts of non-specific 1:3 complexes detected due to excess ligand added in solution. Similarly, we observe resolved MS signals for

apo, 1:1, 1:2, 1:3 and 1:4 Con A tetramer-ligand complexes, with negligible evidence of any non-specific interactions and having integrated intensity values (over all charge states) that correlate well with expected K_D values (32,33). Furthermore, we note that the relative intensities of the bound states observed favor higher ligand occupancies more strongly for lower protein charge states, as observed previously (53), an observation most-likely linked to differences in the kinetic and internal energies of the Con A ions as a function of charge state. In general, manosyl carbohydrate ligands, which are ranked in increasing size and binding affinity (top to bottom), show concomitant increases in the bound population observed when ligand concentration is kept constant. Under our conditions, we found that the M3 ligands (504Da) bound to the Con A tetramer represented the practical limits of our MS resolving power, as shown in Figure 2-2B. Similarly, excess M5G2 (1,235Da) binding can cause MS overlap between the 19+ and 20+ signal clusters in Figure 2-2B. By tuning the molar ratio between ligand and protein used during our experiments, we were able to optimize our MS signal to the extent where such overlaps were minimized.

Table 2-1. Carbohydrate ligands, their correlated abbreviations, molecular mass, and dissociation constant (K_D) relative to the Concanavalin A tetramer.

Structure	Abbr.	Mw(Da)	K_D(μM)*
α-D-Man*p*-(1→6)-D-Man*p* ↑3 ↑1 α-D-Man*p*	M3	504.4	2.97
α-D-Man*p*-(1→6)-α-D-Man*p*-(1→6)-D-Man*p* ↑3 ↑3 ↑1 ↑1 α-D-Man*p* α-D-Man*p*	M5	828.7	2.85 1.51
β-D-Glc*p*NAc-(1→2)-β-D-Man*p*-(1→6)-D-Man*p* ↑3 ↑1 β-D-Glc*p*NAc-(1→2)-β-D-Man*p*	G2M3	910.8	0.71
α-D-Man*p*-(1→6)-α-D-Man*p*-(1→4)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc ↑3 ↑1 α-D-Man*p*	M3G2	910.8	0.77
α-D-Man*p*-(1→6)-α-D-Man*p*-(1→6)-α-D-Man*p*-(1→4)-β-D-Glc*p*NAc-(1→4)-β-D-Glc*p*NAc ↑3 ↑3 ↑1 ↑1 α-D-Man*p* α-D-Man*p*	M5G2	1235.1	0.32

* All K_D values are derived from previous data (53). In cases where multiple K_D values are reported for the same carbohydrate ligand, they are reported from Mandal *et al*(61).

CIU data was recorded for each of the signals observed in Figures 2A and 2B, and the CIU stabilities associated with each of these ligand bound states are summarized in Figures 2C and 2D for Con A dimers and tetramers respectively, displayed as a normalized laboratory frame collision energy averaged overall all charge states observed (45). The CIU outputs shown are averaged from multiple charge states and normalized relative to the CIU stabilities recorded for apo Con A. Results show that when carbohydrates interact with either dimeric or tetrameric Con A, the stability of the assembly is generally enhanced. In addition, our CIU stabilities do not possess a strong correlation with either the molecular mass or the solution-phase K_D values for the ligands assessed (Figure S2-1). The stability enhancements observed in our CIU

measurements cover a broad a range. For example, by comparing apo and 1:4 tetramer-ligand complexes, we measure a stability enhancement of 101.5 eV (laboratory frame energy)forM5G2 complexes when compared to those comprised of M3-bound Con A. It is worth noting that all of the carbohydrate ligands tested in this report contain a tri-mannose core structure previously observed to interact with high affinity with Con A (58; 61). As reference, we tested a number of ligands that lacked this tri-mannose structure and observed no evidence of ligand binding, in a similar fashion to previous data (56; 58) .

In order to ascertain the information content of our IM-MS and CIU data relative to the stabilities of the Con A-ligand complexes obtained through conventional MS measurements, collision induced dissociation (CID) stabilities were also recorded for all signals observed in Figures 2A and 2B and compared to those generated from CIU. As discussed previously (34; 35), CID stabilities can be extracted from MS data in a similar manner to the CIU stability values that are extracted from IM data (see Figure 2-1), and correspond to the collision energy required to dissociate 50% of the bound ligand from the intact Con A-ligand assembly. For example, CID and CIU stabilities are recorded and compared for Con A-M3G2 complexes in Figure 2-3, and these data are representative of all similar comparisons that we conducted comprehensively throughout our Con A-ligand complex dataset (data not shown). While uniform CID stabilities are recorded for all ions, we observe significant differential effects on protein stability by CIU, varying by 13% over the apo to 1:4 Con A tetramer-ligand complexes detected. Similar disparities between CID and CIU stability values have been observed for tetrameric transthyretin (TTR)-thyroxine complexes, with CIU results indicating significant stability differences both

between bound states and mutant forms of the protein, whereas CID only detected stability differences upon protein mutation (35) .
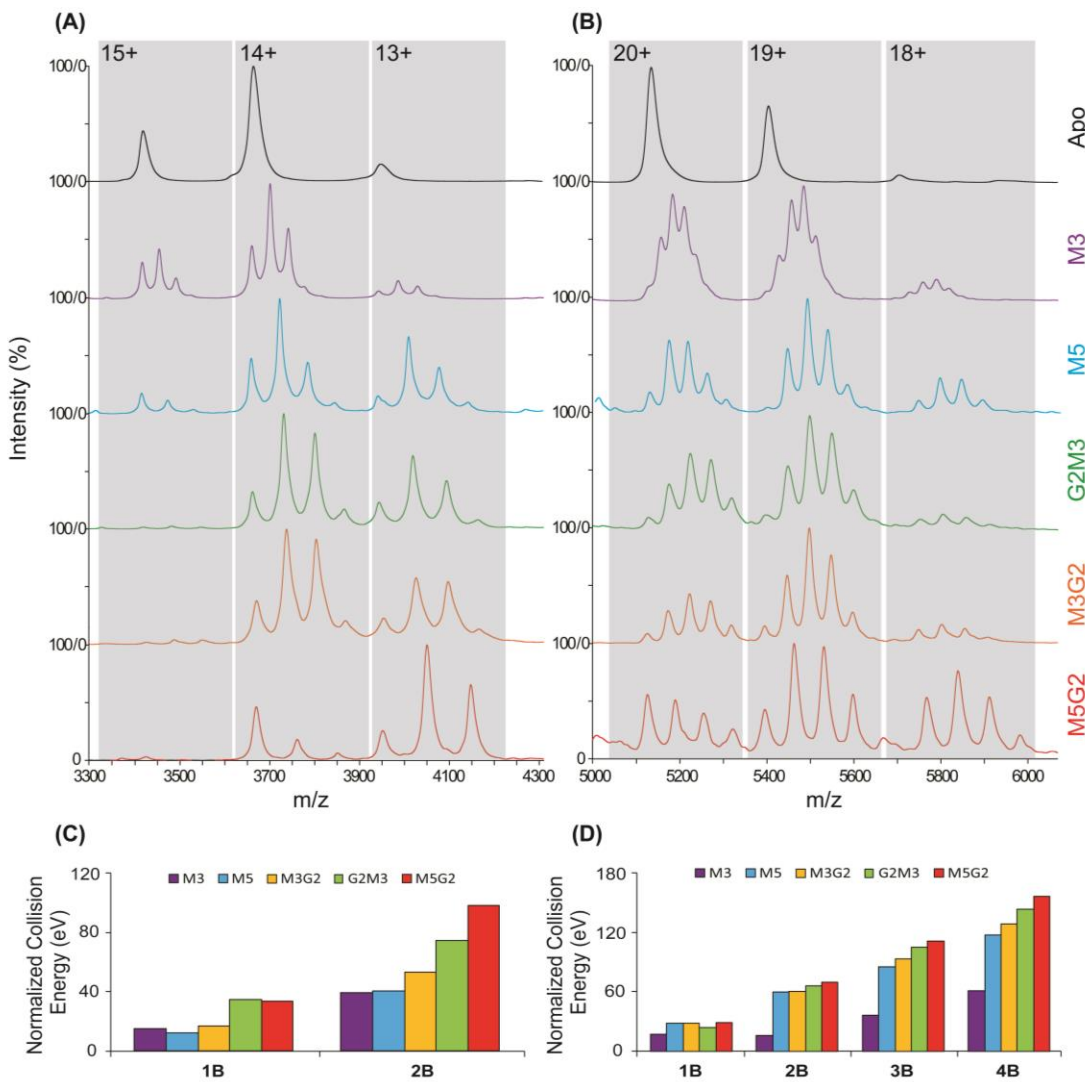
Figure 2-2. MS and CIU results of Con A incubated with all five ligands. MS results for (A) dimeric and (B) tetrameric Con A incubated with all five carbohydrate ligands shown in Table 2-1. CIU stability responses associated with each ligand bound states, across all protein charge states, are summarized in (C) and (D) for Con A dimers and tetramers, in terms of a laboratory collision energy (eV) normalized to that of the apo state. The presented CIU outputs are averaged from multiple charge states and normalized relative to the CIU stabilities recorded for apo Con A.

In addition to detecting stability enhancements upon ligand binding that are not apparent using CID, CIU detects significant differences in the stability conferred to Con A upon ligand binding that, upon close inspection, appears non-linear with respect to Con A ligand occupancy. Figure 2-4A shows CIU stability values for M3 and M5 bound to the 20+ charge state of the Con A tetramer. As discussed above, increasing levels of ligand occupancy enhances the stability of the resulting Con A complex. However, Figure 2-4B, which plots the collision voltage differences recorded between adjacent Con A bound states for M3 and M5, and reveals significant non-linear increases in stability for M5 that are not apparent for M3. For example, upon transitioning from a 1:1 to a 1:2 Con A tetramer: ligand complex, M3 binding allows the assembly to survive for an additional 0.9 V, while M5 binding adds 1.4 V of stability enhancement, despite both ligands conferring nearly identical stability increases upon transition from apo to 1:1 complex forms. Similarly to 1:2 complexes, M5 binding shifts CIU stability by 2.2 V, whereas 1.1 V of stability are added to CIU data for M3, when 1:4 Con A tetramer : ligand complexes are considered. Conversely, 1:3 complexes generated with either ligand generate similar enhancements in tetramer CIU stability. Both of the stability differences cited above, for 1:2 and 1:4 complexes, are outside of the computed standard deviation derived error bars (ranging between 0.1-0.3V) for our relative CIU comparison plots. Overall, the data presented in Figure 2-4B clearly supports a positively cooperative stability enhancement in the Con A tetramer upon binding M5, congruent with known Con A structure and ligand binding mechanisms (58). Similar data analysis was performed in M5G2 and M3G2, also revealing evidence of cooperative stabilization (Figure S2-2), leaving M3 as the only carbohydrate ligand for which no cooperative enhancements in CIU stability are detected.
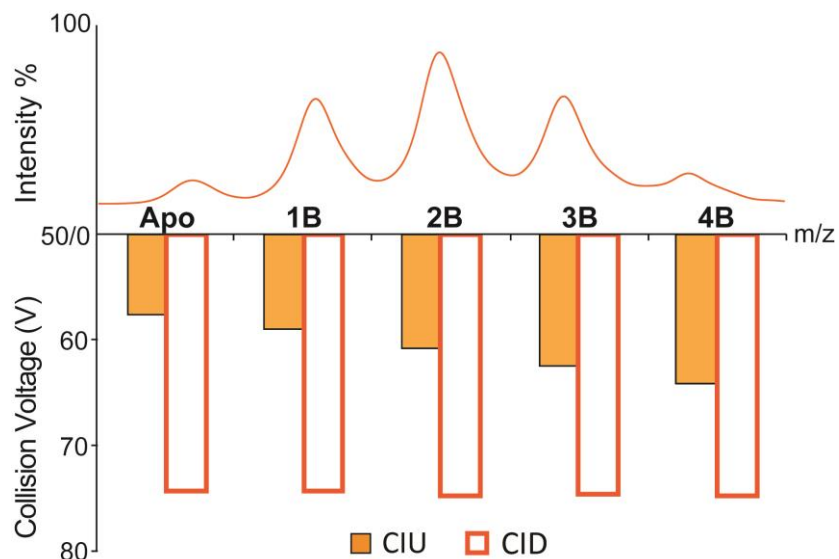
Figure 2-3. MS/CIU/CID results of Con A bound with M3G2. Representative MS results (top), and histograms (bottom) revealing the collision energy required for 50% unfold/dissociation of the 20^+ConA complexes (as labeled), each having different numbers of bound M3G2. The ability to detect differential stabilization upon ligand binding, and thus any cooperative stabilization effect, is unique to our CIU measurements, as no significant differences in MS bound state intensity or CID stability are detected.

In order to compare our gas-phase CIU results with direct measurements of cooperative binding in solution, we performed ITC assays on ~500μM Con A samples incubated with three carbohydrate binders (M3, M5, G2M3). Figure S2-3 plots the enthalpy change recorded by ITC against the number of sample injections performed, which is correlated with carbohydrate ligand concentration. Recorded isotherms for all three Con A-carbohydrate complexes exhibit a 'U' shaped profile, instead of a sigmoid, indicative of positive binding cooperativity (58). Further attempts were made to differentiate and rank the cooperative binding observed. For example, we plotted the fractional saturation ratio of Con A against the free carbohydrate ligand concentration (data not shown), but this analysis resulted in similar values for the relative cooperativity of all

ligands tested. Typically, the magnitude of the binding cooperativity detected is strongly reliant upon ligand binding affinity(63; 64). The ligands chosen for our ITC screen all exhibit similar K_D values relative to Con A, in agreement with this general observation (see Table 2-1). CIU results do not detect any cooperative stabilization upon M3 binding to Con A, despite evidence of a cooperative binding mechanism in solution. This observation may be due to the relatively small mass (504.4 Da) and weak affinity (K_D=2.97 μM) of M3 relative to the other carbohydrates studied here. Alternatively, cooperative stabilization of Con A might not necessarily be linked with cooperative binding, as few examples exist in the literature where both values are probed simultaneously through experiment. Regardless, CIU and ITC results both detect cooperative stabilization and binding, respectively for the other carbohydrate ligands studied here, providing clear evidence of a potential correlation between gas-phase Con A structure and stability and solution-phase protein function.

Still deeper analysis of CIU data provides a more comprehensive picture of the stabilization mechanism adopted by Con A- carbohydrate ligand complexes, as well the potential limitations surrounding the detection of the cooperative stabilization effects described above. Figure 2-4C shows a comparison of the relative differences in CIU stability recorded for 19+ Con A complex ions in comparison with 20+ ions, where the latter ions and their stabilities are described in detail above (Figure 2-4A and 4B). While strong cooperative stabilization of Con A is observed for M5 bound complexes when 20+ ions are analyzed, little evidence of positive cooperativity is observed in the CIU data for 19+ Con A-M5 complex ions. We rationalize this result based on the likely differential ion temperature and energetics of the 19+ and 20+ Con A complexes. It has been observed previously that the CIU and collisional remodeling of protein complexes is

highly charge state dependent [35; 44; 45; 50]. As such, it is not surprising that cooperative stabilization cannot be detected throughout all Con A-ligand complex charge states, as these ions likely possess different threshold energies for collisional unfolding and do so via different mechanistic pathways. Additional pathway details are available through the use of CIU 'fingerprints' which track the size of the Con A complex as a function of the collision voltages used to initiate CIU. These data (Figure S2-4) strongly indicate that the presence of the bound ligand does not alter the unfolding pathway of the Con A tetramer, when compared with control fingerprints acquired for the Apo protein. Instead, global stabilization of Con A relative to CIU (for both the 19+ and 20+ ions) occurs through an increased stability of the most compact form of the assembly (Figure S2-5 and S6). Since such compact forms of the protein are more-closely linked to the native state structure of Con A (35), such results link more closely the cooperative stabilization effects observed to potential analogous stabilization upon Con A-carbohydrate binding in solution.
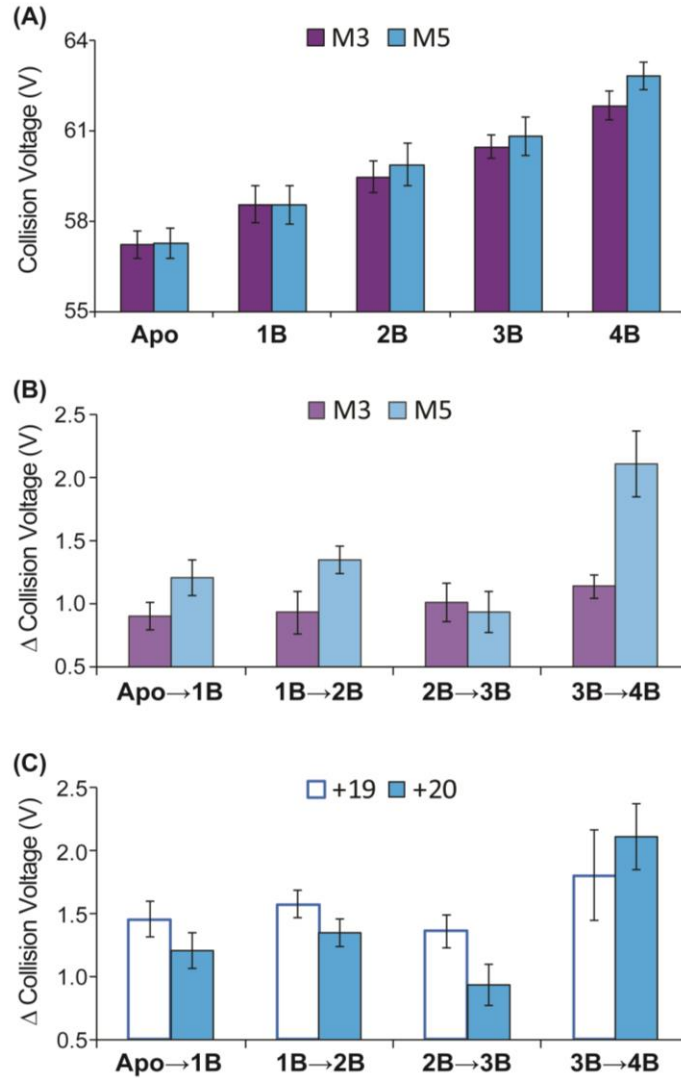
Figure 2-4. (A) CIU stability responses for ligand M3 (purple) and M5 (blue) bound to the 20^+ charge state of Con A tetramer. (B) Collision voltage differences recorded between adjacent Con A bound state for ligand M3 and M5. Data indicates significant non-linear stability increment for M5 that are not apparent for M3. (C) CIU-based collision voltage difference values for M5 bound Con A complexes recorded for 19+ (open blue box) and 20+ (filled blue box) ions. The cooperative stabilization effect observed is clear for those ions having a higher overall charge, but not apparent for those of lesser charge state.

2.4 Conclusions

Here, we present evidence of the cooperative stabilization of a 103 kDa lectin protein tetramer upon carbohydrate ligand binding. We utilize CIU, a novel MS-based methodology that has only been applied in a few cases, to make this assessment, and show that such stabilization cannot easily be detected using solution phase technologies (*e.g.* ITC) or by MS alone. Indeed, such stability measurements require the separation of resolved ligand bound states of the protein complex in order to individually address the stability of each state in isolation. The IM-MS and CIU methods described here are uniquely able to accomplish this, and generate stability data that can potentially be used to inform energy calculations aimed at assessing the details of such cooperative effects. We note a general agreement between our data, and cooperative binding data acquired in solution here, as well our general agreement with previous reports (53; 58; 61). We also note that weak correlations are found between CIU stability and ligand K_D values or molecular mass, demonstrating the orthogonality of CIU stability data and its likely dependence upon complex structure, as well as the mechanistic details of the CIU process.

It is clear, however, that caveats exist for the trends reported here. First, cooperativity is most apparent in higher charge states for the Con A complexes observed. Previous data has linked lower charge states more closely to native-state protein structures (35; 65), but has also discovered clear CIU dependencies upon protein charge state that connect to the overall mechanism of activation-initiated gas-phase protein unfolding and remodeling. In addition, small ligands (*e.g.* M3) do not exhibit cooperative stabilization effects in our CIU dataset, but do

exhibit cooperative binding in solution (Figure S2-2). This discrepancy likely stems directly from the CIU process, which may discriminate against small, weak binders that cannot remain bound to the protein in large numbers following collisional activation. Despite these minor caveats, however, the CIU data shown here adds substantially to the growing list of applications that such gas-phase experiments have found, which currently includes: stability assessments of protein-cation/anion complexes (45; 66), the discovery of conformationally-selective kinase inhibitors (50), and the assignment of membrane protein-lipid binding modes (44). It is clear that future work will continue to expand the CIU experiment to more experiments that utilize its unique ability to extract protein stability values from complex protein mixtures, using relatively small amounts of sample.

2.5 References

1. Williams DH, Stephens E, O'Brien DP, Zhou M (2004) Understanding noncovalent interactions: Ligand binding energy and catalytic efficiency from ligand-induced reductions in motion within receptors and enzymes. Angew Chem-Int Edit 43:6596-6616.

2. Eaton WA, Henry ER, Hofrichter J, Mozzarelli A (1999) Is cooperative oxygen binding by hemoglobin really understood? Nature Structural & Molecular Biology 6:351-358.

3. Pantoliano MW, Horlick RA, Springer BA, Vandyk DE, Tobery T, Wetmore DR, Lear JD, Nahapetian AT, Bradley JD, Sisk WP (1994) MULTIVALENT LIGAND-RECEPTOR BINDING INTERACTIONS IN THE FIBROBLAST GROWTH-FACTOR SYSTEM PRODUCE A COOPERATIVE GROWTH-FACTOR AND HEPARIN MECHANISM FOR RECEPTOR DIMERIZATION. Biochemistry 33:10229-10248.

4. Wyman J (1964) Linked functions and reciprocal effects in hemoglobin: a second look. Adv Protein Chem 19:91.

5. Li R, Botchan MR (1993) The acidic transcriptional activation domains of VP16 and p53 bind the cellular replication protein A and stimulate in vitro BPV-1 DNA replication. Cell 73:1207-1221.

6. Myers RM, Rio DC, Robbins AK, Tjian R (1981) SV40 gene expression is modulated by the cooperative binding of T antigen to DNA. Cell 25:373-384.

7. Martin A, Baker TA, Sauer RT (2005) Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. Nature 437:1115-1120.

8. Katz DH, Hamaoka T, Benacerraf B (1973) Cell interactions between histoincompatible T and B lymphocytes II. Failure of physiologic cooperative interactions between T and B lymphocytes from allogeneic donor strains in humoral response to hapten-protein conjugates. The Journal of experimental medicine 137:1405-1418.

9. McGhee JD, von Hippel PH (1974) Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. Journal of molecular biology 86:469-489.

10. Zhao Y-L, Wu Y-D (2002) A theoretical study of β-sheet models: is the formation of hydrogen-bond networks cooperative? Journal of the American Chemical Society 124:1570-1571.

11. Murphy KP, Freire E (1992) Thermodynamics of structural stability and cooperative folding behavior in proteins. Advances in protein chemistry 43:313-361.

12. Stam JC, Sander EE, Michiels F, van Leeuwen FN, Kain HE, van der Kammen RA, Collard JG (1997) Targeting of Tiam1 to the plasma membrane requires the cooperative function of the N-terminal pleckstrin homology domain and an adjacent protein interaction domain. Journal of Biological Chemistry 272:28447-28454.

13. Kindon H, Pothoulakis C, Thim L, Lynch-Devaney K, Podolsky DK (1995) Trefoil peptide protection of intestinal epithelial barrier function: cooperative interaction with mucin glycoprotein. Gastroenterology 109:516-523.

14. Looger LL, Dwyer MA, Smith JJ, Hellinga HW (2003) Computational design of receptor and sensor proteins with novel functions. Nature 423:185-190.

15. Mizoue LS, Chazin WJ (2002) Engineering and design of ligand-induced conformational change in proteins. Curr Opin Struct Biol 12:459-463.

16. Duke TAJ, Le Novere N, Bray D (2001) Conformational spread in a ring of proteins: A stochastic approach to allostery. Journal of molecular biology 308:541-553.

17. Miller DW, Dill KA (1997) Ligand binding to proteins: The binding landscape model. Protein Sci 6:2166-2179.

18. Feng J, Walter NG, Brooks CL (2011) Cooperative and Directional Folding of the preQ(1) Riboswitch Aptamer Domain. Journal of the American Chemical Society 133:4196-4199.

19. Alexander P, Fahnestock S, Lee T, Orban J, Bryan P (1992) THERMODYNAMIC ANALYSIS OF THE FOLDING OF THE STREPTOCOCCAL PROTEIN-G IGG-BINDING DOMAINS B1 AND B2 - WHY SMALL PROTEINS TEND TO HAVE HIGH DENATURATION TEMPERATURES. Biochemistry 31:3597-3603.

20. Freire E (1999) The propagation of binding interactions to remote sites in proteins: Analysis of the binding of the monoclonal antibody D1.3 to lysozyme. Proc Natl Acad Sci U S A 96:10118-10122.

21. Street TO, Krukenberg KA, Rosgen J, Bolen DW, Agard DA (2010) Osmolyte-induced conformational changes in the Hsp90 molecular chaperone. Protein Sci 19:57-65.

22. Dyachenko A, Gruber R, Shimon L, Horovitz A, Sharon M (2013) Allosteric mechanisms can be distinguished using structural mass spectrometry. Proceedings of the National Academy of Sciences 110:7235-7239.

23. DeChancie J, Houk K (2007) The origins of femtomolar protein-ligand binding: hydrogen-bond cooperativity and desolvation energetics in the biotin-(strept) avidin binding site. Journal of the American Chemical Society 129:5419-5429.

24. Marcos E, Crehuet R, Bahar I (2011) Changes in Dynamics upon Oligomerization Regulate Substrate Binding and Allostery in Amino Acid Kinase Family Members. PLoS Comput Biol 7.

25. Rosengarth A, Rosgen J, Hinz HJ, Gerke V (2001) Folding energetics of ligand binding proteins II. Cooperative binding of Ca2+ to annexin I. Journal of molecular biology 306:825-835.

26. Doyle ML, Gill SJ, Cusanovich MA (1986) Ligand-Controlled Dissociation Of Chromatium-Vinosum Cytochrome-C'. Biochemistry 25:2509-2516.

27. Loo JA (1997) Studying noncovalent protein complexes by electrospray ionization mass spectrometry. Mass Spectrom Rev 16:1-23.

28. Robinson CV, Chung EW, Kragelund BB, Knudsen J, Aplin RT, Poulsen FM, Dobson CM (1996) Probing the nature of noncovalent interactions by mass spectrometry. A study of protein-CoA ligand binding and assembly. Journal of the American Chemical Society 118:8646-8653.

29. Niu S, Rabuck JN, Ruotolo BT (2013) Ion mobility-mass spectrometry of intact protein-ligand complexes for pharmaceutical drug discovery and development. Curr Opin Chem Biol 17:809-817.

30. Hyung SJ, Ruotolo BT (2012) Integrating mass spectrometry of intact protein complexes into structural proteomics. Proteomics 12:1547-1564.

31. Cheng KW, Wong CC, Wang MF, He QY, Chen F (2010) IDENTIFICATION AND CHARACTERIZATION OF MOLECULAR TARGETS OF NATURAL PRODUCTS BY MASS SPECTROMETRY. Mass Spectrom Rev 29:126-155.

32. Wang WJ, Kitova EN, Klassen JS (2003) Influence of solution and gas phase processes on protein-carbohydrate binding affinities determined by nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry. Anal Chem 75:4945-4955.

33. Sun JX, Kitova EN, Wang WJ, Klassen JS (2006) Method for distinguishing specific from nonspecific protein-ligand complexes in nanoelectrospray ionization mass spectrometry. Anal Chem 78:3010-3018.

34. Ruotolo BT, Hyung SJ, Robinson PM, Giles K, Bateman RH, Robinson CV (2007) Ion Mobility–Mass Spectrometry Reveals Long-Lived, Unfolded Intermediates in the Dissociation of Protein Complexes. Angewandte Chemie International Edition 46:8001-8004.

35. Hyung S-J, Robinson CV, Ruotolo BT (2009) Gas-phase unfolding and disassembly reveals stability differences in ligand-bound multiprotein complexes. Chemistry & biology 16:382-390.

36. Benesch JL, Ruotolo BT (2011) Mass spectrometry: come of age for structural and dynamical biology. Curr Opin Struct Biol 21:641-649.

37. Strickland EC, Geer MA, Tran DT, Adhikari J, West GM, DeArmond PD, Xu Y, Fitzgerald MC (2013) Thermodynamic analysis of protein-ligand binding interactions in complex biological mixtures using the stability of proteins from rates of oxidation. Nature protocols 8:148-161.

38. Zhang Q, Chen J, Kuwajima K, Zhang HM, Xian F, Young NL, Marshall AG (2013) Nucleotide-induced conformational changes of tetradecameric GroEL mapped by H/D exchange monitored by FT-ICR mass spectrometry. Sci Rep 3.

39. van Duijn E, Bakkes PJ, Heeren RM, van den Heuvel RH, van Heerikhuizen H, van der Vies SM, Heck AJ (2005) Monitoring macromolecular complexes involved in the chaperonin-assisted protein folding cycle by mass spectrometry. Nature Methods 2:371-376.

40. Monod J, Wyman J, Changeux J-P (1965) On the nature of allosteric transitions: a plausible model. Journal of molecular biology 12:88-118.

41. Changeux JP (2012) Allostery and the Monod-Wyman-Changeux Model After 50 Years. Annu Rev Biophys 41:103-133.

42. Hopper JT, Oldham NJ (2009) Collision induced unfolding of protein ions in the gas phase studied by ion mobility-mass spectrometry: the effect of ligand binding on conformational stability. Journal of the American Society for Mass Spectrometry 20:1851-1858.

43. Zhu F, Lee S, Valentine SJ, Reilly JP, Clemmer DE (2012) Mannose7 glycan isomer characterization by IMS-MS/MS analysis. Journal of the American Society for Mass Spectrometry 23:2158-2166.

44. Laganowsky A, Reading E, Allison TM, Ulmschneider MB, Degiacomi MT, Baldwin AJ, Robinson CV (2014) Membrane proteins bind lipids selectively to modulate their structure and function. Nature 510:172-175.

45. Han L, Hyung S-J, Mayers JJ, Ruotolo BT (2011) Bound anions differentially stabilize multiprotein complexes in the absence of bulk solvent. Journal of the American Chemical Society 133:11358-11367.

46. Kanu AB, Dwivedi P, Tam M, Matz L, Hill HH (2008) Ion mobility–mass spectrometry. Journal of Mass Spectrometry 43:1-22.

47. Jung JE, Pierson NA, Marquardt A, Scheffner M, Przybylski M, Clemmer DE (2011) Differentiation of Compact and Extended Conformations of Di-Ubiquitin Conjugates with Lysine-Specific Isopeptide Linkages by Ion Mobility-Mass Spectrometry. Journal of the American Society for Mass Spectrometry 22:1463-1471.

48. Han L, Ruotolo BT (2013) Traveling-wave ion mobility-mass spectrometry reveals additional mechanistic details in the stabilization of protein complex ions through tuned salt additives. International Journal for Ion Mobility Spectrometry 16:41-50.

49. Han L, Hyung S-J, Ruotolo BT (2013) Dramatically stabilizing multiprotein complex structure in the absence of bulk water using tuned Hofmeister salts. Faraday discussions 160:371-388.

50. Rabuck JN, Hyung S-J, Ko KS, Fox CC, Soellner MB, Ruotolo BT (2013) Activation state-selective kinase inhibitor assay based on ion mobility-mass spectrometry. Anal Chem 85:6995-7002.

51. Lis H, Sharon N (1986) Lectins as molecules and as tools. Annual review of biochemistry 55:35-67.

52. Bittiger H, Schnebli HP (1976) Concanavalin A as a Tool. Wiley.

53. ávan Dongen WD, Albert J (2000) Binding of selected carbohydrates to apo-concanavalin A studied by electrospray ionization mass spectrometry. Analyst 125:583-589.

54. Lotan R, Nicolson GL (1979) Purification of cell membrane glycoproteins by lectin affinity chromatography. Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes 559:329-376.

55. Mandal DK, Kishore N, Brewer CF (1994) Thermodynamics of lectin-carbohydrate interactions. Titration microcalorimetry measurements of the binding of N-linked carbohydrates and ovalbumin to concanavalin A. Biochemistry 33:1149-1156.

56. Dam TK, Roy R, Das SK, Oscarson S, Brewer CF (2000) Binding of multivalent carbohydrates to concanavalin A and Dioclea grandiflora lectin - Thermodynamic analysis of the "multivalency effect". Journal of Biological Chemistry 275:14223-14230.

57. Edelman GM, Cunningham BA, Reeke GN, Becker JW, Waxdal MJ, Wang JL (1972) The covalent and three-dimensional structure of concanavalin A. Proceedings of the National Academy of Sciences 69:2580-2584.

58. Chervenak MC, Toone EJ (1995) Calorimetric analysis of the binding of lectins with overlapping carbohydrate-binding ligand specificities. Biochemistry 34:5685-5695.

59. Zeng X, Andrade CA, Oliveira MD, Sun X-L (2012) Carbohydrate–protein interactions and their biosensing applications. Analytical and bioanalytical chemistry 402:3161-3176.

60. Bhattacharyya L, Brewer CF (1990) Isoelectric focusing studies of concanavalin A and the lentil lectin. Journal of Chromatography A 502:131-142.

61. Mandal DK, Brewer CF (1993) Differences in the binding affinities of dimeric concanavalin A (including acetyl and succinyl derivatives) and tetrameric concanavalin A with large oligomannose-type glycopeptides. Biochemistry 32:5116-5120.

62. Liener I. 2012. The lectins: properties, functions, and applications in biology and medicine, Elsevier.

63. Houtman JC, Brown PH, Bowden B, Yamaguchi H, Appella E, Samelson LE, Schuck P (2007) Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: application to adaptor protein complexes in cell signaling. Protein Sci 16:30-42.

64. Dam J, Schuck P (2005) Sedimentation velocity analysis of heterogeneous protein-protein interactions: sedimentation coefficient distributions c (s) and asymptotic boundary profiles from Gilbert-Jenkins theory. Biophysical journal 89:651-666.

65. Pagel K, Hyung S-J, Ruotolo BT, Robinson CV (2010) Alternate dissociation pathways identified in charge-reduced protein complex ions. Anal Chem 82:5363-5372.

66. Han L, Hyung SJ, Ruotolo BT (2012) Bound cations significantly stabilize the structure of multiprotein complexes in the gas phase. Angewandte Chemie 124:5790-5793.

2.6 Supporting Information
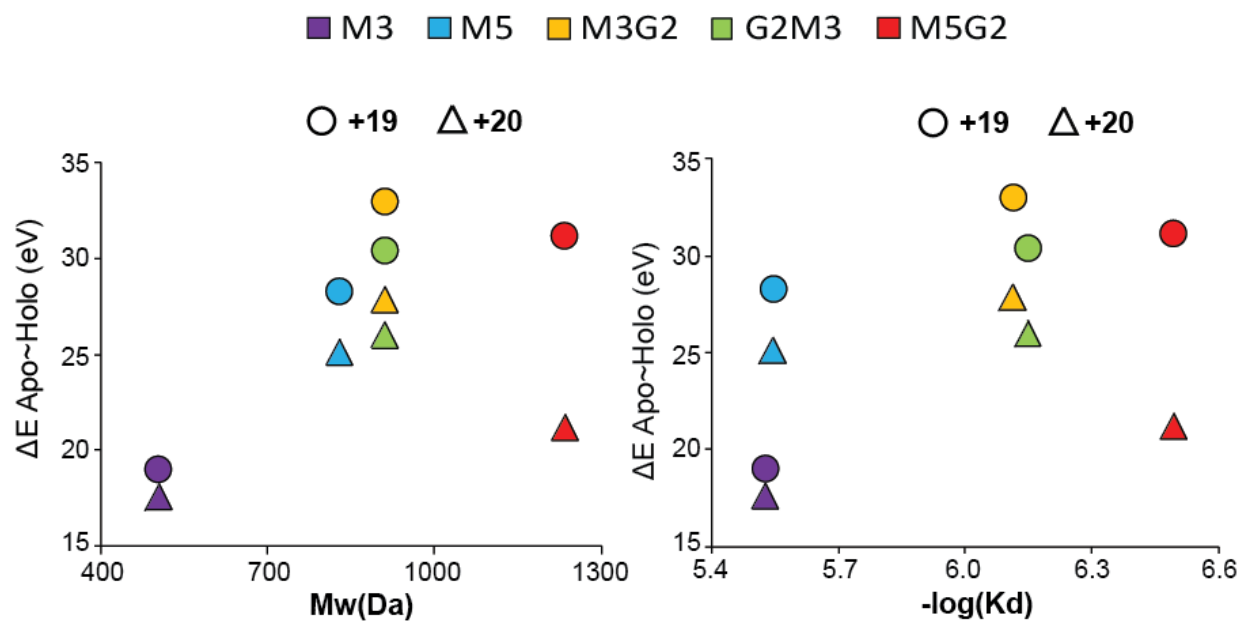
Figure S2-1. Plot showing the first energy leap (Apo - Holo) upon different sugar ligand binding, as a function of **(left)** Mw and **(right)** pKd. Results indicate a mild positive correlation in that, larger/tighter ligand tends to provide better stabilization effect; Within the each ligand binding scenario, lower charge state (+19) consistently provides larger stabilization compared with higher charge state (+20)
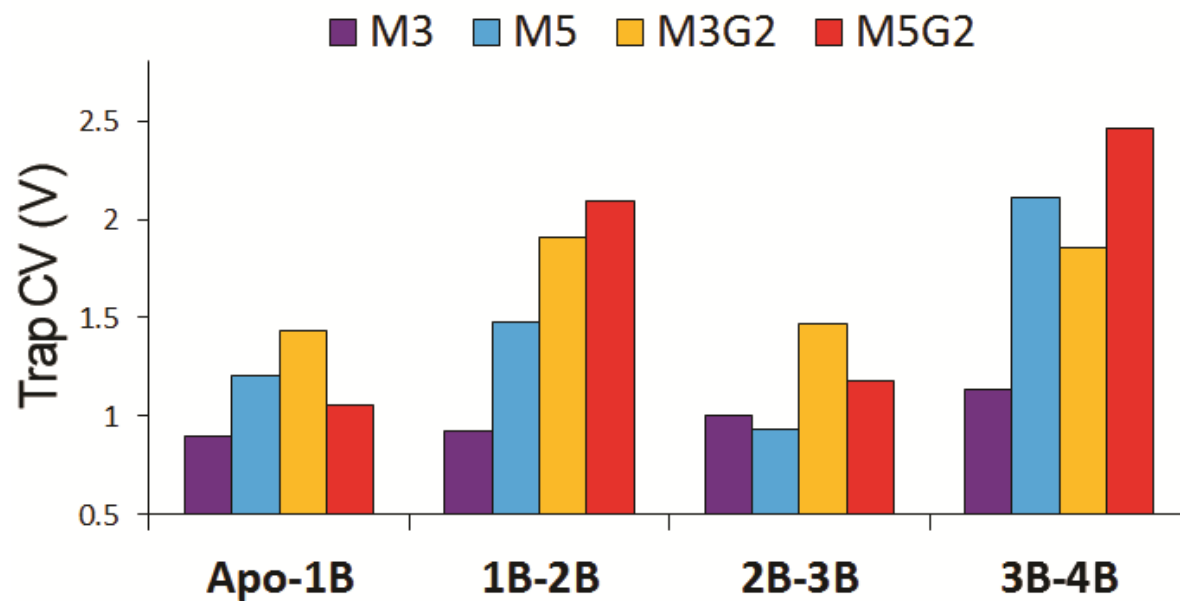
Figure S2-2 Collision voltage differences recorded between adjacent Con A bound state for ligand M3, M5, M3G2 and M5G2. Non-linear stability increments were observed for all the investigated ligands other than the smallest ligand M3. Complete CIU response data for ligand G2M3 is not available, due to insufficient signal at 20+ charge state. Mild cooperative stabilization effect can still be observed based on 19+ data.
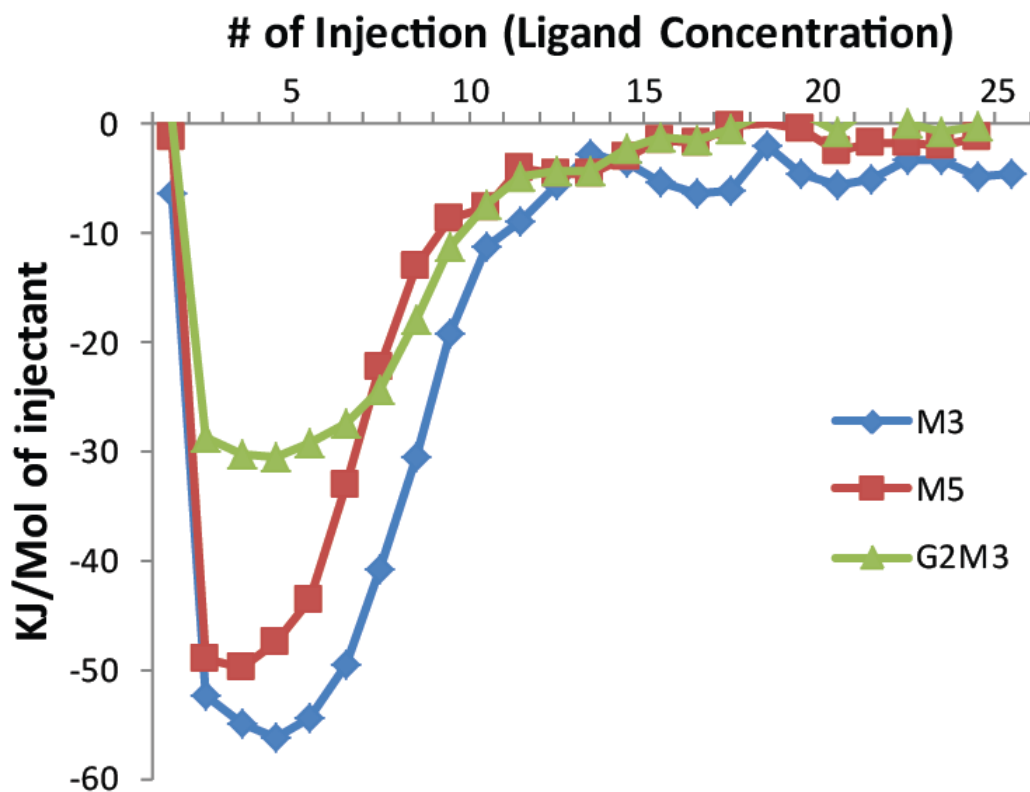
Figure S2-3. Isothermal titration calorimetry data of Con A bound to ligand M3/M5/G2M3. As Con A has 4 identical but dependent binding sites, the fact that the first few injections are increasing in magnitude, instead of decreasing, making the overall isotherm in a non-sigmoidal, U-shaped curve, is the critical evidence revealing positive binding cooperativity in all three cases.
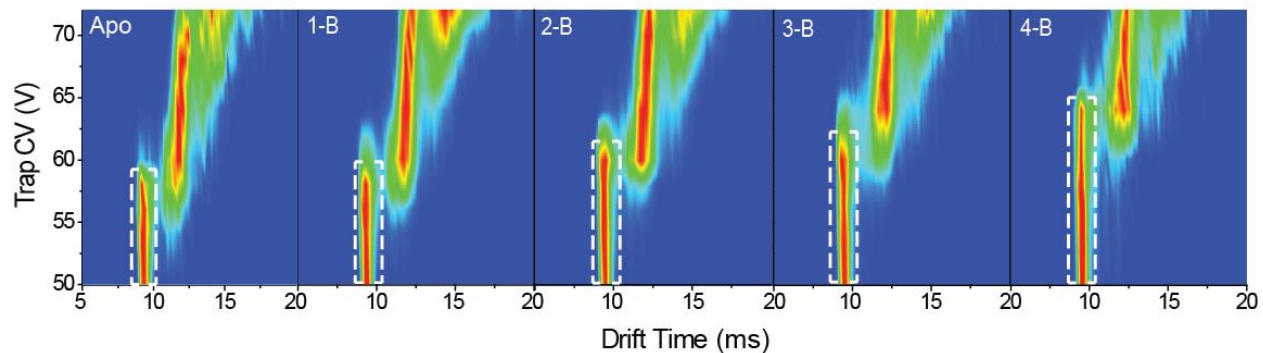
Figure S2-4. Representative CIU fingerprints shown for Con A bound to ligand M5 in four stoichiometries at 20^+ charge state. Results show that, upon activation in the gas-phase, ligand-bound Con A tetramer uniformly unfolds into at least 2 structural intermediate forms over the displayed voltage range. The dashed white boxes highlight the most-compact form of the protein, which persists over larger voltage ranges in cases where more ligands are bound, indicating that ligand binding primarily stabilizes this native-like conformation, rather than other activated structural intermediates.
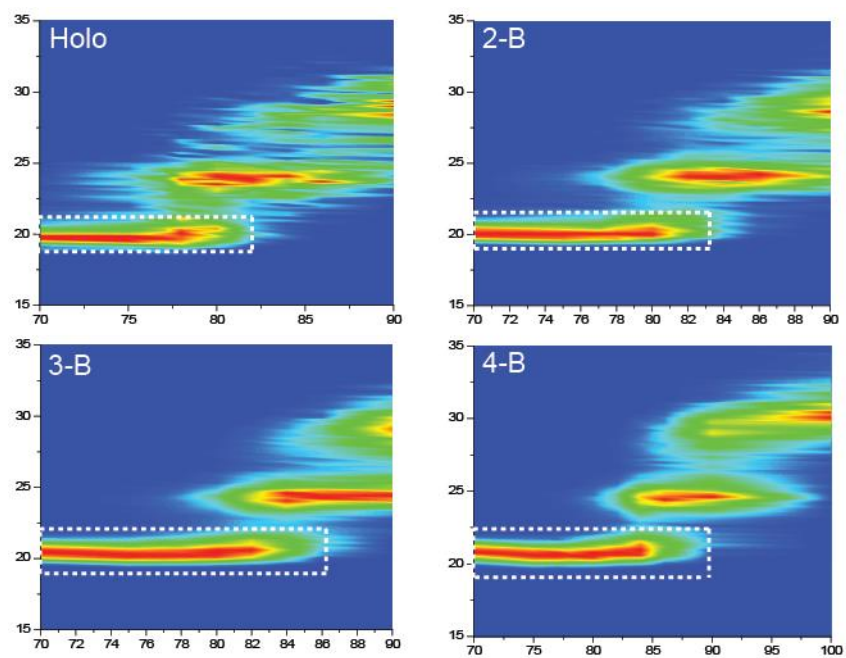
Figure S2-5 Representative CIU fingerprints shown for Con A bound to ligand M5G2 in four stoichiometries at 19^+ charge state. Consistent trends were observed in larger carbohydrate ligand, at lower charge state.
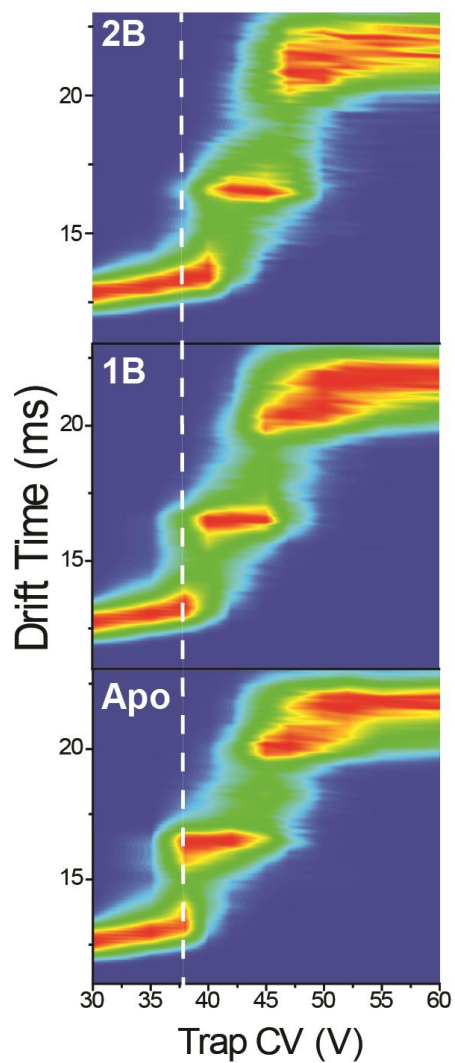
Figure S2-6 Representative CIU fingerprints shown for Con A dimer bound to ligand M3 in two stoichiometries at 15⁺ charge state. Consistent trends were observed in the smallest carbohydrate ligand M3, for Con A dimer as well.

Chapter 3

Development of a Detailed Molecular Model for the Collision Induced Unfolding of Multiprotein-Ligand Complexes

3.1 Introduction

Proteins often function in the context of large assemblies in order to regulate the various biological processes necessary to life.[1, 2] Thus, protein complexes are critically important targets for drug discovery and development.[3, 4]Mass spectrometry (MS) is a critical component of many drug development efforts worldwide, where it is used to assess small molecule binding to target substrates and the pharmacokinetics of drug action. However, MS is rarely used to screen for the structural consequences of drug binding, and more time-consuming technologies capable of projecting atomic models of protein-drug interactions are utilized instead. As the intended phenotypes of drug interactions trend toward those that target critical

conformation changes within the bound substrates, in many cases binding to proteins having no known enzymatic active site, developing rapid tools capable of screening subtle conformational changes within protein-protein and protein-ligand complexes is acritical goal.[5]

Currently, there are many different MS approaches able to recover information from intact protein complexes. As shown in Figure 3-1, intact protein-ligand complexes can be mass analyzed following nESI, revealing binding stoichiometries and affinities,.[6-9]On the other hand, the collisional activation experiments have also been reported as a means of recovering useful structural information from intact assemblies.[10-13]While much is known regarding the composition of the products ions generated upon protein complex collision induced dissociation (CID), much less is known regarding the intermediate protein structural forms produced during the CID process.

It is worth noting that in the context of the first observation of protein complex CID, protein unfolding events were invoked to describe the details of the dissociating process[14]. During protein complex CID, charges are partitioned asymmetrically on the eventual product ions observed, and thus an asymmetrically unfolded transition state was invoked for such experiments early in the development of the field. For example, Jurchen and co-workers provided convincing evidence suggesting that the origin of such asymmetric charge partitioning within protein

homodimers is linked directly to the foldness of the transition state, resulting monomeric product ions that carry away the majority of the precursor charge.[15, 16] Moreover, high transition-state entropies measured during blackbody infrared dissociation experiments on homo-pentameric protein complexes described by Klassen *et al* also indicate that unfolding events take place during the CID transition state.[14]IM-MS data for protein complexes following collisional activation has been used, as well, to directly observe collision induced unfolding (CIU) in protein complexes, and link these forms in general to CID product ion formation[10]. In addition to basic arguments of involving Coulombic repulsion and the increased surface area of the transition state, it has also been proposed that the apparent gas phase basicity of the elongated proteins is enhanced relative to that in the folded state, rendering it energetically favorable to transfer protons asymmetrically relative to product ion mass.[15]However, despite the details above, there are a number of persistent, lingering questions surrounding the nature and consequences of the unfolded CID transition state. For example, what are the chief driving forces for the unfolding process? Does multiprotein unfolding follow a single monomer unfolding mechanism, or do multiple subunits unfold simultaneously? To what extent to native-like small molecule binding sites remain intact within the unfolding proteins, and how would such information be used to interpret CIU data for such systems?
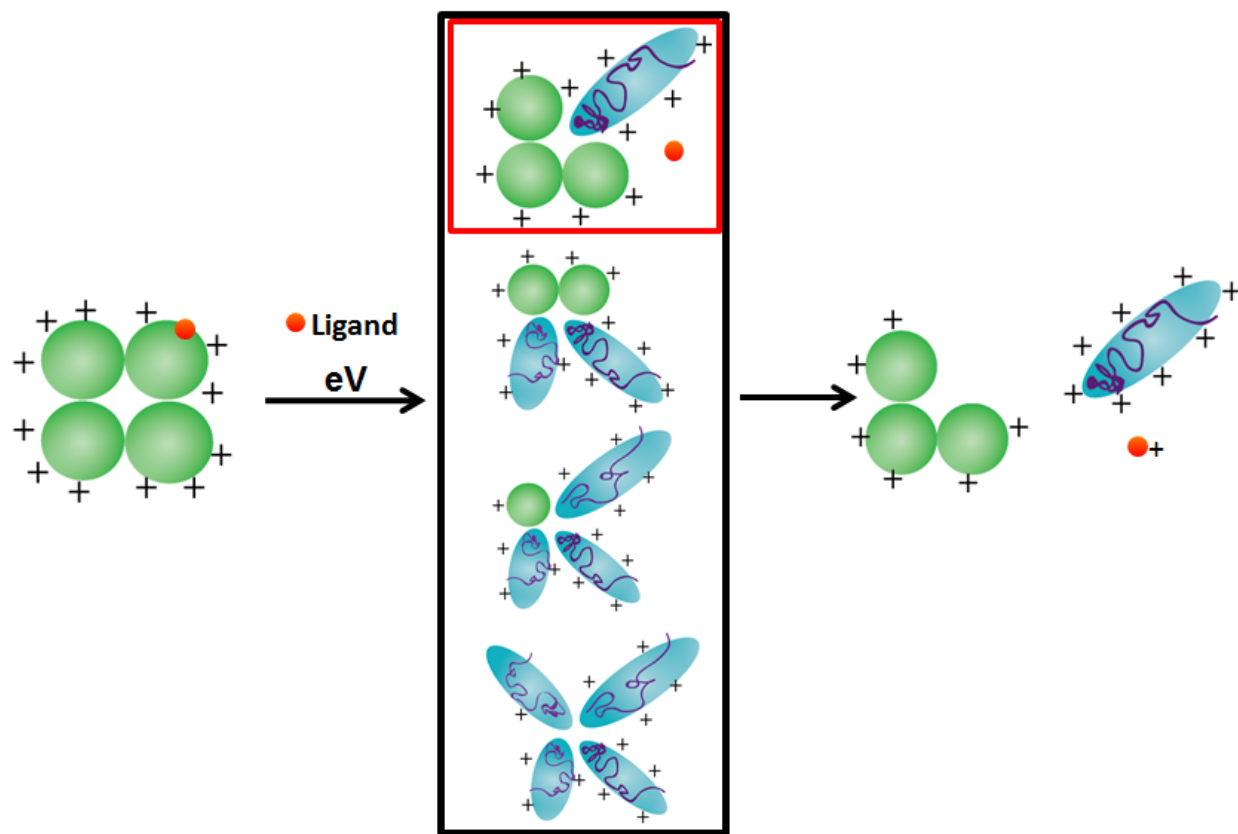
Figure3-1. Schematic illustration of the multi-protein (shown as a homo-tetramer) unfolding and dissociation process. Native protein subunits are labeled in green, unfolded conformers are in blue, and a small molecule binder is shown in red. Two distinct unfolding mechanisms are proposed: A. Charges migrates onto a single monomer following thermal unfolding, and this triggers a broader unfolding event (highlighted in red box); B. Multiple subunits unfolding simultaneously to achieve maximum transition state entropy, with mixed charge migration effects.

To address these questions, we attempt to use IM-MS, in conjunction with molecular modeling

(MM) approaches, to provide enhanced fundamental understanding of the CIU process for .the

tetrameric sugar binding protein Concanavalin A (Con A). Con A is a lectin manosyl carbohydrate-binding protein tetramer, with well-studied sequence and structure.[17, 18] Con A contains one carbohydrate binding site per protein subunit, with each monomer consisting of 237 amino acids (Mw = 25.7kDa), arranged into two anti-parallel β-sheets. While the biological unit of the complex is a tetramer, the assembly has an established pH-dependent equilibrium with a dimeric form, with the dimer dominating below pH 5.6 and at low temperatures[19]. Previous modeling efforts that have aimed to elucidate the details of protein complex CIU and CID have either used coursed-grained models or have not completely accounted for inter-protein charge migration [20, 21]. We employ experimental CIU fingerprints of the intact Con A complex [10, 22] along with Coulombic unfolding data for Con A monomers, to surmount these challenges and create full atomic models of the CID transition states for the first time. We find a strong correlation between the Coulombic unfolding of Con A monomers, and the CIU of intact Con A tetramer, which suggests a single monomer unfolding mechanism. Furthermore, when MM models of Coulombically unfolded Con A monomer are docked onto a folded Con A trimer, strong agreement is found between the measured Con A sizes achieved during CIU of the intact complex. In order to validate this agreement, we develop a new calibration curve that converts between projected model areas and full Monte Carlo, physics-based estimates of ion size. Finally, we quantify the amount of charge transfer needed to achieve each of the four unfolded states observed during Con A tetramer CIU, discovering that the majority of charge transfer takes place during the initial unfolding phase of a single Con A monomer within the assembly.

These results are discussed in the context of the CID mechanism for protein complexes, as well as emerging needs for CIU data analysis.

3.2 Experimental Methods

3.2.1 Ion Mobility-Mass Spectrometry

Con A(jack bean) and m-NBA (3-Nitrobenzyl alcohol)were purchased from Sigma (St. Louis, MO).Con A samples (~10uL) were analyzed using quadrupole ion mobility time-of-flight mass spectrometry (Q-IM-ToF MS) instrument (Synapt G2 HDMS, Waters, Milford, MA). Complex ions were generated using a nano-electrospray ionization(nESI) source and optimized to allow transmission of protein assemblies. The capillary voltage of the nESI source was typically held around 1.6 kV, with the source operating in positive mode. The sampling cone was operated at ~90V. The traveling-wave IM separator was operated at an N_2 pressure of approximately 3.5 mbar, using a 40V wave amplitude traveling at 800-1000 ms^{-1}, in order to generate IM separation. In terms of the solution disruption experiment, Con A sample was incubated with 5% m-NBA, a classic super-charging reagent,[23] to disrupt the monomer-monomer interface of Con A and generated charge amplified monomeric units.
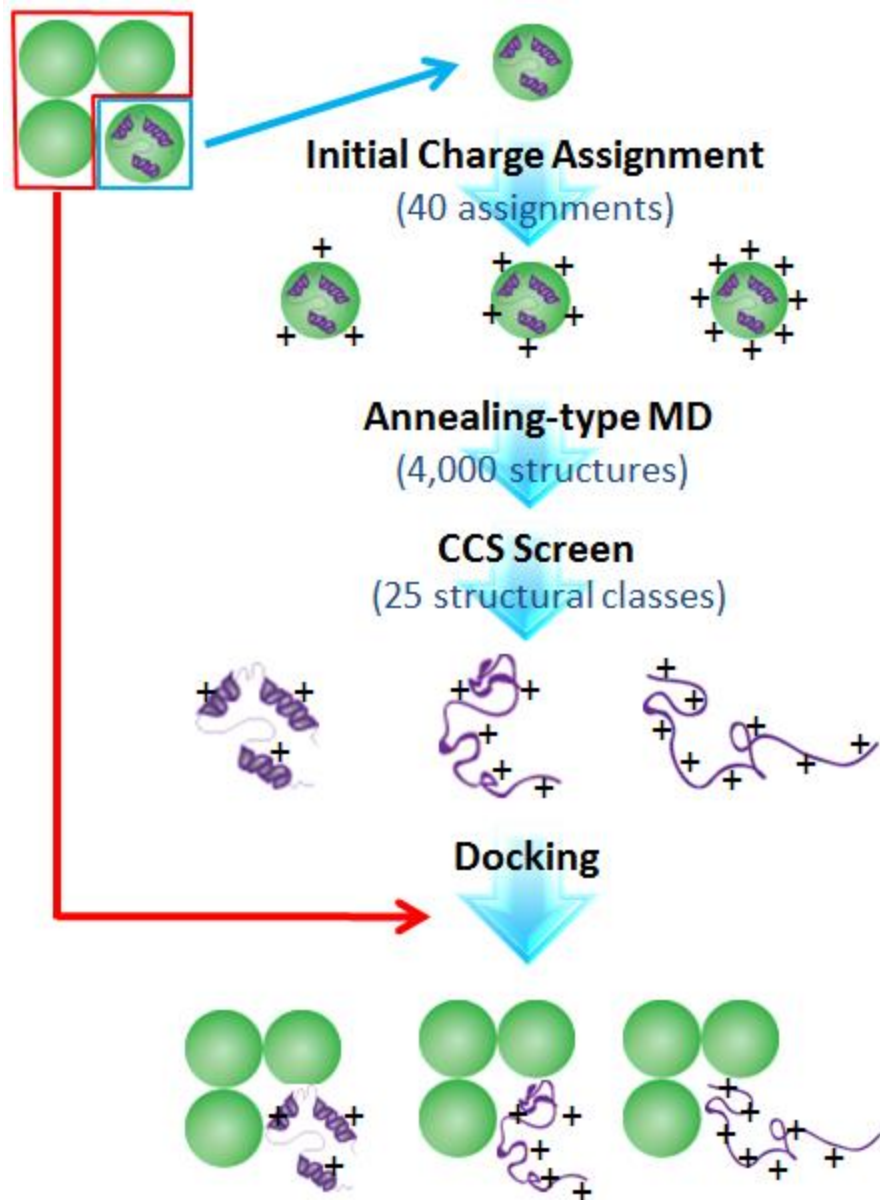
Figure3-2. Schematic workflow for computational modeling protein complex CID and CIU. Con A monomer is isolated from the crystal structure tetramer (PDB: 1VAL), and subject to initial charge assignment; the charged monomer models are then further interrogated by an annealing-type MD simulation, and filtered against experimental CCS measurements. Model structures with low force field energy values and CCSs that agree with experimental values are highlighted, and docked back onto the remaining Con A trimer, yielding the artificial tetrameric model, which is compared to CIU data for the intact assembly.

Intact protein samples were prepared in 100 mM ammonium acetate at pH7 to a concentration of 10 μM following buffer exchange. CIU fingerprints, a contour plot that tracks the size change of the interested protein ions as a function of collision energy, were generated based on the 20^+ and 21^+ charge states of Con A tetramer. All mass spectra were calibrated externally using a solution of cesium iodide (100 mg/mL) and analyzed using MassLynx 4.1 and Driftscope 2.0 software. CCS measurements were externally calibrated using a previously published database of known values in helium, using values for peptides and proteins that bracket the likely collision cross section (CCS) and ion mobility values of the Con A ions.

3.2.2 Computational Work Flow

Figure 3-2 illustrates the computational working protocol of this chapter. Briefly, simulated monomer structure models for Con A were isolated from the intact tetrameric complex X-ray crystal structure. (PDB ID: 1VAL[24]) We started with assigning charges onto the basic residues (Lys, Arg, His and N-terminus) within monomers in a manner such that charge positions were distributed evenly over the surface of the protein structure, in an effort to minimize Coulombic repulsion. 40 different charge assignments, representing monomer charge state ranging from 10^+ to 18^+, were generated. Molecular dynamics simulations were then performed using the

GROMACS[25] software package (ver. 4.5.5) utilizing the GROMOS96 53a6 force field.[26]The charged assigned models were placed in a cubic simulation box with the minimum distance between the simulated molecules and the box wall being 1.0 nm. The system was energy-minimized by the steepest descent approach, and then went through and NVT and subsequently an NPT simulation. The resultant models were subjected to simulated annealing where the system temperature was ramped from 300 to 800K, then cooled back to 300K in a periodic manner within 200 pico-seconds for 5×10^4 iterations over the course of a 20 nano-second simulation (namely 1×10^7 total time steps), in order to allow generated models to escape from local energy minima and enhance equilibration. Simulated structures were saved every 1000 time steps, yielding 10,000 saved structures throughout the simulation. LINCS algorithm was used to constrain the bond lengths, and the particle mesh Ewald method was employed to account for long range electrostatic interactions. Temperature was maintained using the modified Berendsen method.[27] After performing 20 ns of simulation, every 100th 300K structure was exported, yielding 100 models for each charge assignment, 4,000 structures in total. From these structures, compact models that both agreed well with experimental monomeric CCS measurements (as calculated using a scaled projection approximation (PA) method in MOBCAL, with scaling factors described in detail in the text below referenced against the trajectory method (TM), also as implemented in MOBCAL), and possessed the lowest final potential energy with respect to the simulation, were highlighted to construct the artificial docking model.[28, 29] This step yielded 25 monomeric Con A models (as not every simulation would generate theoretical CCS that agree well with experimental measurements) with different extent of unfolding, and

were further docked back onto an intact Con A trimer (indicated by red arrow) using HEX 8.0.0 software package.[30] The CCS of resultant artificial tetramer models were computed using the scaled PA method as mentioned above. These CCS values were compared with experimental tetrameric Con A CCS at various collisional energy levels, as presented on the CIU fingerprints. According to figure 3-7, a scaling factor of 25.38% was used throughout the modeling process, and such scaling factor was generated from thirteen Con A monomer model structures having different degrees of unfolding, ranging from monomers extracted from the tetramer X-ray structure, to near string-like unfolded models.

3.3 Results and Discussion

3.3.1 Depicting the mechanism of multi-protein unfolding using charge state as the only variable

Denatured Con A samples were subjected to IM-MS analysis, resulting in signals for highly unfolded monomeric Con A features detected at multiple charge states, ranging from 9^+ to up to 20^+, as shown in figure 3-3(left). The significant elongation of the monomer signal in the drift time dimension indicates the subunits are highly unfolded. Figure 3-3 (right panel) shows the centroid CCS values for unfolded Con A as function of monomer charge state were. The data show that that for most Con A charge states larger than 10^+, multiple conformers families are observed, with four broader groupings of conformations visible over all charge states analyzed.

These groupings are visible as pseudo-linear trends within the data, with the third of these trends, starting at 2800 $Å^2$, representing a broader array of CCS values than the other three. By comparing the upper and lower limit of the CCS distribution detected, our data shows that Con A monomer can be enlarged by more than a factor of 2 through charge amplification from 9^+ to 18^+.
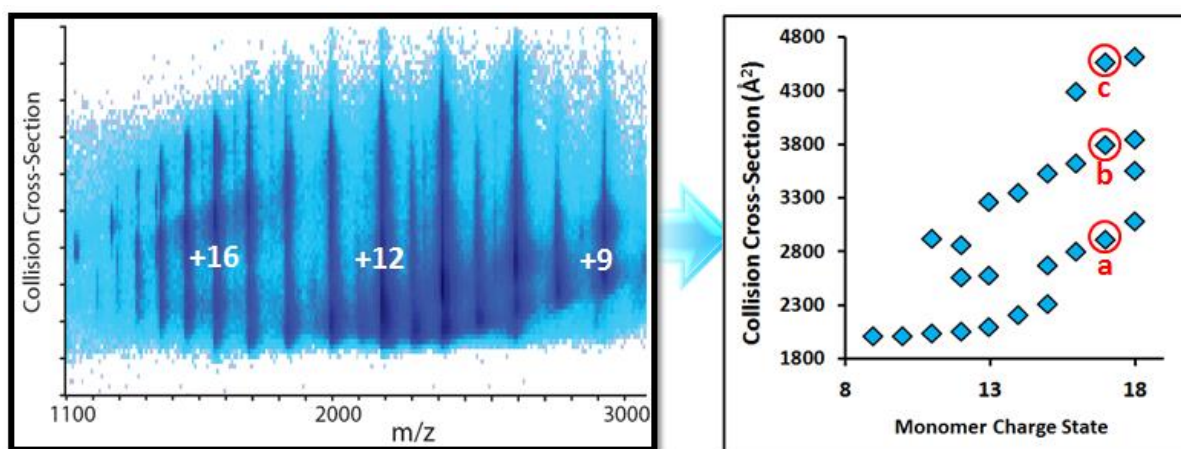
Figure3-3.(Left) IM-MS data for denatured Con A monomers reveals highly unfolded monomeric structures at various charge states. **(Right)** plot of Con A ion CCS as a function of protein charge states. Four distinct groupings of conformational families are observed, with representative CCS marked with a, b and c highlighted in red for discussion in figure 3-4.

Following the Con A monomer data acquisition, experimental CCS values were used to filter low energy structural generated through simulated annealing. Model monomer structures, having

different degrees of unfolding, were then docked back onto the remaining trimer structure, yielding completed tetramer models containing a single unfolded subunit. As highlighted in Figure 3-3, CCSs corresponding to the conformer families marked a, b and c were selected to filter monomer models. The corresponding representative low energy models that match the CCSs of a, b and c, are shown in Figure 3-4 (left panel). The CCS values of the resultant tetramer models, also shown in Figure 3-4(right panel), were calculated using the scaled PA method as described above.
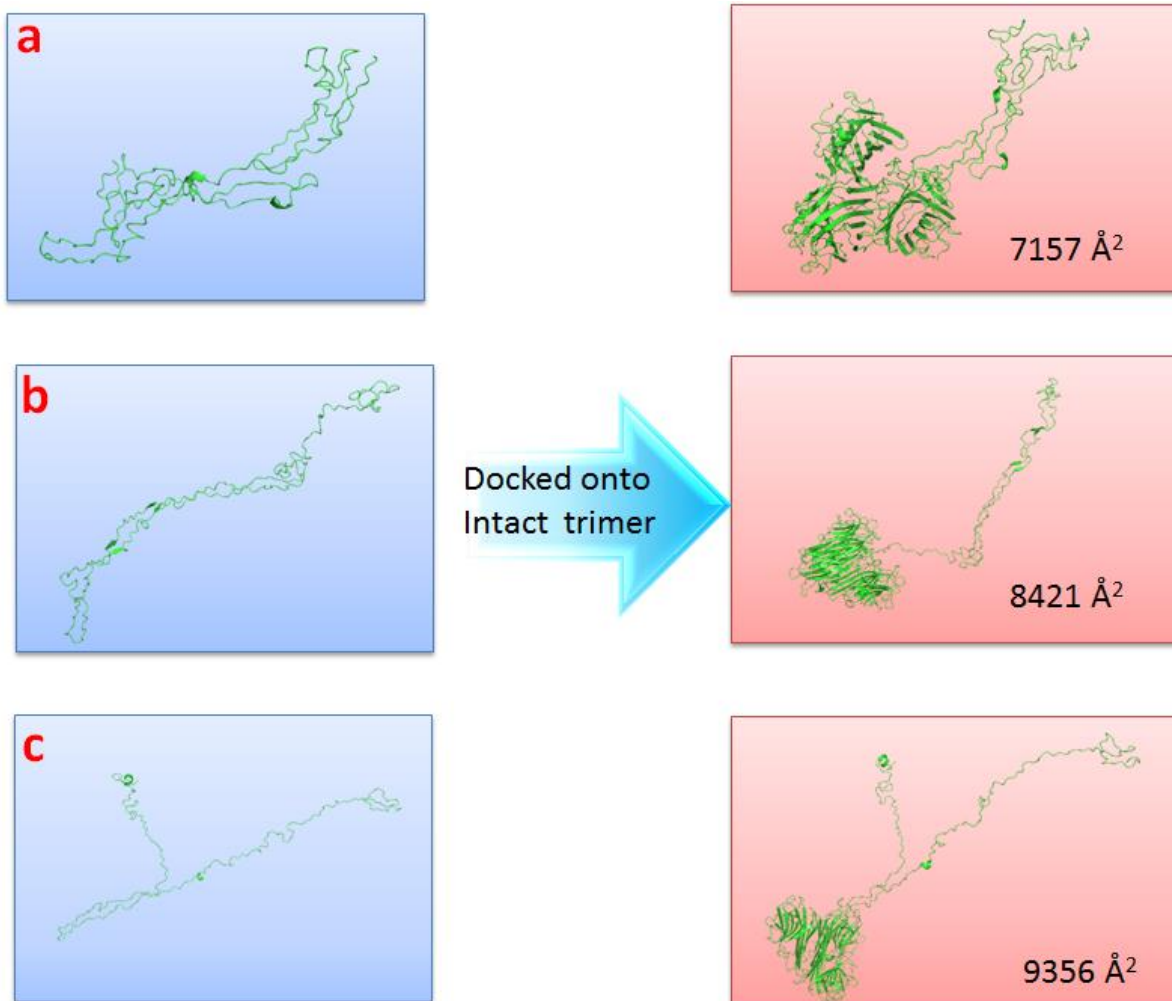
Figure 3-4. Model monomeric structures a, b and c and their docked tetrameric models. **(Left)** Atomic model structures corresponding to CCS values marked as a, b and c in figure 3-3 at 17^+ charge state are shown, exhibiting increasing degree of unfolding. **(Right)** The resultant tetramer model with docked a, b and c monomers onto the intact Con A trimer.

As noted above, Con A monomer Coulombic unfolding yields four total CCS groupings, categorized by linear trends observed in Figure 3-3. Figure 3-5 compares these four groupings of Con A monomer conformations (left panel) to CIU fingerprint for Con A tetramer acquired for the 20^+ charge state (right panel). Remarkably, four main conformational transitions are observed for the intact Con A tetramer , labeled as feature I, II, III (a and b) and IV; highly similar to the Coulombically unfolded Con A monomer data (color coded, in Figure 3-5). It is worth noting that the complexity of feature III in the intact tetramer CIU data is also reflected in Con A monomer data, as the analogous CCS grouping exhibits the broadest array of structural forms. In addition, computational models constructed from docking low energy Con A monomer models that match CCS values for monomer groups I, II, III, or IV onto folded trimers, agree well with the correspondingly labeled CIU feature for the intact tetramer. . Such a trend was not specific to only the 20^+ Con A tetramer charge state, data shown in Figure 3-6 exhibits a similar correlation between denatured Con A CCS values and CIU transitions for the 21^+ intact tetramer. Overall, these results strengthen the general hypothesis linking multi-protein unfolding events to charge migration events from the complex to a single monomer.
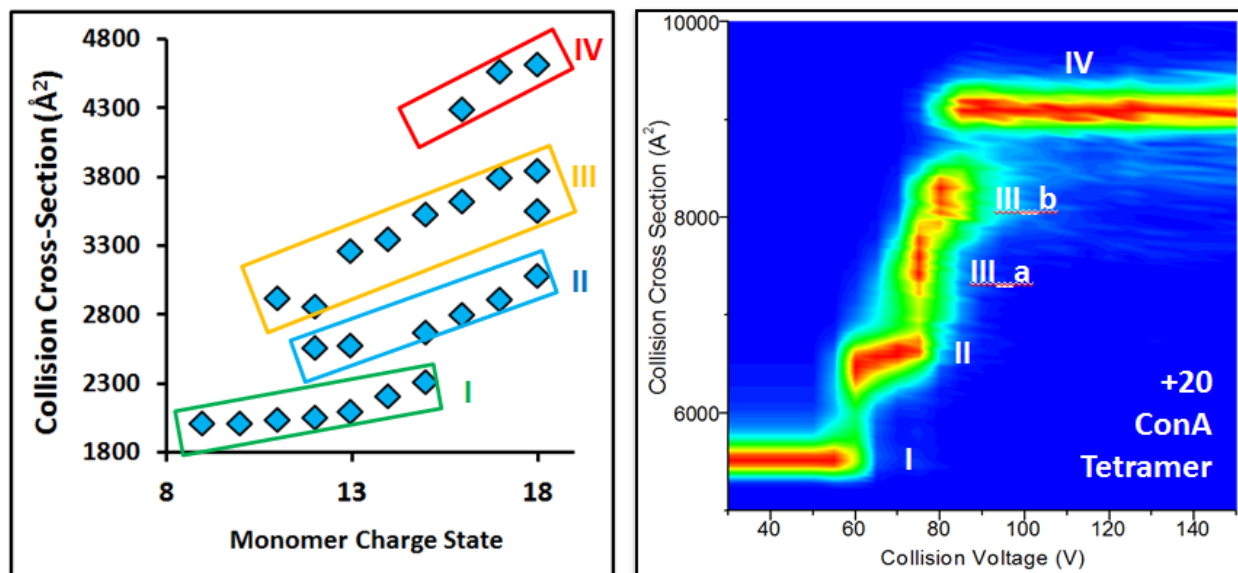
Figure 3-5.(Left) A Plot of Con A monomer CCS as a function of protein charge state, in comparison to 20^+ Con A tetramer. The CCS distributions observed are grouped into four families and labeled as I, II, III and IV (highlighted in green, blue, yellow and red, respectively) **(Right)** CIU fingerprint for the 20^+ Con A tetramer charge state. Four discrete structural families are presented, and labeled as feature I, II, III (including III-a and III-b) and IV.

In order to comprehensively calculate the models output from the workflow shown in Figure 3-2, rapid, scaled projection type calculations needed to be performed. Despite this throughput necessity, typical scaling factors applied commonly to such calculation output for IM-MS data analysis had not been applied to asymmetrically-unfolded proteins, and instead validated on compact, globular protein systems [28]. Furthermore, such scaling factors are only appropriate when comparing X-ray data with IM-MS output, and do not provide information on the potential

gas-phase structures formed [28, 31]. As such, we undertook an effort to generate a new scaling factor that effectively converts PA-type CCS estimates into TM values, which are most appropriate for assigning models to potential gas-phase structures. As shown in Figure 3-9, the CCSs of above docking models were computed with a scaling factor of 18.4% (for 20^+) and 25.8% (for 21^+), respectively, derived, in part, by computing the CCS of thirteen selected models that represent different levels of subunit unfolding, using both the PA and TM methodology. A linear trend line was generated through regression analysis having an excellent correlation coefficient ($R^2 = 0.9346$), resulting in a slope matching the upper end of the conversion factors that scaled from CIU fingerprints for the 20^+ and 21^+ data. These scaling coefficients were generated based on the assumption that the observed most compact native-like structure should represents the CCS of Con A's crystal structure, and are thus only valid for unfolded Con A structures. Because the TM approach requires ~10 hours to finish a CCS calculation for a single monomer model and >30 hours of CPU time for a single tetramer model), and PA calculations take only seconds to complete per model structure, the scaling factors described here enhance the throughput of our modeling workflow by a factor of 4 to 5 orders of magnitude.
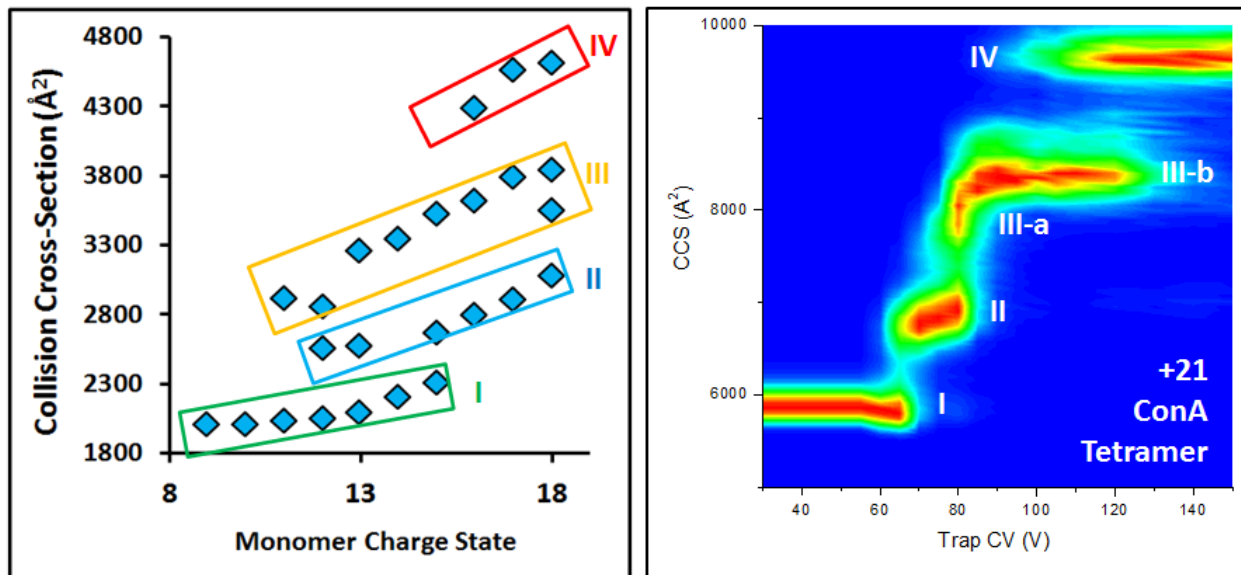
Figure 3-6.(Left) A plot of Con A monomer CCS as a function of protein charge state, in comparison to 21^+ Con A tetramer. The CCS distributions are further categorized into four groupings, and labeled as I, II, III and IV (highlighted in green, blue, yellow and red, respectively) **(Right)** A CIU fingerprint for the 21^+ Con A tetramer. Four discrete structural families are presented, and labeled as I, II, III (including III-a and III-b) and IV.
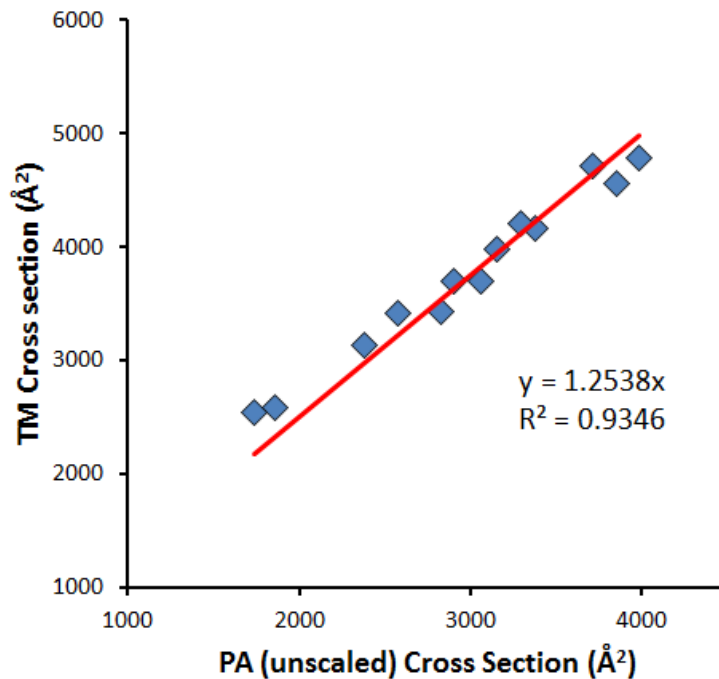
Figure 3-7. Thirteen Con A monomer model structures having different degrees of unfolding, ranging from monomers extracted from the tetramer X-ray structure, to near string-like unfolded models. CCSs for each model are shown, calculated by both the TM and PA (unscaled) approaches. A linear regression model (shown in red trend line) reveals a scaling factor of 25.38%.

3.3.2 Investigating the charge pooling effect and intactness of ligand binding pocket

. Based on the strong agreement between Coulombic monomer unfolding and tetramer CIU for Con A, we worked to leverage the models discussed above in order to quantify the amount of charge migration required to generate each unfolded state observed in Con A tetramer CIU. Figure 3-8 illustrates schematically the charge transfer values estimated from our model for

20$^+$tetramer Con A ions. This model assumes that charges are evenly distributed over each Con A subunit prior to CIU, so that each subunit possesses 5 positive charges on average. Our models suggest, surprisingly, that it takes between 6 and 11 (30% to 55% of total charge) extra charges transferred onto one Con A subunit to trigger the first CIU transition observed in the, 20$^+$ Con A fingerprint (*i.e.,* I→ II). However, the II→III transition requires the transfer of only 3 additional charges on average. Detailed values for the charges required to affect additional CIU transitions in Con A are delineated in table 3-1. These striking results are in good agreement with typical CID experiments for multi-protein ions, where the highly unfolded ejected monomer tends to take more than 50% of the total charge present on the precursor ion prior to activation. Such trends were similar for the 21$^+$ Con A models constructed, which are listed in table 3-2.
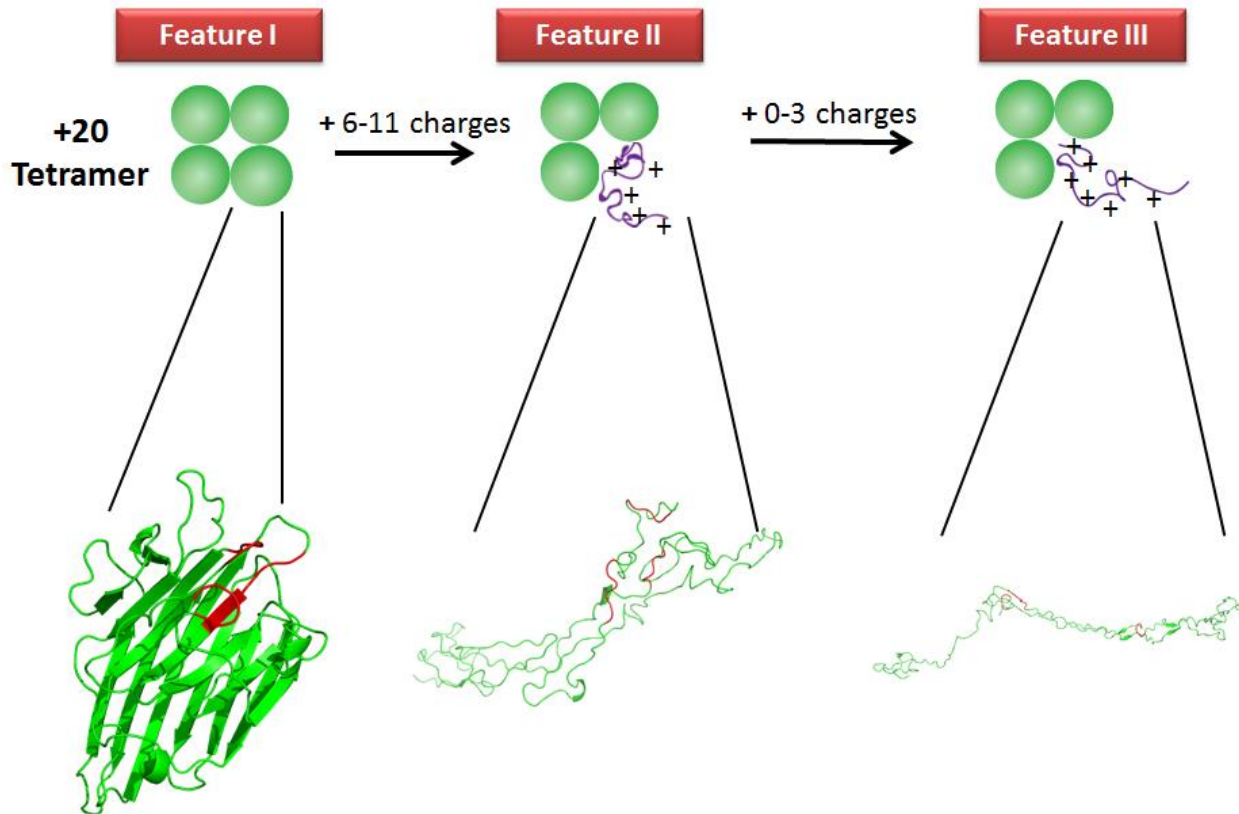
Figure 3-8.(Top) A schematic diagram outlining the charge transfer required to initiate CIU transitions in the 20^+ Con A tetramer**(Bottom)** Con A models showing detailed local structure at the tri-mannose sugar binding site (red) as each unfolding state is achieved.

Increasingly, CIU assays are being used to evaluate the consequences of small molecule binding in the context of target structure and stability, often using features achieved at large activation voltages to differentiate binding modes or structural features [18, 22, 32, 33]. In order to evaluate the integrity of the sugar binding pocket during Con A CIU, we highlight the tri-mannosyl sugar binding site within the Con A monomer models (red) shown in Figure 3-8.[34] Our modeling

suggests that the I→II CIU transition likely destroys most of the binding pocket structure, and that all local structure is disrupted in the transition II→III, as the unfolded monomer achieves a nearly string-like form at this stage. These observations forecasts that any ligand-relevant information in Con A CIU is most likely to be found in transition I→II, and this prediction is strongly supported by our previous analysis of CIU data where evidence of cooperative ligand stabilization was extracted exclusively from this transition.

Table 3-1. Detailed tabular data of charge migration summary for 20^+ Con A.

Con A 20^+, 5 charges/monomer		
# charges needed to achieve	**% charge transferred**	**feature**
11-16 (6-11 transferred)	(30%-55%)	II
14-18 (9-13 transferred)	(45%-65%)	III
>16	(>60%)	IV

Charges are assumed to be evenly distributed on each subunit in native state.

Table 3-2. Detailed tabular data of charge migration summary for 21^+ Con A.

Con A 21^+, 5.25 charges/monomer		
# charges needed to achieve	**% charge transferred**	**feature**
11-15 (5.75-9.75 transferred)	(27%-46%)	II
15-18 (9.75-12.75 transferred)	(46%-61%)	III
>17	(>56%)	IV

Charges are assumed to be evenly distributed on each subunit in native state.

3.4 Conclusions

Here, the underlying mechanism of multi-protein unfolding is systematically studied by IM-MS and MM approaches. The sugar-binding tetrameric protein Con A is used as a model system for this process, mostly due to its well-studied sequence and structure, in both solution and in the gas phase [18, 19]. We observe a strong positive correlation between monomeric Coulombic unfolding and tetrameric CIU. This provides strong evidence that multi-protein unfolding events are initiated primarily by charge migration from the complex to a single monomer. In addition, we investigate how many charges are transferred in order to achieve the various features observed in Con A CIU. Surprisingly, our models suggest that the majority of the charge transfer (up to 55%) takes place during the first of 3 main unfolding transitions in Con A tetramer, and that this event likely disrupts most of the sugar binding pocket structure within the protein. All these results are supported by our previous studies of Con A tetramer, and provide critical, predictive insights for understanding the fundamental mechanism of multi-protein unfolding.

To further evaluate multiprotein unfolding, we plan to expand this work to other multi-protein systems, ranging from dimeric proteins like β-lactoglobulin, to other tetramers like avidin, and transthyretin, as well as and larger protein multimers. Experimental and theoretical validation of our basic workflow and scaling factors developed here from multiple protein systems will provide broader insights into protein complex CIU.

3.5 References

1. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.

2. Heck, A.J. and R.H. van den Heuvel, *Investigation of intact protein complexes by mass spectrometry.* Mass spectrometry reviews, 2004. **23**(5): p. 368-389.

3. Congreve, M., C.W. Murray, and T.L. Blundell, *Keynote review: Structural biology and drug discovery.* Drug discovery today, 2005. **10**(13): p. 895-907.

4. Hofstadler, S.A. and K.A. Sannes-Lowery, *Applications of ESI-MS in drug discovery: interrogation of noncovalent complexes.* Nature Reviews Drug Discovery, 2006. **5**(7): p. 585-595.

5. Niu, S., J.N. Rabuck, and B.T. Ruotolo, *Ion mobility-mass spectrometry of intact protein–ligand complexes for pharmaceutical drug discovery and development.* Current Opinion in Chemical Biology, 2013. **17**(5): p. 809-817.

6. Ruotolo, B.T., et al., *Evidence for macromolecular protein rings in the absence of bulk water.* Science, 2005. **310**(5754): p. 1658-1661.

7. Ruotolo, B.T., et al., *Ion mobility–mass spectrometry analysis of large protein complexes.* Nature Protocols, 2008. **3**(7): p. 1139-1152.

8. Wang, W., E.N. Kitova, and J.S. Klassen, *Influence of solution and gas phase processes on protein-carbohydrate binding affinities determined by nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry.* Analytical chemistry, 2003. **75**(19): p. 4945-4955.

9. Liu, L., et al., *Hydrophobic protein− ligand interactions preserved in the gas phase.* Journal of the American Chemical Society, 2009. **131**(44): p. 15980-15981.

10. Hyung, S.-J., C.V. Robinson, and B.T. Ruotolo, *Gas-phase unfolding and disassembly reveals stability differences in ligand-bound multiprotein complexes.* Chemistry & biology, 2009. **16**(4): p. 382-390.

11. Barrera, N.P. and C.V. Robinson, *Advances in the Mass Spectrometry of Membrane Proteins: From Individual Proteins to Intact Complexes*, in *Annual Review of Biochemistry, Vol 80*, R.D. Kornberg, et al., Editors. 2011, Annual Reviews: Palo Alto. p. 247-271.

12.   Aebersold, R. and D.R. Goodlett, *Mass spectrometry in proteomics.* Chemical Reviews, 2001. **101**(2): p. 269-295.

13.   Sobott, F., et al., *A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies.* Analytical Chemistry, 2002. **74**(6): p. 1402-1407.

14.   Felitsyn, N., E.N. Kitova, and J.S. Klassen, *Thermal decomposition of a gaseous multiprotein complex studied by blackbody infrared radiative dissociation. Investigating the origin of the asymmetric dissociation behavior.* Analytical chemistry, 2001. **73**(19): p. 4647-4661.

15.   Jurchen, J.C. and E.R. Williams, *Origin of asymmetric charge partitioning in the dissociation of gas-phase protein homodimers.* Journal of the American Chemical Society, 2003. **125**(9): p. 2817-2826.

16.   Jurchen, J.C., D.E. Garcia, and E.R. Williams, *Further studies on the origins of asymmetric charge partitioning in protein homodimers.* Journal of the American Society for Mass Spectrometry, 2004. **15**(10): p. 1408-1415.

17.   Bittiger, H. and H.P. Schnebli, *Concanavalin A as a Tool.* Wiley, 1976.

18.   Niu, S. and B.T. Ruotolo, *Collisional unfolding of multiprotein complexes reveals cooperative stabilization upon ligand binding.* Protein Science, 2015.

19.   Bhattacharyya, L. and C.F. Brewer, *Isoelectric focusing studies of concanavalin A and the lentil lectin.* Journal of Chromatography A, 1990. **502**: p. 131-142.

20.   Fegan, S.K. and M. Thachuk, *A charge moving algorithm for molecular dynamics simulations of gas-phase proteins.* Journal of Chemical Theory and Computation, 2013. **9**(6): p. 2531-2539.

21.   Han, L., et al., *Affinities of recombinant norovirus P dimers for human blood group antigens.* Glycobiology, 2013. **23**(3): p. 276-285.

22.   Rabuck, J.N., et al., *Activation state-selective kinase inhibitor assay based on ion mobility-mass spectrometry.* Analytical chemistry, 2013. **85**(15): p. 6995-7002.

23.   Sterling, H.J., et al., *Electrothermal supercharging of proteins in native electrospray ionization.* Analytical chemistry, 2012. **84**(8): p. 3795-3801.

24.   Martin, A.C., *Mapping PDB chains to UniProtKB entries.* Bioinformatics, 2005. **21**(23): p. 4297-4301.

25.   Pronk, S., et al., *GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit.* Bioinformatics, 2013: p. btt055.

26. Oostenbrink, C., et al., *A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force‐field parameter sets 53A5 and 53A6.* Journal of computational chemistry, 2004. **25**(13): p. 1656-1676.

27. Berendsen, H.J., et al., *Molecular dynamics with coupling to an external bath.* The Journal of chemical physics, 1984. **81**(8): p. 3684-3690.

28. Benesch, J.L. and B.T. Ruotolo, *Mass spectrometry: come of age for structural and dynamical biology.* Current opinion in structural biology, 2011. **21**(5): p. 641-649.

29. Larriba, C. and C.J. Hogan Jr, *Ion mobilities in diatomic gases: measurement versus prediction with non-specular scattering models.* The Journal of Physical Chemistry A, 2013. **117**(19): p. 3887-3901.

30. Ghoorah, A.W., et al., *Protein docking using case‐based reasoning.* Proteins: Structure, Function, and Bioinformatics, 2013. **81**(12): p. 2150-2158.

31. Bush, M.F., et al., *Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology.* Analytical chemistry, 2010. **82**(22): p. 9557-9565.

32. Zhong, Y., L. Han, and B.T. Ruotolo, *Collisional and Coulombic Unfolding of Gas‐Phase Proteins: High Correlation to Their Domain Structures in Solution.* Angewandte Chemie, 2014. **126**(35): p. 9363-9366.

33. Laganowsky, A., et al., *Membrane proteins bind lipids selectively to modulate their structure and function.* Nature, 2014. **510**(7503): p. 172-175.

34. Loris, R., et al., *A structure of the complex between concanavalin A and methyl-3, 6-di-O-(α-D-mannopyranosyl)-α-D-mannopyranoside reveals two binding modes.* Journal of Biological Chemistry, 1996. **271**(48): p. 30614-30618.

Chapter 4

Ion Mobility-Mass Spectrometry Reveals Evidence of Specific Complex Formation between Human Histone Deacetylase 8 and Poly-r(C)-binding Protein 1

4.1 Introduction

Histone deacetylases (HDACs) play a key role in regulating transcription and many other cellular processes by catalyzing the hydrolysis of ε-acetyle-lysine residues.[1, 2]Over the past decade, tremendous interest has been centered on these enzymes due to their promise as targets for therapeutic development in the context of a variety of diseases, including asthma, cancer and inflammatory lung diseases.[3-6] Understanding the fundamental role of histone acetylation and deacetylation in the basic processes surrounding gene expression is thus critically important in treating various diseases, as well as for our understanding of basic biochemistry.[7, 8]

There are 18 known HDAC enzymes divided phylogenetically into four classes: class I (HDAC1-3, and HDAC8), class II (HDAC4-7 and HDAC9-10), class III (sirtuins 1-7), and class IV (HDAC11).[3] Among these, histone deacetylase 8 (HDAC8) serves as our scientific focus in this work (Figure 4-1). This enzyme has been directly linked to acute myeloid leukemia and the development of the actin cytoskeleton via its native enzymatic activity.[6, 9] Upon its discovery, HDAC8 was validated as a Zn(II)-dependent metalloenzyme;[10]however, *in vitro* activity and binding affinity assays suggest that Fe could also serve as a native metal cofactor *in vivo*.[11]A recent systematic investigation suggests that HDAC8 can be activated by either Zn(II) or Fe(II), depending on the local cellular environment of the enzyme.[11]

In cells, the majority of class I HDACs execute their biological function as a part of a multi-protein complex[12-14]and it is proposed that HDAC8 may operate similarly, forming complexes that alter the metal selectivity to adjust the activity of HDAC8. The mechanism by which HDAC8 recognizes and incorporates the cognate metal ions (zinc or iron) in the cell is largely unknown. Metal incorporation for a number of metallo-proteins is facilitated by metallo-chaperones[15, 16]and while there are currently no zinc-specific metallo-chaperones identified, several potential iron-specific metallo-chaperones are being investigated for roles in iron homeostasis, particularly in the assembly of Fe-S clusters.[17-20] In 2008, the Philpott lab reported that poly r(C)-binding protein 1 (PCBP1), a cellular iron storage protein, can function as a cytosolic iron chaperone in the delivery of iron to ferritin.[21]

.

The first global protein interaction network for 11 HDACs in human CEM T-cells (leukemic cell line) revealed HDAC8 interacting with multiple members of the cohesin complex[12] associated with sister chromatid segregation during mitosis.[21] Moreover, this analysis suggests that HDAC8 may also interact with the PCBP family of iron-metallo-chaperones, despite lingering controversy on the subject..[12]

The PCBP family consists of four homologous RNA-binding proteins (PCBP2, PCBP3, and PCBP4) that are ubiquitously expressed in the mammalian cytosol and nucleus.[22]Human PCBP1 was recently identified as an iron chaperone for human ferritin (Figure 4-1)[12] and functional assays in yeast indicate that PCBP1 facilitates the incorporation of iron into ferritin through a direct protein-protein interaction.[23]Most recently, *in-vivo* co-immunoprecipitation assays revealed the formation of the HDAC8-PCBP1 complex in cells, indicating that PCBP1 and HDAC8 are physically interacting independent of cellular iron concentrations, although the specificity and strength of this interaction has yet to be determined.
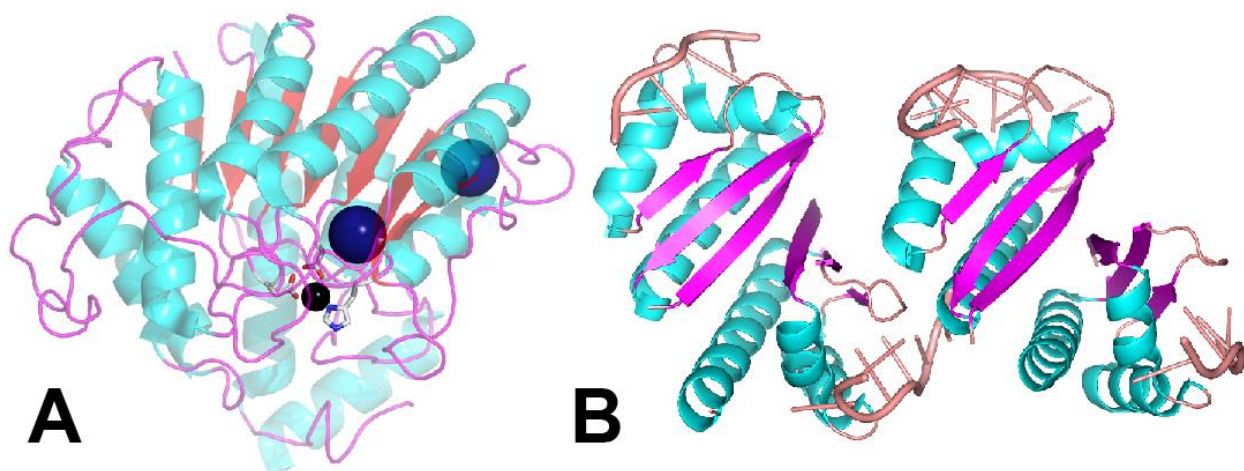
Figure4-1. Crystal structure of HDAC8 and PCBP1 **(A)** Crystal structure of HDAC8. (PDB: 2V5X) Black sphere indicates divalent metal (Zn or Fe) binding, the two blue spheres indicates monovalent metal (K or Na) binding **(B)** Truncated crystal structure of PCBP1 (PDB: 3VKE). The presented structure represents 96% of the full sequence, with all four biological assembly units in the model.

Over the past two decades, ESI-MS has emerged as a key technology for the identification and quantification of protein-ligand interactions *in vitro*.[24-27]. High throughput assays of binding affinity of protein-small ligand complexes [26, 28-30] have been applied to the investigation of systems which are not accessible by conventional techniques, such as protein-glycolipid systems.[29]The major advantages of the ESI-MS approach includes simplicity (no labeling or immobilization required), speed (data can be acquired in minutes), and selectivity (protein assemblies and mixtures can be further analyzed by techniques coupled to MS, for example ion mobility spectrometry).[27]The introduction of ion mobility (IM), which separates ions according to their size and charge on the millisecond timescale, in tandem with MS further enables the acquisition of such binding information by highlighting conformer-specific small

molecule interactions [31, 32], as well as offering an enhanced ability to deconvolute signals for target oligomers [33-36].

While ESI-MS has proven exceptionally useful for quantification of protein-small molecule K_D values, there are many key challenges to the wide application of the methodology. For example, one of the key assumptions in the interpretation of the results referenced above is that the apo and bounded protein possess nearly identical spray and detection efficiencies. The validity of such an assumption is strongly supported by protein-small molecule K_D data currently available, as the overall accessible surface area of the binding target does significantly upon small molecule attachment.[26, 28] In contrast, very few applications have been made for protein-protein K_D estimations by direct MS method. Direct ESI-MS measurements have led to accurate self-dissociation constants for both the β-lactoglobulin and hemoglobin homo-dimer,[37]as well as the complexes involved in the Hsp 90 interaction network.[38]Quantifying protein binding constants by ESI-MS for such stable assemblies is already challenging; however, moving to the analysis of weaker complexes necessarily leads to a background of artifact protein complex borne from the ESI process itself. When protein concentrations are increased in order to favor the formation weaker protein-protein complexes, the ESI droplet formation process can capture more than one biological unit, thus forcing the formation of non-specific complexes through the solvent evaporation and ionization process. As interest in weak protein-protein interactions, and their role in formation of transient signaling complexes, intensifies, it is clear that there is a need for new ESI-MS strategies for the detection and quantification of such assemblies.

Here we explore, for the first time, the binding affinity of a protein-protein system by ESI-MS. We conduct systematic IM-MS investigations of the HDAC8 and PCBP1 complex, in conjunction with Co-IP, metal binding activity and kinetics studies performed by our collaborators, to address three major questions: 1) Is the interaction between HDAC8 and PCBP1 specific? **2)** Can we quantify the strength of this interaction via the direct ESI-MS method? **3)** How does metal association (Zn^{2+} or Fe^{2+}) affect the binding behavior of this complex? Our ESI-MS data provides strong evidence that such assemblies are indeed specific, their binding strength can be quantified by ESI-MS if specific control experiments are performed, and that metal ions alter the binding strength of the HDAC8:PCBP1 assembly to a small, yet measurable extent. This data is discussed both in the specific context of HDAC8 activity, as well as the ability of the ESI-MS techniques described here to broadly quantify the binding strengths of weak protein-protein interactions.

4.2 Experimental Methods

4.2.1 Expression and purification of HDAC8 and PCBP1

Recombinant His_6-HDAC8 was expressed in BL21(DE3) *E. coli* transformed with pHD4 and purified as previously described and concentrated to 2-12 mg/mL . Metal-free HDAC8 was generated by dialyzing purified HDAC8 twice into 500 mL of 25 mM MOPS (pH 7.0), 1 mM EDTA for 12-14 hr at 4 °C, followed by buffer exchange into 25 mM MOPS (pH 7.5), 0.1 mM EDTA and finally 25 mM MOPS (pH 7.5) for 12-14 hr at 4 °C. Where necessary, anaerobic

conditions were achieved using either the captair pyramid glove-bag filled with argon or nitrogen or an anaerobic chamber (Coy, Grass Lake, MI)

Chemically competent BL21(DE3) cells were transformed with pCDF encoding His6SUMO-tagged PCBP1.[22] Starter cultures (10 mL) were inoculated by addition of a single colony and grown in LB media with 100 μg/ μl streptomycin for 5 hr at 37 °C and 250 rpm. The starter culture was diluted (1:1000) into 2 L of LB with appropriate antibiotics in a 6 L culture flask. The temperature was decreased to 15 °C at induction. Expression was induced by addition of 0.1 mM IPTG, and the cells were incubated overnight at 15 °C. Cells from the 2 L growth were harvested, re-suspended in 30 ml buffer A (20 mM Tris [pH 7.9], 250 mM NaCl, 30 mM imidazole, 10% glycerol, and 2.5 mM TCEP) with one EDTA-free protease inhibitor cocktail tablet (Roche). Resuspended cells were lysed using a microfluidizer. The extract was centrifuged at 6000 g for 30 min and the supernatant loaded onto a nickel column. Protein fractionation was carried out using a linear gradient in buffer A from 30 mM to 500 mM imidazole, to buffer B (20 mM Tris [pH 7.9], 250 mM NaCl, 500 mM imidazole, 10% glycerol, 2.5 mM TCEP), with PCBP1 eluting at 110–230 mM imidazole. The His6 -SUMO tag was cleaved by incubating with 1 unit/μl Saccharomyces cerevisiae SUMO protease (Life Technologies) overnight at 4 °C. The protein was buffer exchanged using dialysis with buffer A before passing over the nickel column a second time to separate the tag from the untagged protein. The protein was dialyzed overnight at 4 °C first against buffer A containing 1 mM EDTA to remove metals and then against buffer A 95 to remove EDTA. Finally, the protein was fractionated on a PD-10 column to remove any remaining EDTA. Apo-PCBP1 was concentrated to 100 – 200μM, flash-frozen and stored at ~80°C, prior to IM-MS investigations.

116

4.2.2 IM-MS experiments

All experiments were performed on a Synapt G2 ESI quadrupole-ion mobility- time-of-flight (Q-IMS-TOF) mass spectrometer (Waters, Milford, MA), equipped with a nanoflow ESI source. Mass spectra were collected under positive ion mode using cesium iodide for calibration. A capillary voltage of 1.68kV was applied and sampling cone voltage and source temperature maintained at 50V and 20 ^{o}C during signal acquisition. Backing pressure was set at 7-8 mbar. Quadrupole profile was set as: M_1*= 2000, M_2*= 5000, dwell $time_1$= 2%, ramp $time_{1-2}$= 98%, to ensure near uniform transmission efficiency. To optimize the mass resolution, trap collision voltages ranging from 30-50V were applied, with argon collision gas at a pressure of 2.56 $\times10^{-2}$ mbar. The ion mobility measurements were carried out using N_2 as the mobility buffer gas, at a pressure of 3.5 mbar. Data acquisition and processing were carried out using MassLynxV4.1 software. Protein samples were subjected to a 500 mM ammonium acetate buffer exchange protocol, to produce a final concentration range of 2 -18 μM, prior to MS investigations. Positive control experiments were carried out using Ferredoxin-NADP$^+$ reductase and Ferredoxin protein (Sigma, St. Louis, MO, USA). For metal substitution experiments, either 5 μM $Zn(NO_3)_2$ or 5 μM $(NH_4)_2Fe(SO_4)_2$ with 250 μM ascorbic acid were used as a source of Zn^{2+} or Fe^{2+}, respectively.

4.2.3 Binding affinity (K_D) calculation by ESI-MS

The binding affinity, often used as the dissociation constant K_D, for a given protein-ligand interaction, in our case is interaction between two different proteins P_1 and P_2, is determined from the ratio (R) of the total abundance (Ab) of bounded and free protein ions, as measured by ESI-MS for solutions of known initial concentrations of protein (the larger protein $[P_1]_0$) and ligand (namely the smaller protein $[P_2]_0$). For a typical 1 to 1 protein-protein interaction as shown in equation (1), K_D is calculated using equation (2):

$$P_1 P_2 \leftrightharpoons P_1 + P_2 \qquad\qquad (1)$$

$$K_D = \frac{[P_1]_{eq}\,[P_2]_{eq}}{[P_1 P_2]_{eq}} = \frac{[P_2]_0}{R} - \frac{[P_1]_0}{1+R} \qquad\qquad (2)$$

Where R is given by equation (3):

$$\frac{[P_1 P_2]_{eq}}{[P_1]_{eq}} = \frac{Ab(P_1 P_2)}{Ab(P_1)} = R \qquad\qquad (3)$$

4.3 Results and Discussion

4.3.1 Characterization of HDAC8 and PCBP1 interactions

Figure 4-2 shows the mass spectra acquired from a solution of HDAC8 and PCBP1 and demonstrates the presence of a new species of approximately 80kDa, attributed to the formation of a HDAC8 - PCBP1 heterodimer (42.5kDa + 37.5kDa) which remains stable under mildly collisional activation conditions. The identity of the HDAC8 - PCBP1 heterodimer (H-P) was confirmed by MSMS of the 18^+ H-P assembly precursor ion at m/z 4452. In addition to the H-P complex, HDAC8-HDAC8 (H-H) and PCBP1-PCBP1 (P-P) homo-dimer, even trimeric complexes at higher PCBP1 to HDAC8 molar ratio were observed. (as shown in Figure 4-2, blue shaded region).
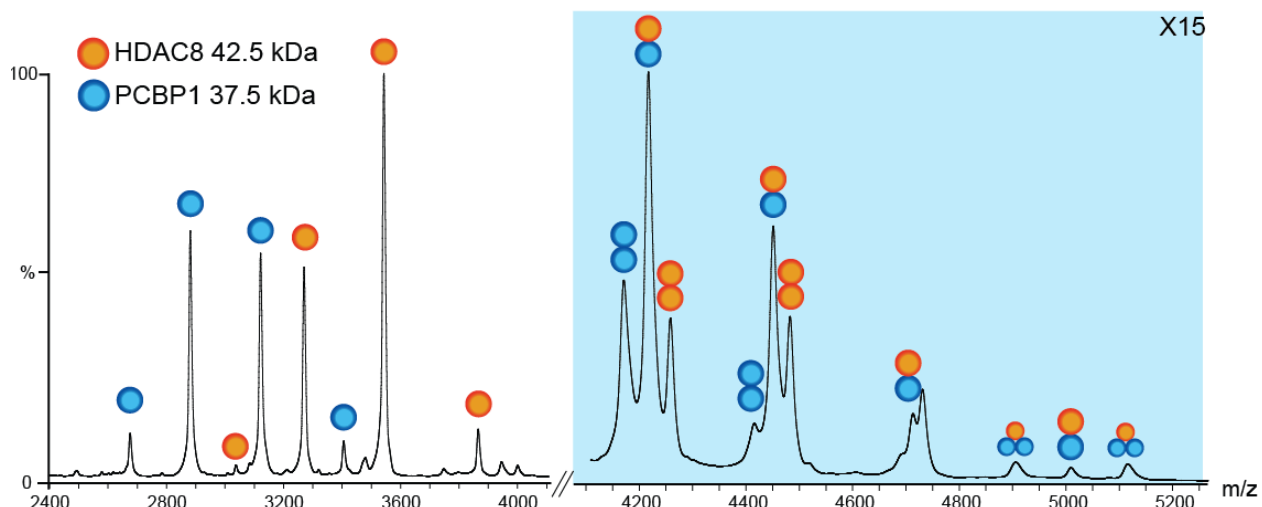
Figure4-2. Native MS data of HDAC8 incubated with PCBP1. Individual protein ions (orange for HDAC8, blue for PCBP1) and the HDAC8-PCBP1 complex are observed. Low levels of homo-dimer and trimer signals are also evident but these dissociate quickly upon collisional activation.

To rule out the possibility that the H-P dimer is a gas phase artifact, carbonic anhydrase II (CAII, molecular mass 29kDa) was employed as a negative control, since this protein is well studied

and does not bind with HDAC8 or PCBP1. Figure 4-3 shows the mass spectra from a 1:1 solution of CAII in the presence of PCBP1 and HDAC8, demonstrating that no CAII-PCBP1 or CAII-HDAC8 hetero-dimer complexes are generated and indicating that HDAC8 binds to PCBP1 in a specific manner.
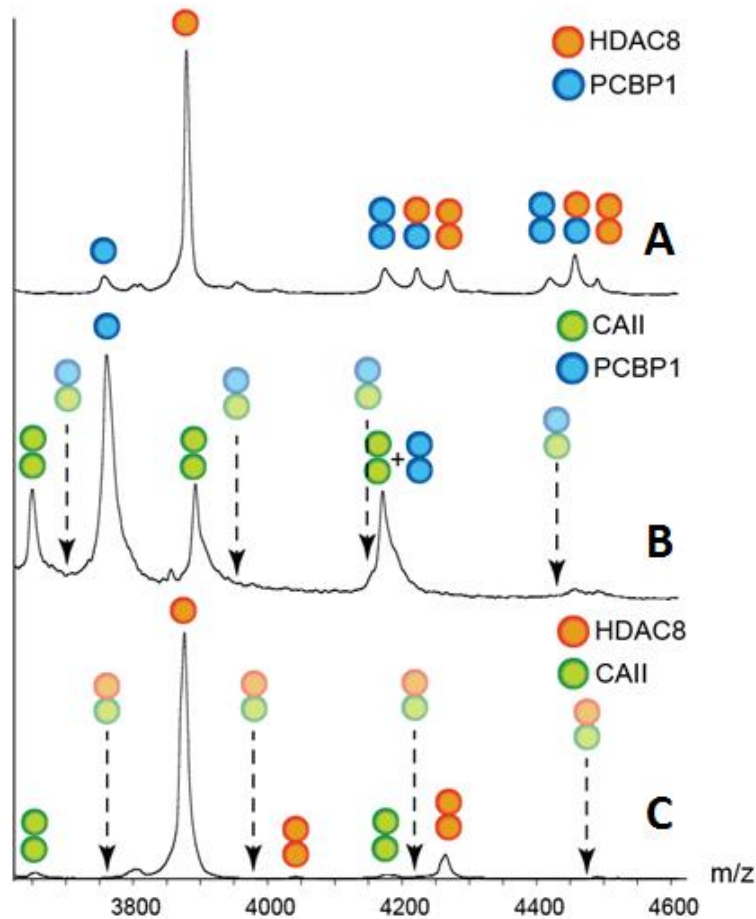
Figure4-3. Negative control MS data with CAII **(A)** HDAC8 and PCBP1 form a heterodimer complex. **(B)** CAII and PCBP1 and **(C)** CAII and HDAC8 spectra reveal no hetero-dimer complex formation (as indicated by the dashed arrows, representing the projected m/z values for the hetero-dimer assemblies)

4.3.2 Estimating Protein-Protein K_D Values for the HDAC8:PCBP1 Complex

In order to overcome the challenges associated with quantifying weak protein K_D values by ESI-MS, we constructed a workflow based on the above-described direct ESI-MS method of protein-protein binding constant determination, we worked to verify key assumptions implicit in the ESI-MS K_D quantification workflow for HDAC8 and PDCB1 specifically, and also build controls that validated the approach for weak protein-protein complexes in general.

Table 4-1. The binding affinity (K_D) values for (Apo/Zn/Fe) HDAC8 - PCBP1 complex.

	$K_D(\mu M)$	# of replicas
(Apo-Apo)	68±30	4
(Zn-Apo)	60±18	3
(Fe-Apo)	44±29	5

The first step in our workflow involves the buffer exchange of analyte protein into an ammonium acetate buffer system appropriate for native MS. Previous reports have indicated that the solution concentration of proteins can be significantly reduced during buffer exchange[39], and as such we worked to quantify any losses experienced for the proteins studied here. We find that for a typical exchange into 500mM ammonium acetate buffer using widely-available centrifugal ion exchange columns (see Experimental Methods section for details), the resultant concentration of HDAC8 tends to decrease by 30-60%, and that of PCBP1 can decrease by up to 80% percent. Such significant drops in protein concentration can result in a 5 fold deviation in ESI-MS K_D measurements, if not controlled. Based on this data, we built a post-buffer exchange concentration measurement step into our ESI-MS protocol for determining HDAC8:PCBP1 K_D values.

In order to control for differential ESI signal intensities as a function of both PCBP1 and HDAC8 concentrations in solution, we carried out a series of carefully-controlled calibration experiments. Figure 4-4 shows a series of ESI-MS intensity values for the two proteins referenced above, charted against their post-buffer exchange concentrations in solution. Our results suggest that both HDAC8 and PCBP1 exhibit similar ionization and detection efficiencies, especially in the primary range from which we extract K_D values (5-15 µM). As its solution concentration is unknown, we cannot present the ESI signal intensities for the intact complex between these two proteins on Figure 4-4; however, from the data shown it is likely the assembly follows similar intensity trends.
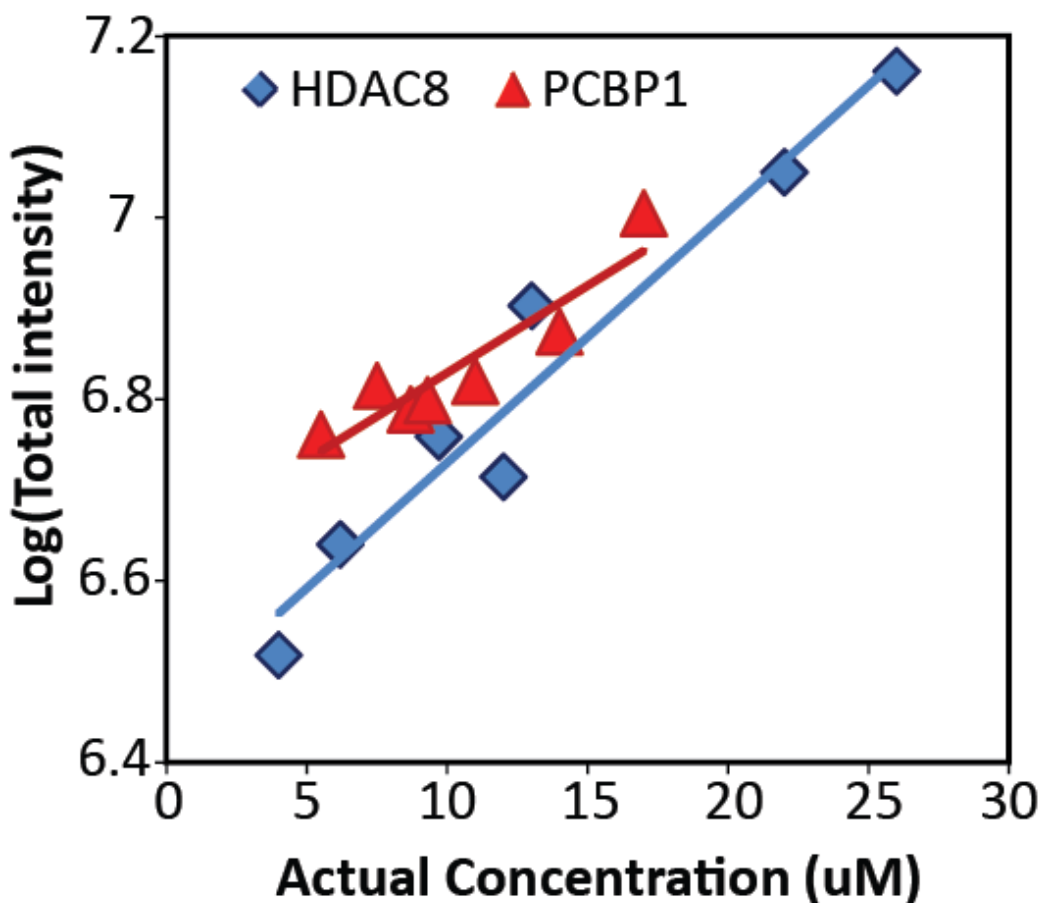
Figure4-4. A plot of the logarithm of ESI-MS intensity for HDAC8 (blue diamonds) and PCBP1 (red triangles), as a function of protein solution concentration following buffer exchange. Linear trend lines from regression analysis are shown,. Taken together, these two trend-lines indicate that ESI-MS intensity can be used to confidently determine HDAC8 and PCBP1 concentrations in solution, and thus the K_D value associated with their complex formation.

In some cases, mild collisional activation is used to dissociate weakly associated artifact assemblies that can be formed at high concentration during ESI. Such experiments must be tightly controlled in terms of their application, however, as they can also act to skew the apparent ESI-MS K_D values of real complexes. In order to evaluate the influence of collisional activation on HDAC8:PCBP1 complexes, we performed a series of systematic CIU/CID experiments,

finding no significant dissociation using up to 140V of activating potential in the ion trap preceding the IM separator in our instrument platform, indicating that the complex formed is highly stable and an unlikely source of K_D error. In addition, we altered the nESI flow rate in an effort to test altered droplet size and rate of production generated any variances in nESI-MS signal intensities for HDAC8 or PCBP1 ions. Results show that any ESI artifact assemblies are too weakly associated to generate any spurious signals for our ESI-MS measurements. In order to ensure that our ESI-MS K_D values were robust, however, we acquired multiple datasets over a wide range of assays require concentrations, and report average values over that range.

Table 4-2. The binding affinity (K_D) values for HDAC8 - PCBP1 complex measured by direct ESI-MS method.

HDAC8 conc. (µM)	PCBP1 conc. (µM)	K_D (µM)
6	13	21
		28
		29
10	10	60
		82
		72
18	15	56
		59
		63
5	5.5	78
		101
		112

Twelve measurements (four different molar ratios, each with three replicas) are presented. Protein concentrations are measured by nanodrop from buffer exchanged solution.

Based on the discussion and data above, herein we report K_D measurements that cover four different molar ratios, each performed in triplicate, (Table 4-2)for the HDAC8 - PCBP1 complex, yielding an average K_D value of 68 $\pm$ 30 µM on average (Table 4-1). The error in these measurements likely comes from 1) errors of the ESI-MS methodology generated from the sample preparation, ionization, transmission and detection processes and 2) batch-to-batch sample difference that may influence protein structure, and thus binding behavior. Ultimately, advancements in IM-MS instrument design and measurement protocols should lead to even more precise K_D values. Currently, it is clear those measurements of protein-protein K_D values that lie within the micro-molar range are challenging using traditional technologies (ITC, SPR, *etc*), and that our method likely provides one of the only routes to an accurate HDAC8:PCBP1 K_D value.

In order to determine either methodology error or lot-to-lot variance yields a larger contribution to the observed error for the reported K_D results, systematic statistical analysis was performed based on the measurements listed in Table 4-2. Briefly, four groups of K_D measurements, each contains three observations (replicas) are listed in the following two groups:

Group I: A (21, 28, 29), B (56, 59, 63)

Group II: A (60, 72, 82), B (78, 101, 112)

With the knowledge that measurements in Group I and Group II were performed with different batch of protein sample, we thus assume different batch of sample has different sample variance whereas the same batch of sample generate same variance, then we assume observations from group IA and IB follow normal distribution $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_1^2)$, respectively. Similarly, we assume the observations from Group IIA and IIB follow $N(\mu_3, \sigma_2^2)$ and $N(\mu_4, \sigma_2^2)$, respectively. As we switch to a new batch of protein sample for Group II data, we want to demonstrate that σ_2^2 is statistically significantly larger than σ_1^2, And at 95% level, we want to construct a lower bound of the ratio σ_2^2/σ_1^2

First, we performed the one sided test. The null hypothesis is set to be H_0: $\sigma_1^2 = \sigma_2^2$ while the alternative hypothesis is $\sigma_1^2 < \sigma_2^2$. We have the estimator $\hat{\sigma}_1^2$ for σ_1^2 which is 15.67 from Group IA and IB samples. Similarly we calculate $\hat{\sigma}_2^2 = 211.16$ from Group IIA and IIB observations. Under the null hypothesis, the ratio $f := \hat{\sigma}_1^2/\hat{\sigma}_2^2$ follows an F distribution with degrees of freedom 4 and 4. Results show the p-value is 0.0136 smaller than 0.05, which means at 95% confidence level we believe that the variance of the Group I is smaller than that of the Group II.

Next we would like to construct a 95% level confidence interval for the ratio of σ_2^2 and σ_1^2. As the 0.05 percentile of $F(4,4)$ distribution is 0.15635. This means that if we perform another test with

H_0: $\eta\sigma_1{}^2 = \sigma_2{}^2$ and H_a: $\eta\sigma_1{}^2 < \sigma_2{}^2$, then under the null hypothesis, the ratio $f_\eta := \eta\hat{\sigma}_1^2/\hat{\sigma}_2^2$ follows an F-distribution with degrees of freedom 4 and 4. So in order to not reject this null hypothesis, f needs to be at least 0.15635. Solving for $\eta\hat{\sigma}_1^2/\hat{\sigma}_2^2 = 0.15635$ we have $\eta = 2.11$. This means at 95% confidence interval, $\sigma_2{}^2/\sigma_1{}^2$ is at least 2.11. In other words, with 95% confidence, $\sigma_2{}^2$ is at least 2.11 times of $\sigma_1{}^2$. Such statistical analysis suggests switching to new batch of protein sample may lead to non-negligible error for K_D estimation, therefore same batch of analyte sample is strongly recommended.

4.3.3 ESI-MS Reveals the Role of Metal Ions in HDAC8:PCBP1 Complex Formation

In order to further investigate the influence of metal-containing HDAC8 or PCBP1 on complex formation, samples containing either Zn^{2+} or Fe^{2+} containing HDAC8, as well as Fe^{2+} bounded PCBP1, were screened for complex formation and structural effects by IM-MS. HDAC8 binds tightly to zinc (K_D = 5 pM) and relatively weakly to iron (K_D = 0.2 μM). In the case of PCBP1, three iron binding sites are available, and ITC assays reveals a 0.9 μM affinity to the first site, and 5.8 μM for the remaining two. Our results, shown in Figure 4-5, suggest that, as observed in Figure 4-2, by using optimized buffer conditions and properly tuned instrument parameters, formation of apoHDAC8:PCBP1 complexes can be observed in significant excess relative to any homoligomer formation by either protein. This observation holds for metal-containing HDAC8 and PCBP1 forms, verified by a binding activity assay, as the mass shifts associated with metal binding for both proteins are too small to resolve by native MS experiments.
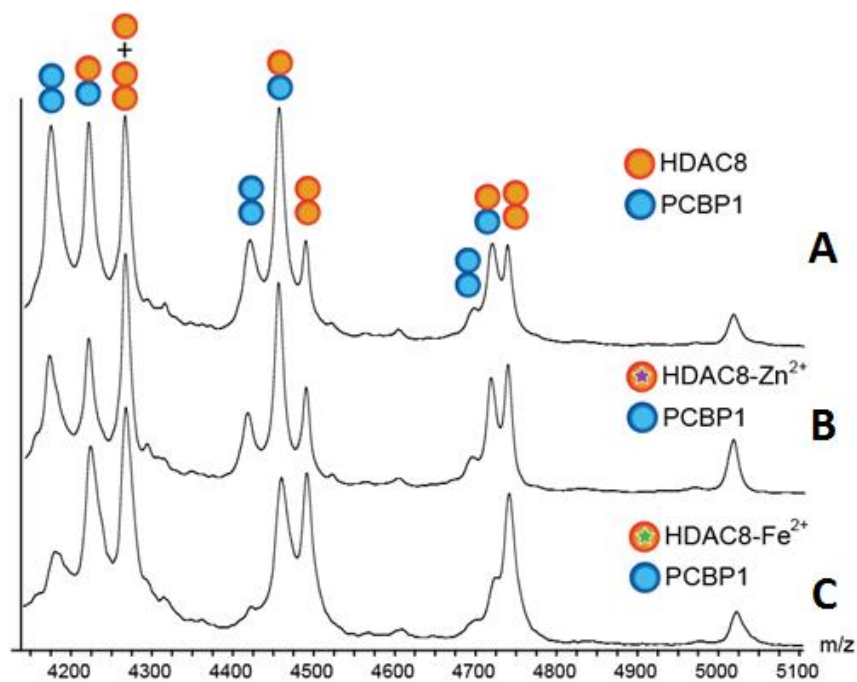
Figure4-5. MS data of **(A)**apo- **(B)** Zn^{2+} bounded (purple star) **(C)** Fe^{2+} bounded (green star) HDAC8 incubated with PCBP1. Hetero-dimer of HDAC8 and PCBP1 are detected in all cases, independent of metal substitution. However, quantitative K_D analysis suggests iron mildly promotes the binding between HDAC8 and PCBP1, and behave as effective as the natural metal binder zinc.

Quantitative ESI-MS K_D measurements further reveal a dissociation constant for the zinc-containing HDAC8 – PCBP1 complex $60 \pm 18\mu M$, while the KD for iron-containing HDAC8 – PCBP1 assemblies is $44 \pm 29\mu M$ (see table 4-3 for details). Data from Table 4-1 indicates that Fe(II) mildly promotes the binding between HDAC8 and PCBP1, while noting that the complex K_D values recorded for this assembly are within error of those generated for Zn(II) containing HDAC8 forms. Given that Zn(II) has long been viewed as the natural metal center for the

HDAC8 active site *in vivo*,[10] our results strongly suggest that Fe(II) generates equivalently competent HDAC8 proteins in the context of PCBP1 binding.

Table 4-3. The binding affinity (K_D) values for (Zn/Fe) HDAC8 - PCBP1 complex, as well as the Ferredoxin-NADP$^+$ reductase – Ferredoxin complex measured by direct ESI-MS method

Protein 1	Conc. (µM)	Protein 2	Conc. (µM)	K_D(µM)	Avg(µM)	Stdev(µM)
(Zn)HDAC8	9	PCBP1	8	77	60	18
	4.5		5.5	71		
	4.5		5.5	34		
(Fe)HDAC8	3	PCBP1	2	15	44	29
	4.5		2.3	21		
	5.5		5.5	28		
	8.5		8	70		
	8.5		8	88		
Ferredoxin Reductase	2.2	Ferredoxin	4	10	16	8
	4.1		7.5	10		
	3.1		11.2	27		

Protein concentrations are measured by nanodrop from buffer exchanged solution. Averaged K_D values and standard deviations are listed in the right-most two columns.

4.3.4 Validating ESI-MS Protein-Protein K_D Measurements Through Additional Control Experiments

As the KD values for the PCBP1: HDAC8 complexes reported here were not known previously, and the ESI-MS methodology has rarely been used to quantify the strength of weak protein-protein contacts as we have described, we undertook additional control experiments to evaluate the ability of our ESI-MS methodology to measure accurately the K_D of a similar protein complex of known binding strength. . Following a systematic screen of the structure-based benchmark library for protein-protein binding affinity,[40] summarized by Kastritis *et al*, we selected the Ferredoxin-NADP$^+$ reductase (32kDa) – Ferredoxin (10.5kDa) complex (which we represent as the FNR – FDX complex here) as our positive control system. The FNR – FDX

complex has a known binding affinity of 3.57 μM, measured by difference absorption spectroscopy, [41] and is similar to HDAC8-PCBP1 complexes in terms of both K_D molecular mass of the interacting partners (the intact complex has a mass of 43 kDa). Our results, listed in table 4-3, reveal a binding affinity of (16$\pm$8) μM, in good agreement with the literature value. We also note that the absolute precision of our ESI-MS KD value for FNR-FDX is also improved when compared with those extracted for HDAC8 – PCBP1, suggesting that the strength of the interaction observed in the latter case may also be a quantifiable source of error in our experiments. Overall, however, these results clearly suggest the direct ESI-MS method described here, when coupled with carefully controlled experimental conditions and control experiments can be used to assess the accurate binding affinities for weak protein-protein interactions.

4.4 Conclusions

In this chapter, the strength of the interactions formed between HDAC8 and PCBP1 were systematically investigated by ESI-MS. Our data suggests that HDAC8 and PCBP1 interact with each other in a specific manner, as confirmed by tandem MS experiments and comparisons with homo-oligomers also observed. Negative control experiments that tested the ability ofcarbonic anhydrase II, a protein with no known affinity for either protein but possessing similar surface area to both HDAC8 and PCBP1, to form complexes with either of the two binding partners discussed above. These experiments revealed no detectible carbonic anydrase II complexes with either HDAC8 or PCBP1, further supporting the specificity of the HDAC8-PCBP1 complex and

the ability of the ESI-MS method to quantify the strength of weakly associated protein complexes. By optimizing buffer conditions and carefully adjusting instrument settings, binding affinities of a weakly-associated hetero-complex can be measured by directly ESI-MS, and we report a K_D value of 68 $\pm$ 30 μM for the HDAC8-PCBP1 dimer. In order to better understand how bound metals affect the structure and binding behavior of the HDAC8 and PCBP1, Zn^{2+}/Fe^{2+} doped HDAC8 were incubated with PCBP1 (with and without Fe^{2+}), and demonstrate that the above-noted dimer forms in a manner independent of metal association. As for the strength of the interactions detected, our measured K_D values suggests that Fe(II) mildly promoting the formation of the HDAC8-PCBP1 dimer.

The accuracy of ESI-MS-based K_D measurements in general for similar hetero-protein dimers was further evaluated by measurements on the FNR - FDX protein complex. Our ESI-MS measurements yield a K_D value the literature value, providing support the general accuracy and precision of this approach from such studies.[26, 27, 30]Future work will aim to produce a comprehensive, robust methodology with high accuracy and precision for the measurement of K_D values for weak protein assemblies. For example, IM separation can be folded into the existing workflow described here to provide additional confidence in identifying heterodimer charge states when the two proteins are close in mass. Moreover, IM offers the capability for monitoring any changes in protein conformational upon complex formation, which when coupled with accurate K_D values would be a powerful approach for validating many key therapeutic targets. [32]Overall, the data presented here predicts a bright future for the application of similar ESI-MS toward weakly-associated proteins that currently represent a bottleneck in the detailed evaluation of many key processes in molecular biochemistry.

4.5 References

1. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications.* Nature, 2000. **403**(6765): p. 41-45.

2. Shiio, Y. and R.N. Eisenman, *Histone sumoylation is associated with transcriptional repression.* Proceedings of the National Academy of Sciences, 2003. **100**(23): p. 13225-13230.

3. Gregoretti, I., Y.-M. Lee, and H.V. Goodson, *Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis.* Journal of molecular biology, 2004. **338**(1): p. 17-31.

4. Remiszewski, S., *Recent advances in the discovery of small molecule histone deacetylase inhibitors.* Current opinion in drug discovery & development, 2002. **5**(4): p. 487-499.

5. Lane, A.A. and B.A. Chabner, *Histone deacetylase inhibitors in cancer therapy.* Journal of Clinical Oncology, 2009. **27**(32): p. 5459-5468.

6. Barnes, P., I. Adcock, and K. Ito, *Histone acetylation and deacetylation: importance in inflammatory lung diseases.* European Respiratory Journal, 2005. **25**(3): p. 552-563.

7. Marmorstein, R., *Structure of histone deacetylases: insights into substrate recognition and catalysis.* Structure, 2001. **9**(12): p. 1127-1133.

8. Grozinger, C.M. and S.L. Schreiber, *Deacetylase enzymes: biological functions and the use of small-molecule inhibitors.* Chemistry & biology, 2002. **9**(1): p. 3-16.

9. Minucci, S. and P.G. Pelicci, *Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer.* Nature Reviews Cancer, 2006. **6**(1): p. 38-51.

10. Finnin, M.S., et al., *Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors.* Nature, 1999. **401**(6749): p. 188-193.

11. Gantt, S.L., S.G. Gattis, and C.A. Fierke, *Catalytic activity and inhibition of human histone deacetylase 8 is dependent on the identity of the active site metal ion.* Biochemistry, 2006. **45**(19): p. 6170-6178.

12. Joshi, P., et al., *The functional interactome landscape of the human histone deacetylase family.* Molecular systems biology, 2013. **9**(1): p. 672.

13. Eide, D.J., *Zinc transporters and the cellular trafficking of zinc.* Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 2006. **1763**(7): p. 711-722.

14. Colvin, R.A., et al., *Cytosolic zinc buffering and muffling: their role in intracellular zinc homeostasis.* Metallomics, 2010. **2**(5): p. 306-317.

15.   Fukada, T., et al., *Zinc homeostasis and signaling in health and diseases.* JBIC Journal of Biological Inorganic Chemistry, 2011. **16**(7): p. 1123-1134.

16.   Beyersmann, D. and A. Hartwig, *Carcinogenic metal compounds: recent insight into molecular and cellular mechanisms.* Archives of toxicology, 2008. **82**(8): p. 493-512.

17.   Costello, L. and R. Franklin, *Novel role of zinc in the regulation of prostate citrate metabolism and its implications in prostate cancer.* The Prostate, 1998. **35**(4): p. 285-296.

18.   Costello, L.C., et al., *Evidence for a zinc uptake transporter in human prostate cancer cells which is regulated by prolactin and testosterone.* Journal of Biological Chemistry, 1999. **274**(25): p. 17499-17504.

19.   Franklin, R.B., et al., *hZIP1 zinc uptake transporter down regulation and zinc depletion in prostate cancer.* Molecular cancer, 2005. **4**(1): p. 32.

20.   Johnson, D.C., et al., *Structure, function, and formation of biological iron-sulfur clusters.* Annu. Rev. Biochem., 2005. **74**: p. 247-281.

21.   Shi, H., et al., *A cytosolic iron chaperone that delivers iron to ferritin.* Science, 2008. **320**(5880): p. 1207-1210.

22.   Bulteau, A.-L., et al., *Frataxin acts as an iron chaperone protein to modulate mitochondrial aconitase activity.* Science, 2004. **305**(5681): p. 242-245.

23.   Jensen, L.T. and V.C. Culotta, *Role of Saccharomyces cerevisiae ISA1and ISA2 in iron homeostasis.* Molecular and cellular biology, 2000. **20**(11): p. 3918-3927.

24.   Katta, V. and B.T. Chait, *Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry.* Journal of the American Chemical Society, 1991. **113**(22): p. 8534-8535.

25.   Loo, R.R.O., et al., *Observation of a noncovalent ribonuclease S-protein/S-peptide complex by electrospray ionization mass spectrometry.* Journal of the American Chemical Society, 1993. **115**(10): p. 4391-4392.

26.   Wang, W., E.N. Kitova, and J.S. Klassen, *Influence of solution and gas phase processes on protein-carbohydrate binding affinities determined by nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry.* Analytical chemistry, 2003. **75**(19): p. 4945-4955.

27.   Kitova, E.N., et al., *Reliable determinations of protein–ligand interactions by direct ESI-MS measurements. Are we there yet?* Journal of the American Society for Mass Spectrometry, 2012. **23**(3): p. 431-441.

28.   El-Hawiet, A., et al., *Quantifying labile protein–ligand interactions using electrospray ionization mass spectrometry.* Journal of the American Society for Mass Spectrometry, 2010. **21**(11): p. 1893-1899.

29.   Han, L., et al., *Protein–Glycolipid Interactions Studied in Vitro Using ESI-MS and Nanodiscs: Insights into the Mechanisms and Energetics of Binding.* Analytical chemistry, 2015. **87**(9): p. 4888-4896.

30.   Han, L., et al., *Identifying carbohydrate ligands of a norovirus P particle using a catch and release electrospray ionization mass spectrometry assay.* Journal of The American Society for Mass Spectrometry, 2014. **25**(1): p. 111-119.

31.   Rabuck, J.N., et al., *Activation state-selective kinase inhibitor assay based on ion mobility-mass spectrometry.* Analytical chemistry, 2013. **85**(15): p. 6995-7002.

32.   Soper, M.T., et al., *Amyloid-β–neuropeptide interactions assessed by ion mobility-mass spectrometry.* Physical Chemistry Chemical Physics, 2013. **15**(23): p. 8952-8961.

33.   Niu, S. and B.T. Ruotolo, *Collisional unfolding of multiprotein complexes reveals cooperative stabilization upon ligand binding.* Protein Science, 2015.

34.   Bornschein, R.E., . *Ion Mobility-Mass Spectrometry Reveals Highly-Compact Intermediates in the Collision Induced Dissociation of Charge-reduced Protein Complexes.* Development of Charge Manipulation Nanoelectrospray Ion Mobility-Mass Spectrometry Techniques for Multiprotein Complex Analysis, 2015: p. 58.

35.   Hyung, S.-J., C.V. Robinson, and B.T. Ruotolo, *Gas-phase unfolding and disassembly reveals stability differences in ligand-bound multiprotein complexes.* Chemistry & biology, 2009. **16**(4): p. 382-390.

36.   Ruotolo, B.T., et al., *Evidence for macromolecular protein rings in the absence of bulk water.* Science, 2005. **310**(5754): p. 1658-1661.

37.   Liu, J. and L. Konermann, *Protein–protein binding affinities in solution determined by electrospray mass spectrometry.* Journal of the American Society for Mass Spectrometry, 2011. **22**(3): p. 408-417.

38.   Ebong, I.-o., et al., *Heterogeneity and dynamics in the assembly of the heat shock protein 90 chaperone complexes.* Proceedings of the National Academy of Sciences, 2011. **108**(44): p. 17939-17944.

39.   Hernández, H. and C.V. Robinson, *Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry.* Nature protocols, 2007. **2**(3): p. 715-726.

40.   Kastritis, P.L., et al., *A structure‐based benchmark for protein‐protein binding affinity.* Protein Science, 2011. **20**(3): p. 482-491.

41.   Sancho, J. and C. Gómez-Moreno, *Interaction of ferredoxin-NADP+ reductase from Anabaena with its substrates.* Archives of biochemistry and biophysics, 1991. **288**(1): p. 231-238.

Chapter 5

Collision Induced Unfolding Identifies Small Molecule Binding Signatures for the Integral Membrane Protein Translocator Protein (TSPO)

5.1 Introduction

Membrane proteins are responsible for a variety of biological functions and physiological responses vital to the survival of organisms, ranging from the transport of chemical signals, to cancer development and progression. [1, 2] Based on this importance, such as semblies are of great pharmaceutical significance in a general sense, representing more than half of all current therapeutic targets.[3, 4]As one of the important transporter membrane protein complexes discussed above, the 18 kDa translocator protein (TSPO), a mitochondrial outer membrane protein dimer that plays a pivotal role in cholesterol and porphyrin transport, apoptotic signaling and cancer development, serves as the focal point of the work described here. [5, 6]

TSPO was originally known as the peripheral-type benzodiazepine receptor (PBR), as it was first discovered as a target for such drugs in peripheral tissues.[7, 8] In 2006, the protein was given its current name to better represent emerging understanding regarding its structure, sub-cellular roles and tissue-specific functions.[5] In addition to the above described functions, TSPO is also associated with many other cellular roles, including the biosynthesis of steroid hormones, cell proliferation, anion transport regulation of mitochondrial functions and immuno-modulation.[9-14] accordingly, numerous potential clinical applications of TSPO modulation have been proposed, in broadly-defined areas ranging from oncology to neuropharmacology and the treatment of neurodegenerative diseases.[15-18]

Recently, the crystal structures of WT TSPO from *Rhodobacter sphaeroides* and a mutant, A139T, (crystal structure shown in Figure 5-1) that mimics the human polymorphism associated with psychiatric disorders and reduced pregnenolone production, were reported, revealing its dimeric conformation and topology for the first time.[6] The *Rs*TSPO currently serves as the best-characterized bacterial homolog in the TSPO family. Previous studies suggest highly conserved functions between bacterial and mammalian organisms, strongly-indicating the value of the *R.sphaeroides* homolog as a model system to investigate structure-function relationships in TSPO.[19-21]It is worth noting that within the A139T X-ray structure, a cluster of electron density found within a cavity between TM-I and TM-II and beneath the LP1region, was

observed and did not fit any known components in the crystallization medium, suggesting the existence of a previously unknown native endogenous ligand bound to the dimer. While the identity of this unknown TSPO binder has been the subject of speculation and putative assignments based on the X-ray data alone, no other data is currently available to directly confirm the identity of this endogenous TSPO ligand, in part due to a lack of appropriate methodologies capable of detailed membrane protein ligand binding analysis.[6]

While many challenges in data interpretation and collection persist, remarkable advances in X-ray and NMR have enabled the structural analyses of host of individual membrane proteins and small complexes.[22, 23]In addition, recent advancements in Cryo-Electron Microscopy (EM)have yielded near atomic resolution of membrane protein complexes prepared in detergent micelles.[24] However, localizing small binders within high resolution atomic structures is still a challenge, as it is often difficult to distinguishing the specific binders within a background of detergent molecules and other buffer components present.[25, 26]

Mass spectrometry (MS) has made profound contributions to the structural characterization of membrane proteins, through methodologies ranging from liquid chromatography (LC)-MS based proteomics to chemical labeling strategies aimed at structure determination.[27-30] MS technologies that detect intact multiprotein membrane protein complexes ionized directly from

carefully-controlled detergent (or lipid) clusters and micelles have more recently emerged, and offer label free analysis of protein complex stoichiometry and ligand binding properties [27, 28]. Through the addition of ion mobility (IM), which is capable of separating membrane protein ions according to their size and charge in the gas phase, and collision induced unfolding (CIU), which can measure shifts in membrane protein stability upon small molecule binding, such 'native MS' experiments have been used to construct models of massive ATPase assemblies and assess the role of lipids in stabilizing ion channel complexes [29]. Many challenges persist, however, in the development of such native IM-MS experiments. Chief among these revolve around the difficulties in performing tandem MS experiments on membrane protein complexes in order to assign the precise composition and ligand identities for isolated assemblies, as such ions are typically encapsulated within extremely polydisperse detergent micelles during their transit through the MS filter proximal to the ion source, rendering any attempts at meaningful precursor ion isolation futile.

Here, we present new IM-MS data for multiple forms of *Rs*TSPO dimers suggesting that careful analysis of CIU data can be used to identify small molecule binders to membrane proteins. We systematically screen eight non-ionic detergents for their ability to preserve compact, dimeric of *Rs*TSPO (wild type and mutants in a manner congruous with predicted IM data from the available X-ray structure. In line with X-ray data,[6]we observe an endogenous ligand bound to the A139T mutant protein in apo, 2:1, and 2:2 stoichiometries. CIU stability assays of these

complexes indicate a larger than expected stability value for 2:1 complexes. Through the construction of detailed control experiments using the A138F TSPO construct that can be purified without endogenous ligand but is still competent for small molecule binding, we find strong correlations between CIU stabilities recorded for TSPO bound to a phosphoglycerol lipid and those observed for the unknown endogenous ligand, strongly supporting its assignment as a similar structure, We conclude by discussing our CIU stability profiling, and put these results in context for both continued studies of TSPO, as well as membrane protein structure in general.
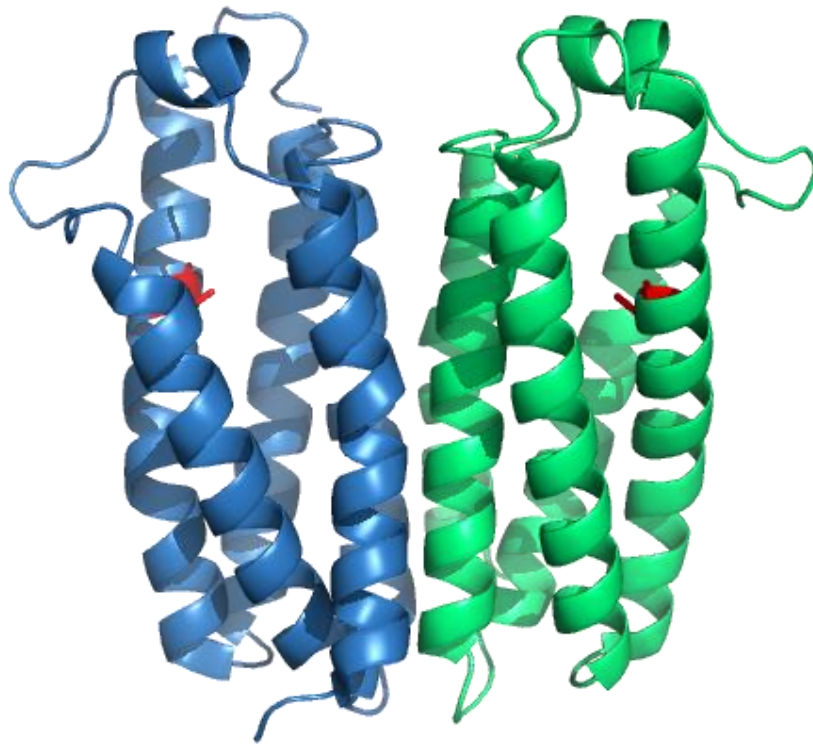
Figure 5-1. Crystal structure of *Rs*TSPO A139T dimer. (PDB: 4UC2) Two monomeric units are labeled in blue and green, respectively. The position of the A139T mutation is labeled in red and shown in sticks.

5.2 Experimental Methods

5.2.1 Materials

Wild type (WT) *Rs*TSPO, as well as its various mutants were expressed and using established protocols.[6, 19] All TSPO samples were generated at a concentration of 70-110 μM, in 50mM Tris, 150mM NaCl, 0.2% decylmaltoside (DM) buffer. Nonionic detergent samples were purchased from Anatrace (Maumee, OH), see Figure 5-2 for details (detergent names, CMC values, and chemical structures). PG lipid (1,2-dihexadecanoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (sodium salt)) was purchased from Avanti Polar Lipids (Alabaster, Alabama). PpIX, (Protoporphyrin IX, 3,7,12,17-tetramethyl-8,13-divinyl-2,18-porphinedipropionicacid disodium salt) was purchased from Sigma-Aldrich (St. louis, MO).
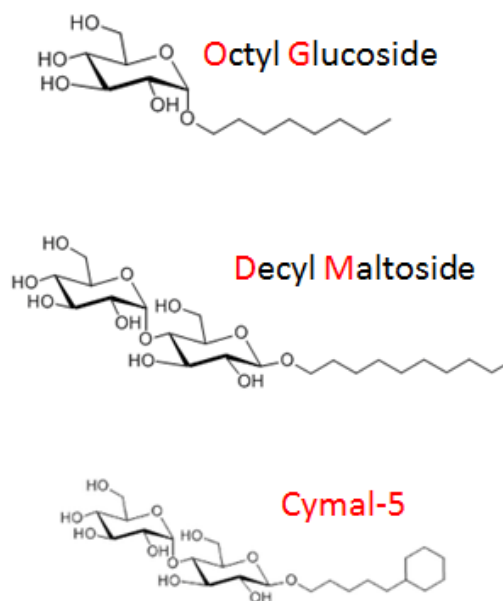
Figure 5-2. List of screened detergents with their abbreviations and CMC values **(Left)** Abbreviations, CMC values to TSPO membrane protein of the detergents screened are listed in left table. Abbreviations stand for: DM, n-Decyl-β-D-Maltopyranoside; NM, n-Nonyl-β-D-Maltopyranoside; OM, n-Octyl-β-D-Maltopyranoside; NG, n-Nonyl-β-D-Glucopyranoside; OG, n-Octyl-β-D-Glucopyranoside; Cymal-5, 5-Cyclohexyl-1-Pentyl-β-D-Maltoside;$C_{10}E_5$, Octaethyleneglycolmonododecylether; $C_{10}E_8$, Pentaethyleneglycolmonodecylether. **(Right)**Representative molecular structure of OG, DM and Cymal-5. The structure of CxEy is in the form of $CH_3(CH_2)_{x-1}$-O-$(CH_2CH_2O)_yH$

5.2.2 Detergent screen and buffer exchange protocol

In order to alter buffer conditions to include detergents of interest and reduce salt content, , Amicon® Ultra-0.5 centrifugal filters (Merck Millipore Ltd. Tullagreen, Carrigtwohill, Co.

Cork, IRL) were pre-rinsed with MS-compatible buffer containing 200mM ammonium acetate (pH 8.0) and detergent of interest at 2x its critical micelle concentration (CMC) . Bio-spin centrifugal columns (accuSpin Micro 17R, Fisher Scientific) were set at 4°C, and a speed of 2,000 x *g* for 10 minutes. An activation wash was performed three times prior to the introduction of TSPO containing sample solutions (50-100 μL) into centrifugal tubes. Amicon tubes were filled with 400-450 μL of rinsing buffer for 20-30 minutes, allowing sufficient time for detergent interaction. The resultant solution underwent another 5-10min spin at the above same conditions as described above. A 5-10 μL aliquot of the resulting   supernatant was subjected to nESI-IM-MS analysis.

5.2.3 IM-MS and CIU Measurements

All experiments were performed on a Synapt G2 nESI- ion mobility separation- time-of-flight (nESI-IMS-TOF) mass spectrometer (Waters, Milford, MA). Capillary voltage was typically kept at ~1.6kV to allow for stable ion current and source backing pressure was set around 7.5-8.9 mbar to obtain optimum transmission efficiency for membrane proteomicelle complexes. Sampling and Cone voltages were set to their maximum values (200V and 10V respectively) and the source temperature was set at 40°C to achieve maximum desolvation and enhanced MS resolving power. Within the traveling-wave IM region, trap and transfer voltages were carefully tuned, in an effort to achieve complete detergent removal and release the ionized protein assemblies from within the gas-phase micelles. IM measurements were carried out using

N$_2$asbuffer gas, at a pressure around 2.1 mbar. IM wave velocity and wave height was set at 700 m/s and 32.5V, respectively. Experimental collision cross sections (CCSs), an orientationally average ion-neutral collision integral, were determined following the method described previously using cytochrome c (equine heart), β-lactoglobulin (bovine milk, both monomer and dimer), avidin (egg white) and Concanavalin A (jack bean) standards.[30, 31] Theoretical CCSs were calculated based on PDB structures by the projection approximation approach from Driftscope software. A scaling factor of 1.15 was applied to correlate experimental and computational CCS measurements for large biomolecules.[32]Monomer model extraction from crystal structure was performed by PyMol. All experimental data acquisition and processing were carried out using MassLynxV4.1 software.

5.3 Results and Discussion

5.3.1 Optimal Detergents for the IM-MS Analysis of TSPO

As shown in Figure 5-2, eight non-ionic detergents were systematically screened with the above described IM-MS protocol in section 5.2.3, in an effort to achieve optimal ionization and analysis conditions for both WT and various mutant forms of RsTSPO. For these experiments, DM served as our control detergent, due to its ubiquity in both native MS and structural biology in general. IM-MS results for DM containing TSPO samples, however, exhibited poor quality, low resolving power MS data at maximum acceleration voltage values, , indicating that DM

binds to TSPO to tightly for facile release in the gas phase. In similar fashion, NM, OM and NG produced IM-MS data outcomes for TSPO similar to DM. In contrast, TSPO samples prepared using the detergent cymal-5 produced well-resolved MS data for both monomeric and dimeric forms of the protein when large activation voltages (trap CV>170V, transfer CV>70V) were employed. A closer inspection of the IM data recorded for this experiment reveals that while high MS resolving power is achieved, TSPO undergoes extensive, uncontrolled CIU making the comprehensive separation of monomer and dimer TSPO forms impossible, indicating insufficient stabilization of the TSPO structure by cymal-5. In contrast, TSPO samples containing OG result in well resolved IM-MS data for all protein variants studied here, requiring only 80V of accelerating potential in the ion trap in order to generate well-resolved, IM-MS data for the A139T and A138F mutants specifically.

In general, we find that surfactants comprised of longer alkyl chains and larger saccharide head groups tend to provide greater stabilization of TSPO, binding more tightly to the gas-phase ions created, and thus requiring larger accelerating potentials for complete desolvation Previous reports have identified polyoxyethylene glycols, such as $C_{10}E_5$ nor $C_{12}E_8$, and *n*-dodecyl-β-D-maltopyranoside (DDM) as optimal detergents for the general IM-MS analysis of membrane proteins. [27, 33] Neither of these classes of surfactant proved optimal for our TSPO studies, further indicating the necessity to optimize such variables on a protein-by-protein basis. In the case of DM, NM, OM and NG bound complexes, our instrument likely the necessary activation

capabilities to completely dissociate the bound detergent, hence yielding unresolved IM-MS data.

,

5.3.2 IM-MS detects native-like TSPO monomer/dimer structures in the gas phase

Figure 5-3 shows the MS data overlaid with an IM-MS contour plot acquired for the TSPO A138F variant, using the optimized OG detergent and activation voltages described above. Signals corresponding to both TSPO monomers (red, 18kDa) and dimers (green, 36kDa) are observed, in accordance with expectations derived from other biophysical data [6, 19]. Data acquired for WT TSPO indicates significant protein unfolding under these conditions, resulting in overlapped IM data that makes the annotation of dimer and monomer signals within this dataset challenging. IM-MS data for the A138F variant reveals resolved 4^+ and 5^+ monomer signals which are well resolved from 8^+ and 10^+ dimer signals respectively by IM. Similarly clear separation of monomer/dimer signals are observed in the case of the A139T variant (Figure 5-3), which exhibits excellent MS resolving power and monomodal IM drift time populations for the majority of the signals recorded.
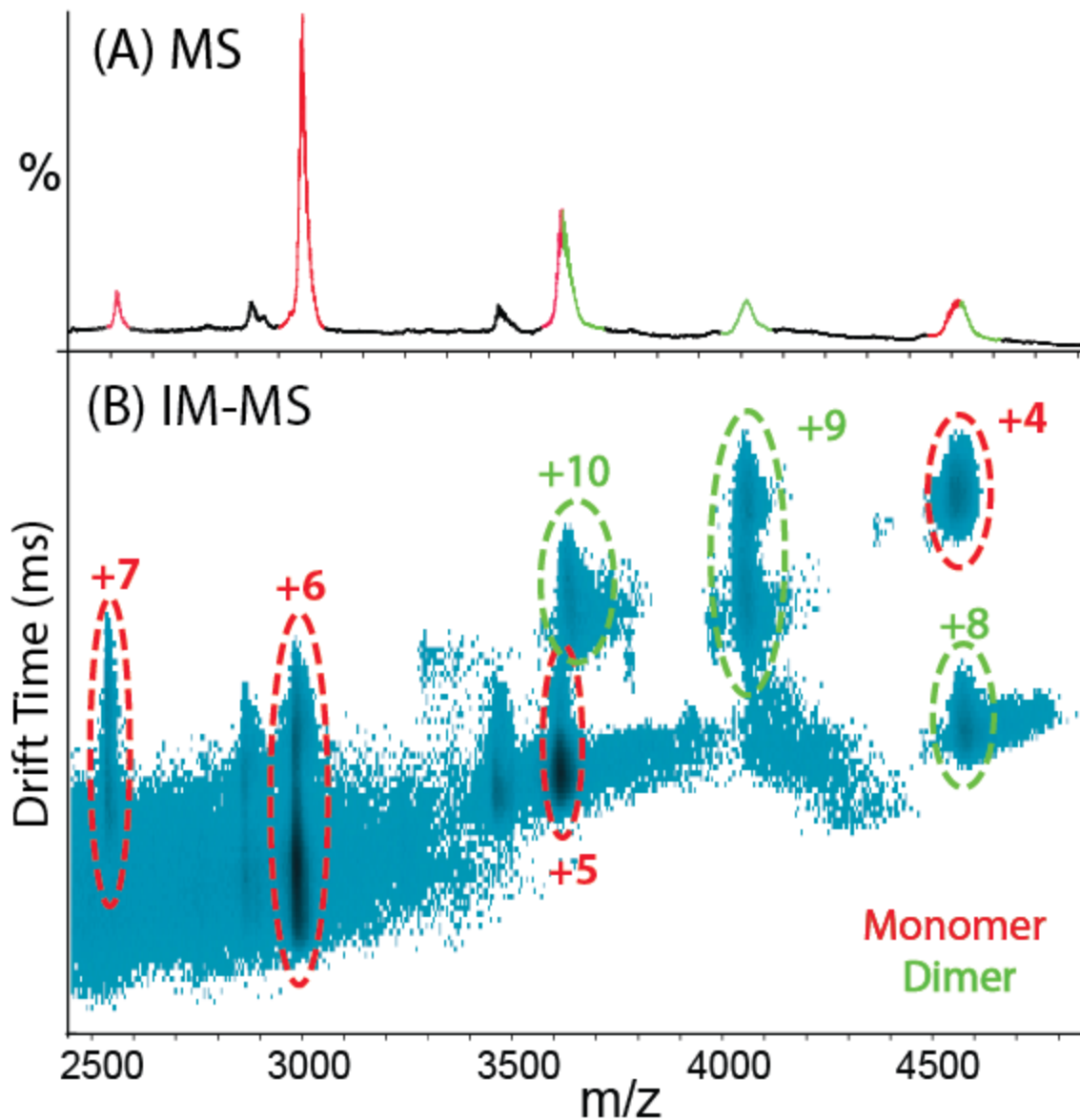
Figure 5-3.(A)MS and **(B)**IM-MS results for the TSPO A138F variant membrane protein monomer (18KDa, in red), and dimer (36KDa, in green), collected at a Trap Collision Voltage of 190V, a Transfer Voltage of 70V, and using 2xCMC (~1%) of Octyl Glucoside (OG) as a detergent.

In an effort to assess the correlation between native-state and gas-phase TSPO structures, we generated CCS estimates based on the TSPO X-ray structure. Using the dimeric TSPO A139T variant as a starting point (PDB ID: 4UC2), we compute a theoretical CCS of 2742Å^2, which agrees well with our experimental CCS values obtained for 9$^+$ dimeric TSPO (2751Å^{2}While an X-ray structure for TSPO monomer is not currently available CCS values for the two non-equivalent monomers found within the X-ray dimer are 1755Å^2 and 1714Å^2. A comparison between these model values and experimental CCSs indicates the presence of compacted (4$^+$), native-like (5$^+$), as well as highly unfolded (6$^+$ and 7$^+$) TSPO structures in the gas phase. Overall, our observations suggest carefully tuned collisional activation allows for the complete desolvation of detergent from the proteomicelle, while generating protein ions that retain a strong correlation with expected native-like protein structures.

5.3.3 Native MS Analysis of a Disease-associated TSPO Variant

As discussed above, the A139T TSPO variant is strongly associated with reduced pregnenolone production, which is in turn correlated with depression and neurodegenerative disease progression [5, 6]. Figure 5-4 shows a representative dataset for A139T TSPO, acquired using the optimized sample preparation and instrument conditions described above. In this dataset, we observe strong signals that correlate to TSPO monomer and dimer (marked with yellow

triangles), as observed in Figure 5-3, as well as new groups of signals having molecular masses that shift by between 720 and 780 Da relative to Apo TSPO, which we putatively identify as an unknown endogenous ligand bound exclusively to the dimeric form of the protein. While weak signals in our IM-MS dataset can be observed for higher binding stoichiometries, the strongest and most consistent signals correspond to 2 specific binding events (1 per TSPO monomer). Taken together, this data suggests the detection of a specific endogenous ligand extracted from the TSPO cellular environment upon protein purification, and determining both its identity and impact on protein structure/stability would clearly illuminate the role TSPO plays in various disease states.

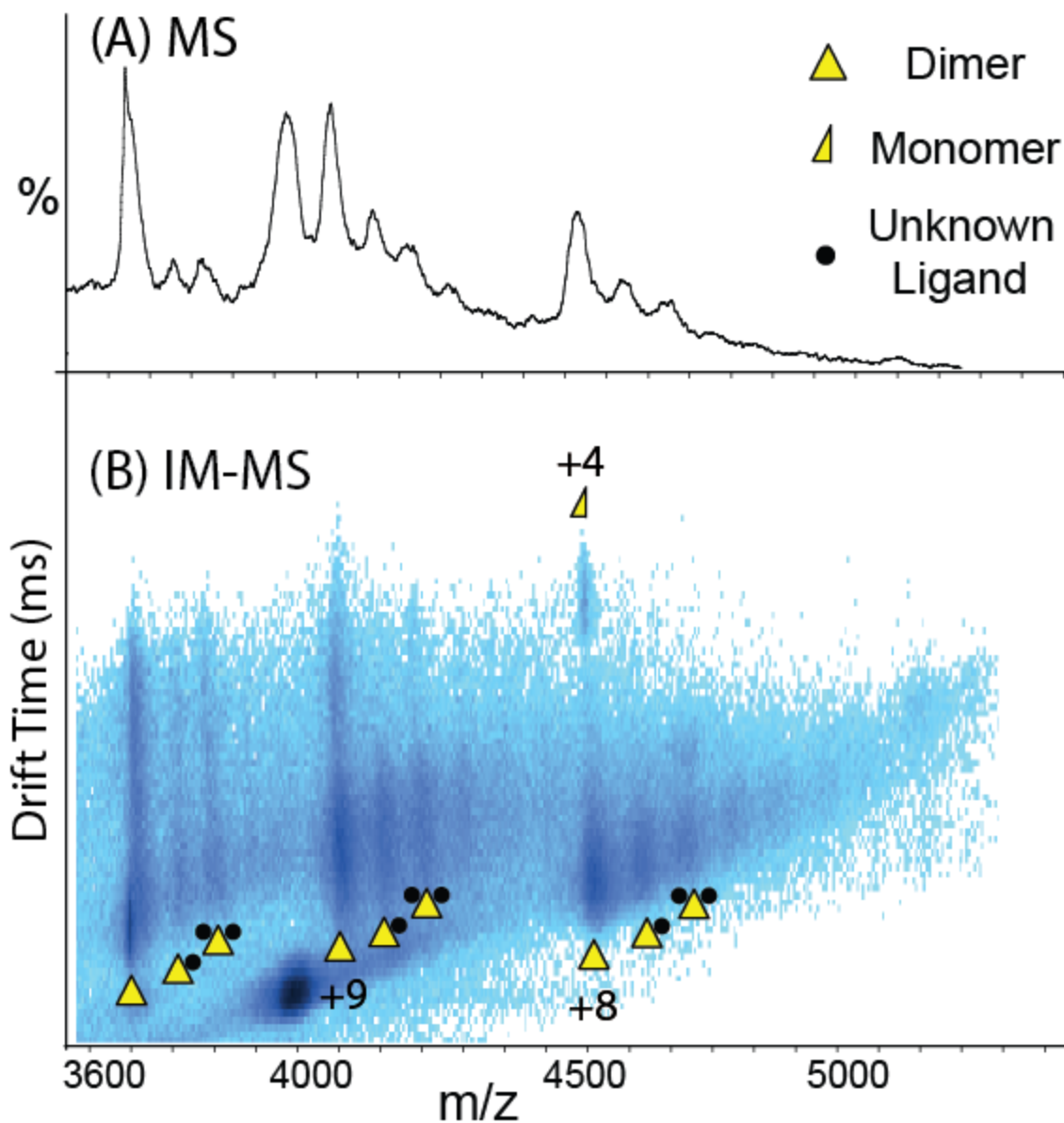**Figure 5-4.(A)** MS and **(B)**IM-MS data for the TSPO A139T variant monomer (18kDa, in half-triangle), and dimer (36kDa, in complete triangle), collected at a Trap Collision Voltage of 150V, a Transfer Voltage of 70V, using 2xCMC (~1%) of OG as the detergent system. An endogenous unknown ligand (molecular weight 720-780Da, indicated using black circles) is observed bound in 2:1 and 2:2 stoichiometries to the dimeric A139T form.

Prior to this study, native MS workflows provided few, if any, options for the identification of previously unknown small molecules bound to a membrane protein such as TSPO. Direct measurements of bound ligand masses are obscured by the low resolving power and mass accuracy that is often encountered in native MS data for membrane protein complexes, resulting from incomplete adduct removal [27, 34]. Protein denaturation of soluble protein-ligand complexes typically leads to ligand release following binding conformation [35], but such experiments would unquestionably result in the development of insoluble aggregates for membrane protein samples. Q-ToF methods developed for soluble protein-ligand complexes, such as MS/MS experiments that aim to select signals for intact complex ions and eject ligand ions upon collisional activation for precise MS analysis, also cannot be typically employed for membrane protein complexes, as such assemblies are encapsulated within heterogeneous detergent micelles during their transmission through the quadrupole [27]. As such, any ions observed in the ToF region of the instrument likely do not possess similar m/z values in the Q region of the instrument, making the direct application of tandem MS experiments exceptionally challenging. Alternative bottom-up strategies for identifying ligand binding would likewise be obscured by chemical noise from potentially unknown populations of co-eluting, but non-interacting, small molecules. In sum, there is a clear need to develop native IM-MS methods that enable the identification of unknown small molecules when bound to intact membrane protein complexes. .
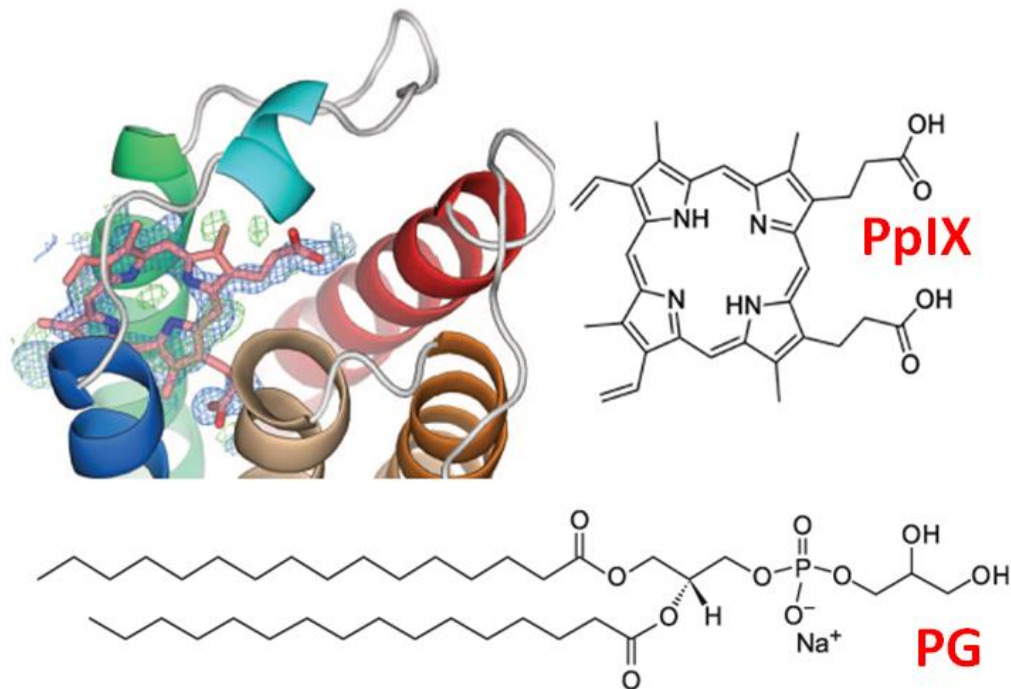
Figure 5-5. Electron density cluster found in the crystal structure of TSPO A139T mutant and proposed structure to the endogenous ligand **(Top)** Electron density cluster (in blue) found in the crystal structure of TSPO A139T mutant. A porphyrin ring structure (pink) was suggested as a tentative assignment for the unknown endogenous ligand. **(Right and Bottom)** Detailed structure of the two small molecules proposed for the putative identities of the TSPO A139T endogenous ligand: Protoporphyrin IX (PpIX, Mw=606Da) and phosphatidyl-glycerol lipid (PG, Mw=745Da)

In an effort to build such a native MS methodology for membrane protein ligand identification, we carried out s set of experiments that spiked excess amounts of candidate molecules for the unknown endogenous TSPO ligand into A139T samples. Similar to methods commonly employed in chromatography, following the addition of selected ligand, MS signal intensities are monitored. Increases in the intensity values for pre-existing bound states under these conditions

would be interpreted as evidence that the added ligand matches the unknown endogenous ligand, whereas the appearance of new MS features would preclude the possibility that the added ligand corresponds to the unknown TSPO binder. Key assumptions in this approach include: 1) That added ligands can penetrate the micelles that encapsulate TSPO in solution, 2) That the endogenous ligand binding pattern observed in Figure 5-4 does not represent the maximum extent of ligand binding under these conditions and 3) That if the new ligands are introduced that do not match the endogenous ligands already bound, that ligand displacement will occur. Given the measured molecular mass range for the unknown endogenous ligand, X-ray modeling efforts with the A139T mutant (Figure 5-5)[6], and previous reports of TSPO binders [19], we selected both phosphatidyl-glycerol lipid (PG, Mw=745Da), and B). Protoporphyrin IX (PpIX, Mw=606Da) for the spiking experiments described above (Figure 5-5).[34, 36, 37] The MS results for our spiking experiments, shown in Figure 5-6, reveal not only the apo A139T protein peak (yellow triangles), but also multiple signals corresponding to either PG or PpIX binding. Similar signal intensities and binding stoichiometries are observed when PpIX is added to A139T TSPO samples, and decreased signal intensities are observed upon the addition of PG, with no experiments providing either new MS-resolved features or, with no increases in MS intensity values. We interpret these results as evidence that are primary assumptions in the context of this experiment are likely not valid. In the case of PpIX, if the small molecule penetrates the pre-existing micelle, then clearly it cannot generate a concentration gradient sufficient to increase signal content, or perhaps lack the ability to exchange with pre-existing, more strongly-bound ligands. For PG, if ligand exchange or binding takes place, excess lipid

must also incorporate into the micelle, altering TSPO stability and decreasing our ability to observe ligand binding overall. Taken together with our inability to resolve any significant differences in the intact mass of the bound species observed, it is not possible based on this data alone to identify the endogenous TSPO binder.
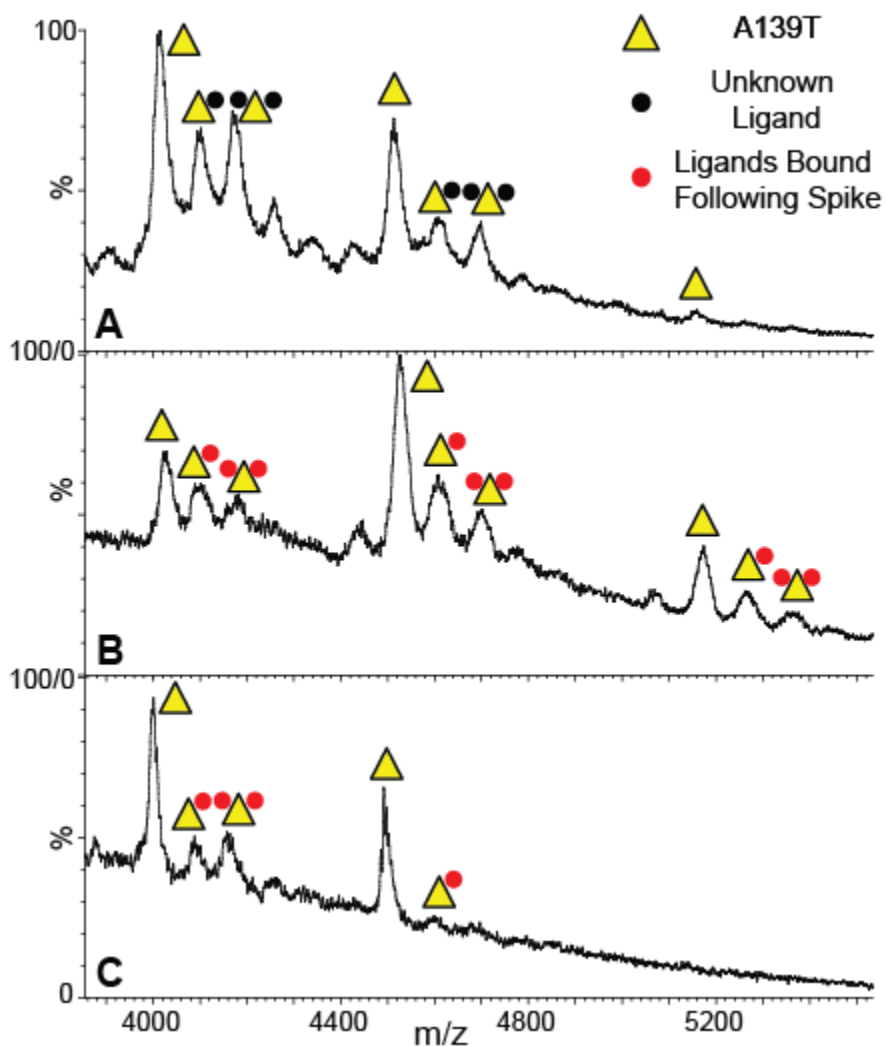
Figure 5-6.MS data of native A139T TSPO variant bound to an unknown endogenous ligand (**A**), as well as for A139T samples incubated with PpIX (**B**) and PG (**C**). The calculated mass for apo dimeric protein are 36119±16 Da (**A**), 36186±3 Da (**B**), 36015±7 Da (**C**), respectively. Ligand bound species are observed in all cases, but their identity cannot be conclusively identified based on the lack of any increases in MS signal intensity or the appearance of new MS features. Consequently, this data also cannot be used to identify the unknown TSPO binding partner shown in the top panel.

5.3.4 CIU Identifies an Unknown Endogenous Ligand Bound to TSPO

In a further attempt to identify the endogenous binding partner for A139T TSPO, we initiated a series of CIU experiments, aiming to develop differentiating gas-phase stability profiles for PG and PpIX ligands bound to TSPO [38]. Data shown in Figure 5-7(A), acquired using the minimal accelerating voltage required to sufficiently desolvate TSPO ions , reveals signals for both single and doubly bound TSPO dimers; whereas larger accelerating voltages, shown in Figure 5-7(B), result in the gas-phase unfolding and ligand dissociation of TSPO ions. A cursory comparison of Figures 5-7A and B reveals atypical CIU stabilities for TSPO complexes attached to endogenous ligands, characterized by an increased stability upon binding the first ligand, and then a gas-phase stability decrease upon binding the second. Figure 5-7C compares this pattern of CIU stability at a fixed acceleration voltage (120V) with identical data acquired for TSPO samples incubated with PG and PpIX. This dataset reveals that PG incubated samples exhibit similar shifts in CIU stability when compared with TSPO bound to the unknown endogenous ligand, while samples incubated with PpIX do not.

154

While the data in Figure 5-7C strongly suggests that the unknown endogenous TSPO binding partner is a PG-like lipid, we worked to confirm this observation through a series of further control experiments. Figure 5-7D shows extensive CID data that monitors the transition between the 2:1 TSPO-ligand complex state and the apo state, acquired over a broad range of accelerating voltages, for A139T variant ions generated from samples with added small molecule binders, as discussed above. As for CIU stability datasets acquired at a single voltage as a function of binding stoichiometry, the CID stabilities of 2:1 TSPO-ligand complex ions bound to the endogenous ligand, match the stabilities for samples incubated with PG, which decay from 65-70% bound to ~40% bound over the voltage range probed, and not PpIX, which retains a 75-80% bound ratio at all instrument settings. In order to test if the CIU and CID stability pattern agreement between PG and the unknown A139T TSPO ligand, we incubated the A138F TSPO variant, which can be isolated without endogenous ligands, with both PpIX and PG for stability measurements. We observe binding of both of these ligands to the A128F variant, and are able to record their CID stabilities, which are also plotted for comparison within Figure 5-7D. These data confirm the stability based assignments made with A139T, and further confirm that a PG-like lipid is the likely endogenous binding partner of observed for this form of TSPO dimer. .
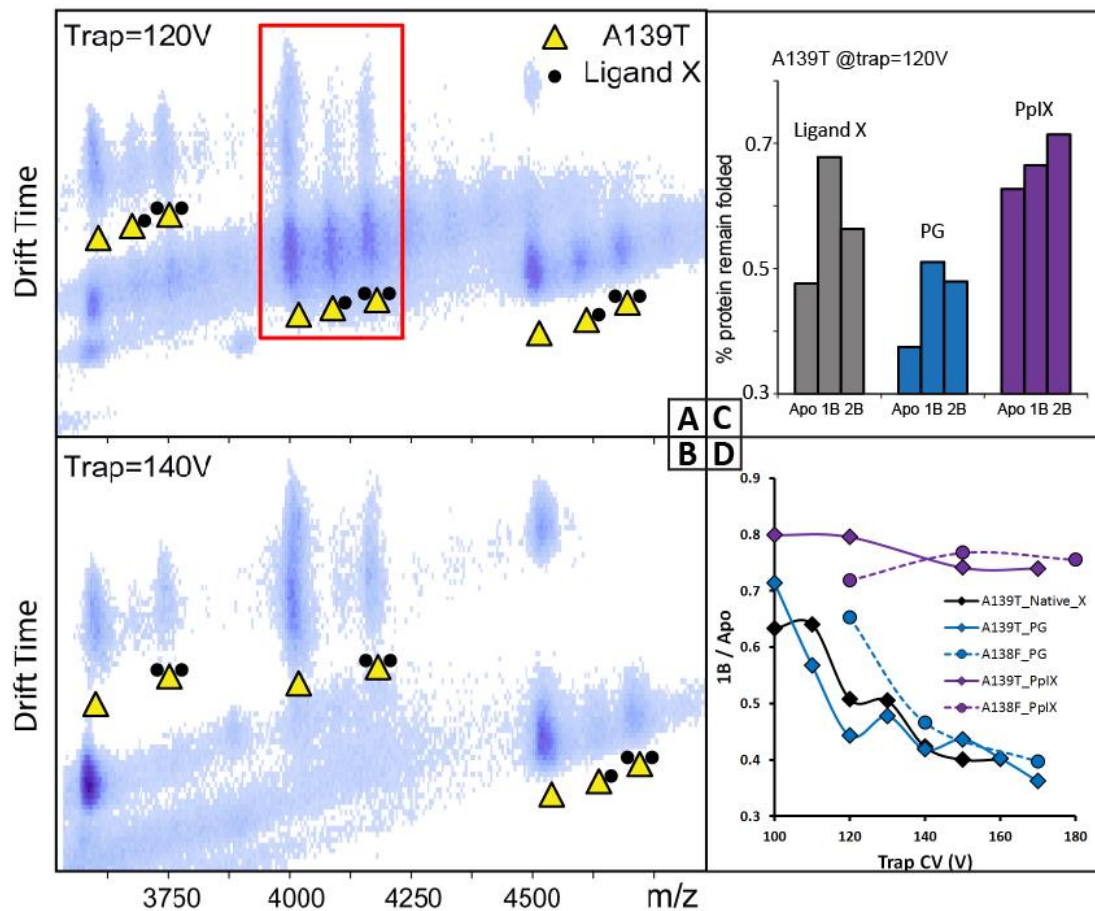
Figure 5-7. CIU/CID analysis of the identity of endogenous unknown ligand. Representative CIU data for endogenous A139T-ligand complex ions at trap collision voltages of 120V(**A**) and 140V(**B**). Yellow triangles represent apo A139T protein ions, black circles represent TSPO-ligand complexes .(**C**)A histogram of CIU stabilities for ions within the red box in (A), showing the percentage of protein signal that remains in its compact state at 120V for Apo, singly (1B) and doubly bound (2B) bounded species. Native A139T bound to the unknown endogenous ligand, and A139T incubated with PG and PpIX are shown in grey, blue and purple, respectively. (**D**)A plot of the gas phase CID stabilities, showing the ratio of 1B/Apo species as a function of trap CV. Native A139T are listed in black, PG and PpIX bounded cases are labeled in blue and purple. Control test with A138F mutant are listed in dashed line.

5.4 Conclusions

Here, the 18-kDa TSPO dimer from *Rhodobacter sphaeroides, as well as its* disease-associated variant forms have been analyzed by IM-MS and CIU assays. Our detergent compatibility screen suggests that Cymal-5 is compatible with wild type *Rs*TSPO, but OG has proven to be the optimum surfactant for IM-MS analysis of all TSPO variants. Trends in our data indicate that the optimum detergent for IM-MS membrane protein analysis should, on the one hand, be able to solubilize and stabilize the membrane protein in solution, as well as survive the nESI process while retaining the non-covalent interactions within the intact membrane protein. In addition, once present in the gas-phase, the surfactant should be easily shed following collisional activation of the proteomicelle. Our data demonstrates that surfactants possessing larger saccharide head groups and longer alkyl chain tend to provide increased membrane stabilization, and that OG is the most 'balanced' yet compatible surfactant for TSPO IM-MS analysis. When compared to other native-MS reports [27, 28, 34, 39], our data strongly suggests that optimal detergent selection is highly protein specific, as the balance referenced above likely depends strongly on both the nature of the detergent the proteomicelle formed.[27, 33]

Our IM-MS data also provides a direct probe of membrane protein ligand binding, and how such events alter protein structure and stability. Optimal buffer and detergent choice leads to the recovery of native-like TSPO CCS values for both dimers and monomers. In addition, for the first time, we utilize a combination of CIU and CID stability data to identify an unknown endogenous ligand bound to a disease-associated variant of the TSPO dimer. Our data strongly

supports a PG-like structure for this unknown binder, a class of molecule observed to interact with TSPO in previous work [19], when compared to a porphyrin binder that was previously suggested by X-ray data[6]. The CIU and CID methods discussed here are among the few techniques available for annotating such ligand binding events for intact membrane protein complexes, with other potential MS workflows likely involving electron-mediated dissociation techniques [40], and has the potential to enable a wide array of therapeutic screening work involving myriad membrane protein targets. Furthermore, the knowledge of specific, endogenous lipid attachment occurs preferentially with the A139T variant, which is linked pregnenolone production enables the construction of new models of TSPO function in this context. Future work will focus on screening additional TSPO binders and lipids to assess the specificity of the CIU/CID fingerprints developed in this report, as well as analyzing further TSPO variant forms. .

5.5 References

1.	Lappano, R. and M. Maggiolini, *G protein-coupled receptors: novel targets for drug discovery in cancer.* Nature reviews Drug discovery, 2011. **10**(1): p. 47-60.

2.	Almén, M.S., et al., *Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin.* BMC biology, 2009. **7**(1): p. 50.

3.	Dorsam, R.T. and J.S. Gutkind, *G-protein-coupled receptors and cancer.* Nature reviews cancer, 2007. **7**(2): p. 79-94.

4.	Overington, J.P., B. Al-Lazikani, and A.L. Hopkins, *How many drug targets are there?* Nature reviews Drug discovery, 2006. **5**(12): p. 993-996.

5.	Papadopoulos, V., et al., *Translocator protein (18kDa): new nomenclature for the peripheral-type benzodiazepine receptor based on its structure and molecular function.* Trends in pharmacological sciences, 2006. **27**(8): p. 402-409.

6.	Li, F., et al., *Crystal structures of translocator protein (TSPO) and mutant mimic of a human polymorphism.* Science, 2015. **347**(6221): p. 555-558.

7.	Braestrup, C., R. Albrechtsen, and R. Squires, *High densities of benzodiazepine receptors in human cortical areas.* 1977.

8.	Gavish, M., et al., *Enigma of the peripheral benzodiazepine receptor.* Pharmacological reviews, 1999. **51**(4): p. 629-650.

9.	Yeliseev, A.A. and S. Kaplan, *TspO of Rhodobacter sphaeroides A STRUCTURAL AND FUNCTIONAL MODEL FOR THE MAMMALIAN PERIPHERAL BENZODIAZEPINE RECEPTOR.* Journal of Biological Chemistry, 2000. **275**(8): p. 5657-5667.

10.	O'Hara, M.F., et al., *Mitochondrial transduction of ocular teratogenesis during methylmercury exposure.* Teratology, 2002. **65**(3): p. 131-144.

11.	Ostuni, M., et al., *Functional characterization and expression of PBR in rat gastric mucosa: stimulation of chloride secretion by PBR ligands.* American Journal of Physiology-Gastrointestinal and Liver Physiology, 2004. **286**(6): p. G1069-G1080.

12.	Faure, J.-P., et al., *Evidence for a protective role of trimetazidine during cold ischemia: targeting inflammation and nephron mass.* Biochemical pharmacology, 2003. **66**(11): p. 2241-2250.

13.	Larcher, J.-C., et al., *Effects of peripheral benzodiazepines upon the O 2 consumption of neuroblastoma cells.* European journal of pharmacology, 1989. **161**(2): p. 197-202.

14.   Torres, S.R., et al., *Anti-inflammatory effects of peripheral benzodiazepine receptor ligands in two mouse models of inflammation.* European journal of pharmacology, 2000. **408**(2): p. 199-211.

15.   Papadopoulos, V., et al., *Peripheral-type benzodiazepine receptor in neurosteroid biosynthesis, neuropathology and neurological disorders.* Neuroscience, 2006. **138**(3): p. 749-756.

16.   Faure, J.P., et al., *Polyethylene glycol reduces early and long-term cold ischemia-reperfusion and renal medulla injury.* Journal of Pharmacology and Experimental Therapeutics, 2002. **302**(3): p. 861-870.

17.   Ducis, I., L.-O.B. Norenberg, and M.D. Norenberg, *The benzodiazepine receptor in cultured astrocytes from genetically epilepsy-prone rats.* Brain research, 1990. **531**(1): p. 318-321.

18.   Leonelli, E., et al., *Ro5-4864, a synthetic ligand of peripheral benzodiazepine receptor, reduces aging-associated myelin degeneration in the sciatic nerve of male rats.* Mechanisms of ageing and development, 2005. **126**(11): p. 1159-1163.

19.   Li, F., et al., *Characterization and modeling of the oligomeric state and ligand binding behavior of purified translocator protein 18 kDa from Rhodobacter sphaeroides.* Biochemistry, 2013. **52**(34): p. 5884-5899.

20.   Korkhov, V.M., et al., *Three-dimensional structure of TspO by electron cryomicroscopy of helical crystals.* Structure, 2010. **18**(6): p. 677-687.

21.   Esser, C., W. Martin, and T. Dagan, *The origin of mitochondria in light of a fluid prokaryotic chromosome model.* Biology Letters, 2007. **3**(2): p. 180-184.

22.   White, S.H., *Biophysical dissection of membrane proteins.* Nature, 2009. **459**(7245): p. 344-346.

23.   Nygaard, R., et al., *The dynamic process of β 2-adrenergic receptor activation.* Cell, 2013. **152**(3): p. 532-542.

24.   Kim, J., et al., *Subnanometre-resolution electron cryomicroscopy structure of a heterodimeric ABC exporter.* Nature, 2015. **517**(7534): p. 396-400.

25.   Wang, H., J. Elferich, and E. Gouaux, *Structures of LeuT in bicelles define conformation and substrate binding in a membrane-like context.* Nature structural & molecular biology, 2012. **19**(2): p. 212-219.

26. Sinning, S., et al., *Binding and orientation of tricyclic antidepressants within the central substrate site of the human serotonin transporter.* Journal of Biological Chemistry, 2010. **285**(11): p. 8363-8374.

27. Laganowsky, A., et al., *Mass spectrometry of intact membrane protein complexes.* Nature protocols, 2013. **8**(4): p. 639-651.

28. Barrera, N.P., et al., *Micelles protect membrane complexes from solution to vacuum.* Science, 2008. **321**(5886): p. 243-246.

29. Levin, M., et al., *Asymmetries in H+/K+-ATPase and cell membrane potentials comprise a very early step in left-right patterning.* Cell, 2002. **111**(1): p. 77-89.

30. Ruotolo, B.T., et al., *Ion mobility–mass spectrometry analysis of large protein complexes.* Nature Protocols, 2008. **3**(7): p. 1139-1152.

31. Bush, M.F., et al., *Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology.* Analytical chemistry, 2010. **82**(22): p. 9557-9565.

32. Benesch, J.L. and B.T. Ruotolo, *Mass spectrometry: come of age for structural and dynamical biology.* Current opinion in structural biology, 2011. **21**(5): p. 641-649.

33. Laganowsky, A., et al., *Membrane proteins bind lipids selectively to modulate their structure and function.* Nature, 2014. **510**(7503): p. 172-175.

34. Bechara, C. and C.V. Robinson, *Different modes of lipid binding to membrane proteins probed by mass spectrometry.* Journal of the American Chemical Society, 2015. **137**(16): p. 5240-5247.

35. Perozzo, R., G. Folkers, and L. Scapozza, *Thermodynamics of protein–ligand interactions: history, presence, and future aspects.* Journal of Receptors and Signal Transduction, 2004. **24**(1-2): p. 1-52.

36. Demmers, J.A., et al., *Electrospray ionization mass spectrometry as a tool to analyze hydrogen/deuterium exchange kinetics of transmembrane peptides in lipid bilayers.* Proceedings of the National Academy of Sciences, 2000. **97**(7): p. 3189-3194.

37. Barrera, N.P., et al., *Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions.* Nature Methods, 2009. **6**(8): p. 585-U49.

38. Rabuck, J.N., et al., *Activation state-selective kinase inhibitor assay based on ion mobility-mass spectrometry.* Analytical chemistry, 2013. **85**(15): p. 6995-7002.

39.   Calabrese, A.N., et al., *Amphipols outperform dodecylmaltoside micelles in stabilizing membrane protein structure in the gas phase.* Analytical chemistry, 2014. **87**(2): p. 1118-1126.

40.   Xie, Y., et al., *Top-down ESI-ECD-FT-ICR mass spectrometry localizes noncovalent protein-ligand binding sites.* Journal of the American Chemical Society, 2006. **128**(45): p. 14432-14433.

Chapter 6

Conclusions and Future Directions

6.1 Conclusions

Here we have demonstrated using collisional unfolding of multiprotein complexes reveals cooperative stabilization upon small molecule binding. (Chapter 2). We use IM-MS and CIU to individually assess the stabilities of Concanavalin A bound states. Our results show that the cooperative stabilization effect, as revealed from CIU assays, cannot easily be detected by solution phase technologies (For example, ITC) or by MS alone. Indeed, this type of stability measurements require the separation of resolved ligand bound states of the protein complex to individually address the stability of each state in isolation. The IM-MS and CIU methods described above are uniquely able to accomplish this and generate stability data that can potentially be used to inform energy calculations aimed at assessing the details of such cooperative effects.[5]

By combining IM-MS and molecular modeling approaches, the underlying mechanism of multi-protein unfolding wassystematically investigated. (Chapter 3) The sugar-binding tetrameric

protein Concanavalin A (Con A) was used as a model system for this process, our results revealed a strong positive correlation between monomeric Coulombic unfolding and tetrameric CIU process. This provides strong evidence that multi-protein unfolding events are initiatedprimarily by charge migration from the complex to a single monomer. Moreover, the number of charges that are transferred in order to achieve the various features observed in Con A CIU were investigated based on our docking simulations. Surprisingly,molecular modeling assays suggested that the majority of the charge transfer (up to 55%) takes place during the first of 3 main unfolding transitions in Con A tetramer, and that this event likely disrupted most of the sugar binding pocket structure within the protein. All these resultswere supported by our previous studies of Con A tetramer, and provide critical, predictive insights for understanding the fundamentalmechanism of multi-protein unfolding.

Probing the binding strength of protein assemblies is critically important in understanding their interaction and cellular functions. The interactions between HDAC8 and PCBP1 were systematically investigated by IM-MS. (Chapter 4) Our data suggest that HDAC8 and PCBP1 interact with each other in a specific manner, as confirmed by tandem MS and negative control experiments with carbonic anhydrase II. With optimized buffer conditions and careful adjusted instrument tuning, we for the first time expanded the utility of the classic ESI-MS method onto assessing the binding affinity for HDAC8 - PCBP1 hetero-complex, and reported the K_D of H-P complex 68 ± 30 μM.The non-negligible variances of the measured K_D results are majorly

generated from two sources: A) error of the ESI-MS method itself and B) batch differences of protein sample. Systematic statistical analysis suggests that with 95% confidence, the variances of the second batch measurements are at least 2.11 times big as that of the first batch. Such results demonstrate the fact switching to different batch of protein sample may generate major deviation for K_D calculation, which in turn confirms the usefulness of our methodology. Analytically speaking, in ideal cases, sufficient sample supply with equal quality is strongly recommended for this type of binding affinity estimation. In addition, as for the metal substitution (with Zn^{2+}/Fe^{2+}) experiments,our data suggested that Fe^{2+} mildly promotes the binding between HDAC8 and PCBP1. Given the scenario that Zn^{2+} has long viewed as a natural binder to HDAC8 in cell,[6, 7] our results suggested that iron, at the very least, performed similarly to zinc as an HDAC8 binder *in vitro*.

In Chapter 5, the TSPO dimer from *Rhodobactersphaeroides,* as well as its disease-associated variant forms, has been analyzed by IM-MS and CIU assays. According to our detergent compatibility screen test,OG has proven to be the optimum surfactant for IM-MS analysis of all TSPO variants. In addition, our IM-MS data also provided a direct probe of membrane protein ligand binding, and elucidated the extent to which such events alter protein structure and stability. Optimal buffer and detergent choice lead to the recovery of native-like TSPO CCS values for both TSPO dimers and monomers. In addition, for the first time, we utilized a combination of CIU and CID stability data to identify an unknown endogenous ligand bound to a

166

disease-associated variant of the TSPO dimer.Our data strongly supported a PG-like lipid structure for the unknown binder.The CIU and CID methods employed in this workwere among the few techniques available for annotating such ligand binding events for intact membrane protein complexes, with other potential MS workflows likely involving electron-mediated dissociation techniques [8], and has the potential to enable a wide array of therapeutic screening work involving myriad membrane protein targets.

6.2 Future Directions

6.2.1 CIU of large multi-dentate protein-ligand assemblies

The underlying mechanism of protein CIUwas systematically evaluated in Chapter 2, as well as in other previous studies. That previous work, however, was mainly focused on proteins bound to a four or less small molecule ligands. In the future, the CIU mechanism for even larger multi-proteinsystems, that exhibit larger-order multi-dentate binding behavior should be probed.RuBisCo (ribulose-1,5-bisphosphate carboxylase oxygenase), an assembly thought to be themost abundant protein in nature, would serve as a excellent model system for such mechanistic studies.[9]The complex has a molecular mass of 550kDa,consisting of 16 subunits - 8 large and 8 small. Rubisco is the key enzyme of both the carbon reductive pathway (*i.e.,* the Calvin cycle) and of photorespiration. It catalyzes the carbon-fixing reaction in which a molecule of CO_2 is added to the enediol form of ribulose-1,5-bisphosphate (RuBP), which is a double

phosphate ester of the ketose (ketone-containing sugar) called ribulose, with a molecular weight of 310 Da.[10-12]It would be of great interest is to map the well-established CIU protocol for small protein complexes onto the 550 kDa assembly, and attempt to detect the influence of ligand binding and the unfolding pathway(s) observed.
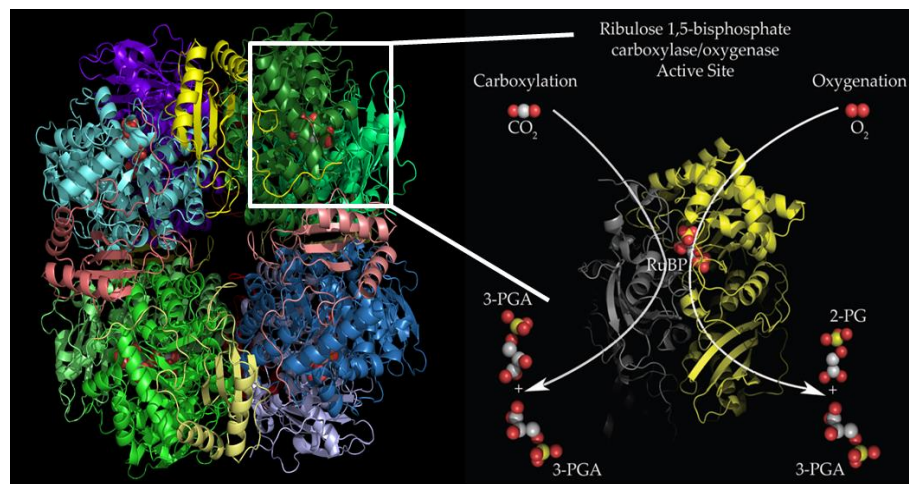

Figure 6-1.(**Left**)X-ray structure of hexadecameric Rubisco.(PDB: 8RUC) (**Right**)Theactive binding site of Rubisco. The basic catalytic unit of all Rubisco molecules is a dimer of large catalytic subunits in which each monomer interacts with the N-terminus of the second monomer. The active site thus contains residues from primarily the C-terminus of the one monomer, but also includes residues from the N-terminus of the second monomer, conferring two active sites per dimer.

6.2.2 An alternative protein model system for studying the details of multi-protein CIU.

The mainhypothesis that work shown in chapter 3 attempts to validate states that multi-protein unfolding events areinitiated by charge migration from the complex to a single monomer.Chapter 3 uses the Con A tetramer to systematically investigate this claim, but in order to evaluate the universality of our findings, additional model systems must be similarly evaluated. Multi-protein systems, ranging from dimeric proteins like β-lactoglobulin, to other tetramers like avidin or transthyretin,extending to even larger protein multimers could all be used in such studies. Ideally such systems would cover a broad range of protein folds, domain archetypes, and sequence bacisities in order to establish the true breadth of the predictions generated by our models.

Another factor that could be validated in a potentially more rigorous way, is the scaling factor used to related PA CCS calculations to X-ray model structures. As mentioned in Chapter 3, a value of 1.15 was developed an extensively validated for globular protein models, and our work in chapter 3 creates a new set of scaling factors that relate unfolded models to their corresponding TM CCS estimate. There are two routes toward broadly evaluating the generality of the scaling coefficientswe discuss in Chapter 3. The first is to use TM methods to calibrate the PA results more broadly,using advanced computing; the other is toutilize next-generation CCS calculators. For example, IMPACT[13] provides ultra-high speed CCS computing based on a projection method, but this method is well benchmarked to convert such values to those that emulate TM computations for compact globular proteins and complexes .

6.2.3 Searching for Ligand-specific CIU Fingerprints in the TSPO-LAF Variant

We studies both WT and disease-associated variants of *Rs*TSPO in Chapter 5. The cholesterol binding behavior of TSPO would be an potentially interesting and impactful area of study. This would also be highly challenging using the variants from Chapter 5, as it has been reported that *Rs*TSPO shows more than 1000-fold lower affinity for cholesterol binding than mammalian TSPO.[14] Therefore another variant, *Rs*TSPO-LAF, in which the three residues (A136-T137-A138) that precede the cholesterol binding motif are replaced with the human sequence LAF, has been constructed in order to understand how such binding events could affect TSPO structure and stability. [15, 16]Preliminary datashown in figure 6-2 indicates that the dimeric TSPO structure for this variant is largely disrupted, and protein existsprimarly as highly unfolded monomer under our optimized detergent conditions. In addition, putative mass assignments and preliminary CIU assays reveal the evidence of yet another native endogenous ligand bound to the protein. To fully understand the structure and ligand binding behavior of this LAF variant, further experiments, focused on optimizing detergent and IM-MS analysis condition would need to be conducted. Following this work, controls and workflows simlar to those described in Chapter 5 could be used to decipher the identity of unknown binder.In addition, a variety of other TSPO ligands, including cholesterol, curcumin, Gossypol, retinoic acid, as well as the diagnostic drug PK11195, (1-(2-Chlorophenyl)-N-methyl-N-(1-methylpropyl)-3-

isoquinolinecarboxamide)[15]could be screened to generate comprehensive CIU/CID profiles capable of discriminating a range of different TSPO binders. Once completed, such a dataset could be further utlized to perform medium or high-throughput assays aimed at developing TSPO binding inhibitors that may be impactful in the treatment in human diseases ranging from psychological disorders to cancer.
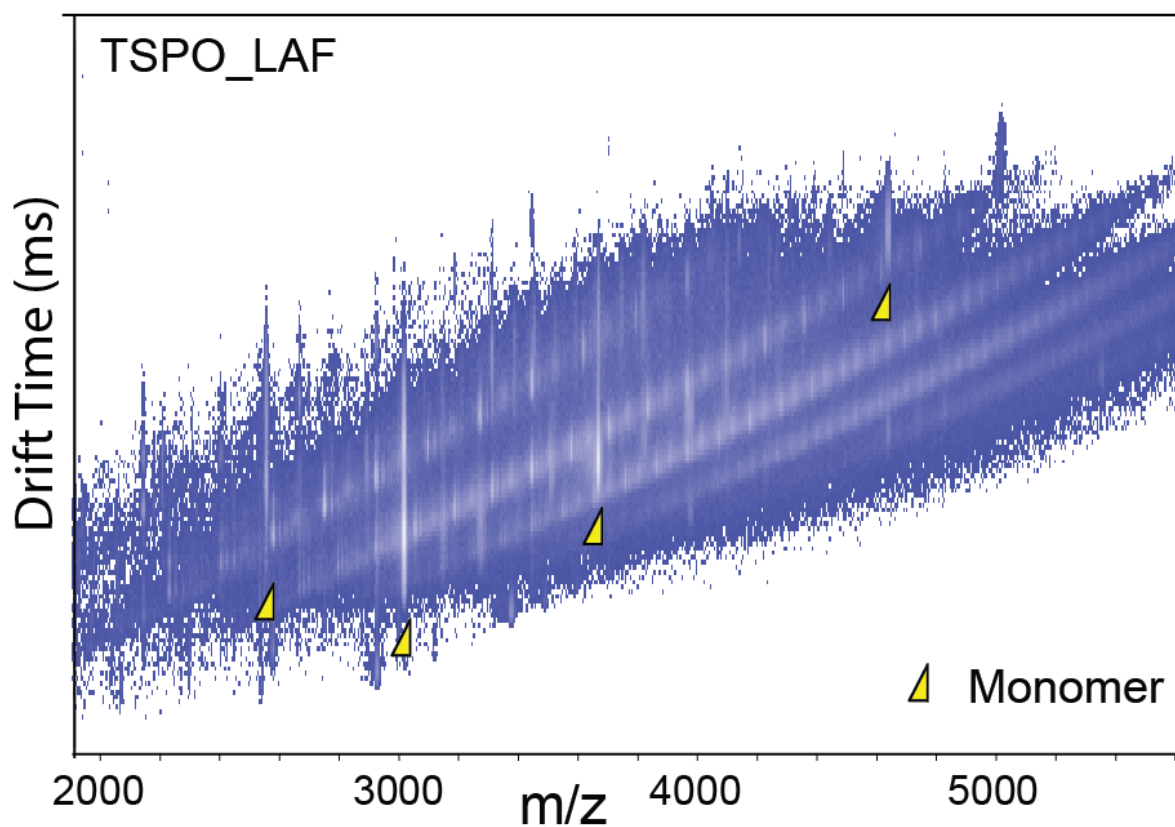

Figure 6-2. Preliminary IM-MS data for TSPO LAF mutant in the presence of cholesteryl hemi-succinate (CHS) at Trap voltage 140V, with 2xCMC of OG as the detergent. The protein dimer structure is largely disrupted and exists in its monomeric form under these conditions.

6.3 References

1.   Niu, S., J.N. Rabuck, and B.T. Ruotolo, *Ion mobility-mass spectrometry of intact protein–ligand complexes for pharmaceutical drug discovery and development.* Current opinion in chemical biology, 2013. **17**(5): p. 809-817.

2.   Geoghegan, K.F. and M.A. Kelly, *Biochemical applications of mass spectrometry in pharmaceutical drug discovery.* Mass spectrometry reviews, 2005. **24**(3): p. 347-366.

3.   Hyung, S.-J., C.V. Robinson, and B.T. Ruotolo, *Gas-phase unfolding and disassembly reveals stability differences in ligand-bound multiprotein complexes.* Chemistry & biology, 2009. **16**(4): p. 382-390.

4.   Hopper, J.T. and N.J. Oldham, *Collision induced unfolding of protein ions in the gas phase studied by ion mobility-mass spectrometry: the effect of ligand binding on conformational stability.* Journal of the American Society for Mass Spectrometry, 2009. **20**(10): p. 1851-1858.

5.   Niu, S. and B.T. Ruotolo, *Collisional unfolding of multiprotein complexes reveals cooperative stabilization upon ligand binding.* Protein Science, 2015.

6.   Finnin, M.S., et al., *Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors.* Nature, 1999. **401**(6749): p. 188-193.

7.   Barnes, P., I. Adcock, and K. Ito, *Histone acetylation and deacetylation: importance in inflammatory lung diseases.* European Respiratory Journal, 2005. **25**(3): p. 552-563.

8.   Xie, Y., et al., *Top-down ESI-ECD-FT-ICR mass spectrometry localizes noncovalent protein-ligand binding sites.* Journal of the American Chemical Society, 2006. **128**(45): p. 14432-14433.

9.   Ellis, R.J., *The most abundant protein in the world.* Trends in Biochemical Sciences, 1979. **4**(11): p. 241-244.

10.   Knight, S., I. Andersson, and C.-I. Brändén, *Crystallographic analysis of ribulose 1, 5-bisphosphate carboxylase from spinach at 2· 4 Å resolution: Subunit interactions and active site.* Journal of molecular biology, 1990. **215**(1): p. 113-160.

11.   Badger, M.R. and G.H. Lorimer, *Interaction of sugar phosphates with the catalytic site of ribulose-1, 5-bisphosphate carboxylase.* Biochemistry, 1981. **20**(8): p. 2219-2225.

12.   Blayney, M.J., S.M. Whitney, and J.L. Beck, *NanoESI mass spectrometry of Rubisco and Rubisco activase structures and their interactions with nucleotides and sugar phosphates.* Journal of the American Society for Mass Spectrometry, 2011. **22**(9): p. 1588-1601.

13.   Marklund, E.G., et al., *Collision Cross Sections for Structural Proteomics.* Structure, 2015. **23**(4): p. 791-799.

14.   Li, F., et al., *Identification of a key cholesterol binding enhancement motif in Translocator protein 18 kDa.* Biochemistry, 2015. **54**(7): p. 1441-1443.

15.   Li, F., et al., *Characterization and modeling of the oligomeric state and ligand binding behavior of purified translocator protein 18 kDa from Rhodobacter sphaeroides.* Biochemistry, 2013. **52**(34): p. 5884-5899.

16.   Li, F., et al., *Crystal structures of translocator protein (TSPO) and mutant mimic of a human polymorphism.* Science, 2015. **347**(6221): p. 555-558.