Development of Ion Mobility-Mass Spectrometry as a High-throughput Approach for Structural Genomics

by

Yueyang Zhong

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2014

Doctoral Committee

Assistant Professor Brandon T. Ruotolo, Chair
Professor Charles L. Brooks III
Professor Kristina Håkansson
Professor Janet L. Smith

Acknowledgement

As the journey pursuing my Ph.D. is reaching the end, I would like to take this opportunity to thank all those who supported and helped me during the past years.

First and foremost, I would like to express my deepest gratitude to my Ph. D. advisor Dr. Brandon T. Ruotolo, for his great guidance and enormous patience during the past years. This work could not become even close to the way it is now without his encouragement and critical suggestions. I am always inspired by Brandon's insight and passion in science. His diligence and profound knowledge set him a role model for me as a scientist. It has been truly a privilege to work with Brandon.

I would like to give great thanks to my dissertation committee members for their valuable advices during my candidacy, data meeting and dissertation. In addition to that, I would like to thank Dr. Kristina Håkansson for guiding me into the world of mass spectrometry during the class and rotation, Dr. Charles L. Brooks III for providing solid computational assistance, and Dr. Janet L. Smith for offering her expertise in protein structures.

It is my great pleasure to work in the Ruotolo lab with a group of passionate and intelligent people. I would like to thank my past and present colleagues, Dr. Suk-Joon Hyung for training me using ion mobility-mass spectrometry, Linjie Han for enormous help and inspiring discussion on my experiments, Joe Eschweiler for helping with

computational working during the past months, Jessica Rabuck for offering Abl and Src proteins and delicious birthday cake, Dr. Billy Samulak for her great advices on my dissertation, Russ Bornschein for helpful discussions and sharing parenting moments, Molly Soper for insightful suggestions on all of my presentations, and Shuai Niu for teaching me using linux-based software. I sincerely appreciate their kindness and friendship.

I would also like to thank my collaborators who allow me to expand our technique to various systems, including Dr. Jun Feng from Brooks group, Dr. Stephen Ragsdale and his post-docs Ireena Bagai, Angela Fleischhacker and Andrea Morris Spencer, Dr. Hangtian Song from Hakansson group, Dr. Carol Fierke and her students Xin Liu and Michael Howard, and Dr. Robert Hausinger and his student Mark Farrugia from Michigan State University. I would also like to acknowledge my rotation advisors, Dr. Mark Banaszak Holl and Dr. David Lubman for their scientific training during my first year. I would also like to extend my thanks to my funding supports, primarily from national institute of health (NIH) and University of Michigan chemistry department.

Beyond the scope of research, I am thankful to my dear friends who share laugh and tears with me and cheer me up as always in my graduate life. Special thanks to Dr. Bo Peng, Si Yang, Junyoon Lee and Dr. Jing Nie for the moments we spent together and to Dr. Mou-Chi Cheng and Shu-Lan Hung for taking care of me like a family member. At last but not least, this dissertation could not be accomplished without the endless encouragement and support from my beloved family. I want deliver my sincere appreciation to my parents, my husband and my baby daughter for their unconditional love.

Table of Contents

Acknowledgement ii
List of Figures viii
List of Tables x
List of Abbreviations xi
Abstract xiv
Chapter 1. Introduction 1
1.1 Tools for Protein Structural Studies 5
1.2 Mass Spectrometry 7
1.2.1 Electrospray Ionization (ESI) 9
1.2.2 Quadrupole Mass Filters 10
1.2.3 Time of Flight (ToF) Mass Analyzers 11
1.3 Mass Spectrometry Study for Protein Structures and Interactions 12
1.3.1 Mass Spectrometry of Intact Non-Covalent Protein Complexes 12
1.3.2 Hydrogen-Deuterium Exchange (HDX) and Oxidative Foot-Printing 13
1.3.3 Chemical Crosslinking 14
1.4 IM-MS Technology 15
1.4.1 Drift –Tube (DT) 17
1.4.2 Differential Mobility Analyzer (DMA) 18
1.4.3 Traveling-Wave (T-wave) 18
1.5 IM-MS Data Interpretation 23
1.5.1 MS Data Analysis 23
1.5.2 IM Data Analysis 24
1.6 Optimizing the Information Content of IM-MS for Structural Proteomics 26
1.6.1 Gas Phase Activation 26

1.6.2 Solution Phase Disruption 28
1.6.3 Protein Structure Stabilization 30
1.7 Integrating Bottom-up Proteomics with Intact Protein IM-MS 31
1.8 Summary 34
1.9 References 37
Chapter 2. Characterizing the Resolution and Accuracy of a Second-Generation Traveling-Wave Ion Mobility Separator for Biomolecular Ions 45
2.1 Introduction 45
2.2 Experimental 49
2.2.1 Sample Preparation 49
2.2.2 Synapt G2 Instrumentation 50
2.2.3 Data Collection and Analysis 51
2.3 Results and Discussion 52
2.3.1 Trends in Drift Time Resolution on the Second-Generation Traveling-Wave Platform 52
2.3.2 Optimizing Conditions for Maximum Cross-section Accuracy 56
2.3.3 The Influence of Ion Heating on Traveling Wave Resolution and Cross Section Accuracy 58
2.3.4 Collision Cross-Section Resolution Compared with Drift Time Resolution for Traveling-Wave Instruments 61
2.4 Conclusions 64
2.5 Supplemental Information 67
2.6 References 70
Chapter 3. Robotically-Assisted Titration Coupled to Ion Mobility-Mass Spectrometry Reveals the Interface Structures and Analysis Parameters Critical for Multiprotein Topology Mapping 72
3.1 Introduction 72
3.2 Experimental 76
3.2.1 Sample Preparation 76
3.2.2 Ion Mobility-Mass Spectrometry 77
3.2.3 X-ray and IM Data Comparisons and Computational Analysis 79
3.3 Results and Discussion 80
3.3.1 Evaluating the Fidelity of IM-MS Derived Protein Topology 82

3.3.2 Solution Disruption followed by IM-MS Reveals the Details of Protein-Protein Interface Chemistry 89
3.4 Conclusions 94
3.5 Supplemental Information 97
3.6 References 106
Chapter 4. The Collisional and Coulombic Unfolding of Gas-Phase Proteins is Highly-Correlated to Their Domain Structure in Solution 109
4.1 Introduction 109
4.2 Experimental 113
4.2.1 Sample Preparation 113
4.2.2 IM-MS Instrument and Methods 115
4.2.3 Data Analysis 115
4.3 Results and Discussion 116
4.3.1 Estabilising a General Correlation between Gas-Phase Unfolding and Protein Domain Number 119
4.3.2 Broadening the Unfolding Dataset using Proteins that are Predominantly α-helical or β-sheet 122
4.3.3 Selecting Optimal Charge States for CIU Protein Domain Analysis 126
4.4 Conclusions 127
4.5 Supplemental Information 131
4.6 References 137
Chapter 5. Applying IM-MS to Protein Complexes with Unknown Topology and Interface Structure 141
5.1 Introduction 141
5.2 Studying the Heme Oxygenase-2 (HO-2) and NADPH-Cytochrome P450 Reductase (CPR) interaction with IM-MS 142
5.2.1 Background Information about Heme Degradation 142
5.2.2 Methods and Materials 144
5.2.3 Results and Discussion 145
5.3 Topology Construction for Constructing an IM-MS-derived Topology for the Human Mitochondrial Ribonuclease P (RNase P) subcomplex 152
5.3.1 Background Information on Human Mitochondrial RNase P 152
5.3.2 Methods and Materials 153
5.3.3 Results and Discussions 154

5.4 Conclusions and Future Directions 157
5.5 References 159
Chapter 6. Conclusions and Future Directions 161
6.1 Conclusions 161
6.2 Future Directions 162
6.2.1 Refine Protein-Protein Interface Study with Defined Protein Complexes 162
6.2.2 Extend the Application on Assemblies with Unknown Architectures, with the Assistance from other Techniques 163
6.2.3 Antibody Disulfide Mapping 165
6.3 References 167

List of Figures

Figure 1-1 The challenge of structural genomics: converting protein interaction networks into protein structures. 4
Figure 1-2 A schematic diagram of the Synapt G2, quadrupole-ion mobility-time-of-flight mass spectrometry instrument used in these studies. 8
Figure 1-3 Ion mobility-mass spectrometry data acquisition and basic principles. 16
Figure 1-4 A schematic diagram illustrating key differences between the first and second generations of Synapt instrumentation, focusing on the TW ion guides and regions directly surrounding the TWIM separator. 21
Figure 1-5 A high-throughput screening process to discover optimal solution conditions for protein complex topology mapping by IM-MS. 28
Figure 1-6 A flow diagram for an integrated IM-MS structural proteomics workflow. 33
Figure 2-1 Charting the resolution of the second-generation traveling-wave ion mobility separator on the Synapt G2. 53
Figure 2-2 The goodness of fit for calibration curves that allow for the conversion between IM drift time measurements and collision cross-section accuracy depend primarily upon wave velocity, and do not have a strong wave height dependence. 57
Figure 2-3 Exploring the influence of ion heating on the accuracy of collision cross section measurements using a traveling wave ion mobility device. 59
Figure 2-4 Achieving optimum collision cross-section resolution. 63
Figure S. 2-1 Traveling wave drift time resolution versus *[WH]/[WV]*, shown at each *[WH]* measured. 67
Figure S. 2-2 Data for calibration curve exponential factor X 68
Figure S. 2-3 Collision cross section resolution of peptide ions at different pressures. 69
Figure 3-1 IM-MS data collection and analysis procedures. 80
Figure 3-2 Evaluating the fidelity of IM-MS derived protein topology. 82
Figure 3-3 Comparison between the IM distributions and centroids recorded for intact tetramer ions generated under control (200mM NH_4Ac, upper panel) and disruption conditions (lower panel). 85
Figure 3-4 AVD topology models from IM-MS data are brought into close agreement with X-ray when disruption conditions incorporating high ionic strength are used. 87
Figure 3-5 Solution disruption followed by IM-MS reveals the details of protein-protein interface chemistry. 91
Figure S. 3-1 A comparison between the m/z and intact mass recorded for tetramer ions generated under control (200mM NH_4Ac, upper panel) and optimized disruption conditions (lower panel) discussed in Figure 3-3. 100
Figure S. 3-2 The positions of the disruption conditions shown in Table S.3-1 within the 2D titration maps shown in Figure 3-5A. 102

Figure S. 3-3 Comparing X-ray and IM-MS data provides critical insight into the structural deformations observed in AVD and CAT subcomplexes. 103
Figure S. 3-4 Correlations between IM-MS disruption experiments using a three-component titration strategy and salt bridge strength calculations from X-ray structure data 104
Figure S. 3-5 Correlations between our IM-MS titration experiments (X axis) and X-ray structure data (Y axis), shown as a scatter plot with fitted trend lines. Dashed lines indicate the 95% confidence interval for each fit. 105
Figure 4-1 A diagram illustrating IM-MS data collection and analysis procedures. 117
Figure 4-2 Collisional unfolding (top) and Coulombic unfolding (bottom) data of poly-ubiquitin domains linked in a linear fashion. 120
Figure 4-3 X-ray structure (top), collisional unfolding (middle) and Coulombic unfolding (bottom) data for predominantly α-helical proteins. 123
Figure 4-4 X-ray structure (top), collisional unfolding (middle) and Coulombic unfolding (bottom) data predominantly β-sheet proteins. 125
Figure 4-5 Optimal charge states for CIU protein domain analysis 127
Figure S. 4-1 Coulombic unfolding data for proteins discussed in the main text of this chapter, including 1-4 ubiquitins, 3 primarily α-helix proteins and 3 primarily β-sheet proteins. 133
Figure S. 4-2 Collisional unfolding data of non-linear di-ubiquitins, K48 and K63 linked, at charge states 7^+ and 8^+. 134
Figure S. 4-3 Crystal structure (left), collisional unfolding (middle) and Coulombic unfolding (right) data for NADPH-Cytochrome P450 Reductase (CPR). 134
Figure S. 4-4 Crystal structures (left), collisional unfolding (middle) and Coulombic unfolding data (right) for SCOP/CATH inconsistent proteins, Abelson protein tyrosine kinase and Transferrin. 135
Figure S. 4-5 Crystal structure (top, left), Coulombic unfolding (bottom left) and CIU (right) data of Biliverdin Reductase (BVR). 136
Figure 5-1 2D titration experiments on the HO-2+CPR complex. 147
Figure 5-2 NH_4Ac required for dissociation at 10% DMSO versus the number of polar contacts for both calibrants (black) and HO-2 1-288/CPR complex. 148
Figure 5-3 Titration experiment increasing the NH_4Ac concentration at fixed DMSO content in solution for HO-2 1-288/CPR (black) and HO-2 1-248/CPR complexes (red). 149
Figure 5-4 HO-2 structures with predicted C-terminal (243-288). 150
Figure 5-5 Docking simulation for optimized HO-2/CPR complex model using ZDOCK. 151
Figure 5-6 Ion mobility-mass spectrometry (IM-MS) analysis of MRPP1·MRPP2 complexes. 155
Figure 5-7 Analysis of the modeling output for the MRPP1/2 heterohexamer. 157
Figure 6-1 Simulated working model of urease activation. 164
Figure 6-2 Collisional unfolding data of IgG heavy chain, with intact disulfide bonds or reduced disulfide bonds. 166

List of Tables

Table S. 3-1 Solution conditions optimized for protein complex disruption, used in Figure 3-3. 101
Table S. 4-1 Detailed information for proteins used in this chapter, including protein names, organisms, masses, PDB codes, secondary structure content (α helix and β sheet percentage) and domain numbers defined by SCOP (red) and CATH (green). 131

List of Abbreviations

3D	3 dimensional
Abl	Abelson protein tyrosine kinase
AC	Alternating current
ACN	Acetonitrile
ADH	Alcohol dehydrogenase
ALD	Aldolase
AUI	Advanced user interface
AVD	Avidin
βGL	β-galactosidase
BVR	Billverdin reductase
CAT	Catalase
CATH	Class, architecture, topology and homologous superfamily
CCS	Collision cross section
CG	Coarse Grain
CID	Collision induced dissociation
CIU	Collision induced unfolding
CON/ConA	concanavalin A
CPR	NADPH-dependent cytochrome P450 reductase
CRM	Charge residue model
CXL	Chemical cross linking
DC	Direct current
DMA	Differential mobility analyzer
DMS	Differential mobility spectrometry
DMSO	Dimethyl sulfoxide

DNA	Deoxyribonucleic acid
DT	Drift time/Drift tube
EM	Electron microscopy
ESI	Electrospray ionization
FAIMS	Field asymmetric waveform ion mobility spectrometry
FTICR	Fourier transform ion cyclotron resonance
FWHM	Full width at half maximum
GDH	Glutamate dehydrogenase
GFP	Green fluorescent protein
GST	Glutathione S-transferase
HDX	Hydrogen/deuterium exchange
HO	Heme oxygenase
IEM	Ion evaporation model
IM	Ion mobility
IMS	Ion mobility spectrometry
kDa	Kilo dalton
LC	Liquid chromatography
m/z	Mass-to-charge ratio
MALDI	Matrix-assisted laser desorption ionization
MMA	Mass measurement accuracy
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
mt	Mitochondrial
nESI	Nano electrospray ionization
NH4Ac	Ammonium acetate
NMR	Nuclear magnetic resonance spectroscopy
PA	Projection approximation
PDB	Protein data bank
PKI	Pyruvate kinase
PSA	Projection superposition approximation

Q	Quadrupole
RF	Radio frequency
RNase P	Ribonuclease P
SAXS	Small angle x-ray scattering
SCOP	Structural classification of proteins
SD	Standard deviation
SID	Surface induced dissociation
ToF	Time of flight
Trap CE	Trap collisional energy
tRNA	Transfer Ribonucleic acid
TROSY	Transverse relaxation optimized spectroscopy
TTR	Transthyretin
T-wave/TW	Traveling wave
UBQ	Ubiquitin
WH	Wave height
WL	Wave length
WV	Wave velocity

Abstract

Ion mobility-mass spectrometry measurement of proteins and multi-protein complexes is a tool of rapidly-growing importance in structural biology. However, many challenges remain in its development, including: optimizing protein size/shape measurements, developing methods for protein-protein interface characterization, and constructing high-throughput platforms for multiprotein topology determination. This thesis focuses on these challenges by developing new methods for ion mobility-mass spectrometry protein structure characterization.

First, the performance characteristics of a second-generation travelling-wave ion mobility separator are assessed, focusing on those parameters that lead to the collection of high-accuracy, high-precision measurements of protein size. The conditions for high accuracy protein size measurements are significantly different from those optimized for separation resolution, indicating that a balance between these two metrics must be attained for traveling wave ion mobility separations of biomolecules.

Second, in order to enable the high-throughput structural analysis of protein complexes and their subcomplexes, ion mobility-mass spectrometry is coupled with automated robotic sampling of carefully-titrated solution conditions. By altering solution ionic strength in concert with dimethyl sulfoxide content, the data collected shows that simple two-dimensional solvent screens are sufficient to disrupt protein-protein interfaces for a

broad array of complex structures and folds. Ion mobility measurements captured for both intact assemblies and subcomplexes matched expected values from available X-ray data in all cases save two, where extreme disruption conditions were employed. Strong correlations between the disruption agents and chemical nature of interfacial interactions are observed.

A key challenge for protein ion mobility measurements of intact proteins is accessing local, domain-level structure information. In a third set of experiments, gas-phase protein unfolding data were acquired for a range of monomeric proteins. The unfolding of multi-domain proteins, using either collision-induced unfolding or Coulomb-associated stretching, revealed a strong, positive correlation with known protein domain structures in solution.

In a final set of experiments, the protocols developed here using model protein systems are applied to two multiprotein complexes of currently unknown structure: the heme oxygenase-2/cytochrome P450 reductase dimer and the human mitochondrial ribonuclease P hetero-hexamer. In these cases, IM-MS aided by elucidating either the chemical nature of the protein contacts or coarse-grained topologies of the complexes studied.

Chapter 1. Introduction

Protein structures play key roles in protein functions. From primary sequence to quaternary structure, deviation at any level can cause malfunctions that lead to severe disease.[1] Proteins rarely act in isolation, resulting in massive, dynamic multiprotein machines and networks that govern most critical cellular processes.[2] As such, the structural characterization of these large-scale multi-component protein complexes has been a key goal of structural biology for decades.[3] By the same token, protein complexes comprise most of the drug targets currently sought in treatment efforts for many human diseases. More recently, attention has turned to high-throughput technologies capable of assessing the structure and topology of many complexes in a serial fashion. The ambitious goals set in such studies include the structural annotation of large protein networks, thus enabling a three-dimensional view of vast webs of interacting proteins.[4] High-throughput versions of X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy have been important and pioneering tools in this process, and seek to directly convert isolated proteins and complexes into atomic-resolution representations of the interacting proteins involved in such networks (Figure 1-1).[5] Such technologies currently dominate the landscape of structural genomics and proteomics research.

While this direct approach has been highly successful for a large number of protein complexes and networks, the vast majority of multiprotein systems provide significant

challenges for X-ray and NMR-based approaches. For example, although detection limits for both technologies continue to improve, relatively large amounts of protein are still required to acquire usable data. In addition, as the complexity of the protein network under investigation increases, so do many parameters that complicate NMR and X-ray analysis, such as the increased presence of protein flexibility, heterogeneity, and polydispersity.[3,5] These properties are found in abundance within membrane-associated protein complexes[6] - a class of protein assemblies that are among the most-sought after as therapeutic targets.[7] Further, since neither technology seeks to separate components during analysis, both require highly purified samples. These and other challenges draw into dramatic relief the necessity for alternative approaches to solving multiprotein topology in a high-throughput framework.[3,5,8] To supplement X-ray and NMR protein structure data, information from lower-resolution tools, with less-stringent sample requirements, are complementary methods in structural analysis for protein complexes such as small angle X-Ray scattering (SAXS) or cryo-electron microscopy (cryo-EM), are frequently integrated to provide models of protein complexes. Rather than atomic details, these techniques focus on the shape, topology, and connectivity of interacting protein components[8].

Mass spectrometry (MS), and more-recently ion mobility-MS (IM-MS), of intact complexes is emerging as one of many such alternative approaches in the field of structural proteomics.[3,5,9-15] In these experiments, an alternate route to a final high-resolution map of the protein complex is sought, where protein connectivity and topology are defined in stages, prior to detailed homology modeling or partial atomic-resolution data is used to complete the process (Figure 1-1).[16,17] For many years, MS technology,

driven by the needs of proteomics and other biology-related application areas, has been honed as a high-throughput methodology. This makes MS a more obvious choice for high-throughput data production than many other tools capable of producing protein structure data. Further, MS of intact protein complexes has been highly successful in determining the details of protein-protein interaction within assemblies.[14,18-20] Such two-dimensional contact maps (Figure 1-1) are generated by exhaustive measurements of proteins and sub-complexes produced upon the disruption of intact assemblies.[17,18,21] Such disruption experiments can be carried out both in solution [21,22] and in the gas-phase,[23,24] and both approaches will be covered in some depth here. In this way, three-dimensional models are built by integrating distance and size constraints from IM-MS measurements, along with other structural biology datasets and molecular modeling.[16,25]

IM-MS, as the chief technology used to generate data in this thesis, will be discussed in detail, including topics such as: the evolution of IM-MS technology, data interpretation, methods of generating enhanced IM-MS data for protein complexes, and the integration of IM-MS with established proteomics protocols in efforts to define a structural picture of protein interaction networks. IM-MS will also be discussed broadly in the context of the analysis of intact proteins and their complexes.

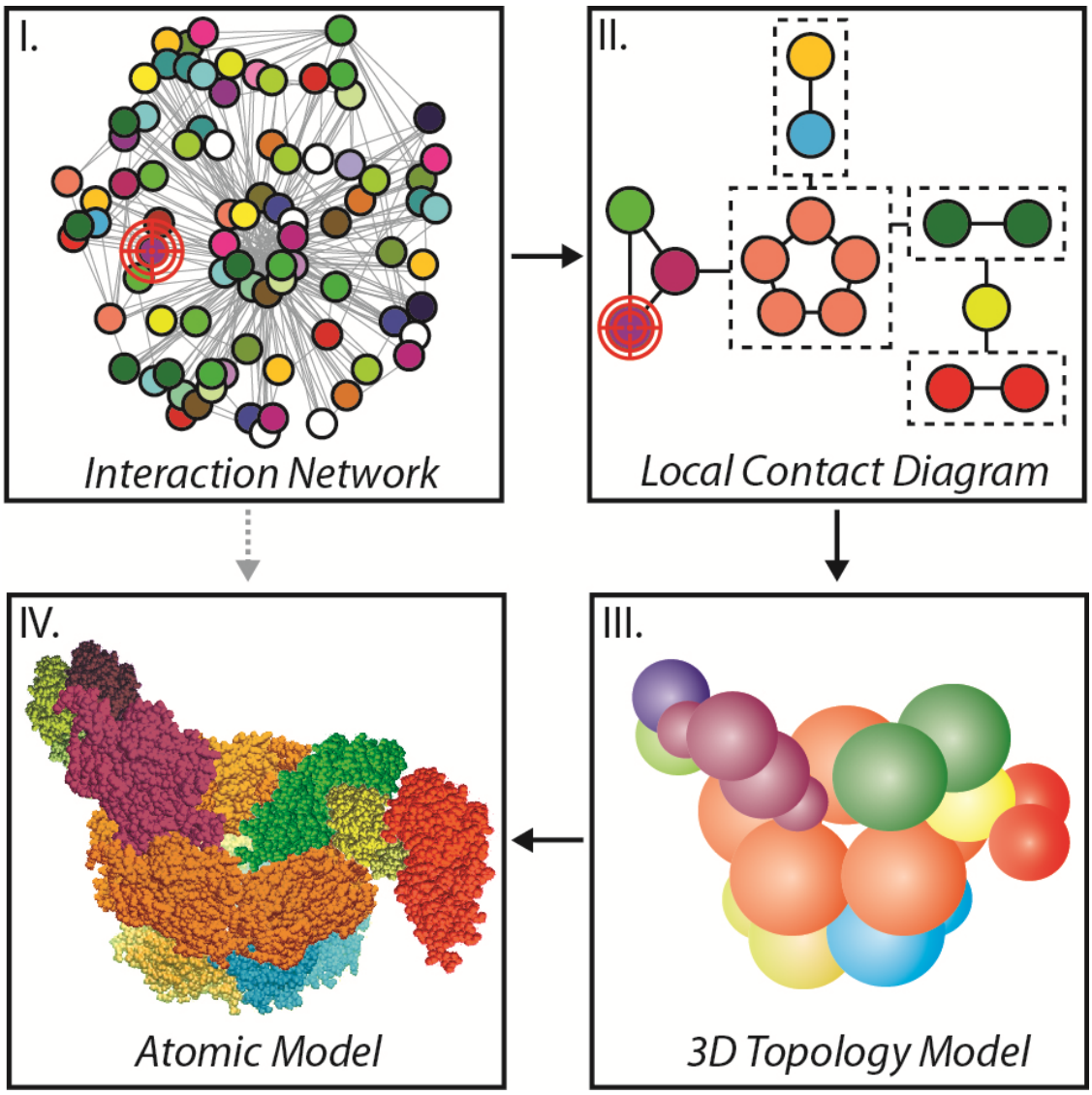

Figure 1-1 The challenge of structural genomics: converting protein interaction networks into protein structures.
High-throughput structural genomics efforts currently underway rely heavily on technologies that can convert an isolated multi-protein complex directly into a structural model of atomic resolution (dashed grey arrow). The stringent sample requirements involved result in relatively high failure-rates for such experiments. An alternate approach (black arrows) uses MS data of intact complexes to generate a contact map, integrates IM data and other constraints to build a 3D topology model, and utilizes homology modeling or other forms of local constraint to generate the final atomic model for multi-protein complexes. This IM-MS approach is projected to be a more-universal, more-sensitive, higher-throughput alternative to contemporary structural biology technologies.

1.1 Tools for Protein Structural Studies

Besides mass spectrometry and related technologies, which will be covered in Chapter 1.3 and throughout this thesis, mature structural tools like X-ray crystallography, NMR spectroscopy, cryo-electron microscopy. and small angle X-ray scattering have played key roles in protein structural studies. Combined, these techniques cover a broad scope of structural data, from atomic level detail to multiprotein-scale size and shape, with each technology possessing unique advantages and shortcomings that drive their overall complementarity.

X-ray crystallography is a method for determining the arrangement of atoms within a crystal using a beam of X-rays and their diffraction.[26] A three-dimensional picture of the density of electrons within the crystal is derived from the angles and intensities of the diffracted X-rays, from which the mean positions of the atoms in the crystal can be determined, as well as protein interactions. Since the first high resolution structures of proteins were solved in the 1950s,[27] X-ray crystallography has become the benchmark for protein structural analysis, accounting for 85% of the entries in the Protein Data Bank (PDB) (PDB, http://www.pdb.org/). Although this technique has developed significantly since its inception, from small molecules to intact ribosome complexes, many challenges remain, including membrane proteins, low-purity samples, and dynamic macromolecular machines.

NMR spectroscopy works by aligning nuclei in a strong external magnetic field, and then uses an orthogonal pulsed magnetic field to measure the magnitude of nuclear magnetic moments, using deviations in the observed frequencies of their nuclear spins to deduce

atomic geometries.[28] NMR approaches are frequently used to determine the structure and dynamics of protein structures in solution, in contrast to X-ray analyses, which are comprised of static structures. Each peak in an NMR spectrum represents a nucleus in a different chemical environment. One dimensional NMR can generate structural information for small proteins, however, for larger proteins, multi-dimensional NMR technologies are typically used to reach optimized resolution of nuclear spin states and connectivity.[29] Examples of these approaches include transverse relaxation-optimized spectroscopy (TROSY)[30] and isotope labeling approaches[31]. When a target protein binds to another protein or a ligand, changes in the chemical environment at the binding site can be observed by NMR.[32,33] NMR protein analysis is typically frustrated by low concentration, low-purity samples containing very large protein complexes.[3,5]

Cryo-EM analysis of large proteins and complexes typically generates intermediate-resolution structure information (~10 Å).[34] This allows shape and topology definition to be obtained for large complexes in a native-like state. In a cryo-EM experiment, a beam of electrons is aimed at a flash-frozen sample in vitreous ice. Electrons scattered by protein samples in this manner can be converted into a two-dimensional image. Single-particle analysis, as well as electron tomography with multiple tilted views of the same object, can reveal the shape and symmetry of an assembly.[5] Following the construction of class average structures and average electron densities from EM data, individual proteins with known atomic structures can be docked and integrated into the determined topologies if the locations of individual chains are known. Although cryo-EM data are often obtained at a resolution of 10Å, several EM structures have been extended to 3.5Å resolution.[35,36] While small proteins and samples of low purity still provide challenges for

EM analysis, the technology is able to provide structures of high-detail for a broad range of large, labile assemblies.[37]

Finally, SAXS enables the determination of shapes of high-concentration, purified biopolymers in solution by measuring X-ray scattering patterns.[38] Unlike crystal structures, SAXS collects the average signal over all protein conformations and orientations for molecules in solution, which results in lower resolution structure data.[39] Advantages of SAXS include the ability to analyze large protein complexes (as large as 580 kDa) when compared to NMR, and avoids the damage associated with the high-energy electrons used in cryo-EM.[40]

1.2 Mass Spectrometry

As discussed earlier, mass spectrometry (MS) and related technologies have provided dramatic insights into the composition and structure of multiprotein complexes. MS detects the mass to charge ratio (m/z) of an analyte after ionization. Structural approaches in MS are notable for their high sensitivity, unlimited mass range, high speed and high tolerance to sample impurities.[13,14,41] The key components of MS often include ionization source, analyzer and a detector. Typical ionization sources used to analyze biological samples include electrospray ionization (ESI)[42-44] and matrix-assisted laser desorption ionization (MALDI).[45-47] The mass analyzer, the key component of MS, has evolved to include many devices that make use of a range of physical principles for m/z separation in time and space. These include quadrupole (Q) filters.[48,49] ion traps,[50-52] time-of -flight (ToF),[53,54] orbitrap,[55-57] and Fourier transform ion cyclotron resonance (FTICR).[58-60]

Between the ionization source and the ultimate mass analyzer, many specialized components, tailored to enable different experimental needs, are typically mounted, and can include ion separators, guides, and additional mass filters.[61-63] Only those components, ion sources and analyzers used during the prosecution of this thesis will be covered in detail below. All data shown in this thesis are generated on a second-generation Q-ToF instrument incorporating a traveling-wave IM separation device (Synapt G2, Waters, Milford, MA, USA) as shown in Figure 1-2.

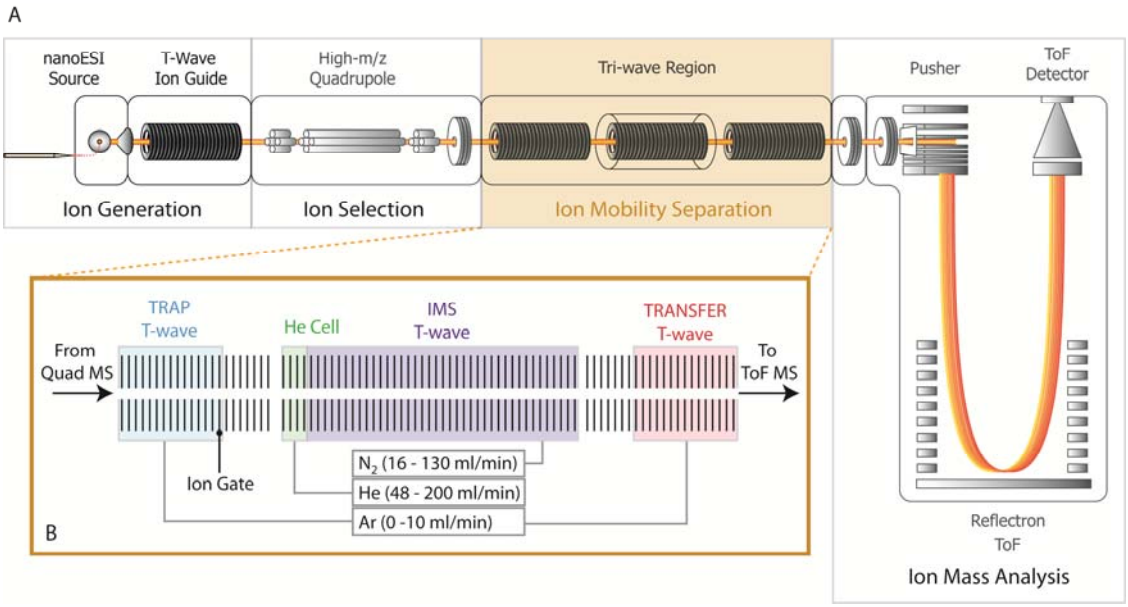

Figure 1-2 A schematic diagram of the Synapt G2, quadrupole-ion mobility-time-of-flight mass spectrometry instrument used in these studies.
(A) A general schematic of the complete instrument indicating the four main regions of instrument operation: ion generation (using a nESI ion source), ion selection (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 traveling-wave (T-wave) ion guides dominate this section of the instrument, and include the ion trap (blue), ion mobility (purple), and the transfer (light 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.

1.2.1 Electrospray Ionization (ESI)

ESI is a soft ionization method used to produce gas phase ions for labile biomolecules.[64] In ESI, the analytes are present in a dilute solution and passed through a conductive emitter. An electric field is typically applied between the emitter outlet and a counter-electrode. This high voltage causes charge accumulation at the emitter tip, forming a Taylor cone, which then works to aerosolize the solution into charged droplets at atmospheric pressure, typically assisted with gas flow and increased temperature. As the solvent within the aerosolized droplets evaporates, the Rayleigh charging limit is reached, prompting the droplets to fission into smaller droplets, eventually giving rise to charged analytes.

Two mechanisms govern the ionization process for ESI. For smaller analytes, the ions are emitted directly from the droplet surface, typically described as the Ion Evaporation Model (IEM).[65,66] While for larger analytes, like intact proteins, successive fission events result in a nano-scale droplet from which all solvent evaporates, leaving behind a charged analyte.[67] This process is typically referred to as the Charge Residue Model (CRM).[68] ESI characteristically produces multiply-charged analytes, and minimal ion activation or fragmentation. ESI charging through a CRM mechanism is highly correlated with protein surface area.[69] 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.[70]

To solve these problems, nano-electrospray ionization (nESI) was developed,[71] using tip 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. The extremely low sample consumption makes nESI compatible with nano-scale LC.[72] 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.[73,74] Overall, nESI tends to provide ionziation and desolvation conditions that are conducive to maintaining fragile protein structures glycan structures[73] and non-covalent protein complexes[75,76] as discussed in section 1.3.1 and elsewhere in this thesis.

1.2.2 Quadrupole Mass Filters

Quadrupole filters consist of 2 pairs of parallel conductive rods positioned perpendicularly to each other. Alternating voltages (at RF frequencies), along with a DC voltage offset, is applied to each pair of rods. When the ions are traveling between the rod sets in the axial direction, they will oscillate longitudinally. Each combination of AC and DC voltages will only allow certain m/z values to pass through. Others will hit the rods before reaching the exit aperture. When operated in an RF-only mode, the rods act as an ion guide that pass all ions beyond certain m/z value determined by the RF amplitude

applied. A modified quadrupole mass filter is housed in the Synapt G2 IM-MS system, capable of selecting ions up to 32,000 m/z. Increasing the quadrupole m/z from typical values (<4,000), involves decreasing the operating RF frequency that drives the quadrupole rods.[77,78] Selected ions can be subjected to collision induced dissociation (CID) or collision induced unfolding (CIU) following selection in the quadrupole.

1.2.3 Time of Flight (ToF) Mass Analyzers

ToF analyzers are popular due to their relatively high m/z resolution (>50,000 in high-performance instruments), high mass accuracy (< 1ppm), high sensitivity, large dynamic range and fast (μs) response times.[53,79,80] When the ions enter the ToF source, they are subjected to an electric field designed to accelerate all ions. Thus, ions of like charge, starting position and energy prior to acceleration will, upon exiting the ToF source, possess equal kinetic energies as defined in Equation 1.1 below:,

$$E_p = zU = E_k = \frac{1}{2}mv^2 \tag{1.1}$$

where E_p is the electric field potential energy, z is the charge of the ion, U is the electric potential difference (voltage), E_k is the ion kinetic energy, m is the mass of the ion, and v is the ion velocity when exiting the electric field. With a fixed travel distance d, the time required for an accelerated ion to reach the ToF detector t can be determined by Equation 1.2 below:

$$t = \frac{d}{\sqrt{2U}}\sqrt{\frac{m}{z}} \tag{1.2}$$

In ToF MS, ion flight time is measured in an evacuated tube, and is used along with known values of *d* and *U* to enable the determination of ion m/z. One or two ion mirrors (reflectrons) can be used to double or quadruple the flight distance respectively, and correct for small differences in initial ion energy or position in the ToF source, in order to improve analyzer resolution. ToF is often coupled with IM in IM-MS instruments because of its fast response time, enabling many of ToF spectra to be acquired for a single IM peak.[81-83]

1.3 Mass Spectrometry Study for Protein Structures and Interactions

With soft ionization methods, fast measurements and versatile activation methods, MS has become an important tool in the characterization of many types of analytes. MS, often in combination with complementary techniques, is capable of assigning protein complex connectivity, structural characteristics and interactions. There are several MS methods used in the analysis of multiprotein complexes. For example, MS of intact protein complexes, hydrogen-deuterium exchange (HDX), oxidative foot printing, and chemical cross-linking are all utilized in this context. IM, a relative new comer to the protein structure field and the focus of this thesis, will be discussed in detail in section 1.4.

1.3.1 Mass Spectrometry of Intact Non-Covalent Protein Complexes

By using soft ionization sources, mass spectrometry is able to analyze non-covalent interactions between proteins or protein and ligands in the gas phase. This allows complex stoichiometry, subunit connectivity and ligand specificity to be studied in

unprecedented detail. From the original observations of intact protein-ligand complexes by ESI-MS over twenty years ago, the mass and complexity range of the protein complexes that can be analyzed intact have increased exponentially, to include intact ribosomes (>2 MDa)[84] and virus capsids (16 MDa)[85]. Affinity-based purification coupled to high-sensitivity MS detection has provided some of the clearest depictions of the macromolecular protein networks operating on the cellular level in terms of their connectivity, leading eventually to include sophisticated quantitative proteomics protocols.[86,87] Dissociation techniques available, both in the gas phase and in solution , allow even further study of protein complex stoichiometry, subunit connectivity and ligand specificity.[88] The fast response time of MS detection also enables the investigation of dynamic protein complex systems with high degrees of polydispersity, i.e. amyloid-beta protein complexes.[89]

1.3.2 Hydrogen-Deuterium Exchange (HDX) and Oxidative Foot-Printing

HDX and oxidative labeling are being used with increasing frequency prior to MS in order to assess dynamic structural changes within a host of multiprotein systems.[90-94] While the implementation of most MS experiments involves the proteolytic digestion of denatured proteins following the steps taken within the protocol to capture native protein-protein contacts or structure information,[95,96] HDX is frequently used to determine which parts of a protein are exposed to solvent and which are inaccessible due to of protein structure, folding or higher-order interactions.[91,92,97,98] Labile hydrogens on the surface of proteins can be replaced with deuterium upon exposure to a D_2O-containing solvent.

After a sample is incubated in deuterated solvent for a specific amount of time, exchange is quenched by reducing solution pH and temperature, followed by proteolysis and analysis of the resulting peptides by chromatography coupled to MS. HDX-MS data analysis is typically challenging due to back-exchange and spectral complexity,[99] but residue-level exchange levels can be quantified by using electron-based tandem MS technologies.[100,101] Oxidative labeling, unlike HDX, generates highly stable labels through exposure of photons that excite water or hydrogen peroxide to create ·OH radicals, which then covalently modify protein surface residues.[94,102,103] Then following proteolysis, separation, and MS analysis, protein interfaces and solvent exposure can be assessed by comparison with control samples. Like HDX-MS, oxidative footprinting experiments by MS are hampered by data interpretation and informatics challenges resulting from highly-complex labeling patterns, but residue level labeling can be assessed through modern tandem MS experiments.[93,104]

1.3.3 Chemical Crosslinking

Crosslinking protocols have been in use since the early 1950s.[105] Crosslinking in combination with MS detection has increased in popularity over the past decade, both to study protein tertiary structure and protein-protein interactions.[106-108] The crosslinkers, typically with the length between 10 to 20 Å, covalently link residues on protein chains that are within interaction distance with each other. Proteolytic digestion followed by MS is typically used to identify the increased mass associated with peptides having undergone the cross-linking reaction. Several modified peptide versions are possible, including those that contain un-reacted sites, dead-end cross-links, and completely cross-

linked peptides. This sample complexity, and the informatics challenges that result, present the greatest challenges associated with cross-linking data analysis by MS. Despite these challenges, chemical cross-linking has been extended to larger protein systems[109-112] and *in vivo* measurements[113,114] in recent years.

1.4 IM-MS Technology

Ion mobility (IM) separation, when coupled to MS, enables 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.[88] In addition, IM-MS provides signal-to-noise, spectral deconvolution, and tandem MS benefits unique to the analysis of the multiply-charged protein complex ions produced by nESI.[115] Following appropriate sample purification, separation, and ionization,[18] 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 collision cross section (CCS) than more compact protein ions of similar mass (Figure 1-3).[116] 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 multiprotein complexes, takes three basic forms: drift tube (DT)-type, differential mobility analyzer (DMA)-type, and travelling-wave (T-wave)-type instruments. All of these technologies have both strengths and weaknesses for the analysis of multiprotein assemblies.[117,118] Other types of IM technology have yet to be

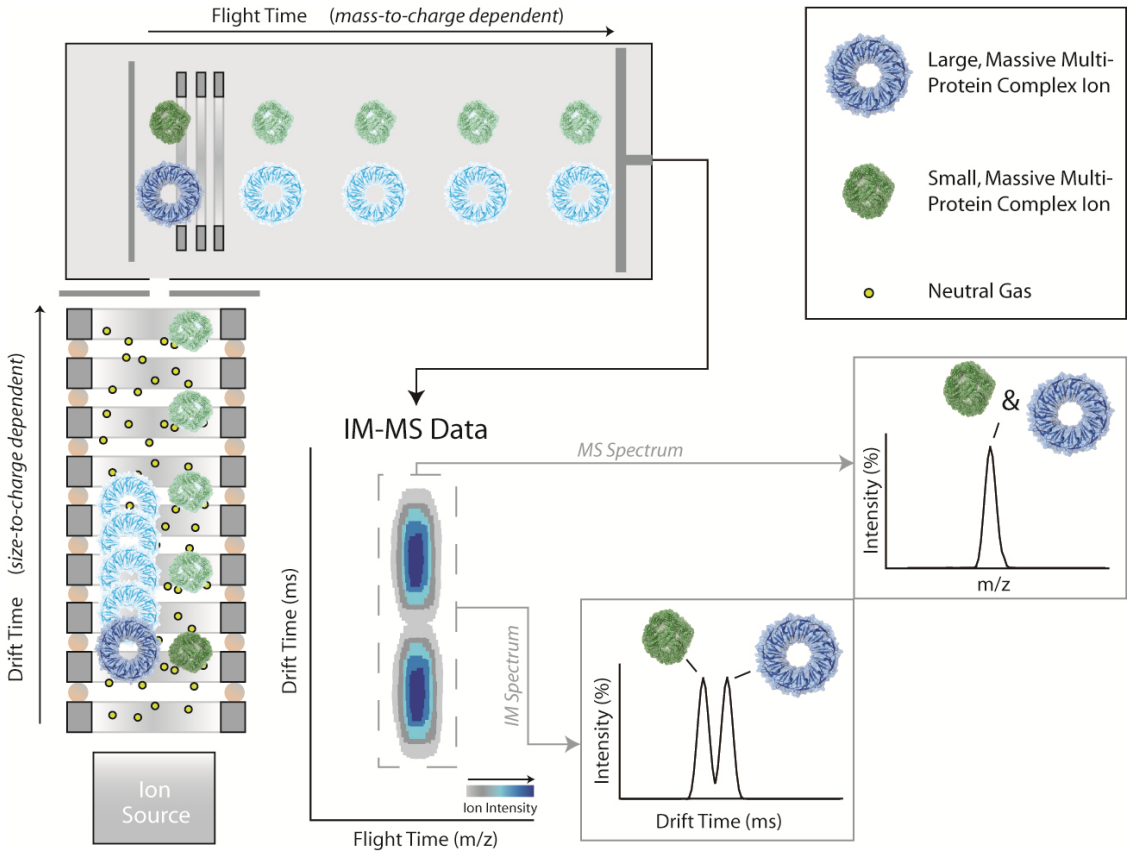

Figure 1-3 Ion mobility-mass spectrometry data acquisition and basic principles. Ions are generated at the ion source (lower left), and are allowed to drift in an ion guide filled with neutral gas molecules under the influence of an electric field. The ions migrate through this region according to their 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). A key for the diagram is shown, upper right.

used in the analysis of proteins and their complexes, and as such, will not be discussed in detail here. For example, high-field asymmetric waveform ion mobility spectrometry (FAIMS) and differential mobility spectrometry (DMS), separate ions based on observed differences in ion mobility as a function of separation field strength. [119,120,,121] While both are commercially available, it is currently not possible to use these technologies to reliably measure ion CCS. Consequently, FAIMS and DMS are typically not used in

experiments aimed at determining multiprotein structure. Other high-resolution IM technologies, such as overtone ion mobility spectrometry (OMS) [122,123] and cyclotron-type drift tube analyzers[124] have yet to be applied to large protein size measurement, but offer exciting opportunities for detailed IM measurements on such systems in the future.

1.4.1 Drift –Tube (DT)

Many of the first IM separations performed on intact proteins and small complexes were conducted using DT-type IM-MS instrumentation.[118,125] DT-IMS measures ion CCS directly and provides high IMS resolving power in research-grade instruments. Such analyzers typically comprise a series of ring electrodes upon which a fixed axial gradient of the electric potential is constructed. A direct proportionality between drift time and CCS is generally used to convert measurements into ion size information. While early versions of these devices suffered from significant ion losses and poor sensitivity when used for the analysis of biomolecules, contemporary implementations of the technology use ion guides either during or after IM separation in order to re-focus ions and preserve limits of detection.[126-129] Furthermore, tandem DT analyzers have been used to assess micro-conformational states within the structural envelope of gas-phase protein ions, and have demonstrated that such structural populations can be stable on the millisecond timescale, thus enabling detailed assessments of gas-phase protein structure.[130,131]

1.4.2 Differential Mobility Analyzer (DMA)

DMA-type analyzers have also been utilized to measure the size of a range of protein complexes.[132] These devices function by introducing ions into a region filled with inert gas, often consisting of two plates having offset apertures for ion entrance and exit. A voltage is then scanned and the resulting ion trajectory between the entrance and exit aperture is dependent upon ion CCS. In contrast to DT type IMS, which measures a size-proportionate ion drift time, DMA analyzers measure the voltage required for ions to exit the analyzer and be transmitted for MS analysis. Ion mobility is inversely proportional to this voltage, and converting such information into ion CCS is typically performed through a rigorous calibration methodology.[133,134] In an analogous fashion to quadrupole mass filters, DMA analyzers can 'scan' over all ion mobilities or 'select' and transmit ions having a narrow range of mobilites for more detailed analyses.[118,134,135] In recent experiments, the proteins investigated using this technology have appeared more compact in the gas-phase than previously thought,[133] although a detailed structural analysis of these data, and assessment of their broader implications, has yet to appear. Significant databases of protein size, as well as more-detailed treatments of multiprotein structure, have appeared based on DMA measurements,[136-138] although many of these measurements are performed in the absence of MS measurements in tandem.

1.4.3 Traveling-Wave (T-wave)

The majority of IM-MS datasets for multiprotein complexes have been generated on IM-MS instruments using T-wave IM analyzers. T-wave IM 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.[77,139] 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.[140] An important feature of this process is that, due to the nature of the separation mechanism employed, T-wave drift times are most often calibrated using standard CCS values for protein complexes rather than calculated directly from drift time measurements.[115] Apart from being the only IM analyzer currently incorporated into commercially available, high-sensitivity IM-MS instrumentation for ion size measurement in wide distribution, T-wave analyzers offer some modest advantages in terms of separation resolution.[141-143]

Often defined in terms of the centroid arrival time of the IM peak normalized to the IM peak width ($t/\Delta t$), 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).[124,144,145] Because of the physical principles involved in T-wave IM separation, drift time is correlated to CCS through an exponential relationship,[115] which results in a T-wave drift time axis that is effectively 'stretched' relative to those achieved on DT analyzers. This relationship enables T-wave separators to achieve 40-60 CCS resolution ($CCS/\Delta CCS$) using comparatively shorter devices, operating at lower fields and pressures, than DT devices of equivalent dimensions.[142,143,146]

The specific instrument involved used in this thesis is the Synapt G2 as shown in Figure 1-2.[142] Currently, two different instrument generations exist for Synapt instrumentation. The major differences between the first and second generation of the instrument are illustrated in Figure 1-4. The 'tri-wave' region of the instrument (Figure 1-2 B) consists of three main traveling-wave ion guide sections, enclosed from the main vacuum system, coupled to smaller traveling-wave ion guides that operate within the main vacuum system for increased conductance of the gas load within the enclosed regions of the instrument. This increased conductance between sections is an improvement over the first-generation instrument geometry. Subsequent to ion transport through the quadrupole mass analyzer, the initial traveling-wave ion guide encountered functions as an ion trap, where confining RF voltages act to contain the ions radially, and a DC gate field contains the ions axially. The first 9 cm of the device is enclosed and pressurized with argon gas to axialize the ions and store them prior to release into the next stage of the instrument. No traveling waves are active in this area of the instrument during ion trapping. After the gate field is released, the ions travel through a 4 cm traveling-wave ion guide to the next enclosed region of the instrument, pressurized with helium gas.

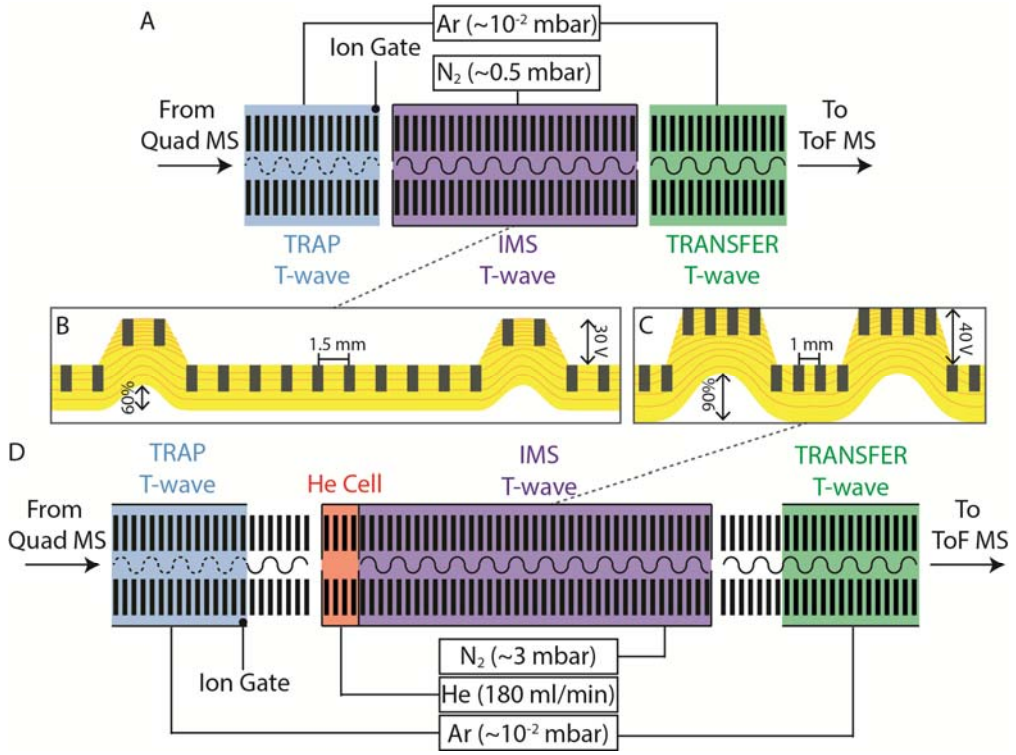

Figure 1-4 A schematic diagram illustrating key differences between the first and second generations of Synapt instrumentation, focusing on the TW ion guides and regions directly surrounding the TWIM separator.

(A) Geometry for the 'tri-wave' region of the first generation instrument, consisting of three main TW ion guide sections: the ion trap region (blue), the TWIM separator (purple), and the ion transfer region (green). The dashed waves shown in the ion trap TW guide indicate that no traveling-waves are active in this region during the ion trapping portion of the instrument cycle. **(B)** Detailed view inside the first-generation TWIM separator. The total device is comprised of 122 ring electrodes (185 mm total length). Each electrode is 0.5 mm thick, with 1.5 mm spacers inserted between electrodes. DC pulses (traveling waves) are applied to the electrodes shown, with a 6 electrode pair period for the overall waveform. A maximum of 30V can applied to electrode surfaces, but only 60% of that applied voltage is experienced by ions in the center of the guide. **(C)** A detailed view of the second generation TWIM region. The length of the device is defined by 168 ring electrodes (254 mm total length), having equal dimensions to the first-generation device, but decreased spacing (1mm). In further contrast to the first-generation device, traveling waves are applied to 4 electrodes simultaneously and are spaced by 4 electrodes. A maximum of 40 volts can be applied, and the improved geometry allows for 90% of the applied voltage to be experienced by ions within the guide. **(D)** Schematic diagram detailing the overall configuration of the traveling wave ion guides within the second-generation Synapt instrument. Key differences between this design and the one shown in A include: decreased vacuum conductance in both the ion trap and transfer regions, with regions of differential pumping installed (not shaded) and a short 'helium gate' region (orange) which acts as a buffer between a higher pressure TWIM separation region and the ion trap.

This short ‘helium gate’ region acts as a buffer between the IM separation region, which operates at high pressures of N_2, and the ion trap, which operates at low pressures of Ar. The high pressures of He within this region of the instrument allow ions to be injected against a pressure gradient composed of a less-massive gas; and therefore, the ions experience less collisional activation and heating as a result of the process. In order to ensure that this region remains pressurized with primarily He, the pressures in the helium gate are typically kept at higher levels than IM travelling-wave region, resulting in IM separations carried out in a mixture of He and N_2 gases on the Synapt G2. The 'helium gate' was not present in the first-generation version of the Synapt instrument and is a key difference between the two instrument geometries, allowing the IM separation region in the Synapt G2 to operate at higher pressures compared with the first-generation instrument.

The IM separation region on the Synapt G2 is 25cm long (7cm longer than the first-generation instrument), comprised of 168 plates, each 0.5mm thick and spaced by 1mm. Again, an RF field operates to compress the ions radially, while traveling-waves drive the ions axially to perform ion mobility separation. In the first generation instrument, the traveling waves within this region were generated by applying the DC pulse to plate pairs, resulting in a narrow voltage wave capable of applying only 60% of the applied voltage. The Synapt G2 applies voltage to 4 plates simultaneously, resulting in a broader wave that sustains 90% of the voltage applied.

After IM separation, the ions then encounter another short (4 cm) region of traveling-wave ion guide that operates in the main vacuum system. Following this, a 9cm traveling-

wave device, pressurized with the same Ar flow used for the ion trap region, is used to transfer the separation achieved in the IM separation region, without degradation, to the ToF analyzer for eventual m/z detection. The traveling-waves operating in this region are slow moving with relatively high amplitudes, and act to capture ions having undergone IM separation between waves to limit further diffusion and uncorrelated separation prior to MS analysis. Each enclosed region is connected to a gas-feed that is metered by a mass-flow controller calibrated for the gas utilized in each respective region of the instrument, and normal operating ranges for these controllers and recommended pressures are also shown in Figure 1-2 B and Figure 1-4 respectively, for each of the enclosed traveling-wave ion guide regions.

1.5 IM-MS Data Interpretation

Generating both protein contact and topology information from IM-MS data requires multiple levels of computational analysis.

1.5.1 MS Data Analysis

The first stage in this process is typically focused on the interpretation of MS data derived from both intact complexes and their constituent subcomplexes, as well as protein subunits generated through careful disruption of intact complexes. Obtaining an accurate component list of proteins within the complex, often from LC-MS protein identification experiments, is a critical, challenging stage in the analysis, as the mass measurement accuracy (MMA) achieved for intact proteins and complexes is often much

lower than those achieved for the peptide ions commonly encountered in 'bottom-up' proteomics experiments. This difference is primarily driven by the relative fragility and size of the multiprotein complex ions generated, thus limiting the extent of chemical purification and desolvation that can be used in their MS analysis.[147,148] Often, the MMA achieved is due primarily to the level of desolvation achieved following nESI, with the peak widths and mass shifts recorded for complexes being directly correlated with the amount of non-volatile buffer material adhered to the assembly post-ionization.[149,150] Multiple software approaches have been developed for the automated mass analysis and deconvolution of nESI-MS datasets of protein complexes, enabling the interpretation of MS spectra containing multiple overlapping species in *m/z*, [148,151-153] and in compiling MS datasets to generate protein complex contact diagrams. Further, recent advancements have been made in modeling the relative intensities within MS datasets in a more automated approach to spectral deconvolution.[154]

1.5.2 IM Data Analysis

In the interpretation of the IM data for multiprotein complexes, there are two main stages: the conversion of drift times into CCS values, and their comparison with model protein architectures. The first stage in this two-step process is relatively straightforward for DT and DMA-type IM analyzers, which rely upon highly-robust calibration and simple algebraic relationships to convert primary data to ion sizes respectively. T-wave analyzers, on the other hand, require careful calibration against standards having known CCS values in order to generate accurate results.[115,155] Theoretical treatments of T-wave analyzers have indicated a quadradic relationship between drift time and ion CCS when

operated under a relatively narrow set of conditions.[140] Experimental calibration relationships, however, give a range of exponential values for the drift time / CCS relationship, and thus must be determined experimentally during IM-MS data collection in order to achieve optimum accuracy.[115,142] The accuracy of such measurements has increased dramatically in recent reports, owing primarily to the availability of a larger pool of available standards,[155] such that modern T-wave CCS measurements are equivalent in accuracy and precision to those produced by other IM analyzers. While some reports have indicated that ions can be significantly heated during T-wave IM separation,[156,157] recent results indicate that this effect, if present, does not influence CCS accuracy or precision on second-generation T-wave platforms.[142] Following accurate CCS measurement, model structures are generated *in silico* and their CCS values are estimated computationally for comparison with experiment. IM datasets recently published exhibit a broad correlation between known X-ray structures of complexes and the CCS values recorded for these same assemblies in the gas phase.[11,158] While many approaches exist for estimating the CCS of X-ray and other structural models *in silico* [10], and relating the two datasets can be challenging,[16] no computational approaches currently provide greater accuracy for multiprotein systems than simple, properly scaled, projection approximation (PA) calculations,[158] and such estimates have been effectively used for a number of years to correlate IM measurements with trial multiprotein topologies.[16,23,25] Such methodologies are computationally inexpensive, and have the benefit of supporting coarse-grained protein representations for rapid modeling and topological prototyping.[23,159] Scaling of such estimates is a key component of relating PA calculations to IM measurements, because without a properly tuned scaling factor such calculations

significantly mis-estimate the size of potential model structures.[10] Recent computational results support the potential accuracy of scaled-PA estimates for the IM analysis of proteins and complexes, as well as introducing a new computational approach, termed the projection superposition approximation, or PSA, which features a tunable shape factor and retains much of the computational speed of PA calculations.[160]

1.6 Optimizing the Information Content of IM-MS for Structural Proteomics

The IM-MS protocols for structural proteomics described above rely heavily on measurements of intact protein complexes, as well as subcomplexes and subunits, for the determination of multiprotein topology. Furthermore, proteins are typically represented in IM-MS derived models as spheroids unless additional information on the domain structure or fold of the protein is available from other sources that can be integrated with IM-MS constraints.

1.6.1 Gas Phase Activation

Early experiments relied heavily on tandem MS data, utilizing collision induced dissociation (CID), to disrupt protein complexes in the gas phase for the determination of complex stoichiometry.[21] Recent CID experiments have revealed that the charge states of precursor ions selected for activation and dissociation can dramatically influence both the structure and identity of the product ions formed. Precursor ions having charges unmodified from those produced from standard nESI conditions can occupy

intermediates having undergone both quaternary remodeling and monomer unfolding prior to dissociation.[161-163] By contrast, charge reduced complexes subjected to CID can produce compact and presumably folded product ions.[164] In rare cases, charge amplification has also been observed to enhance the folded character of product ions produced by multiprotein CID.[165] Either extensive charge reduction or amplification, coupled with higher-energy CID, can result in the dissociation of covalent bonds within the complex to produce sequence informative peptide ions from protein termini.[24,164] This last observation suggests the exciting possibility of 'top down'-type protein identification experiments performed from multiprotein precursor ions, and has been duplicated using electron-mediated fragmentation approaches.[166] Surface induced dissociation (SID) technology has recently emerged as an alternative to CID for multiprotein complexes, as it can generate subcomplex and presumably compact subunit product ions following multiprotein complex activation.[167] Taken together with the important advances made recently in CID and electron-based fragmentation, gas-phase protein disruption approaches have the ability to access multiple levels of multiprotein organization to inform greatly protein contact map construction.

Prior to dissociation, protein complexes typically unfold in the gas-phase upon collisional activation.[23] The gas-phase unfolding of intact proteins is typically marked by a multitude of stable intermediate structures that, upon sufficient activation, give way to a string-like conformational state. By careful examination of such collision induced unfolding (CIU) patterns as a function of ion energy, protein stability, ligand-binding specificity and protein tertiary structures can be assessed.[149,168-171]

1.6.2 Solution Phase Disruption

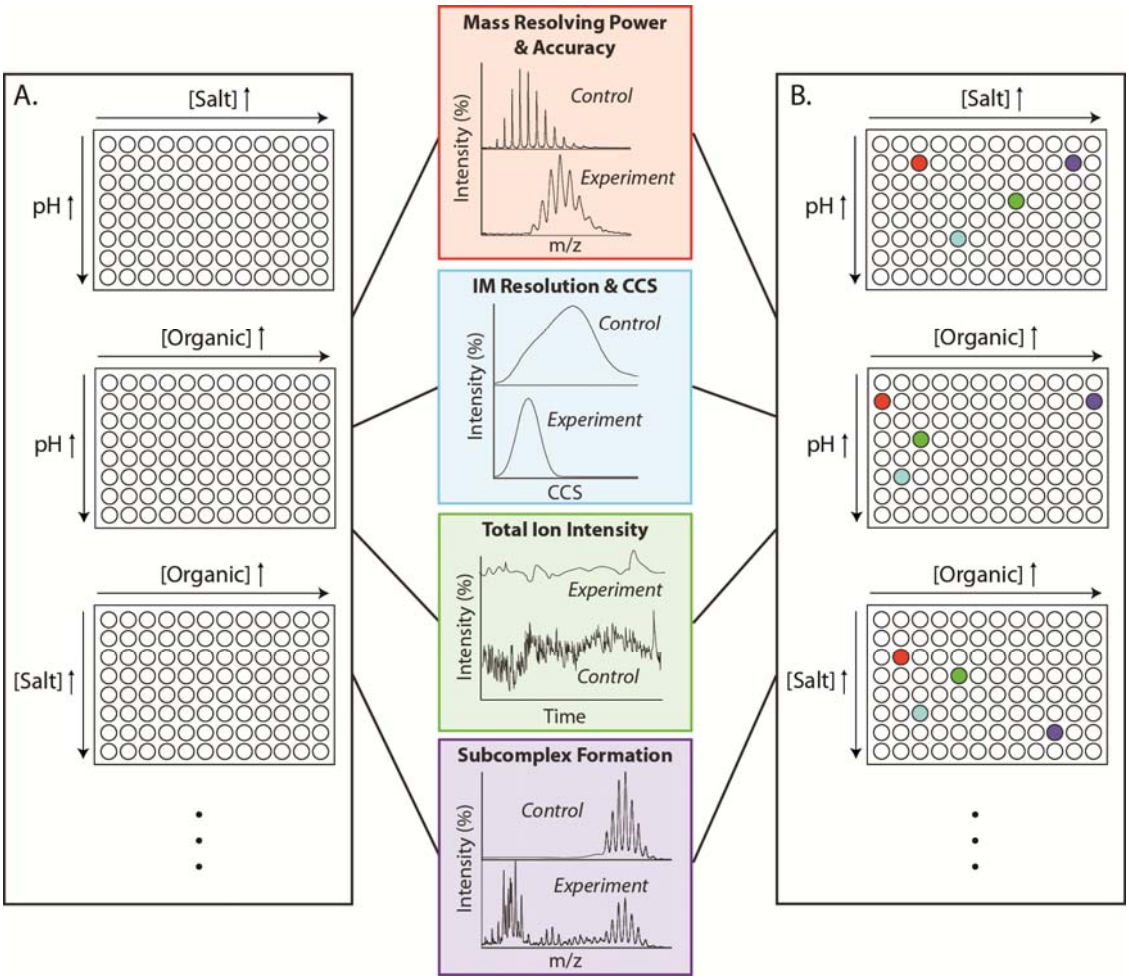

Figure 1-5 A high-throughput screening process to discover optimal solution conditions for protein complex topology mapping by IM-MS.
A two-dimensional screen is developed by varying the composition of solutions in a stepwise fashion over several important variables (e.g., organic content). Ions produced from each solution state are then measured against basic figures of merit: MS resolving power and mass accuracy (red), IM resolution and collision cross-section (blue), total ion intensity (green), and the percentage of current that carries signal for subcomplexes or monomeric proteins (purple). Optimal solutions for each of these classes of information are identified and recorded. In many cases, optimal solution conditions for each figure of merit are mutually exclusive.

Despite the proliferation of gas-phase dissociation and disruption methodologies for multiprotein assemblies, such technologies often do not provide sufficient information on their own to deduce a complete protein network map. As a complementary approach, those studies that have reported high-confidence level protein contact diagrams have optimized conditions in solution prior to nESI in order to generate subcomplexes having largely orthogonal compositions to those produced by the gas-phase methodologies described above. Such disruption experiments often involve small additions of organic solvent, alterations in ionic strength and solution pH in order to elicit the formation of topologically-informative subcomplexes.[21,22,25,172,173] While the phenomenology of this process has been described in multiple reports, and early studies suggested that the evolutionary origins of the protein-protein interface structures within complexes can be used to predict disruption behavior,[174,175] many of the basic principles at work during such multiprotein disruption experiments are still a matter of intense research. Therefore, a relatively exhaustive search of potential solution conditions for optimal disruption, performed in a trial-and-error manner, is usually necessary to develop the information necessary for protein contact map generation (Figure 1-5). Protein disruption data are expected to correlate with the physical and chemical properties of the interacting interfaces between protein subunits. [174] Note that the optimal conditions for subcomplex generation may not overlap with those for other key figures of merit for IM-MS of intact protein complexes. This can result in losses in overall signal intensity, MS resolving power, and IM resolution during a search of solution conditions for subcomplex formation. While currently time consuming, such experiments have been critical in

establishing the assembly dynamics of viral coat proteins and assigning connectivity within multiple heteroprotein complexes of varying size and structure.[25,176,177]

1.6.3 Protein Structure Stabilization

In addition to actively destabilizing the structure and connectivity of protein complexes in both the gas-phase and in solution, stabilizing protein structures is an equally important goal of IM-MS protocols in structural proteomics. Charge manipulation of protein complex ions produced by nESI, especially charge reduction, can be an effective method of protein stabilization in the gas-phase. Recent work has indicated that charge reduction approaches that utilize gas-phase chemistries may be the most effective in the universal structural stabilization of multiprotein complexes.[178] Similarly, small molecule additives can be used to influence evaporative cooling during nESI, and have been shown effective in stabilizing protein-ligand systems.[179] Likewise, small molecules may be added in solution prior to nESI to stabilize the transition between solution and gas-phase for protein complexes. Recent studies focused on Hofmeister-type salt systems have discovered dramatic variations between the stabilizing influences of different salts on gas-phase protein structure compared to solution.[149,180] This work also fits into the solution-screening framework described above (Figure 1-5), where multiple salts and small molecules are titrated against IM measurements to gauge protein structural integrity for topology modeling.

1.7 Integrating Bottom-up Proteomics with Intact Protein IM-MS

Any effort made to structurally assess unknown protein networks by IM-MS incorporates a robust MS-based protein identification workflow in parallel. Most IM-MS workflows for protein complex topology mapping from measurements of intact protein complexes have relied upon a complete list of interacting components, derived from 'bottom-up' and denatured protein-based 'top-down' type approaches for LC-MS protein identification. IM-MS technology also has a long history in both peptide mass mapping and automated LC-type protein identification experiments, which enables IM to contribute to protein component identification as well as protein structure assignment.[181,182] By incorporating IM separation into standard LC-MS experimental workflows, peptide separation capacity can be increased by 1-2 orders of magnitude[118,159,183-185] and highly-parallelized approaches to MS/MS can be implemented to increase sample throughput.[186,187] If these advantages are coupled with the unique structure-based separations[188,189] and chemical tagging methodologies that allow for different chemical classes to be distinguished based on the analysis of 'trend-lines' within IM-MS datasets,[183] a compelling argument can be made for IM-MS workflows that cover both multiprotein interaction topology determination and protein identification. MS experiments that have aimed to deduce protein contact diagrams for multiprotein complexes have incorporated extensive protein identification schemes into their published workflows (Figure 1-6). Most of these involve partitioning samples into 2-3 tracks, where one track incorporates intact MS analysis of the complexes and sub-assemblies generated via disruption (see above), a second track uses 'bottom-up' proteomics to identify the proteins present in the mixture

via mass sequence tags and database searching, and a third (optional) track uses intact protein mass measurements on the denatured protein subunits to evaluate the level of post-translational modification present within the constituent proteins. This last track can be critical to successful contact map generation, especially in cases where multiple subunits within the complex have similar sequence masses. IM measurements on protein complexes have been incorporated into such workflows recently (Figure 1-6). For example, IM measurements were used to determine the connectivity and topology of a trimeric protein subcomplex within the eIF3 heterocomplex.[25] In these experiments, protein dimers that compose the trimer in question were observed for all but one of the possible pairs. Such dimer signals could be absent from the MS measurements for many reasons, but IM measurements identified a linear arrangement of subunits which allowed the model of the entire 13 subunit complex to be restrained with high confidence.

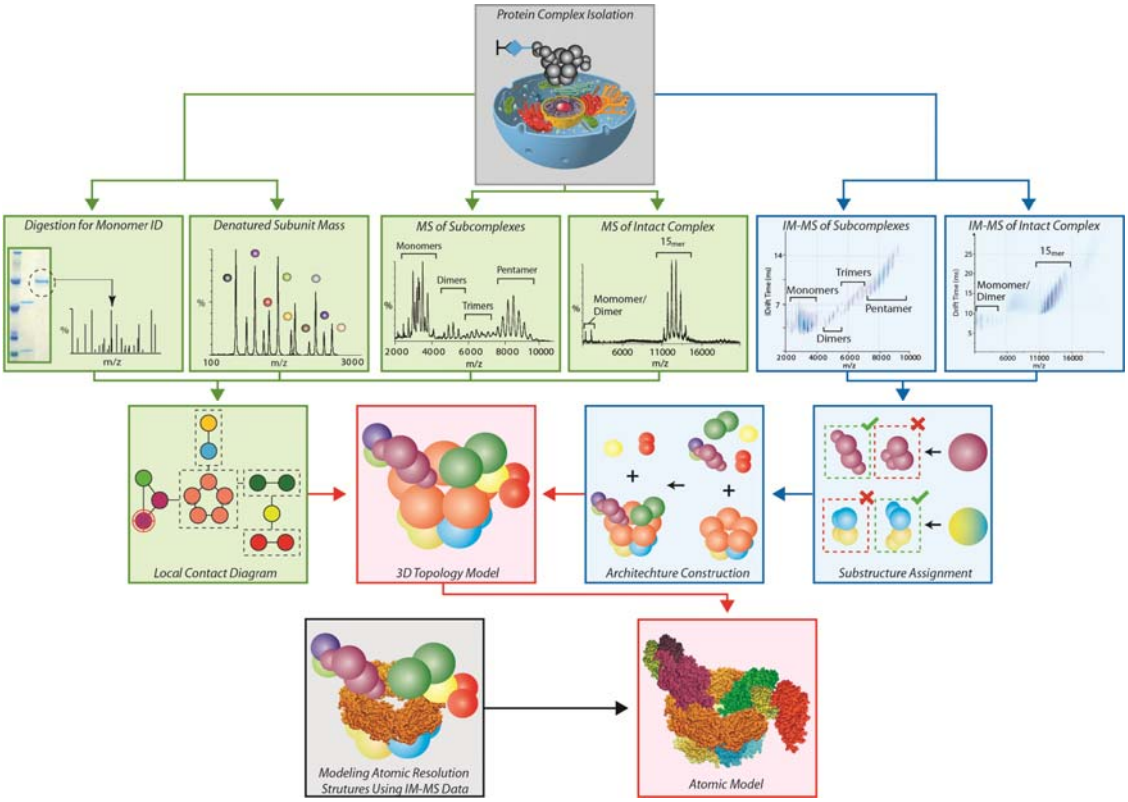

Figure 1-6 A flow diagram for an integrated IM-MS structural proteomics workflow.

After protein complex isolation via either standard affinity purification strategies, or following over-expression and reconstitution of the complex in vitro (Grey box, top), unknown protein samples are split into three channels. One portion of the sample is subjected to denaturation and enzymatic digestion for a combination of 'top-down' and 'bottom-up' type protein identification experiments. These steps are critical for forming an accurate component list for protein contact map generation. A second sample fraction is submitted to MS for intact analysis, where the assembly is dissociated using a combination of solution and gas-phase approaches to deduce protein connectivity. Information from both of the above sample streams (green) is combined to assemble a protein contact map. A third sample fraction, using optimized solution conditions, is submitted for IM analysis and measurement of protein size. Ideally, this step can be performed in parallel with MS analysis of the intact protein complex and subcomplexes created by dissociation. Size information on monomers is used to refine structures of the subunits within in the complex, and various subcomplexes in a step-wise fashion. A cartoon showing how IM data can refine structures for both a multi-domain monomer (purple) and a protein dimer (yellow/blue) are shown, and this information is combined with MS-derived contact map information to provide a complete 3D protein topology (red). This information is then further combined with other sources of information or homology modeling to provide a complete atomic model of the complex of interest.

1.8 Summary

IM-MS has, over the past several years, been used in a number of cases to determine the structures of aggregating peptide and protein systems,[190-192] perform structural studies on a range of protein homo-oligomers,[12,193] to establish the structural details of small hetero-complexes,[16] and to refine MS-derived contact maps for large hetero-assemblies.[22] These achievements have been driven equally by the advent of novel research-grade instruments and commercially-available IM-MS equipment optimized for the analysis of large protein assemblies. In many cases, advances in instrumentation have provided increased IM resolution and MS resolving power.[143] More importantly, modern instrumentation incorporates IM separation devices having high ion transmission efficiency, low limits of detection, that are capable of highly sensitive analyses.[127,129,143] It is clear that integrating IM separation into any MS workflow aimed at characterizing protein mixtures and interaction networks, in either a 'bottom-up' or 'top-down' paradigm, will substantially increase the information content of the resulting analysis.[194]

While the outlook for IM-MS is generally positive, there are a number of challenges that limit its applicability in certain areas of proteomics. A number of these challenges are in the area of computational IM-MS data interpretation. While existing methodologies allow for model structures and experimental constraints to be compared relatively accurately, many sources of potential error still exist in such assessments.[10] Appropriately-scaled versions of *in silico* estimates of model structure CCS values, for example, are relatively

accurate for currently studied protein complexes, however it is unknown if such approaches will continue to be accurate for the myriad of assembly structures predicted within the 3D proteome. A potentially greater challenge for IM-MS computational data interpretation is the rapid and accurate generation of model structures from relatively limited primary information, or solely from IM-MS constraints.[16] In the absence of X-ray or NMR datasets, or when those datasets contain partial information only, algorithms for generating coarse-grained approximations of protein complex structure and topology must likely be substantially improved to begin assessing the structure of large hetero-complexes in a high-throughput fashion.

Experimental challenges for IM-MS measurements revolve around the ability to either stabilize protein structure in the absence of bulk solvent or to disrupt that structure to determine protein connectivity, substructure, and topology. Furthermore, extending the IM-MS information content beyond subunit level resolution is another problem surrounding accurate topology construction. This challenge is amplified in experiments seeking to analyze protein complexes with multi-domains within each subunit. These protein subunits, often having flexible domains, typically require extra stabilization, and cannot be accurately modeled using single coarse-grain spheres.

To overcome the challenges discussed above, several IM-MS experiments have been performed and are reported in this dissertation. This research is divided into 4 research-related chapters aimed with presenting methods and data that circumvent some of the current limitations of IM-MS for protein complex structure analysis.

In chapter 2, the resolution and accuracy CCS measurements on the Synapt G2 are characterized for biomolecular ions, with the purpose of optimizing the instrument performance for accurate multiprotein topology construction. This chapter is published in the journal *Analyst*, 2011 136(17): 3534-41.

In chapter 3, robotically-assisted titration is coupled to IM-MS to reveal the interface structures and analysis parameters critical for multiprotein topology mapping. A protocol for systematically searching for optimized solution conditions for protein-protein disruption is also presented. This chapter is in press in the journal *Analytical Chemistry*, 2013, DOI: 10.1021/ac402276k

In chapter 4, CIU and Coulombic unfolding of monomeric proteins is examined and correlated, for the first time, with protein domain structure in solution.

Chapter 5 presents several datasets where the methods developed in Chapters 2-4 are applied to protein complexes with currently unknown architecture.

In Chapter 6, conclusions and future directions of these newly developed tools for IM-MS are discussed.

The majority content in this introduction was published as a review on *Expert Review of Proteomics*, 2012, 9(1): 47-58.

1.9 References

(1) Lee, D.; Redfern, O.; Orengo, C. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 995.
(2) Alberts, B. *Cell* **1998**, *92*, 291.
(3) Robinson, C. V.; Sali, A.; Baumeister, W. *Nature* **2007**, *450*, 973.
(4) Alber, F.; Kim, M. F.; Sali, A. *Structure* **2005**, *13*, 435.
(5) Sali, A.; Glaeser, R.; Earnest, T.; Baumeister, W. *Nature* **2003**, *422*, 216.
(6) von Heijne, G. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 909.
(7) Pierce, K. L.; Premont, R. T.; Lefkowitz, R. J. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 639.
(8) Steven, A. C.; Baumeister, W. *J. Struct. Biol.* **2008**, *163*, 186.
(9) Wyttenbach, T.; Bowers, M. T. *Annu. Rev. Phys. Chem.* **2007**, *58*, 511.
(10) Barran, P. E.; Jurneczko, E. *Analyst* **2011**, *136*, 20.
(11) Scrivens, J. H.; Scarff, C. A.; Thalassinos, K.; Hilton, G. R. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 3297.
(12) Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. *Science* **2005**, *310*, 1658.
(13) Benesch, J. L. P.; Ruotolo, B. T.; Simmons, D. A.; Robinson, C. V. *Chem. Rev.* **2007**, *107*, 3544.
(14) Heck, A. J. R. *Nat. Methods* **2008**, *5*, 927.
(15) Ruotolo, B. T.; Robinson, C. V. *Curr. Opin. Chem. Biol.* **2006**, *10*, 402.
(16) Politis, A.; Park, A. Y.; Hyung, S. J.; Barsky, D.; Ruotolo, B. T.; Robinson, C. V. *Plos One* **2010**, *5*.
(17) Taverner, T.; Hernandez, H.; Sharon, M.; Ruotolo, B. T.; Matak-Vinkovic, D.; Devos, D.; Russell, R. B.; Robinson, C. V. *Accounts Chem. Res.* **2008**, *41*, 617.
(18) Hernandez, H.; Robinson, C. V. *Nat. Protoc.* **2007**, *2*, 715.
(19) Sharon, M.; Robinson, C. V. *Annu. Rev. Biochem.* **2007**, *76*, 167.
(20) Zhou, M.; Robinson, C. V. *Trends Biochem. Sci.* **2010**, *35*, 522.
(21) Hernandez, H.; Dziembowski, A.; Taverner, T.; Seraphin, B.; Robinson, C. V. *EMBO Rep.* **2006**, *7*, 605.
(22) Leary, J. A.; Schenauer, M. R.; Stefanescu, R.; Andaya, A.; Ruotolo, B. T.; Robinson, C. V.; Thalassinos, K.; Scrivens, J. H.; Sokabe, M.; Hershey, J. W. B. *J Am. Soc. Mass Spectrom.* **2009**, *20*, 1699.
(23) Ruotolo, B. T.; Hyung, S. J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V. *Angew. Chem. Int. Edit.* **2007**, *46*, 8001.
(24) Benesch, J. L. P. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 341.
(25) Pukala, T. L.; Ruotolo, B. T.; Zhou, M.; Politis, A.; Stefanescu, R.; Leary, J. A.; Robinson, C. V. *Structure* **2009**, *17*, 1235.
(26) Hull, A. W. *J. Am. Chem. Soc.* **1919**, *41*, 1168.
(27) Kendrew, J. C.; Parrish, R. G.; Marrack, J. R.; Orlans, E. S. *Nature* **1954**, *174*, 946.
(28) Rabi, I. I.; Zacharias, J. R.; Millman, S.; Kusch, P. *Phys. Rev.* **1938**, *53*, 318.
(29) Folkers, P. J.; Clore, G. M.; Driscoll, P. C.; Dodt, J.; Koehler, S.; Gronenborn, A. M. *Biochemistry* **1989**, *28*, 2601.

(30) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 12366.
(31) Verardi, R.; Traaseth, N.; Masterson, L.; Vostrikov, V.; Veglia, G. In *Isotope labeling in Biomolecular NMR*; Atreya, H. S., Ed.; Springer Netherlands: 2012; Vol. 992, p 35.
(32) Goldflam, M.; Tarragó, T.; Gairí, M.; Giralt, E. In *Protein NMR Techniques*; Shekhtman, A., Burz, D. S., Eds.; Humana Press: 2012; Vol. 831, p 233.
(33) O'Connell, M. R.; Gamsjaeger, R.; Mackay, J. P. *Proteomics* **2009**, *9*, 5224.
(34) Falke, S.; Tama, F.; Brooks Iii, C. L.; Gogol, E. P.; Fisher, M. T. *J. Mol. Biol.* **2005**, *348*, 219.
(35) Liu, J.; Bartesaghi, A.; Borgnia, M.; Sapiro, G.; Subramaniam, S. *Nature* **2008**, *455*, 109.
(36) Wolf, M.; Garcea, R.; Grigorieff, N.; Harrison, S. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 6298.
(37) Adrian, M.; Dubochet, J.; Lepault, J.; McDowall, A. W. *Nature* **1984**, *308*, 32.
(38) Blanchet, C. E.; Svergun, D. I. In *Annual Review of Physical Chemistry, Vol 64*; Johnson, M. A., Martinez, T. J., Eds.; Annual Reviews: Palo Alto, 2013; Vol. 64, p 37.
(39) Blobel, J.; Bernadó, P.; Svergun, D. I.; Tauler, R.; Pons, M. *J. Am. Chem. Soc.* **2009**, *131*, 4378.
(40) Petoukhov, M. V.; Svergun, D. I. *Int. J. Biochem. Cell Biol.* **2013**, *45*, 429.
(41) Berggård, T.; Linse, S.; James, P. *Proteomics* **2007**, *7*, 2833.
(42) Fenn, J.; Mann, M.; Meng, C.; Wong, S.; Whitehouse, C. *Science* **1989**, *246*, 64.
(43) Cole, R. B. *Electrospray Mass Spectrometry - Fundamentals, Instrumentation and Application*; John Wiley: New York, 1997.
(44) Smith, R. D.; Loo, J. A.; Loo, R. R. O.; Busman, M.; Udseth, H. R. *Mass Spectrom. Rev.* **1991**, *10*, 359.
(45) Song, F. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1286.
(46) Kaufmann, R. *J. Biotechnol.* **1995**, *41*, 155.
(47) Mann, M.; Talbo, G. *Curr. Opin. Biotechnol.* **1996**, *7*, 11.
(48) Yost, R. A.; Enke, C. G. *Anal. Chem.* **1979**, *51*, 1251.
(49) Limbach, P. A.; Crain, P. F.; McCloskey, J. A. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 27.
(50) Jonscher, K. R.; Yates Iii, J. R. *Anal. Biochem.* **1997**, *244*, 1.
(51) March, R. E. *Int. J. Mass Spectrom.* **2000**, *200*, 285.
(52) Schwartz, J. C.; Senko, M. W.; Syka, J. E. P. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 659.
(53) Wiley, W.; McLaren, I. H. *Rev. Sci. Instrum.* **1955**, *26*, 1150.
(54) Mamyrin, B.; Karataev, V.; Shmikk, D.; Zagulin, V. *Zh. Eksp. Teor. Fiz* **1973**, *64*, 82.
(55) Rose, R. J.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. R. *Nat. Methods* **2012**, *9*, 1084.
(56) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. *J. Mass Spectrom.* **2005**, *40*, 430.
(57) Perry, R. H.; Cooks, R. G.; Noll, R. J. *Mass Spectrom. Rev.* **2008**, *27*, 661.

(58) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom.Rev.* **1998**, *17*, 1.
(59) Marshall, A. G.; Hendrickson, C. L. *Int. J. Mass Spectrom.* **2002**, *215*, 59.
(60) Heeren, R. M. A.; Kleinnijenhuis, A. J.; McDonnell, L. A.; Mize, T. H. *Anal. Bioanal. Chem.* **2004**, *378*, 1048.
(61) Louris, J. N.; Wright, L. G.; Cooks, R. G.; Schoen, A. E. *Anal. Chem.* **1985**, *57*, 2918.
(62) Mitchell Wells, J.; McLuckey, S. A. In *Methods in Enzymology*; Burlingame, A. L., Ed.; Academic Press: 2005; Vol. Volume 402, p 148.
(63) Sleno, L.; Volmer, D. A. *J. Mass Spectrom* **2004**, *39*, 1091.
(64) Takáts, Z.; Wiseman, J. M.; Gologan, B.; Cooks, R. G. *Anal. Chem.* **2004**, *76*, 4050.
(65) Nguyen, S.; Fenn, J. B. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 1111.
(66) Iribarne, J.; Thomson, B. *J.Chem. Phys.* **1976**, *64*, 2287.
(67) Kebarle, P.; Verkerk, U. H. *Mass Spectrom. Rev.* **2009**, *28*, 898.
(68) Dole, M.; Mack, L.; Hines, R.; Mobley, R.; Ferguson, L.; Alice, M. d. *J. Chem. Phys.* **1968**, *49*, 2240.
(69) Felitsyn, N.; Peschke, M.; Kebarle, P. *Int. J.Mass Spectrom.* **2002**, *219*, 39.
(70) Page, J. S.; Kelly, R. T.; Tang, K.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1582.
(71) Wilm, M. S.; Mann, M. *Int. J. Mass Spectrom. Ion Processes* **1994**, *136*, 167.
(72) Bereman, M.; Williams, T.; Muddiman, D. *Anal. Chem.* **2009**, *81*, 1130.
(73) Bahr, U.; Pfenninger, A.; Karas, M.; Stahl, B. *Anal. Chem.* **1997**, *69*, 4530.
(74) Karas, M.; Bahr, U.; Dülcks, T. *Fresenius. J. Anal. Chem.* **2000**, *366*, 669.
(75) Fitzgerald, M.; Chernushevich, I.; Standing, K.; Whitman, C.; Kent, S. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 6851.
(76) Loo, J.; Loo, R.; Udseth, H.; Edmonds, C.; Smith, R. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 101.
(77) Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H. *Int. J. Mass Spectrom.* **2007**, *261*, 1.
(78) Sobott, F.; Robinson, C. V. *Int. J. Mass Spectrom.* **2004**, *236*, 25.
(79) Guilhaus, M.; Selby, D.; Mlynski, V. *Mass Spectrom.Rev.* **2000**, *19*, 65.
(80) Guilhaus, M. *J. Mass Spectrom.* **1995**, *30*, 1519.
(81) Schröder, D. *Angew. Chem. Int. Edit.* **2011**, *50*, 10490.
(82) Hoaglund, C. S.; Valentine, S. J.; Sporleder, C. R.; Reilly, J. P.; Clemmer, D. E. *Anal. Chem.* **1998**, *70*, 2236.
(83) Henderson, S. C.; Valentine, S. J.; Counterman, A. E.; Clemmer, D. E. *Anal. Chem.***1998**, *71*, 291.
(84) Videler, H.; Ilag, L. L.; McKay, A. R. C.; Hanson, C. L.; Robinson, C. V. *FEBS Letters* **2005**, *579*, 943.
(85) Uetrecht, C.; Versluis, C.; Watts, N. R.; Roos, W. H.; Wuite, G. J. L.; Wingfield, P. T.; Steven, A. C.; Heck, A. J. R. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 9216.
(86) Benesch, J. L. P.; Ruotolo, B. T.; Simmons, D. A.; Robinson, C. V. *Chem. Rev.* **2007**, *107*, 3544.

(87) Shiio, Y.; Aebersold, R. *Nat. Protocols* **2006**, *1*, 139.
(88) Zhong, Y.; Hyung, S.-J.; Ruotolo, B. T. *Expert Rev. Proteomics* **2012**, *9*, 47.
(89) Bernstein, S. L.; Dupuis, N. F.; Lazo, N. D.; Wyttenbach, T.; Condron, M. M.; Bitan, G.; Teplow, D. B.; Shea, J.-E.; Ruotolo, B. T.; Robinson, C. V.; Bowers, M. T. *Nat. Chem.* **2009**, *1*, 326.
(90) Kaltashov, I. A.; Bobst, C. E.; Abzalimov, R. R. *Anal. Chem.* **2009**, *81*, 7892.
(91) Konermann, L.; Pan, J.; Liu, Y.-H. *Chem. Soc. Rev.* **2011**, *40*, 1224.
(92) Wales, T. E.; Engen, J. R. *Mass Spectrom. Rev.* **2006**, *25*, 158.
(93) Kiselar, J. G.; Chance, M. R. *J. Mass Spectrom.* **2010**, *45*, 1373.
(94) Konermann, L.; Stocks, B. B.; Pan, Y.; Tong, X. *Mass Spectrom. Rev.* **2010**, *29*, 651.
(95) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198.
(96) Chait, B. T. *Science* **2006**, *314*, 65.
(97) Engen, J. R. *Anal. Chem.* **2009**, *81*, 7870.
(98) Marcsisin, S. R.; Engen, J. R. *Anal. Bioanal. Chem.* **2010**, *397*, 967.
(99) Iacob, R.; Engen, J. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1003.
(100) Abzalimov, R. R.; Bobst, C. E.; Kaltashov, I. A. *Anal. Chem.* **2013**.
(101) Rand, K. D.; Zehl, M.; Jensen, O. N.; Jørgensen, T. J. D. *Anal. Chem.* **2009**, *81*, 5577.
(102) Roeser, J.; Bischoff, R.; Bruins, A.; Permentier, H. *Anal. Bioanal. Chem.* **2010**, *397*, 3441.
(103) Mendoza, V. L.; Vachet, R. W. *Mass Spectrom. Rev.* **2009**, *28*, 785.
(104) Gau, B.; Garai, K.; Frieden, C.; Gross, M. L. *Biochemistry* **2011**, *50*, 8117.
(105) Alexander, P.; Fox, M.; Stacey, K. A.; Smith, L. F. *Biochem. J.* **1952**, *52*, 177.
(106) Petrotchenko, E. V.; Borchers, C. H. *Mass Spectrom. Rev.* **2010**, *29*, 862.
(107) Sinz, A. *Anal. Bioanal. Chem.* **2010**, *397*, 3433.
(108) Leitner, A.; Walzthoeni, T.; Kahraman, A.; Herzog, F.; Rinner, O.; Beck, M.; Aebersold, R. *Mol. Cell Proteomics* **2010**, *9*, 1634.
(109) Lasker, K.; Förster, F.; Bohn, S.; Walzthoeni, T.; Villa, E.; Unverdorben, P.; Beck, F.; Aebersold, R.; Sali, A.; Baumeister, W. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1380.
(110) Herzog, F.; Kahraman, A.; Boehringer, D.; Mak, R.; Bracher, A.; Walzthoeni, T.; Leitner, A.; Beck, M.; Hartl, F.-U.; Ban, N.; Malmström, L.; Aebersold, R. *Science* **2012**, *337*, 1348.
(111) Trnka, M. J.; Burlingame, A. *Mol. Cell. Proteomics* **2010**, *9*, 2306.
(112) Jennebach, S.; Herzog, F.; Aebersold, R.; Cramer, P. *Nucleic. Acids. Res.* **2012**, *40*, 5591.
(113) Solforosi, L.; Criado, J. R.; McGavern, D. B.; Wirz, S.; Sánchez-Alavez, M.; Sugama, S.; DeGiorgio, L. A.; Volpe, B. T.; Wiseman, E.; Abalos, G.; Masliah, E.; Gilden, D.; Oldstone, M. B.; Conti, B.; Williamson, R. A. *Science* **2004**, *303*, 1514.
(114) Zheng, C.; Yang, L.; Hoopmann, M. R.; Eng, J. K.; Tang, X.; Weisbrod, C. R.; Bruce, J. E. *Mol. Cell. Proteomics* **2011**, *10*.
(115) Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S. J.; Robinson, C. V. *Nat. Protocols* **2008**, *3*, 1139.

(116) Mason, E. A.; McDaniel, E. W. *Transport Properties of Ions in Gases*; John Wiley & Sons: New York, 1988.
(117) Harvey, S. R.; Macphee, C. E.; Barran, P. E. *Methods* **2011**, *54*, 454.
(118) Kanu, A. B.; Dwivedi, P.; Tam, M.; Matz, L.; Hill, H. H. *J. Mass Spectrom.* **2008**, *43*, 1.
(119) Eiceman, G. A.; Karpas, Z. *Ion Mobility Mass Spectrometry*; Taylor and Francis: Boca Raton FA, 2005.
(120) Shvartsburg, A. A. *Differential Ion Mobility Spectrometry: Nonlinear Ion Transport and Fundamentals of FAIMS* CRC, 2008.
(121) Kolakowski, B. M.; Mester, Z. *Analyst* **2007**, *132*, 842.
(122) Kurulugama, R.; Nachtigall, F.; Lee, S.; Valentine, S.; Clemmer, D. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 729.
(123) Valentine, S.; Stokes, S.; Kurulugama, R.; Nachtigall, F.; Clemmer, D. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 738.
(124) Merenbloom, S. I.; Glaskin, R. S.; Henson, Z. B.; Clemmer, D. E. *Anal. Chem.* **2009**, *81*, 1482.
(125) Clemmer, D. E.; Jarrold, M. F. *J. Mass Spectrom.* **1997**, *32*, 577.
(126) Gillig, K. J.; Ruotolo, B.; Stone, E. G.; Russell, D. H.; Fuhrer, K.; Gonin, M.; Schultz, A. J. *Anal. Chem.* **2000**, *72*, 3965.
(127) Gillig, K. J.; Ruotolo, B. T.; Stone, E. G.; Russell, D. H. *Int. J. Mass Spectrom.* **2004**, *239*, 43.
(128) Valentine, S. J.; Koeniger, S. L.; Clemmer, D. E. *Anal. Chem.* **2003**, *75*, 6202.
(129) Tang, K.; Shvartsburg, A. A.; Lee, H. N.; Prior, D. C.; Buschbach, M. A.; Li, F. M.; Tolmachev, A. V.; Anderson, G. A.; Smith, R. D. *Anal. Chem.* **2005**, *77*, 3330.
(130) Koeniger, S. L.; Merenbloom, S. I.; Valentine, S. J.; Jarrold, M. F.; Udseth, H. R.; Smith, R. D.; Clemmer, D. E. *Anal. Chem.* **2006**, *78*, 4161.
(131) Koeniger, S. L.; Merenbloom, S. I.; Clemmer, D. E. *J. Phys. Chem. B* **2006**, *110*, 7017.
(132) Laschober, C.; Kaddis, C. S.; Reischl, G. P.; Loo, J. A.; Allmaier, G.; Szymanski, W. W. *J. Exp. Nanosci.* **2007**, *2*, 291.
(133) Hogan, C. J., Jr.; Ruotolo, B. T.; Robinson, C. V.; Fernandez de la Mora, J. *J Phys Chem B* **2011**, *115*, 3614.
(134) de la Mora, J. F.; de Juan, L.; Eichler, T.; Rosell, J. *Trac-Trends Anal. Chem.* **1998**, *17*, 328.
(135) Hogan, C. J., Jr.; de la Mora, J. F. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 158.
(136) Kaufman, S. L.; Kuchumov, A. R.; Kazakevich, M.; Vinogradov, S. N. *Anal. Biochem.* **1998**, *259*, 195.
(137) John, J. T.; Brian, B.; Joe, T.; Benner, W. H.; Gary, S. *Spectroscopy* **2004**, *18*, 31.
(138) Loo, J.; Berhane, B.; Kaddis, C.; Wooding, K.; Xie, Y.; Kaufman, S.; Chernushevich, I. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 998.
(139) Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2401.
(140) Shvartsburg, A. A.; Smith, R. D. *Anal. Chem.* **2008**, *80*, 9689.
(141) Giles, K.; Wildgoose, J. L.; Langridge, D. J.; Campuzano, I. *Int. J. Mass Spectrom.* **2010**, *298*, 10.

(142) Zhong, Y.; Hyung, S. J.; Ruotolo, B. T. *Analyst* **2011**, *136*, 3534.
(143) Giles, K.; Williams, J. P.; Campuzano, I. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 1559.
(144) Dugourd, P.; Hudgins, R. R.; Clemmer, D. E.; Jarrold, M. F. *Rev. Sci. Instrum.***1997**, *68*, 1122.
(145) Kemper, P. R.; Dupuis, N. F.; Bowers, M. T. *Int. J. Mass Spectrom.* **2009**, *287*, 46.
(146) Roscioli, K. M.; Davis, E.; Siems, W. F.; Mariano, A.; Su, W.; Guharay, S. K.; Hill, H. H. *Anal. Chem.* **2011**, *83*, 5965.
(147) Loo, J. A.; Kaddis, C. S.; Lomeli, S. H.; Yin, S.; Berhane, B.; Apostol, M. I.; Kickhoefer, V. A.; Rome, L. H. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1206.
(148) McKay, A. R.; Ruotolo, B. T.; Ilag, L. L.; Robinson, C. V. *J. Am. Chem. Soc.* **2006**, *128*, 11433.
(149) Han, L.; Hyung, S. J.; Mayers, J. J.; Ruotolo, B. T. *J. Am. Chem. Soc.* **2011**, *133*, 11358.
(150) Freeke, J.; Robinson, C. V.; Ruotolo, B. T. *Int. J. Mass Spectrom.* **2010**, *298*, 91.
(151) Liepold, L.; Oltrogge, L.; Suci, P.; Young, M.; Douglas, T. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 435.
(152) Tseng, Y. H.; Uetrecht, C.; Heck, A. J. R.; Peng, W. P. *Anal. Chem.* **2011**, *83*, 1960.
(153) van Breukelen, B.; Barendregt, A.; Heck, A. J. R.; van den Heuvel, R. H. H. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 2490.
(154) Hernández, H.; Makarova, O. V.; Makarov, E. M.; Morgner, N.; Muto, Y.; Krummel, D. P.; Robinson, C. V. *Plos One* **2009**, *4*, e7202.
(155) Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. *Anal. Chem.* **2010**, *82*, 9557.
(156) Sharon, M.; Michaelevski, I.; Eisenstein, M. *Anal. Chem.* **2010**, *82*, 9484.
(157) Morsa, D.; Gabelica, V.; De Pauw, E. *Anal. Chem.* **2011**, *83*, 5775.
(158) Benesch, J. L. P.; Ruotolo, B. T. *Curr. Opin. Struct. Biol.* **2011**.
(159) Ruotolo, B. T.; McLean, J. A.; Gillig, K. J.; Russell, D. H. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 158.
(160) Bleiholder, C.; Wyttenbach, T.; Bowers, M. T. *Int. J. Mass Spectrome.***2011**, *308*, 1.
(161) Felitsyn, N.; Kitova, E. N.; Klassen, J. S. *Anal. Chem.* **2001**, *73*, 4647.
(162) Jurchen, J. C.; Williams, E. R. *J. Am. Chem. Soc.* **2003**, *125*, 2817.
(163) Benesch, J. L. P.; Aquilina, J. A.; Ruotolo, B. T.; Sobott, F.; Robinson, C. V. *Chem. Biol.* **2006**, *13*, 597.
(164) Pagel, K.; Hyung, S. J.; Ruotolo, B. T.; Robinson, C. V. *Anal. Chem.* **2010**, *82*, 5363.
(165) Boeri Erba, E.; Ruotolo, B. T.; Barsky, D.; Robinson, C. V. *Anal. Chem.* **2010**, *82*, 9702.
(166) Zhang, H.; Cui, W. D.; Wen, J. Z.; Blankenship, R. E.; Gross, M. L. *Anal. Chem.* **2011**, *83*, 5598.
(167) Wysocki, V. H.; Jones, C. M.; Galhena, A. S.; Blackwell, A. E. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 903.
(168) Han, L.; Ruotolo, B. *Angew. Chem. Int. Edit.* **2013**, *52*, 8329.

(169) Han, L.; Hyung, S. J.; Ruotolo, B. *Angew. Chem. Int. Edit.* **2012**, *51*, 5692.
(170) Han, L.; Ruotolo, B. *Int J. Ion Mobil. Spectrom.* **2013**, *16*, 41.
(171) Jessica, N. R.; Suk-Joon, H.; Kristin, S. K.; Christel, C. F.; Matthew, B. S.; Brandon, T. R. *Anal. Chem.* **2013**, *85*.
(172) Zhou, M.; Sandercock, A. M.; Fraser, C. S.; Ridlova, G.; Stephens, E.; Schenauer, M. R.; Yokoi-Fong, T.; Barsky, D.; Leary, J. A.; Hershey, J. W.; Doudna, J. A.; Robinson, C. V. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 18139.
(173) Sharon, M.; Taverner, T.; Ambroggio, X. I.; Deshaies, R. J.; Robinson, C. V. *PLoS. Biol.* **2006**, *4*, 1314.
(174) Levy, E. D.; Boeri Erba, E.; Robinson, C. V.; Teichmann, S. A. *Nature* **2008**, *453*, 1262.
(175) Nooren, I. M.; Thornton, J. M. *EMBO J* **2003**, *22*, 3486.
(176) Uetrecht, C.; Barbu, I. M.; Shoemaker, G. K.; van Duijn, E.; Heck, A. J. R. *Nat. Chem* **2011**, *3*, 126.
(177) Knapman, T. W.; Morton, V. L.; Stonehouse, N. J.; Stockley, P. G.; Ashcroft, A. E. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 3033.
(178) Bornschein, R. E.; Hyung, S. J.; Ruotolo, B. T. *J. Am. Soc. Mass Spectrom.* **2011**.
(179) Bagal, D.; Kitova, E. N.; Liu, L.; El-Hawiet, A.; Schnier, P. D.; Klassen, J. S. *Anal. Chem.* **2009**, *81*, 7801.
(180) Flick, T. G.; Merenbloom, S. I.; Williams, E. R. *Anal. Chem.* **2011**, *83*, 2210.
(181) Bohrer, B. C.; Mererbloom, S. I.; Koeniger, S. L.; Hilderbrand, A. E.; Clemmer, D. E. In *Annual Review of Analytical Chemistry*; Annual Reviews: Palo Alto, 2008; Vol. 1, p 293.
(182) Valentine, S. J.; Plasencia, M. D.; Liu, X. Y.; Krishnan, M.; Naylor, S.; Udseth, H. R.; Smith, R. D.; Clemmer, D. E. *J. Proteome Res.* **2006**, *5*, 2977.
(183) Ruotolo, B. T.; Gillig, K. J.; Stone, E. G.; Russell, D. H. *J. Chromatogr. B* **2002**, *782*, 385.
(184) Ruotolo, B. T.; McLean, J. A.; Gillig, K. J.; Russell, D. H. *J. Mass Spectrom.* **2004**, *39*, 361.
(185) Valentine, S. J.; Kulchania, M.; Barnes, C. A. S.; Clemmer, D. E. *International J. Mass Spectrom.* **2001**, *212*, 97.
(186) Hoadlund-Hyzer, C. S.; Li, J. W.; Clemmer, D. E. *Anal. Chem.* **2000**, *72*, 2737.
(187) Stone, E.; Gillig, K. J.; Ruotolo, B.; Fuhrer, K.; Gonin, M.; Schultz, A.; Russell, D. H. *Anal. Chem.* **2001**, *73*, 2233.
(188) McLean, J. A.; Fenn, L. S. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2196.
(189) Woods, A. S.; Ugarov, M.; Egan, T.; Koomen, J.; Gillig, K. J.; Fuhrer, K.; Gonin, M.; Schultz, J. A. *Anal. Chem.* **2004**, *76*, 2187.
(190) Bernstein, S. L.; Dupuis, N. F.; Lazo, N. D.; Wyttenbach, T.; Condron, M. M.; Bitan, G.; Teplow, D. B.; Shea, J. E.; Ruotolo, B. T.; Robinson, C. V.; Bowers, M. T. *Nat. Chem.* **2009**, *1*, 326.
(191) Ashcroft, A. E. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1087.
(192) Radford, S. E.; Smith, D. P.; Ashcroft, A. E. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 6794.
(193) van Duijn, E.; Barendregt, A.; Synowsky, S.; Versluis, C.; Heck, A. J. R. *J. Am. Chem. Soc.* **2009**, *131*, 1452.

(194) Heck, A. J. R.; Uetrecht, C.; Rose, R. J.; van Duijn, E.; Lorenzen, K. *Chem.l Soc. Rev.* **2010**, *39*, 1633.

Chapter 2. Characterizing the Resolution and Accuracy of a Second-Generation Traveling-Wave Ion Mobility Separator for Biomolecular Ions

2.1 Introduction

Ion mobility (IM) separation is used extensively to assess the structure and identity of a large number of molecular analytes, ranging from atomic species[1] to multiprotein complexes[2,3] and viruses.[4,5] The combination of this separation technology with mass spectrometry (IM-MS) has allowed more detailed and multi-dimensional analysis of both complex mixtures and biomolecules.[6,7] Such hybrid IM-MS instruments have taken multiple forms, and recent advancements have seen improvements in both the IM and MS stages in modern instrumentation.[8-11] Among these hybrid instruments, the quadrupole (Q)-IM-ToF MS instrument known as the Synapt HDMS system (Waters, Milford, MA, USA) uses a traveling-wave ion guide[12] to perform IM separation ($t/\Delta t \approx 10\text{-}15$), coupled to a high-performance time-of-flight MS ($m/\Delta m \approx 20{,}000$) and was released as the first commercial IM-MS platform in 2006.[13] A second-generation version of this platform, the Synapt G2 HDMS was released in 2009, and has demonstrated higher resolution and resolving powers in both the IM and MS dimensions.[14]

While many applications exist for IM-MS instrumentation, our group and others are interested in experiments aimed at determining the structure of multiprotein complexes.[15,16] Multiple datasets acquired on first-generation Synapt instrumentation exist for a number of applications in this general area including: the overall assessment of multiprotein topology,[17,18] the study of host-guest complexes,[19] the characterization of protein-ligand complexes,[20] and the refinement of small sections of multiprotein topology maps.[21] In all of these cases, IM resolution, precision and accuracy are the key figures of merit that determine the extent to which IM separation can be used to refine the structure of biomolecules in a quantitative fashion. Currently, few reports exist detailing the application of the second-generation, higher-performance Synapt G2 instrumentation to similar protein systems, but it is clear that defining the conditions and instrument settings that lead to the highest-precision, highest-accuracy IM-MS data on this platform will be crucial for applications in structural biology.

The factors that influence IM resolution and accuracy have been the subject of detailed study for decades.[22-24] For drift tube-type linear-field IM separators, the arrival time distribution of ions can be approximated by a modified Gaussian function, the total width of which (W_T) depends upon broadening components from multiple sources:

$$W_T = W_D + W_{SC} + W_P + W_{RXN} + W_C \quad (2.1)$$

Equation 2.1 describes the additive influence of diffusion (W_D), space charge (W_{SC}), initial pulse width (W_P), reaction chemistry (W_{RXN}), and any changes in conformation

that occur on a timescale faster than the IM separation performed (W_C) on the total width of the recorded IM arrival time distributions for biomolecules. For cases where the W_T can be approximated by W_D alone, the resolution of IM (R_{DT}) can be expressed as:

$$R_{DT} = \frac{1}{4}\left(\frac{zeEL}{kTln2}\right)^{\frac{1}{2}} \tag{2.2}$$

where z is the ion charge, e is the charge on an electron, E is the field strength, L is the length of the drift tube, k is the Boltzmann constant, and T is drift tube temperature. Experimental evidence has shown that R_{DT} can occupy values greater than 150 for instruments following conventional design practices,[25] although such high R_{DT} values have yet to be demonstrated for large biomolecules. The precision of IM measurements is often reported as a relative variance in collision cross-section (Ω), the ultimate size information extracted from IM separations. Such variances can be as low as 1% and depend upon the reproducibility of the drift time measurement, and the ability to measure drift cell pressure and temperature accurately.

Since traveling-wave ion mobility separators use time-varying potential 'waves' to drive ions through the drift tube, rather than a constant electric field, the resolution, precision, and accuracy achievable using such devices are defined differently. The instrumental factors that influence the resolution of traveling-wave IM separators (R_{TW}) have been described from first-principles.[26] In cases where the width of the arrival time distribution is limited by diffusion alone, IM resolution is defined as:

$$R_{TW} = \frac{1}{4}\left(\frac{2ze[WH][WL]KE}{[WV]kTln2}\right)^{\frac{1}{2}} \quad (2.3)$$

where *K* is the ion mobility, *[WH]* is the wave amplitude or 'wave height' (in units of V), *[WL]* is the wave length, and *[WV]* is the wave velocity (in units of m/s). Therefore, theory predicts that R_{TW} will increase with increasing *K* and *[WH]*, and exhibit an inverse relationship to wave velocity *[WV]*; however, this relationship has yet to be rigorously tested experimentally. Due to the complex nature of ion drift in a traveling-wave IM separator, a calibration curve is typically constructed to convert ion drift times into Ω values,[27] where Ω is related to traveling-wave drift time via an exponential function:

$$\Omega = A\,[t_D'^{X}]z\left[\frac{1}{\mu}\right]^{\frac{1}{2}} \quad (2.4)$$

where t_D' is traveling-wave drift time corrected for time spent outside the drift tube, *z* is the ion charge, μ is the reduced mass of the ion-neutral collision complex, A is a fitting parameter, and X is the exponential calibration factor. Travelling-wave ion mobility data and theory currently indicate that the accuracy of this conversion greatly depends upon the calibration dataset and the instrument settings used for the calibration;[26] however, the complete extent of potential variations in Ω from traveling-wave IM measurements is currently unknown. Note also that the exponential relationship between t_D and Ω leads to different effective resolution values when ions are considered in either drift time or collision cross-section space. For example, for a quadratic relationship between Ω and t_D, the traveling wave Ω

resolution (R_{TW}^{Ω}) would be twice the traveling wave resolution on a drift time axis (R_{TW}^{t}).[26]

In this study, we chart the resolution (in both drift time and collision cross-section space) and Ω accuracy of the second-generation traveling-wave ion mobility separator within the Synapt G2 platform. From these investigations, we are also able to generate general predictions regarding the Ω precision on this instrument platform. Our results indicate good agreement with theory and determine optimal conditions where maximum resolution and accuracy is achieved for biomolecules. We also briefly consider the issue of ion heating during the IM measurement, and find its influence on cross-section accuracy and IM resolution to be negligible for biomolecule measurements made on the second-generation platform utilized in these studies.

2.2 Experimental

2.2.1 Sample Preparation

Peptide SDGRG (Sigma, S3771) and additional peptide mixture (5 μM, Waters, 186002337) were prepared in water/methanol/acetic acid (49/49/2) solutions. Cytochrome c (5 μM, Sigma, C2506) was prepared in 200mM aqueous ammonium acetate solution, and then stored at -20℃. Protein complexes avidin (Sigma, A9275), alcohol dehydrogenase (ADH) (Sigma, A7011) and glutamate dehydrogenase (GDH) (Sigma, G7882) were solublized in 200 mM aqueous ammonium acetate solutions and

stored at -20 ^{o}C. The samples (10 μM concentration of the complex) were then buffer-exchanged using a Micro Bio-Spin 6 column (Bio-Rad, Hercules, CA) equilibrated with 200mM aqueous ammonium acetate just prior to IM-MS analysis.[29]

2.2.2 Synapt G2 Instrumentation

The second-generation Synapt platform has been described in some detail elsewhere,[14] and its general layout and improvement has been discussed in Chapter 1 in Figures 1-2 and 1-4. Although the instrument utilizes an upgraded ToF geometry with improved detection electronics as well as some additions to the ion source region, we will not discuss these aspects of the instrument in any detail here. The 'tri-wave' region of the instrument consists of three main traveling-wave ion guide sections, enclosed from the main vacuum system, coupled to smaller traveling-wave ion guides that operate within the main vacuum system for increased conductance of the gas load within the enclosed regions of the instrument. This increased conductance between sections is an improvement over the first-generation instrument geometry. All data were collected using instrument settings on the Synapt G2 designed to preserve fragile non-covalent interactions within protein structures and between protein-protein interactions. This typically involves elevated pressures in the source region (5-7 mbar), and decreasing all focusing voltages (*e.g.*, cone, extractor, and bias voltages) to their minimum values for ion transmission.[28]

2.2.3 Data Collection and Analysis

To investigate the influence of *[WH]*, *[WV]* and pressure in the IM separation region (*p)* on IM resolution, all other instrument parameters are kept at constant values unless otherwise stated. For example, it was necessary to vary the operating pressure of the ion trap region of the instrument to achieve ion transmission for both peptides and large protein complexes under similar instrument settings (varied from 0 to 10 mL/min Ar flow rate), and our experimental results indicate that altering this operating pressure has no influence on IM resolution or calibration. Typical tuning ranges are: *[WH]* = 16-40 V, *[WV]* = 150-2000 m/s and *p* = 1-4 mbar. The ion release time from the ion trap region can have a substantial influence on resolution, and was kept at a minimum value (< 200 μs) for all experimental data reported here. When increasing the pressure of the IMS cell, we endeavored to keep the He/N_2 constant in terms of gas flow rate, although only limited experiments have been performed so far to investigate optimum gas flow-rate ratios (Figure S.2-3).

As with the experiments aimed at IM resolution, IM calibration experiments targeted the influence of *[WH]*, *[WV]* and *p* on instrument performance. Calibration curves were generated using a previously-described protocol,[27] and using literature Ω values derived for use with the Synapt instrument platform.[29] Gaussian fitting was used to determine centroid drift time values for both standards and unknowns for the calibration results reported here. While we endeavored to keep all but the key experimental variables constant, some parameters were modified by small amounts to achieve ion transmission for calibration experiments in excess of those discussed

above. These included small variations to the helium cell DC bias, trap bias, and overall ToF record length. In all cases, such variations in instrument parameters were found to have no influence on IM calibration under the conditions reported here.

2.3 Results and Discussion

2.3.1 Trends in Drift Time Resolution on the Second-Generation Traveling-Wave Platform

Equation 2.3 predicts that changes in both *[WH]* and *[WV]* will influence the R_{TW}^{t} recorded. In addition, since Equation 2.3 only applies under conditions where ions exposed to relatively low values of field-per-unit pressure (*E/p*), the increase in absolute drift tube pressure should also have an influence on IM separations, as its increase maximizes the total field strength that can be usefully applied to the ions. Therefore, we chose to track the resolution of IM separation on Synapt G2, as a function of the two traveling-wave associated parameters, *[WH]* and *[WV]*, as well as the pressure in the IM separation region, p. Also, we surveyed a number of peptides, proteins and protein complexes in order to cover biomolecules with a range of K values, and to consider the effect of ion mobility itself on IM resolution.

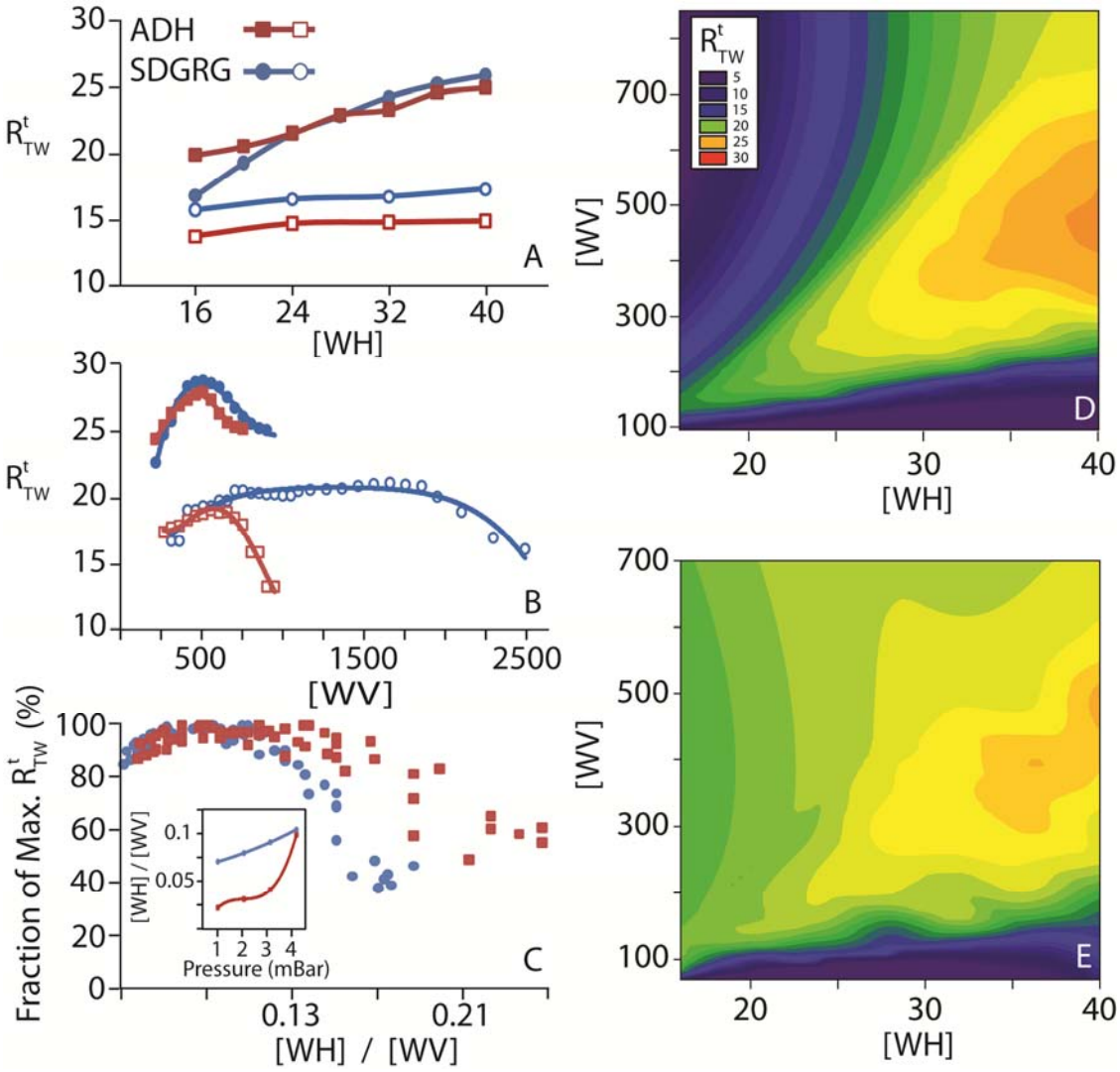

Figure 2-1 Charting the resolution of the second-generation traveling-wave ion mobility separator on the Synapt G2.

(A) A plot of wave height, *[WH]*, versus traveling-wave drift time resolution ($\boldsymbol{R^t_{TW}}$), for both the ADH tetramer (red squares) and the SDGRG pentapeptide (blue circles). Data is shown for two pressures, 4mBar (closed symbols) and 2mBar (open symbols). **(B)** A plot of $\boldsymbol{R^t_{TW}}$ versus traveling wave velocity, *[WV]*. Data is shown for ADH and SDGRG at two pressures in the same manner as A. **(C)** A plot of the *[WH]* / *[WV]* ratio against the fraction of maximum resolution achieved at a given *[WH]* at 4mbar. Data is shown for ADH and SDGRG, and indicates that similar *[WH] / [WV]* ratios bring about maximum $\boldsymbol{R^t_{TW}}$ for both species (~0.1). The inset shown indicates the trend for the *[WH] / [WV]* ratio as a function of pressure in the IM separator. **(D)** Contour plot charting the influence of *[WH]* and *[WV]* on $\boldsymbol{R^t_{TW}}$ (indicated on a color scale) for the SDGRG pentapeptide at 4 mBar pressure in the IM separator. **(E)** A contour plot similar for D, but for ADH tetramer data. Similar *[WH]* and *[WV]* values provide optimum resolution for both ADH and SDGRG ions.

Figure 2-1 shows several slices across the multi-dimensional datasets acquired, where the instrumental and sample-related variables described above were altered and R_{TW} recorded. For the data shown in Figure 2-1, we focus on two ions having disparate *K* values: the SDGRG pentapeptide[30] (490 Da, 1^+) and alcohol dehydrogenase (ADH) tetramer ions (141 kDa, 24^+) as snapshots of our current dataset. First, we examined the performance characteristics of IM separator at a given pressure by examining R_{TW}^t at a range of values of *[WV]* and *[WH]*. In Figure 2-1A, a plot of the maximum R_{TW}^t value achieved at any *[WV]* value is shown versus *[WH]*. The plot exhibits a positive correlation between the two parameters, as predicted by Equation 2.3 for both SDGRG pentapeptide and ADH tetramer ions. Note that the maximum R_{TW}^t for the peptide is observed to exceed that of the protein complex under most *[WH]* values, also as predicted by Equation 2.3. We cannot, however, rule out the influence of enhanced conformational broadening for ADH ions relative to SDGRG ions (W_c, Equation 2.1) as a root cause of this observation.

Figure 2-1B shows similar data where maximum R_{TW}^t (for all *[WH]* values) is plotted against *[WV]*. Here again, the trend observed is in good agreement with that predicted by Equation 2.3. In addition, previously-reported numerical simulations of R_{TW}^t show similar trends as a function of *[WV]*.[26] We interpret the observed decreases in R_{TW}^t at the lowest *[WV]* values reported in Figure 2-1B as related to the greater proportion of ions carried with the traveling wave, rather than undergoing IM separation, under these conditions.

In order to provide a metric for tuning the resolution of the traveling-wave device the

ratio between *[WH]* and *[WV]* was calculated and the maximum resolution afforded under the setting was estimated (Figure 2-1C). Since *[WH]* is positively-correlated with R_{TW}^{t} at all *[WV]* (Figure 2-1A), the data shown in Figure 2-1C are normalized for its influence, and plotted on a relative resolution axis, expressed as a percent of the total resolution achieved at a given *[WH]* value. As with other datasets shown in Figure 2-1, p is held constant at 4mbar. The ideal value of *[WH] / [WV]* for both ADH and SDGRG centers on ~0.1, however the precise value changes slightly depending upon the wave height (e.g., between 0.088 and 0.13, Figure S.2-1). Much larger changes in the ideal *[WH] / [WV]* ratio for R_{TW}^{t} are observed as a function of pressure in the IM separation region (inset, Figure 2-1C). Measured values of p ranging from 1 to 4 mBar provide ratios ranging from 0.025 to 0.1 for ADH tetramer ions, while much less variation is observed for SDGRG ions.

A contour plot, showing the values of *[WH]* and *[WV]* that lead to optimum R_{TW}^{t} for SDGRG at 4 mbar pressure in the IM separator, is shown in Figure 2-1D. Steep drop-offs in R_{TW}^{t} are observed at values of low *[WH]* and high *[WV]* (upper left, Figure 2-1D). We also note that some areas on the plot shown, e.g., high *[WH]* and low *[WV]* do not provide either resolved IM data or ion transmission, and are shown on this plot for purposes of completeness only. The optimum ratio between *[WH]* and *[WV]*, shown explicitly in Figure 2-1C, can also be observed in this data, as a curved area of high R_{TW}^{t} values that extends from *[WH]* = 20, *[WV]* = 150 to *[WH]* = 40, *[WV]* = 500. As discussed above, the maximum R_{TW}^{t} are universally accessed at the highest *[WH]* values.

A similar three-dimensional plot is shown illustrating the influence of *[WH]* and *[WV]* on R_{TW}^{t} for ADH tetramer ions. Upon comparing Figures 2-1D and 2-1E, it is apparent that ADH ions can be transmitted and IM separated over a wider range of *[WH]* and *[WV]* values, and that the R_{TW}^{t} recorded at many *[WH] / [WV]* ratios for ADH exceed those of the SDGRG pentapeptide. This trend is also observable in Figure 2-1A, where the 4 mBar dataset shows a cross-over point between the R_{TW}^{t} values of ADH and SDGRG at low *[WH]* values. The relative 'flatness' of the ADH R_{TW}^{t} trends compared to those recorded for SDGRG is most-likely due to the substantially higher charge state of the ADH ions measured here (24^{+} versus 1^{+}), leading to higher effective separation field strengths experienced for the tetramer ions at all *[WH]* values.

2.3.2 Optimizing Conditions for Maximum Cross-section Accuracy

The discussion above is limited to the factors that influenceR_{TW}^{t}; however, resolution is only one of several contributing factors to high quality IM-MS data. When IM drift times are converted to Ω values, a calibration curve is typically constructed from ion standards with known Ω values (Equation 2.4). The goodness of fit of this calibration curve to the experimental data, along with the error associated with the standards used and the reproducibility of the measurements is a contributing factor to the overall accuracy of Ω values reported from traveling-wave IM measurements. While previous protocols have suggested that maximizing the calibration curve correlation coefficient (R^2) is a necessary step to achieve high-accuracy Ω values from traveling-wave drift

times, little data is currently available on what instrument parameters most-influence this critical metric of Ω accuracy.[27]

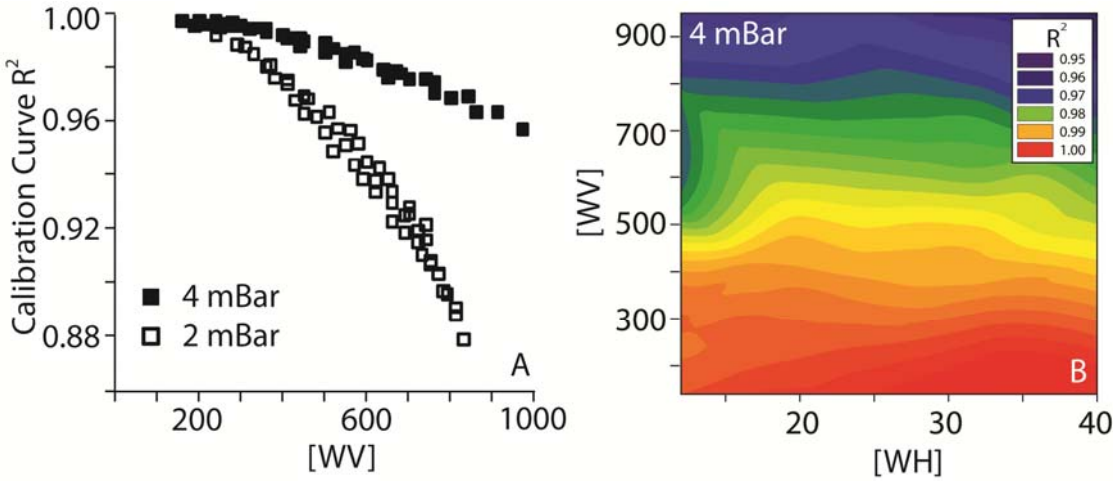

Figure 2-2 The goodness of fit for calibration curves that allow for the conversion between IM drift time measurements and collision cross-section accuracy depend primarily upon wave velocity, and do not have a strong wave height dependence.
(A) A plot of *[WV]*, versus the correlation coefficient for calibration curves recorded for two operating pressures for the IM separator. Decreased accuracy is observed at lower pressure and higher *[WV]* values. **(B)** A 3Dplot of *[WV]*, *[WH]* and calibration curve correlation coefficient (R^2). Again, substantial differences in correlation are observed only as a function of *[WV]*, and not *[WH]*.

Figure 2-2A shows a plot of *[WV]* versus R^2 for approximately 100 calibration curves acquired under two separate pressure conditions, one at 4 mbar, and the other at 2 mbar, recorded across *[WH]* ranging from 12 to 40 V. Peptide, protein, and multiprotein complex ions were used to construct the calibration curves for Figure 2-2, and our calibration dataset also extends to both lower and higher pressures (Figure S.2-1). In general, we observe substantial decreases in R^2, and also Ω accuracy, at both high *[WV]* values and lower pressures on the second-generation traveling wave IM separator on the Synapt G2. At 4 mBar, R^2 remains relatively high across a broad *[WV]* range, and decreases to a value of 0.96 for calibration data collected over all

[WH] at *[WV]* = 1000. By contrast, Ω accuracy decreases more significantly at 2 mBar, where only the lowest *[WV]* values provide suitable Ω accuracy for most IM-MS applications.

In order to investigate the influence of *[WH]* on R^2 at all *[WV]* values, we constructed a three-dimensional plot shown in Figure 2-2B. Here, *[WH]* is plotted against *[WV]*, in a similar fashion to Figures 2-1D and 2-1E, where R^2 values are denoted by the color axis shown. In almost all cases, *[WH]* has no significant influence on the R^2 values determined for IM calibration curves. For example, for calibration curves determined at *[WV]* = 300, R^2 values varied by less than 0.1%. Overall the data suggests that increased Ω accuracy can be achieved by operating the instrument at minimized *[WV]* values and at maximized pressures, in a fashion relatively independent of *[WH]* settings.

2.3.3 The Influence of Ion Heating on Traveling Wave Resolution and Cross Section Accuracy

The apparently independent nature of Ω accuracy as a function of *[WH]*, the main driver in the maximum effective field strength that the ions experience during IM separation, led us to consider more closely the influence of ion heating in our dataset. If ions are thermally excited via energetic collisions with the neutral gas used for IM separation, structural transitions may occur, and degrade both Ω accuracy and R_{TW}^{t} (by increasing W_c, Eqn. 1). Previous theoretical and experimental reports have discussed the potential of ion heating in the traveling-wave IM separation region.[26,31]

It is important to point out, however, that although ion heating may occur at some level during IM separation, threshold ion internal energies are required for biomolecules to undergo structural rearrangements that can be observed by IM. Therefore, increased ion temperature may not always lead to the decreases in IM data quality described above.

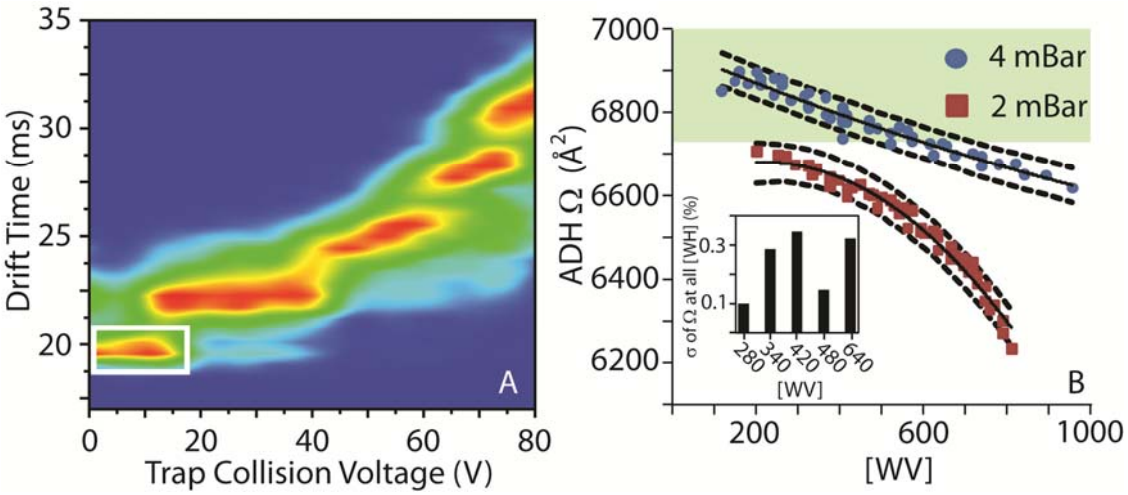

Figure 2-3 Exploring the influence of ion heating on the accuracy of collision cross section measurements using a traveling wave ion mobility device.
(A) The collision-induced unfolding fingerprint for ADH ions, after having undergone activation in the ion trap that precedes the IM separation region on the Synapt G2. The most compact form of ADH (white box) undergoes unfolding at significantly lower trap collision voltages than other multiprotein complexes. **(B)** ADH Ω values are plotted as a function of the *[WV]* used to acquire the data for two pressures, 2 mBar (red squares) and 4 mBar (blue circles). Data is shown using all *[WH]* values available at each *[WV]*, and the prediction interval for the fitted polynomial trend line is computed for each pressure dataset (black dashed line). The inset shown plots the relative standard deviation (σ) for Ω recorded at all *[WH]* values at five selected *[WV]*. The average σ as a function of [WH] is 0.23 %.

In order to investigate the influence of ion heating on the accuracy of Ω values determined by travelling-wave IM drift time measurements, we selected 24^+ ADH tetramer ions and treated them as ions with unknown Ω within our calibration

functions. The selection of ADH ions specifically was motivated by the data shown in Figure 2-3A, a collision induced unfolding (CIU) dataset for 24^+ ADH tetramer ions, where an activating acceleration voltage in the ion trap prior to IM separation is increased and the drift time values occupied by intact tetramer are recorded. The intensity for each feature is represented on a color axis, similar to other 3D plots shown in this chapter. The most compact form of the ADH tetramer persists only until trap collision voltages exceed 15 V. This value is far less than those recorded for smaller peptides, proteins, and protein tetramers under similar conditions.[32,33] Therefore, as the ADH tetramer is prone to relatively facile gas-phase structural rearrangements as a function of ion heating, measurements of the Ω of the ion as a function of *[WH]* should reveal evidence of any adverse effects of activating collisions during IM separation on Ω accuracy for multiprotein systems.

To test this hypothesis we measured Ω values were recorded for ADH tetramer ions as a function of *[WV]* over all *[WH]* values accessible at both 4 mBar and 2 mBar operating pressures in the IM separation (Figure 2-3B) region. Experimental Ω values decrease in our dataset relative to the known ADH Ω values, a type of transition not observed in Figure 2-3A. As observed in Figure 2-2, the primary driver of defects in Ω inaccuracy results from increasing *[WV]*, not *[WH]*. The inset in Figure 2-3B illustrates this point, as do the prediction intervals for the fitted trends shown for the datasets in the main figure. Variations in Ω as a function of *[WH]* for ADH tetramers demonstrate a remarkably small amount of variation under the measurement and calibration conditions used here, with standard deviation (σ) values ranging from

0.1% to 0.3% across the whole dataset at all pressures (Figure 2-3B, Inset). Thus, our data suggests that ion heating is an insignificant factor in the Ω accuracy for ions measured here using second-generation traveling-wave instruments. Data shown in Figure 2-1A also illustrates that maximum R_{TW}^{t} continually increases as a function of *[WH]* and, therefore, IM separation field strength, further indicating the lack of ion heating influence in our current dataset.

2.3.4 Collision Cross-Section Resolution Compared with Drift Time Resolution for Traveling-Wave Instruments

Because traveling-wave IM separators have an exponential relationship between measured drift time and collision cross-section, the resolution on a drift time axis, R_{TW}^{t}, does not equal the effective resolution of the separation on a collision cross-section axis, R_{TW}^{Ω} (Equation 2.4). Since the calibration function is known to change as a function of *[WV]*, *[WH]*, and *p*, IM drift time data must be calibrated, and the exponential function is known to change as a function of *[WV]*, *[WH]*, and *p*, IM drift time data must be calibrated, and the exponential calibration factor, X, determined before R_{TW}^{Ω} can be calculated. Obviously, lower values of X are preferred, as they lead to higher overall R_{TW}^{Ω} values. Figure 2-4A plots X values determined for calibration curves determined at both 4 mBar and 2 mBar as a function of *[WV]* and at all accessible *[WH]* settings. The calibration function exponential factor is observed to decrease as a function of *[WV]* giving the lowest X factors at maximum *[WV]* for both pressures (Figure S.2-2). In addition, X decreases dramatically at lower pressures. R_{TW}^{t} values are also shown in Figure 2-4A over the same *[WV]* range and

illustrate that while R_{TW}^{t} is maximized at low *[WV]*, the calibration curve X factors reaches its optimum (lower) values at larger *[WV]* values.

By combining the recorded X factors and R_{TW}^{t} through Equation 2.4, we have projected effective values of R_{TW}^{Ω} for our data at 4 mBar and 2 mBar (Fig 2-4B, red data points), where both trends reach a maximum value of ~40 for ADH tetramer ions (maximum R_{TW}^{t} under similar conditions was measured to be ~25). While R_{TW}^{t} is maximized at low *[WV]* and high *[WH]*, higher *[WV]* values at maximum *[WH]* is necessary to achieve maximum R_{TW}^{Ω}. However, as shown in Figure 2-2, and overlaid in Figure 2-4B (blue data points), the accuracy of the calibration curve decreases at higher *[WV]* and lower pressures. These conditions of minimum Ω accuracy correspond to those conditions where R_{TW}^{Ω} is maximized. This observation, and the data shown in Figure 2-4B, frames a critical outcome of this work, in that it appears that a balance exists between maximizing R_{TW}^{Ω} and optimizing Ω accuracy on traveling wave-type IM separation devices.

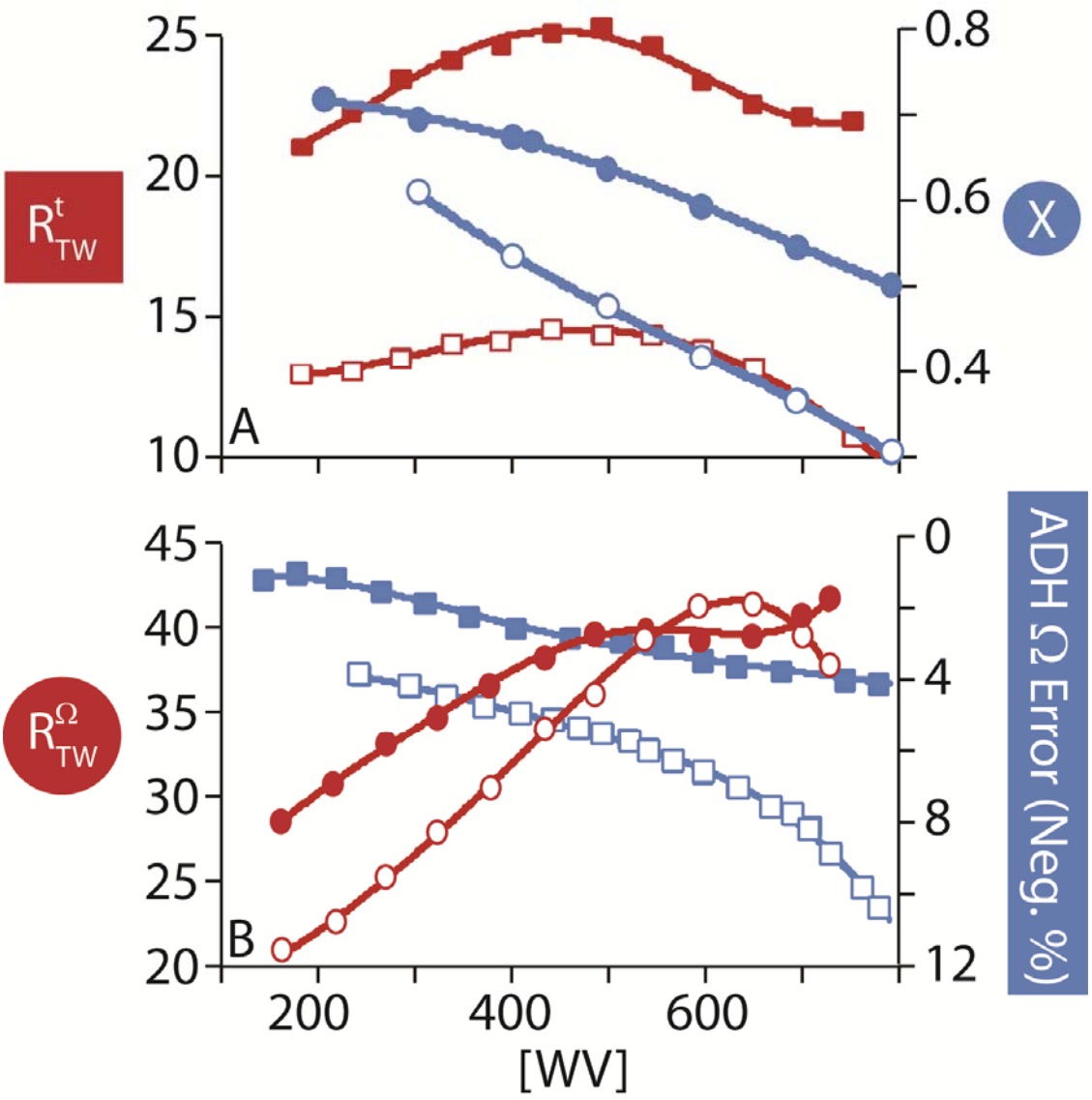

Figure 2-4 Achieving optimum collision cross-section resolution.
(A) A plot of drift time resolution for the traveling-wave (red squares) and calibration function exponential factor, X (blue circles) as a function of wave velocity, *[WV]*. While R^t_{TW} achieves maximum values at *[WV]* = 500, X achieves minimum values at the highest *[WV]* available. Two pressures are shown, 4 mbar (closed symbols) and 2 mbar (open symbols), and decreased R^t_{TW} and X are observed at lower pressures **(B)** By combining R^t_{TW} and X, we can project maximum R^Ω_{TW} values as a function of *[WV]* (red circles). As expected, values are maximized at high *[WV]* values, where X is minimized. Somewhat surprisingly, lower pressures also display high R^Ω_{TW}, due to the lower X factor under these conditions. We also overlay a plot of ADH collision cross-section (Ω) accuracy in this plot to emphasize the point that optimal conditions for separating ions and measuring their Ω values are not identical on traveling-wave IM instruments.

2.4 Conclusions

The data presented here describes the first attempt to generate a working description of both collision cross-section resolution and accuracy and the instrumental factors that influence them on a traveling-wave ion mobility separation device. Four main conclusions can be drawn from our dataset. First, optimum R_{TW}^{t} values are achieved at maximum values of p and *[WH]*, but at intermediate values of *[WV]*, where the ratio between *[WH]* and *[WV]* seems to be a useful value for determining the maximum R_{TW}^{t} at all IM region pressures. Second, the collision cross-section accuracy, which is indicated by the correlation coefficient for the calibration curve used to convert between traveling-wave drift times and collision cross-section, tends to decrease at higher *[WV]* and at lower pressures.

Thirdly, the Ω accuracy we record is observed to be relatively invariant as a function of *[WH]*, indicating that ion heating within the IM separation region does not drive conformational changes for biomolecules under the conditions reported here. Similar results were obtained for R_{TW}^{t}, further indicating the lack of effective ion heating within the second-generation traveling-wave device. This conclusion stands in contrast to previous theoretical and experimental reports that indicated that such heating was both probable and capable of causing conformational transitions in a range of proteins and their complexes.[26,31] It should be noted, however, that previous experimental data was acquired on a first-generation traveling-wave instrument that, while running at lower *[WH]* values, used substantially lower operating pressures within the IM separation region, leading to more-substantial maximum field strengths

encountered by ions undergoing IM separation.

A final conclusion illustrates that while R_{TW}^{t} is maximized at low *[WV]*, R_{TW}^{Ω} is maximized at higher *[WV]* as the exponential factor that determines the conversion between drift time and collision cross section is minimized. These observations frame an apparent balance between tuning the IM separator for either maximum R_{TW}^{Ω} or maximum Ω accuracy. While the maximum value of R_{TW}^{Ω} for multiprotein complexes reported here is ~40, values between 50 and 60 R_{TW}^{Ω} have been recorded in this study for peptide ions (Figure S.2-3). Surprisingly, operating the IM separation region at lower pressures, and thus driving X to lower values, leads to equivalent or superior R_{TW}^{Ω} values when compared with experiments carried out at higher operating pressures (Figure 2-4B). We consider that differences in the composition of the gas within the IM separation region may account for these observations, and our continuing work in this area will test this hypothesis.

Overall, the data presented here creates, for the first time, a formal description of the tuning parameters and protocols necessary for achieving high-accuracy and high-resoltuion IM measurements on widely available commercial IM-MS instrumentation (e.g. Synapt G2). While the potential applications of the database presented here are broad, future work in our group will utilize the IM data presented here to achieve highly accurate and precise Ω values for protein and their complexes for generating topology models of multiprotein assemblies.

Acknowledgements to Kevin Giles (Waters), Matthew F. Bush (Oxford) and Zoe Hall (Oxford) for informative discussions. This work was supported by University of

Michigan Startup funds (BTR).

2.5 Supplemental Information

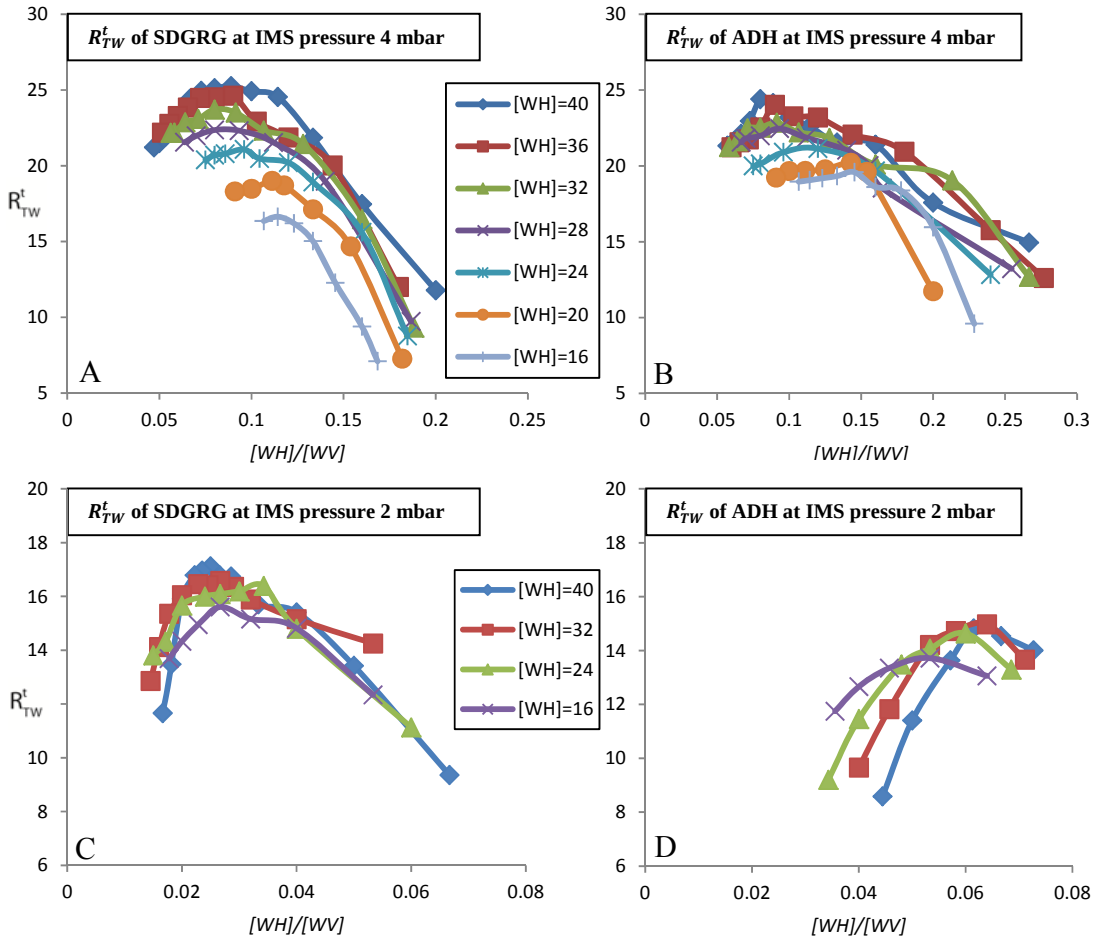

Figure S. 2-1 Traveling wave drift time resolution (R^t_{TW}) versus *[WH]/[WV]*, shown at each *[WH]* measured.

Although each *[WH]* gives different maximum R^t_{TW}, the ideal *[WH]/[WV]*s which give the maximum R^t_{TW} are similar across different *[WH]* (0.088-0.114 for SDGRG at 4mbar **(A)**). Much larger changes in the ideal *[WH]/[WV]* ratio for R^t_{TW} are observed as a function of pressure in the IM separation region. As shown in the figures, ideal *[WH]/[WV]* centers at ~0.1 for SDGRG at 4mbar (A) decreased to ~0.025 for SDGRG at 2mbar **(C)**, and it decreased from ~0.11 for ADH at 4mar **(B)** to ~0.5 for ADH at 2mbar.

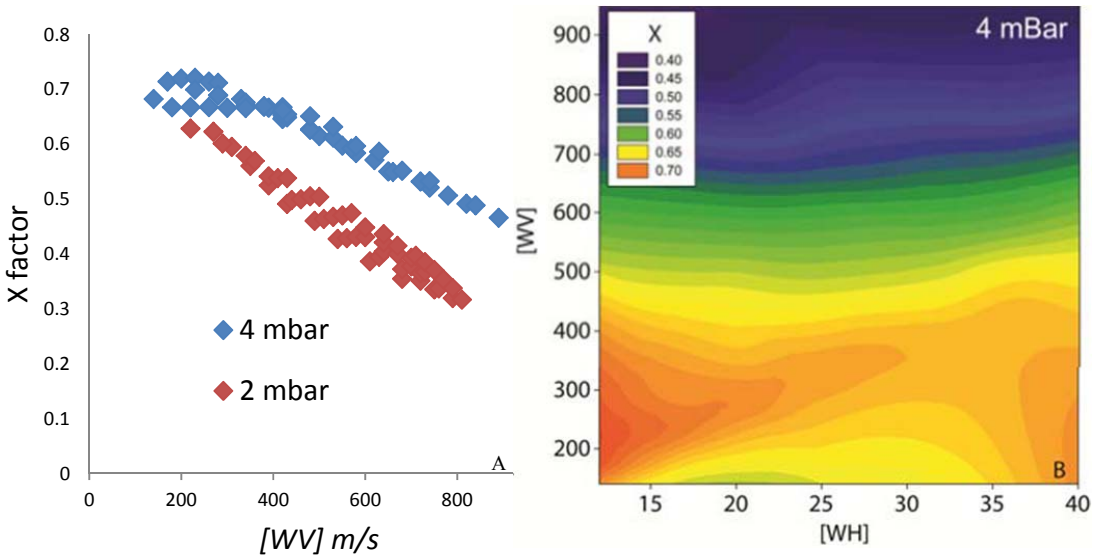

Figure S. 2-2 Data for calibration curve exponential factor X

(A) A plot of *[WV],* versus the calibration function exponential X factor recorded for two operating pressures for the IM separator. **(B)** A 3D plot of *[WV], [WH]* and calibration function exponential X factor. Similar to calibration curve R2 (see details in main article), substantial differences in X factor are observed only as a function of wave velocity, not wave height. There is <1% difference across *[WH]* for both IMS pressure 4mbar and 2mbar, while the *[WV]* gives ~50% difference across its applicable range for both pressures.

Peptide	*m / Da*	*Z*	Ω_{He} / Å^2	Ω_{N2} / Å^2
RASG-1	1000.5	+2	225	331
Angiotensin II	1045.5	+2	245	335

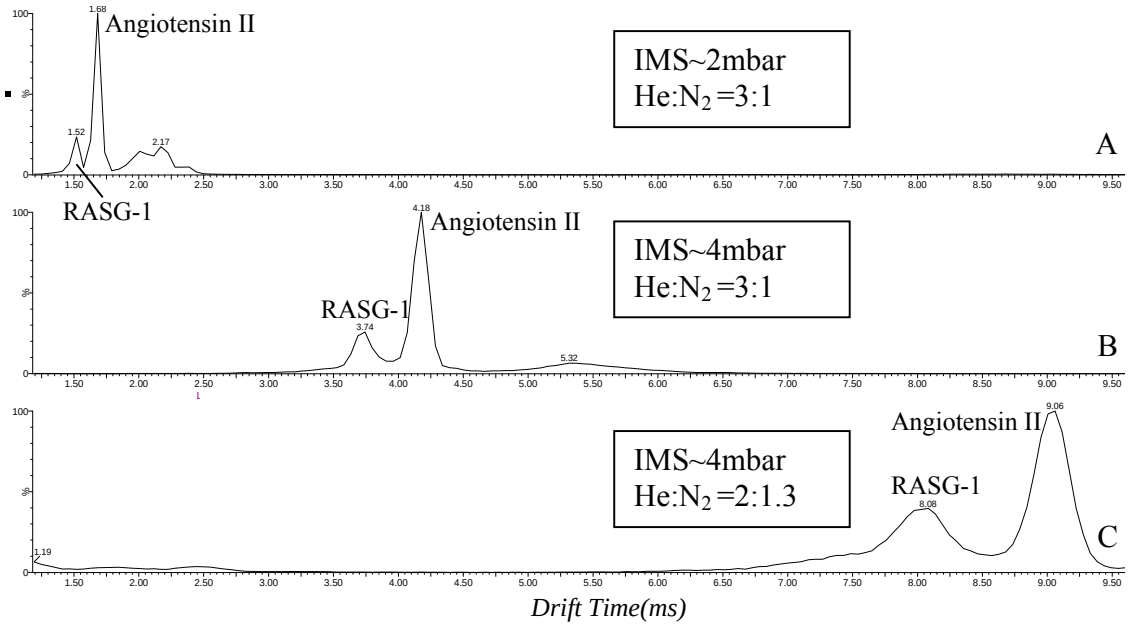

Figure S. 2-3 Collision cross section resolution R_{TW}^{Ω} of peptide ions at different pressures.

Peptides RASG-1 and angiotensin II are separated under different pressures or gas compositions. **(A)** 1.98 mbar, **(B)** 3.86 mbar and **(C)** 3.93 mbar. All 3 data are taken under *[WH]*=40 V and *[WV]*=1000 m/s.

Taking angiotensin II as example, the traveling wave drift time resolution (R_{TW}^{t}) is calculated as 22, 27 and 29 for conditions in (A), (B) and (C) respectively. The higher R_{TW}^{t} generated in (B) and (C) agrees with our conclusion that higher pressure is preferred for higher R_{TW}^{t}. However, with similar pressure, (B) and (C) resulted in significantly different drift time and thus different R_{TW}^{t}, which may indicate the influence of the gas composition on drift time resolution.

Further, calibration function exponential X factor, 0.35, 0.45 and 0.5 for (A), (B) and (C) conditions respectively (extrapolated from calibration experiments discussed in the main article), is combined to convert the drift time resolution into collision cross section resolution R_{TW}^{Ω}. The calculated results are 58 for (A), 60 for (B) and 60 for (C). These values are higher than that achieved for ADH tetramer (~40) and further indicates that gas pressure and composition have little influence on R_{TW}^{Ω}. Instead, optimum values of *[WH]* and *[WV]* are most important. We note, however, as the pirani-type pressure gauge used in this region of the Synapt G2 is both not calibrated for He/N_2 mixtures and relatively inaccurate, the IMS pressure values given must be treated as rough approximations only.

2.6 References

(1) Bowers, M. T.; Kemper, P. R.; VonHelden, G.; VanKoppen, P.A.M. *Science*, **1993**, *260*, 1446.
(2) Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H; Robinson, C. V. *Science*, **2005**, *310*, 1658.
(3) Loo, J. A.; Berhane, B.; Kaddis, C. S.; Wooding, K. M.; Xie, Y. M.; Kaufman, S. L.; Chernushevich, I. V. *J. Am. Soc. Mass Spectrom*., **2005**, *16*, 998.
(4) Hogan, C. J.; Kettleson, E. M.; Ramaswami, B.; Chen, D. R.; Biswas, P. *Anal. Chem*. **2006**, *78*, 844.
(5) Uetrecht, C.; Versluis, C.; Watts, N. R.; Wingfield, P. T.; Steven, A. C.; Heck, A. J. R. *Ang.Chem. Int. Ed.,* **2008**, *47*, 6247.
(6) Bohrer, B. C.; Mererbloom, S. I.; Koeniger, S. L.; Hilderbrand, A. E.; Clemmer, D. E. *Annual Rev. Anal. Chem*., **2008**, *1*, 293.
(7) McLean, J. A.; Ruotolo, B. T.; Gillig, K. J.; Russell, D. H. *Int. J. Mass Spectrom.*, **2005**, *240*,301.
(8) Gillig, K. J.; Ruotolo, B. T.; Stone, E. G.; Russell, D. H. *Int. J. Mass Spectrom.*, **2004**, *239*, 43.
(9) Tang, K.; Shvartsburg, A. A.; Lee, H. N.; Prior, D. C.; Buschbach, M. A.; Li, F. M.; Tolmachev, A. V.; Anderson, G. A.; Smith, R. D. *Anal. Chem*. **2005**, *77*, 3330.
(10) Koeniger, S. L.; Merenbloom, S. I.; Valentine, S. J.; Jarrold, M. F.; Udseth, H. R.; Smith, R. D.; Clemmer, D. E. *Anal. Chem*. **2006**, *78*, 4161.
(11) Kemper, P. R.; Dupuis, N. F.; Bowers, M. T. *Int. J. Mass Spectrom*., **2009**, *287*, 46.
(12) Giles, K.; Pringle, S. D.; Worthington, K. R.; Wildgoose, D. Little, J. L.; Bateman, R. H. *Rapid Comm. Mass Spectrom.*, **2004**, *18*, 2401.
(13) Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T. J. H. Scrivens, *Int. J. Mass Spectrom.*, **2007**, *261*,1.
(14) Wallace, A. *Am. Laboratory*, **2010**, *42*, 13
(15) Benesch, J. L. P.; Ruotolo, B. T.; Simmons, D. A.; Robinson C. V. *Chem. Rev.*, **2007**, *107*, 3544.
(16) Uetrecht, C.; Rose, R. J.; van Duijn, E.; Lorenzen, K.; Heck, A. J. R. *Chem. Soc. Rev*., **2010**, *39*, 1633.
(17) Politis, A.; Park, A. Y.; Hyung, S. J.; Barsky, D.; Ruotolo, B. T.; Robinson, C. V. *PLoS One*, **2010**, *5*, e12080.
(18) Bernstein, S. L. ; Dupuis, N. F.; Lazo, N. D. ; Wyttenbach, T.; Condron, M. M. ; Bitan, G.; Teplow, D. B.; Shea, J. E. ; Ruotolo, B. T. ; Robinson, C. V. ; Bowers, M. T. *Nature Chem.*, **2010**, *1*, 326.
(19) van Duijn, E.; Barendregt, A. ; Synowsky, S.; Versluis, C.; Heck, A. J. R.*J. Am. Chem. Soc*., **2009**, *131*, 1452.
(20) Hyung, S. J. ; Robinson, C. V. ; Ruotolo, B. T. *Chem. Biol.*, **2009**, *16*, 382.
(21) Pukala, T. L. ; Ruotolo, B. T. ; Zhou, M.; Politis, A. ; Stefanescu, R.; Leary, J. A. ; Robinson, C. V. *Structure*, **2009**, *17*, 1235.
(22) Wu, C.; Siems, W. F.; Asbury, G. R. ; Hill, H. H. *Anal. Chem*. **1998**, *70*, 4929.

(23) Wu, C.; Siems, W. F.; Hill, H. H.; Hannan, R. M. *J. Microcolun Sep.*, **1999**, *11*, 251.
(24) Wyttenbach, T.; Kemper, P. R.; Bowers, M. T. *J. Am. Chem. Soc*., **1996**, *118*, 8355.
(25) Dugourd, P.; Hudgins, R. R.; Clemmer, D. E.; Jarrold, M. F. *Rev. Sci. Instrum.* **1997**, *68*, 1122.
(26) Shvartsburg, A. A.; Smith, R. D. *Anal. Chem*. **2008**, *80*, 9689.
(27) Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S. J.; Robinson, C. V. *Nature Prot*., **2008**, *3*, 1139.
(28) Hernandez, H; Robinson, C. V. *Nature Prot*., **2007**, *2*, 715.
(29) Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. *Anal. Chem*., **2010**, *82*, 9557.
(30) Wu, C.; Siems, W.F.; Klasmeier, J.; Hill, H.H. *Anal. Chem.*, **2000**, *72*, 391.
(31) Michaelevski, I.; Eisenstein, M.; Sharon, M. *Anal. Chem*. **2010**, *82*, 9484
(32) Ruotolo, B. T.; Hyung, S. J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V. *Ang.Chem. Int. Ed.,* **2007**, *46*, 8001.
(33) Han, L.; Hyung, S. J.; Mayers, J. J.; Ruotolo, B. T. *J. Am. Chem. Soc.* **2011**, *133*, 11358.

Chapter 3. Robotically-Assisted Titration Coupled to Ion Mobility-Mass Spectrometry Reveals the Interface Structures and Analysis Parameters Critical for Multiprotein Topology Mapping

3.1 Introduction

Proteins form most of the central macromolecular machines that support cellular life through the creation of dynamic multiprotein complexes.[1] By the same token, such complexes comprise most of the important drug targets sought in treatment efforts for human disease.[2,3] Due to their biochemical importance, broad efforts are currently underway to determine the structures of protein complexes on a proteome-wide scale. These efforts typically rely upon high-throughput X-ray crystallography and NMR protocols, with many other technologies operating in a supporting role to ensure high protein purity, stability, and concentrations.[4,5] While this has been a very successful strategy, leading to the determination of many structures for large protein complexes, continued application of these basic approaches has also served to highlight their limitations. For example, such experiments typically require the availability of a sufficient quantity of homogeneous material and significant method development efforts for success.[6] These conditions are often difficult to meet and, thus, the number of

structures of multi-subunit protein complexes deposited in structural databases remains relatively low when compared to monomeric proteins.[7]

Recently, technologies based on mass spectrometry (MS) have provided dramatic insights into the composition and structure of multiprotein complexes.[8,9] Growing significantly from previous yeast two-hybrid screens,[10] affinity-based purification coupled to MS detection has provided some of clearest depictions of macromolecular protein networks in terms of their connectivity,[11] leading eventually to include sophisticated quantitative proteomics protocols.[11] Chemical cross linking (CXL) in conjunction with MS experiments also have expanded our knowledge of both global protein networks and individual macromolecular protein complexes.[11,12] Hydrogen-deuterium exchange (HDX) and oxidative labeling are being used with increasing frequency prior to MS in order to assess dynamic structural changes within a host of multiprotein systems.[13-15] While the implementation of most MS experiments involve the proteolytic digestion of denatured proteins following the steps taken within the protocol to capture native protein-protein contact or structure information,[16] MS experiments that introduce intact protein complexes directly into the instrument can also be used to provide, in many cases, a greater amount of native structure information.[17] Despite the challenges inherent in analyzing large protein ions by MS, such experiments have aided in the development of protein structure models for a range of complexes in advance of more-conventional structural biology approaches.[17-19]

While such MS measurements rely principally upon preserving an entire multiprotein assembly during both sample preparation and detection,[20] recent experiments in this area

have highlighted the utility of partially disrupting the protein-protein contacts within complexes so that various subcomplexes are also detected during the experiment.[19,21,22] Such disruption steps enable the determination of protein complex connectivity and stoichiometry to a level of detail that few other approaches can match.[23,24] Typically, two separate and complementary approaches for subcomplex formation are pursued: gas-phase activation and solution-phase disruption.[21,25-28] The former most-often utilizes collisional activation of gas-phase ions to initiate dissociation, forming individual protein subunit and stripped protein complex product ions.[25,29,30] While an indispensible tool for verifying protein connectivity and stoichiometry, larger subcomplexes are typically not formed in the unimolecular decay reactions that proceed during collision induced dissociation (CID) of multiprotein complex ions, thus somewhat limiting the overall utility of the technique in the context of protein structure modeling efforts.[19] Charge manipulation in combination with collisional activation, as well as alterative activation methods, have shown promise in altering the overall mechanism of protein complex CID to realize increased information content.[31] Current practice, however, dictates significant solution-phase protein complex disruption efforts in order to generate sufficient information to fully-define protein topology and connectivity by intact MS methods.[32] Such disruption is often initiated by altering solvent composition, pH and ionic strength to favor protein subcomplex formation.[32] The subcomplexes formed in such experiments have been shown to reflect the known organizational substructure of complexes,[21,22,33] and when combined with CID, can be used to deduce a complete organizational diagram of protein complexes of previously unknown structure.[34] However, conditions that optimize the ion signal recorded for protein subcomplexes are

often found through a time consuming, trial-and-error based process that can substantially limit the throughput of MS experiments.

Ion mobility (IM) separation,[35] when coupled to MS, enables the collection of protein complex size information, and when this is combined with the connectivity information described above, coarse-grained (CG) or atomic models of the assemblies can be constructed.[32,36] In addition, IM-MS provides signal-to-noise, spectral deconvolution, and tandem MS benefits unique to the analysis of the multiply-charged protein complex ions produced by nESI.[37] Specifically, IM-MS was used to refine a connectivity model for the eukaryotic ribosomal initiation factor 3 complex,[22] provided key structural data for models of rotary adenosine triphosphatases/synthases,[38] pallindromic repeat-associated protein complexes,[39] and utilized to deduce complete models for αB-crystallin[40] and DNA replisome assemblies.[41] In addition, models for the highly-dynamic olgiomers involved in the amyloid-β,[42] norwalk virus,[43] and β-2 microglobulin[44] aggregation/assembly pathways have been constructed using IM-MS. In every case, IM measurements of many inter-related protein complexes, connected either through assembly, equilibrium, or intentional disruption, were used to construct the resultant models. Key assumptions surrounding such IM-MS derived models include: a strong relationship between subcomplex and higher-order assembly quaternary structure, a high degree of structural correlation between gas-phase and native state proteins, and a lack of significant protein subunit conformation changes upon higher-order complex formation or disruption. While most IM-MS models have been cross-validated using other structural biology tools and control samples,[22,39,40] there are many foundational questions that still surround the use of subcomplex information as a constraint for larger related

oligomers, especially when ideal disruption conditions introduce significant amounts of potential chemical denaturants in solution.

Here, we couple automated titration and nESI of protein complex samples to IM-MS in order to deduce the optimum protocols and analysis guidelines necessary for high-throughput protein topology mapping experiments. Our total dataset contains ca. 400 individual IM-MS datasets, thus necessitating an automated approach. We discover that IM-MS measurements of intact multiprotein complexes and subcomplexes can be readily mapped onto known solution structures, with few exceptions. For those exceptions, the distortion of subcomplexes is mirrored in distorted intact complexes, which are also detected by IM-MS. We find strong correlations between the disruption of intact complexes with specific reagents and quantitative values for the different classes of interactions that form the protein-protein interfaces within the intact assembly. We propose that such information can be used in the future to deduce the critical contacts within unknown complexes, and project the utility of IM-MS in structural proteomics and genomics efforts broadly in light of the data presented.

3.2 Experimental

3.2.1 Sample Preparation

Homo-tetrameric protein complexes studied included transthyretin (TTR, human, Sigma P1742), avidin (AVD, egg white, Sigma A9275), concanavalin A (CON, jack bean,

Sigma C2010), alcohol dehydrogenase (ADH, Saccharomyces cerevisiae, Sigma A7011), aldolase (ALD, rabbit muscle, Sigma A2714), pyruvate kinase (PKI, rabbit muscle, Sigma P9136), catalase (CAT, bovine liver, CalBioChem 219001) and β-galactosidase (βGL, E.coli, Sigma G5635). Standards used to construct collision cross-section (CCS) calibration curves include: a mixture of peptides (Waters, product code: 186002337), monomeric cytochrome C (Sigma, C2506), and the protein complexes AVD, CON, ADH, and glutamate dehydrogenase (GDH) (Sigma, G7882). All protein complexes were solubilized in 200 mM aqueous ammonium acetate (NH_4Ac) solutions and stored at -80 ℃. Stock samples (50 μM with respect to the complex, diluted 10 fold afterwards) were then buffer-exchanged just prior to disruption experiments and IM-MS analysis, using a Micro Bio-Spin 30 column (Bio-Rad, Hercules, CA) equilibrated with 200 mM aqueous NH_4Ac. While the final concentration of protein complexes is unknown, control experiments performed with the protein complexes studied here over a range of concentrations (as low as 3 μM) did not alter the oligomeric states detected. The peptide mixture (~2 μM for each peptide) was prepared in a water/methanol/ acetic acid (49/49/2) solution. Stock cytochrome C (5 μM) was prepared in 200 mM aqueous NH_4Ac, and stored at -80 ℃.

3.2.2 Ion Mobility-Mass Spectrometry

Solution composition screens for subcomplex formation were performed using an automatable, chip-based nano-electrospray (nESI) robot, (TriVersa NanoMate, Advion, Ithaca NY), capable of both rapid sampling and mixing. Solution parameters were screened in a two-dimensional fashion to be compatible with robotic sampling of a 96-

well plate. An Advanced User Interface (AUI) module in the control software allowed programmed sequential aspiration, and allowed for protein samples to be quickly dispensed into individual wells and mixed with disrupting agents under highly-controlled temperatures (18 °C). After mixing, chip-based nESI using nano-machined emitters (5 μm) was initiated through the robotic sampling of individual solutions with conductive pipette tips. In general, samples were sprayed immediately after mixing, however, various incubation times were tested for selected protein complexes (e.g., AVD), ranging from minutes to hours, and resulted in no significant change in disruption behavior. Chip based emitters were maintained at 1.3 kV to 1.9 kV. Samples prepared at high ionic strengths or solutions containing large amounts of organic co-solvent generally required lower nESI voltages and higher backing pressures (0.1-1 psi) to avoid salt cluster-related chemical noise and generate stable spray conditions respectively.

IM-MS data were acquired on a Synapt G2 (Waters, Milford, MA) [45] optimized for the study of large multiprotein complexes, as described previously.[46] For the purposes of topology construction, the accuracy of CCS measurements for both intact protein complexes and dissociated products was optimized through highly linear CCS calibration curves ($R^2 > 0.98$), resulting from high pressure (3-4 mbar), low wave height (12-15 V) and low wave velocity (150-200 m/s) travelling wave IM conditions.[46] Initial IM-MS Data processing was performed using Masslynx v4.1, Driftscope v2.0 (both from Waters, Milford, MA), as well as previously-described IM calibration tools.[37,47]

3.2.3 X-ray and IM Data Comparisons and Computational Analysis

Coarse-grained (CG) topology models for the protein tetramers discussed in this chapter were constructed either using a simple symmetry analysis of X-ray structures using in-house python scripts and visualized in PyMol, or using IM-derived CCSs as described below. Protein data bank (PDB) IDs for the X-ray structures used here are given in Supplemental Information. All model structures were evaluated in terms of their agreement with IM data using a version of the program Mobcal modified for the analysis of coarse grained structures.[37] For all model CCS values, a scaled version of the projection approximation (PA) found in Mobcal was used,[48] as such methods have exhibited strong, broad agreement with experimental IM-MS data.[49] For model topologies constructed using IM data, the size of a single sphere was constrained using monomer CCS data. Then, the distance between monomer units was constrained using protein dimer CCS values. Finally, a dimer-dimer distance was constrained using tetramer CCS measurements. The final IM-based model was then compared to the same CG values derived from the X-ray structure. Note that not all tetramers studied in this chapter are planar, and that the above protocol does not evaluate the relative angles between protein subunits.[33,41] To resolve a planar tetramer geometry from that of a tetrahedral structure, an IM CCS resolution in excess of 100 has been projected to be necessary for protein complexes comprised of >10 kDa subunits, whereas the maximum CCS resolution recorded to-date on the Synapt G2 is 50-60.[37,46] A number of quantitative values related to protein structure were extracted from X-ray and IM-MS data, and compared. These values are defined in detail in Supplemental Information.

3.3 Results and Discussion

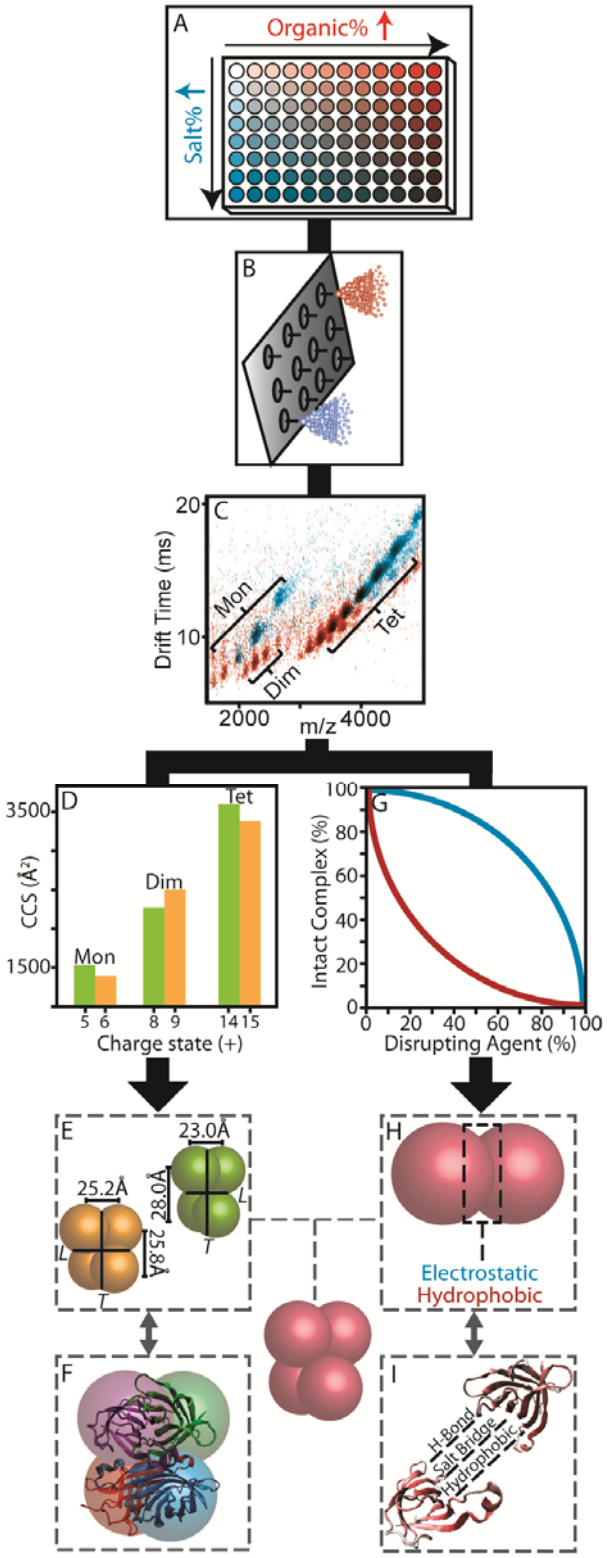

Figure 3-1 IM-MS data collection and analysis procedures.
(A) Disrupting agents and protein complexes are mixed in solvents containing either increased amounts of organic (DMSO, red) or salt (NH_4Ac, blue). **(B)** Each sample is sprayed using a chip-based nESI emitter array. **(C)** An example overlay of two IM-MS plots under high organic (red) and high salt (blue) conditions respectively. **(D)** IM-based size measurements recovered for both intact complexes and subcomplexes as a function of charge state (green and yellow for different charge states). **(E & F)** Protein sizes are converted to distance constraints, and tested against those calculated from X-ray. **(G)** Signal intensities recorded for intact complex ions are plotted against the amount of disrupting agent added in solution. **(H & I)** These data are analyzed with respect to the separate interaction types found within the protein-protein interfaces from X-ray.

Figure 3-1 illustrates the experimental protocol used in this study. Protein complex sample was partitioned into a 96-well plate where different amounts of disrupting agents were robotically added into the existing aqueous solutions (Figure 3-1A). The amounts of these agents were then modified in a stepwise fashion across one axis of the plate. For the complexes studied here, we found that a two-component screen, that varied the DMSO (red wells, Figure 3-1A) and NH_4Ac (blue wells, Figure 3-1A) content of the protein complex-containing solutions was sufficient to disrupt and interrogate all eight protein tetramers, while keeping the total protein concentration constant (5μM). Solutions contained, at most, 67% DMSO by volume and 4M NH_4Ac, beyond which significant protein secondary structure changes may occur.[50] Following the addition of disruption agents, samples were then immediately sprayed through one of the nozzles on a nESI chip (Figure 3-1B). IM-MS data were then recorded (Figure 3-1C), and subtle changes in protein oligomeric state, CCS, and charge state were detected. As observed previously,[18,23] the sub-assemblies observed under disruption conditions reflect the known sub-structure of the intact complexes analyzed. All complexes studied here are tetramers comprised of dimeric subcomplexes, possessed of non-covalent bonds of varying strengths. By quantifying the sizes of the subcomplexes observed by IM-MS (Figure 3-1D) over a range of charge states, CG models of the relative interface sizes within these tetramers were constructed (Figure 3-1E) and compared against X-ray datasets (Figure 3-1F) to detect correlations. In parallel with this analysis track, the relative intensities of the subcomplexes were recorded (Figure 3-1G) and correlated to the influence of specific disrupting agents likely to disproportionately influence selected sub-classes of protein-protein contacts (Figure 3-1H). These data were then also correlated

with X-ray data where specific interactions were precisely measured (Figure 3-1I), allowing us to evaluate the ability of IM-MS to quantify the type of interactions present within unknown protein-protein interfaces.

3.3.1 Evaluating the Fidelity of IM-MS Derived Protein Topology

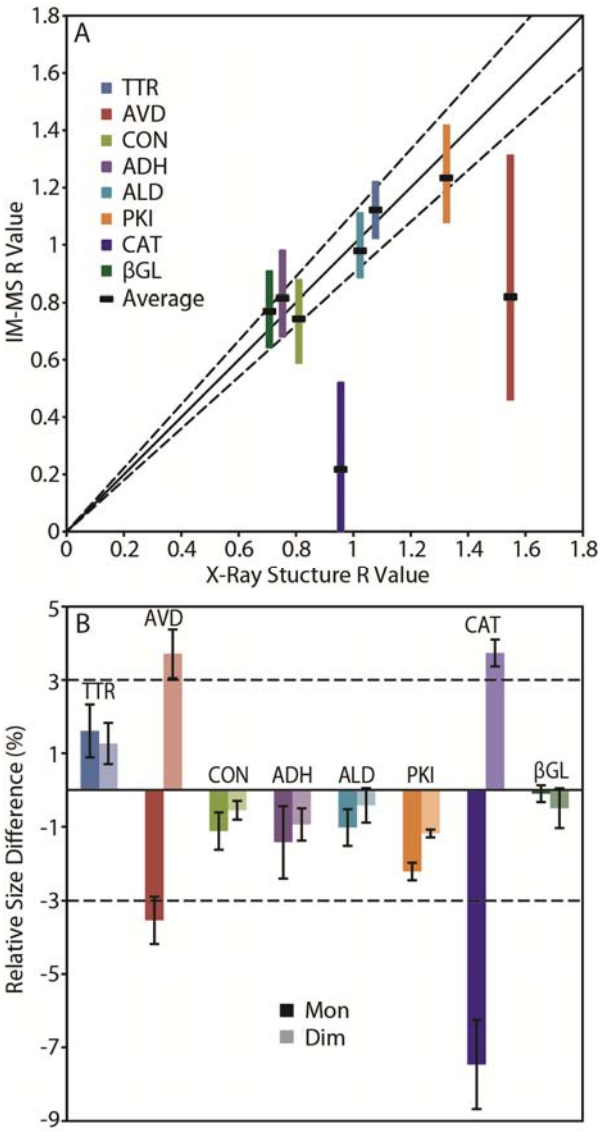

Figure 3-2 Evaluating the fidelity of IM-MS derived protein topology. **(A)** A plot of R values generated from IM-MS data vs. R values from X-ray structures for eight protein tetramers. Color-coded bands indicate the range of R values recovered from IM-MS data, and black bands represent the average. The solid line represents a 1:1 correlation between the IM-MS experiment and X-ray, with the dashed lines representing a 90% confidence interval. **(B)** Relative size differences observed between IM-MS and X-ray data for monomers (dark colors) and dimers (light colors) released during disruption. Dashed lines indicate the ±3% error for IM-MS CCS determinations.

In order to evaluate the agreement between IM-MS derived protein topology measurements and those from X-ray datasets, we generated R values from both classes of data, which track the relative distance between dimers and monomers within the complex (see Supplemental Information). A plot of IM-MS R values against those extracted from X-ray structures reveals a tight correlation between the two for all complexes save two: AVD and CAT (Figure 3-2A). For the other six protein tetramers, average topologies, as well as much of the total spread of possible topologies created if all charge state-based CCS combinations are considered with equal weight, fall within a 90% confidence prediction interval. Conversely, the entire range of R values generated from IM-MS for AVD and CAT fall outside of the same prediction interval in Figure 3-2A. Note that AVD and CAT present larger R value ranges than the other protein complexes studied here, and that such data on its own could be used in broader IM-MS topology construction efforts to identify distortion-prone complexes.

If we evaluate protein CCS measurements used to generate the R values shown in Figure 3-2A in detail, the origins of the topological defects observed in our CAT and AVD datasets becomes apparent (Figure 3-2B). While CCS values recorded for the intact tetramer ions generated under control conditions (200 mM NH_4Ac, pH 6.9) are all within 3% of the expected X-ray CCS values, the disrupted monomer and dimer CCS values determined experimentally vary considerably over the eight protein tetramers studied here. Notably, both CAT and AVD display compacted monomer and enlarged dimer CCS values. Specifically, AVD disruption generates monomers that are, on average, compressed by 3.5% and dimers that possess similarly inflated CCS values when compared to X-ray estimates. CAT dimers are also enlarged by 3.7%, but generate

monomer CCS values upon disruption that are 7.5% compressed when compared to X-ray data. While the data shown in Figure 3-2B provides a clear empirical explanation for the topology defects detected in Figure 3-2A, it provides neither a structural explanation for the same observations nor a robust analytical methodology to avoid using such IM-MS data in the construction of CG structures for unknown protein complexes.

A broader analytical protocol for IM-MS based protein complex topology mapping can be derived from the data shown in Figure 3-3, which tracks the IM arrival time profiles for all eight protein tetramers under two separate conditions: those optimized for protein disruption and control measurements of ions formed from 200 mM NH_4Ac. For six of the eight protein complexes, no significant differences are observed between the IM data recorded under either conditions (<1% deviation), both in terms of the centriods or peak widths of the IM distributions collected (three replicates shown for each dataset). In two cases, however, small yet significant increases (> 1%) are observed in the IM drift time centriod observed for intact tetramer ions. Since the intact masses of AVD and CAT are identical under both disruption and control solution conditions (Figure S.3-1, Supplemental Information), the increase in drift time recorded is likely derived from conformational changes in the intact tetramers. These two complexes, AVD and CAT, correspond precisely to those that produced erroneous protein CG structures in our analysis shown in Figure 3-2. As such, the observation of similar shifts in IM data for intact complexes allows us to detect and remove distorted subcomplex CCS data from our modeling constraints. It is also notable that this strategy can, in theory, be extended to protein complexes of unknown structure, as the IM drift time shifts recorded in Figure 3-3 are not predicated on any knowledge of the tetramer structures studied here.

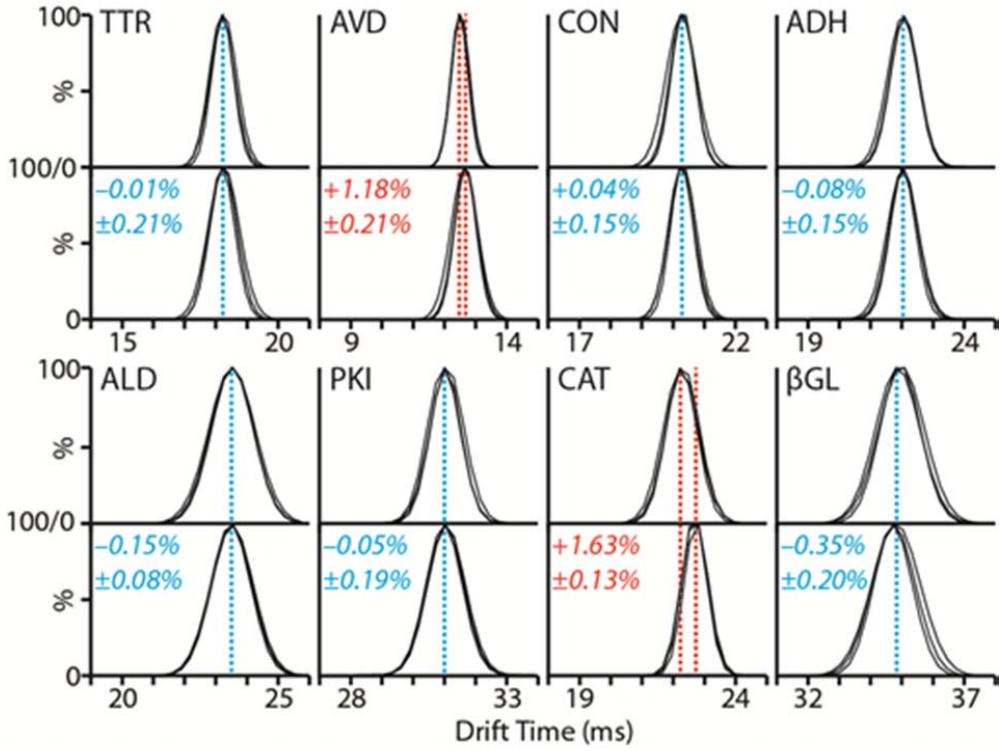

Figure 3-3 Comparison between the IM distributions and centroids recorded for intact tetramer ions generated under control (200mM NH_4Ac, upper panel) and disruption conditions (lower panel).
Three replicates are measured and overlaid for each dataset shown. Blue dashed lines indicate those datasets where no significant shift in tetramer drift time is observed (<1%). Red dashed lines indicate those cases where a drift time shift (≥1%) is observed under disrupting solution conditions when compared to control.

While the criteria we use to construct Figure 3-3 requires similar protein disruption from all tetramers studied, significantly different solution conditions are represented in each of the lower panels in Figure 3-3 (see Table S.3-1 and Figure S.3-2, Supplemental Information), chosen to represent equivalent amounts of protein complex disruption. Despite these differences, we note that the six protein complexes that produce high-fidelity R values when compared to X-ray data do not, under any solution composition probed in this chapter, generate ions with shifted drift times. For example, TTR tetramer ions retain their centroid drift time when generated from solutions comprised of 50%

DMSO and 4M NH_4Ac, while CAT displays the shifted drift time profile shown in Figure 3-3 at 33% DMSO and 2M NH_4Ac. Previously published results for the CON tetramer support these findings, where methanol-based disruption of the tetramer results in both dramatically enlarged intact complexes and distorted monomeric subunits upon disruption.[51] Thus, taken as a complete dataset, the results shown in Figure 3-3 constitute a general analytical framework for detecting those conditions likely to produce deformed protein subcomplexes through disruption, as the swelling of intact protein complex CCS appears to accurately predict the formation of such deformed sub-assemblies.

A detailed analysis of the X-ray data available for the eight protein tetramers studied here provides a structural explanation for the data shown in Figures 3-2 and 3-3. First, the X-ray structures for the monomeric units within the CAT and AVD tetramers are the least spherical of the complexes studied here (Figure S.3-3A, Supplemental Information). Furthermore, AVD and CAT exhibit both the largest number of inter-protein contacts per unit mass (Figure S.3-3B, Supplemental Information) and the most closely-packed monomers (Figure S.3-3C, Supplemental Information) in our field of tetramers. In summary, therefore, AVD and CAT are comprised of relatively non-spherical subunits, possessed of larger than average protein-protein interfaces, and exhibit comparatively tight inter-subunit packing. Given such a description, it is relatively facile to rationalize the data shown in Figures 3-2 and 3-3. AVD and CAT subunits are likely more stabilized by inter-protein contacts than the subunits that make up the other complexes studied here, thus making them more-prone to deformation upon release from their respective tetramers.

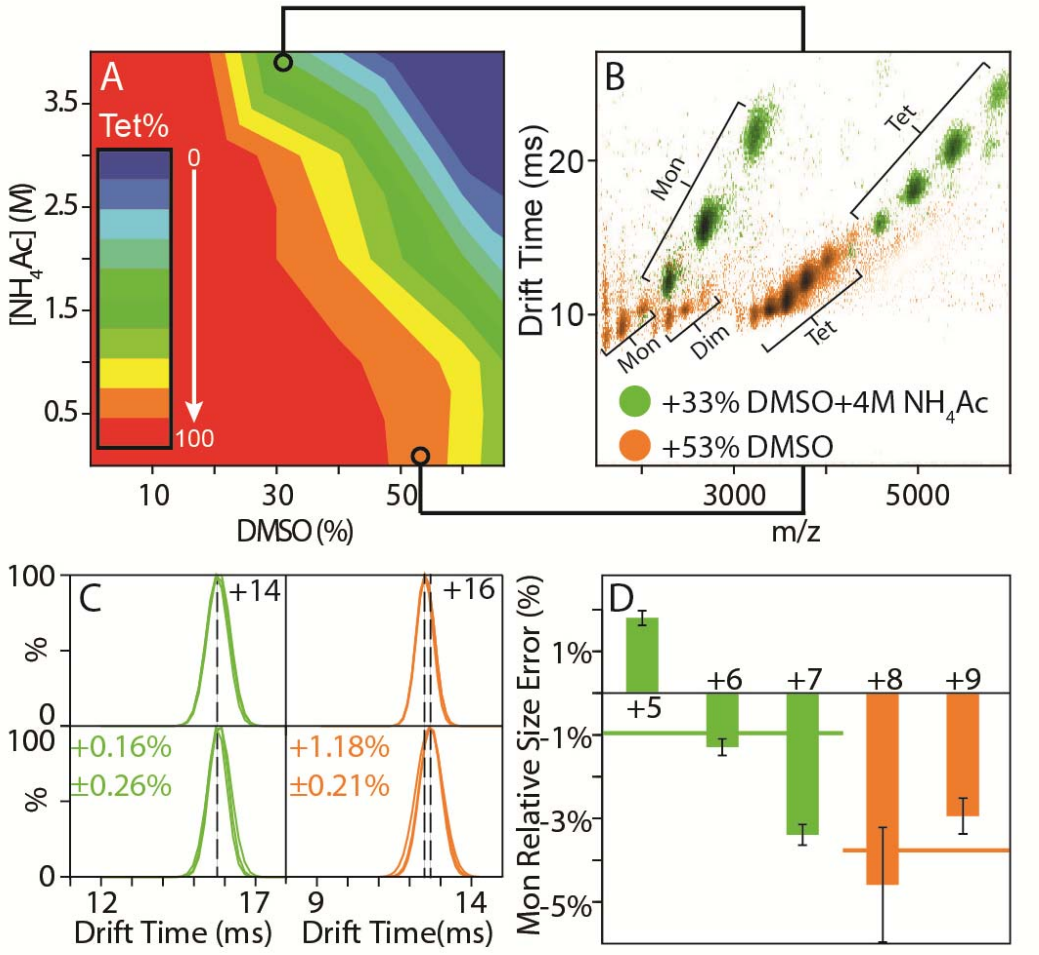

Figure 3-4 AVD topology models from IM-MS data are brought into close agreement with X-ray when disruption conditions incorporating high ionic strength are used.
(A) 2D titration results for AVD using DMSO and NH_4Ac as disrupting agents. **(B)** Overlay of two IM-MS contour plots extracted from the dataset shown in **(A)** (green and orange). **(C)** A comparison of the drift time distributions recorded for the remaining intact tetramer ions formed from the two disruption conditions (lower panels) with those measured under control conditions (200 mM NH_4Ac, upper panels). **(D)** The relative errors in monomer size measured by IM when compared to X-ray data for different AVD monomer charge states derived from the disruption conditions defined in **(B).** Solid lines indicate the average monomer size for each solution condition.

In an effort to find those disruption conditions capable of generating subcomplex CCS

values for AVD and CAT that are in close agreement with X-ray data, we initiated a thorough analysis of all IM-MS data collected in our screens. An example of such an analysis is shown in Figure 3-4, where we use a contour plot to represent the relationship between AVD tetramer signal intensity, DMSO solution fraction, and NH_4Ac concentration (Figure 3-4A). Within this dataset, each of the 49 data points used for its construction represent a complete IM-MS spectrum (Figure 3-4B), and comparison of multiple points within this plot reveals stark differences in the dissociation patterns of the AVD tetramer. Two drift time versus m/z plots are shown, overlaid, in Figure 3-4B, one acquired from a solution containing 53% DMSO with minimal (10 mM) NH_4Ac and the other containing 33% DMSO and 4 M NH_4Ac. Preparing AVD in higher salt-concentration solutions prior to nESI engenders lower than average charge states for both intact complexes and disruption products as well as an absence of any dimer signal when compared to IM-MS data acquired from solutions containing low salt and higher amounts of organic co-solvent. Charge states for the AVD tetramer under control conditions range from 14^+ to 17^+, while charge states from 11^+ to 15^+ are observed under conditions of high salt (green data, Figure 3-4B) and 15^+ to 20^+ when low salt is coupled with large amounts of DMSO in solution (orange data, Figure 3-4B). When the IM drift time profiles from charge states that sufficiently overlap with control data are compared, it is clear that the AVD tetramer shifts its size when large amounts of DMSO are used for protein disruption in the absence of high salt concentrations (Figure 3-4C, orange), as observed in Figure 3-3. However, when a similar comparison is made for disruption conditions that add 4M NH_4Ac, no shift is detected under conditions that lead to complex disruption (Figure 3-4C, green). Analysis of the monomer CCS values from both of these

datasets reveals an agreement with X-ray data to within acceptable errors when disruption is performed under conditions of high salt (Figure 3-4D, green), but not when solutions with added organic solvent are used (Figure 3-4D, orange). A few caveats surround the result presented in Figure 3-4. Most notably, the lack of a dimeric subcomplex disruption product for the AVD tetramer under the high salt conditions presented means that such a result, on its own, does not provide sufficient information to generate the CG structure and R values analyzed in Figure 3-2. In addition, similar conditions that lessen the deformation experienced by disrupted monomers have yet to be discovered for the CAT tetramer. Regardless, the results presented in Figures 3-2, 3-3, 3-4, and S.3-3 provide a general picture that illustrates the overall agreement between IM-MS and X-ray datasets, methodologies that enable the detection of distorted subcomplex formation in unknowns, and approaches that enable the recovery of partial subcomplex CCS data from proteins previously distorted in a more-general screen.

3.3.2 Solution Disruption followed by IM-MS Reveals the Details of Protein-Protein Interface Chemistry

Figure 3-4A indicates that the AVD tetramer responds more strongly to DMSO as a disruption agent than NH_4Ac. For example, by following a vertical line fixed at 10% DMSO concentration in Figure 3-4A, it is apparent that no AVD tetramer disruption occurs during such an NH_4Ac-only titration. This observation could indicate a general correlation between the chemical nature of the protein-protein interfaces within the complex and the disruption of the intact assembly under specific solvent conditions as recorded by IM-MS. In order to pursue this basic premise, we analyzed the

multidimensional DMSO/NH_4Ac titration data for all eight protein tetramers studied in this chapter (Figure 3-5A), and evaluated them for general correlations with X-ray data. Each dataset shown is individually normalized to the tetramer intensity observed for a given complex under control conditions (200 mM NH_4Ac, 0% DMSO), and each contour transition shown represents a 10% relative decrease in tetramer signal intensity observed. Every protein complex represented in Figure 3-5A displays a unique disruption pattern as a function of DMSO and NH_4Ac solution content, and qualitative comparisons between the plots illustrate how such data can be used to differentiate assemblies based on the general strength of their non-covalent protein interactions. For example, TTR, AVD, and CAT are amongst the most disruption resistant of the complexes studied here relative to DMSO/NH_4Ac titration, as evidenced by the relatively large number of total solution conditions probed in our screen where IM-MS data records only intact tetramer for these three cases. Of these three, CAT exhibits the steepest descent from 100% intact tetramer to complete disruption of any of the complexes studied here, once a sufficient threshold amount of disrupting agent is reached. PKI also represents a unique disruption profile, as the tetramer does not generate subcomplex ion current as a function of DMSO added in solution unless a certain threshold of NH_4Ac is present as well. Previous studies have shown that key salt bridges within the PKI protein-protein interfaces exist and stabilize the quaternary structure of the complex.[52] Our IM-MS data agrees well with these findings qualitatively.

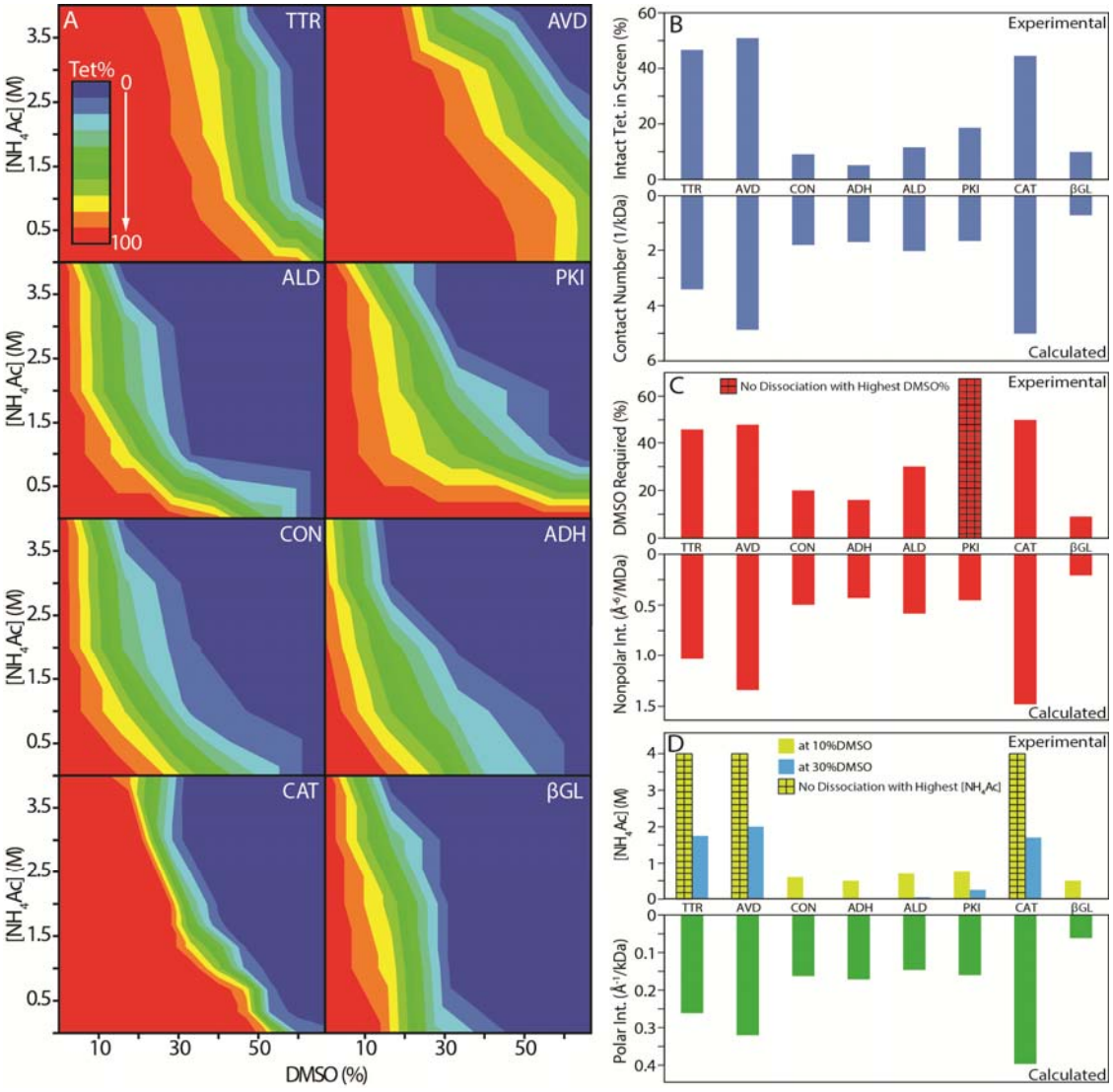

Figure 3-5 Solution disruption followed by IM-MS reveals the details of protein-protein interface chemistry.

(A) 2D titration results obtained for the eight homo-tetrameric protein complexes studied here, where NH_4Ac concentration (M) is plotted against DMSO solution content (%). The colors shown indicate the normalized intensity of the tetramer ion signal recorded under each solution condition, from red (100%) to blue (0%). Correlations between our IM-MS titration experiments (top) and X-ray structure data (bottom) are presented in three histogram plots (B), (C) and (D). **(B)** Normalized intact tetramer intensity integrated over all disruption conditions vs. the number of contacts per unit kDa calculated from X-ray data. **(C)** The DMSO % in solution, without added NH_4Ac, required to initiate tetramer disruption (signal drops by 10%) vs. the estimated average strength of hydrophobic contacts within the protein-protein contacts from X-ray ($Å^{-6}$/MDa). **(D)** NH_4Ac concentration required for tetramer disruption at fixed 10% (yellow) and 30% DMSO

(blue) vs. the estimated average strength of the polar interactions ($Å^{-1}$/kDa) within the protein-protein contacts from X-ray. For all datasets, bars marked with a grid pattern indicate that the complex did not undergo disruption using any amount of the selected agent under the conditions of the screen. Linear correlation coefficients between IM-MS and X-ray datasets are discussed in the text.

In addition to the qualitative analysis above, a quantitative comparison between IM-MS and X-ray datasets are necessary to extend the application of the data shown in Figure 3-5A to unknown protein complexes. We begin by noting the strong correlation between the integrated areas found by IM-MS multi-dimensional titration corresponding to those conditions where tetramers remain completely un-disrupted (red-shaded areas within Figure 3-5A) and the normalized number of inter-protein contacts contained within the X-ray structures for the complexes studied here (Figure 3-5B). These two parameters appear to be linearly correlated in our dataset, having an R^2 value of 0.826, despite the fact that our analysis clearly over-simplifies the complex nature of the underlying disruption mechanism. We extended this analysis to include narrower definitions for the inter-protein contacts and discovered additional strong correlations between IM-MS and X-ray structure data. By keeping the NH_4Ac concentration of protein complex containing solutions at fixed minimum values, and titrating DMSO content, we observe a strong correlation between the amount of tetramer disruption observed by IM-MS and the calculated strength of the inter-chain hydrophobic contacts revealed by X-ray (Figure 3-5C), generating an R^2 value of 0.931 to a linear relationship. We find a somewhat less-strong linear correlation coefficient of 0.797 between the strength of polar inter-protein interactions and tetramer disruption efficiency as a function of NH_4Ac concentration in solutions with fixed DMSO content (yellow data, Figure 3-5D). The weaker correlation we observe for polar interactions in our IM-MS data can be attributed, in part, to the lack

of orthogonality between the types of molecular interactions likely disrupted by DMSO and NH_4Ac. While polar and non-polar interactions can be accessed exclusively by DMSO and NH_4Ac respectively in our screens, both solvents likely affect H-bonds within protein interfaces.[50] In addition, the long-range nature of polar and electrostatic interactions, in contrast to the forces involved in non-polar interactions, may elicit non-native contacts within complexes during disruption, thus degrading the observed correlation.[53,54] There are a number of caveats associated with these correlations, however. For example, NH_4Ac content alone does not dissociate any of the complexes studied here; therefore, a baseline level of DMSO is required to generate the correlations associated with polar interaction strength. Specifically, when DMSO content of the solutions is locked at 10%, and NH_4Ac concentration is ramped, TTR, CAT and AVD do not generate any measurable disruption products over the range of salt probed in this chapter (10mM to 4M). Similarly, by increasing the fixed DMSO value to 30%, CON, ADH and βGL all undergo disruption such that the intact tetramer is 50% depleted. The linear correlation coefficient recorded under this condition is similar to that generated by our data when minimal DMSO is used during the titration (blue data, Figure 3-5D, R^2 = 0.775). Also, PKI is not observed to undergo disruption as a function of DMSO content when the NH_4Ac concentration is kept at 10mM (the maximum DMSO solution content in our screen is 67%). We attribute this observation to the critical salt-bridges known to stabilize the protein-protein contacts in the PKI tetramer, as discussed above. Despite these caveats, we find that the correlations between IM-MS titration data and X-ray protein-protein contact data are surprisingly strong, and likely pave the way for future use

of such quantitative values for different types of inter-chain interactions to constrain models of unknown protein complexes.

3.4 Conclusions

Here, we utilize robotically-assisted multidimensional titration of disrupting agents in solution to probe the ability of IM-MS to produce topology models in agreement with existing X-ray data for a series of eight tetrameric proteins. Our final dataset contains 8 2D screens of protein complex disruption, using both ionic strength and organic content axes, containing ca. 400 individual protein complex IM-MS datasets, each rich with size and mass data and many individual signals. As such, the acquisition of this data was greatly enabled by our automated approach. In addition, the ability to robotically-manipulate samples with great precision facilitates the cross-dataset comparisons discussed throughout this chapter. We demonstrate that R values computed from IM-MS data agree well with the majority of those extracted from X-ray. Clear disagreements are found in the case of the AVD and CAT tetramers, and we rationalize these observations based on the large number of inter-subunit contacts found within these two complexes that likely stabilize their relatively aspherical subunits. Furthermore, a careful analysis of IM-MS data allows deformed protein subcomplexes to be detected, even in the absence of X-ray comparisons, as tetramer CCS increases and broadened R value ranges are recorded in concert with such subcomplex distortions. Also, careful analysis of multi-dimensional titration data enables the collection of IM-MS information under conditions that better match X-ray structures, as we demonstrate in the case of the AVD tetramer. In addition to providing foundational data for future protein topology discovery efforts

using IM-MS, the data presented here provides evidence of solution phase memory effects for proteins and their associated complexes, adding substantially to previous reports.[22,51,55-57] The data shown here is the first to indicate that there are broad sets of solution conditions that are cable of reorganizing protein structure in such a way to influence the resulting gas-phase conformation of intact protein ions, and it is clear that in future IM-MS efforts to constrain protein topology models data similar to those shown in Figures 3-3 and 3-4 should be acquired to direct and evaluate the use of subcomplex size information in protein complex model generation.

Our IM-MS titration data are also used to generate quantitative values associated with the number and type of stabilizing interactions that populate protein-protein interfaces. By determining the ratios of intact tetramers to disrupted subcomplexes over a controlled range of solution conditions, we find strong correlations between IM-MS and X-ray datasets. We estimate the average relative errors for this type of measurement at +/- 27% for quantifying polar interactions, 14% for hydrophobic interactions, and 33% for total numbers of interactions per-unit mass (see Supplemental Information for the details of this error calculation). Such data is potentially highly-complimentary to CXL[12] and HDX[13] MS information. While providing a lower-resolution estimate of protein-protein interface area than these other MS approaches, such IM-MS based multi-dimensional titration results may provide quantitative values for different types of inter-chain contacts within protein assemblies that are not accessible to protease digestion or present at low levels, following appropriate calibration with proteins of known structure.

Acknowledgements to Dr. Jun Feng and Prof. Charles Brooks (University of Michigan) for aiding in the computational analysis presented in this chapter and funding support from National Institutes of Health (1-R01-GM-095832-01).

3.5 Supplemental Information

X-ray and IM-MS Structure analysis. To quantitatively evaluate the agreement between X-ray and IM-MS structure information, we computed a number of values extracted from the CG models constructed from both datasets. Equation S.3.1 defines the parameter R:

$$\mathrm{R} = \frac{Dim-Dim\ Distance}{Mon-Mon\ Distance} \tag{S.3.1}$$

which is a ratio of the distance between the geometric centers of the two dimeric subcomplexes that comprise the tetramer over the distance between the two monomeric units that comprise those dimers. We also computed the relative size differences between crystolographic and IM-measured monomers:

$$\%\ \text{Monomer Size Difference} = \frac{(\frac{Mon}{Tet})_{IM}-(\frac{Mon}{Tet})_{Crystal}}{(\frac{Mon}{Tet})_{Crystal}} \tag{S.3.2}$$

where the monomer size from both datasets is expressed as a fraction of the total complex size. A similar metric is also used in our data to evaluate dimer size agreement between X-ray and IM datasets:

$$\%\ \text{Dimer Size Difference} = \frac{(\frac{Dim}{Tet})_{IM}-(\frac{Dim}{Tet})_{Crystal}}{(\frac{Dim}{Tet})_{Crystal}} \tag{S.3.3}$$

For X-ray data analysis alone, we have developed several useful metrics that we use to evaluate the trends observed in our IM-MS dataset. First, we calculate the asphericity of protein subunits in complexes using Equation S.3.4:

$$\text{Monomer Asphericity Index} = \frac{L_{SD}}{L_{AV}} \tag{S.3.4}$$

where L_{AV} is the average distance between the protein center-of-mass and its surface, and L_{SD} is the standard deviation of all of the length measurements used to define L_{AV}. Larger index values indicate less-spherical monomers. The second method of X-ray structure analysis used here involves counting the number of inter-chain contacts within protein complexes. These values are normalized by the total molecular weight of the proteins studied here to generate a contact-per-unit mass (kDa) value for each tetramer included in this chapter. Cut-off distances for inter-protein contacts were defined as 4 Å for both salt bridges (Lys/Arg/N-terminus with ASP/GLU/C-terminus) and hydrophobic (C-C) interactions and 3.5 Å for hydrogen bonds (polar-polar or charged-polar residue interactions). Total contact number values sum the contacts discovered in all three categories. Where interaction strengths are indicated, these values are computed using either:

$$\text{Polar interaction strength} = \sum_{sum}^{polar} \frac{1}{d} \tag{S.3.5}$$

where d is the distance between contacting residues for both salt bridge and h-bonds in the context of polar-type interactions (equation derived from calculations used to estimate the strength of electrostatic interactions[58]) or:

$$\text{Hydrophobic interaction strength} = \sum_{sum}^{hydrophobic} \frac{1}{d^6} \tag{S.3.6}$$

for apolar interaction strength values, where d is the distance between hydrophobic interacting amino acid residues (equation derived calculations used to estimate the strength of van der Waals interactions[1]). The final X-ray structure analysis method used here is computed using Equation S.3.7:

$$\text{Monomer Proximity Index} = \frac{\sqrt{Size\ of\ Biological\ Dimer}}{Mon{-}Mon\ Distance} \tag{S.3.7}$$

where the larger the monomer proximity index value, the closer the packing of monomeric units within the dimer subcomplexes that comprise the tetramer.

PDB IDs for complexes studied here are: 1ICT for TTR, 3VHM for AVD, 1VAL for CON, 2HCY for ADH, 1ZAH for ALD, 1A5U for PKI, 4BLC for CAT and 1BGL for βGL.

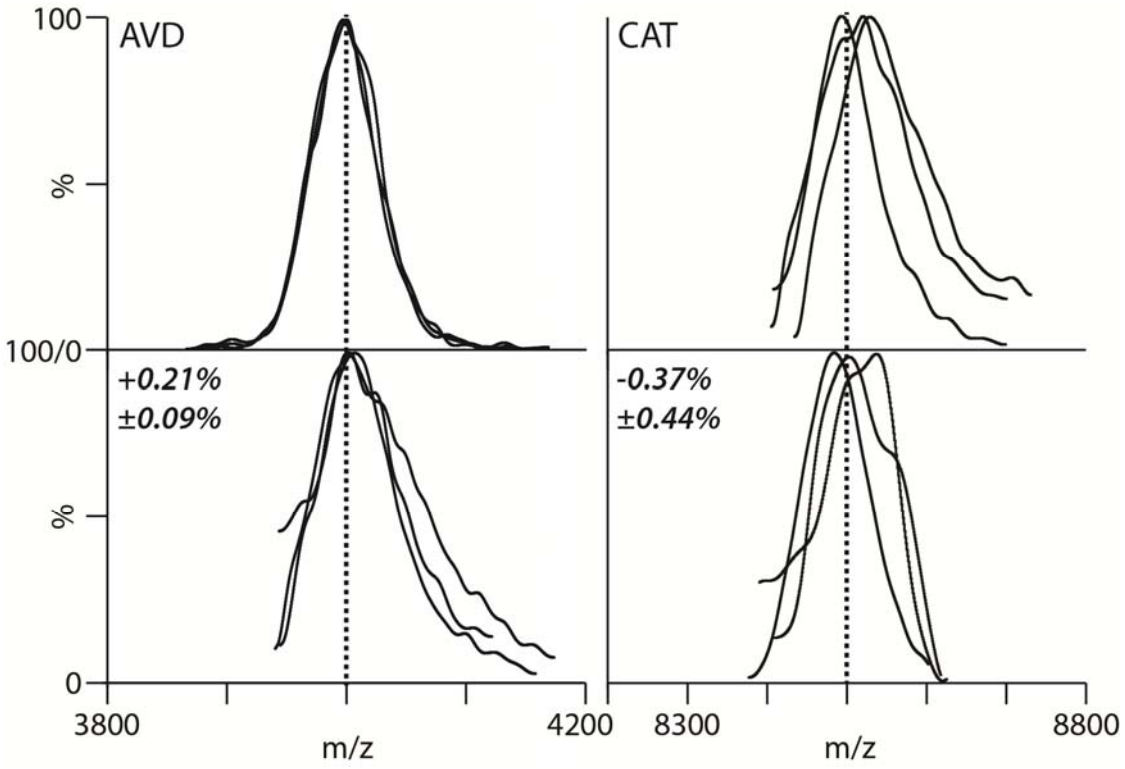

Figure S. 3-1 A comparison between the m/z and intact mass recorded for tetramer ions generated under control (200mM NH_4Ac, upper panel) and optimized disruption conditions (lower panel) discussed in Figure 3-3.
Three replicates are measured and overlaid for each dataset shown. Minimum smoothing is used for the data shown. The charge states displayed are the 16^+ for AVD and the 28^+ for CAT. The actual mass shifts observed for these assemblies under conditions optimized for protein complex disruption were +8.7 Da for AVD and -31.9 Da for CAT, compared to control.

Table S. 3-1 Solution conditions optimized for protein complex disruption, used in Figure 3-3.

Protein complex	Solution Conditions Optimized for Protein Disruption
TTR	47% DMSO, 4M NH_4Ac
AVD	53% DMSO, 10mM NH_4Ac
CON	17% DMSO, 3M NH_4Ac
ADH	17% DMSO, 2M NH_4Ac
ALD	33% DMSO, 1M NH_4Ac
PKI	17% DMSO, 2M NH_4Ac
CAT	33% DMSO, 2M NH_4Ac
βGL	17% DMSO, 3M NH_4Ac

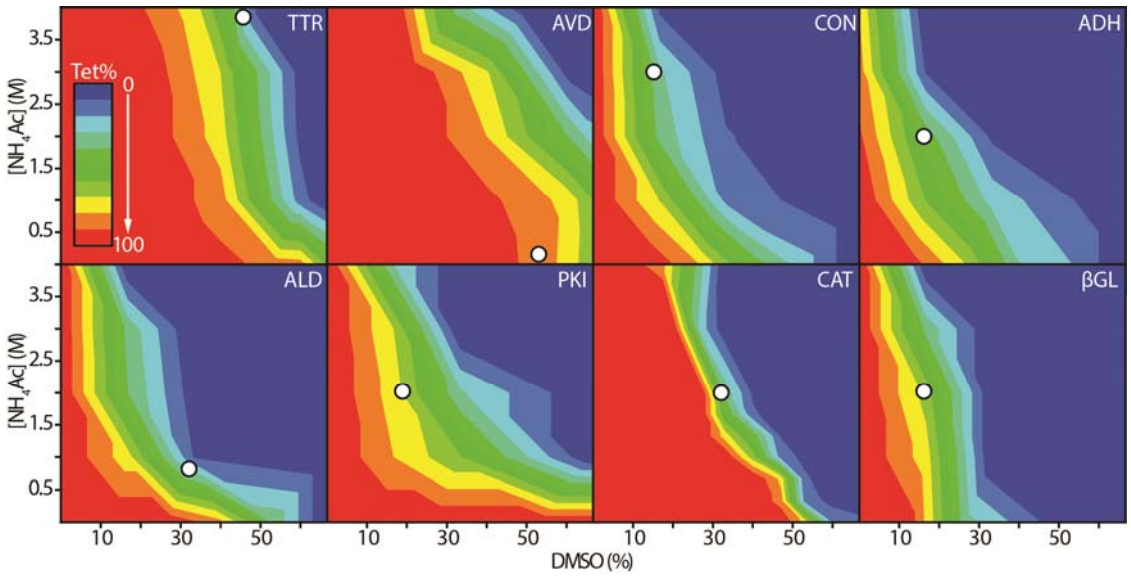

Figure S. 3-2 The positions of the disruption conditions shown in Table S.3-1 within the 2D titration maps shown in Figure 3-5A.
Conditions were selected based on two criteria: 1) That they contain optimized signal for both dimer and monomer ions and 2) that at least 50% of the intact tetramer had undergone dissociation. Conditions optimized for AVD disruption, as shown in Table S.3-1 and discussed as the 'orange dataset' in Figure 3-4, were chosen to maximize dimer signal intensity, even though intact tetramer disruption is not as complete as in the other complexes studied here.

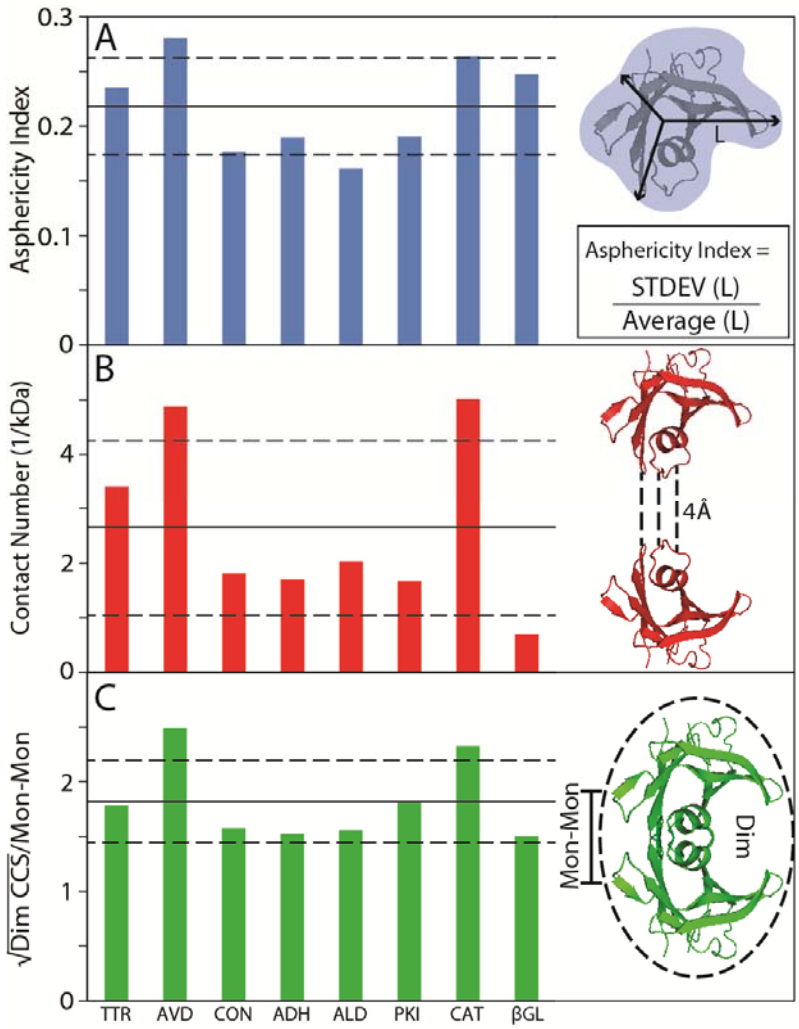

Figure S. 3-3 Comparing X-ray and IM-MS data provides critical insight into the structural deformations observed in AVD and CAT subcomplexes.
(A) Asphericity index values calculated for subunits within each protein complex. Larger values indicate a less globular shape for protein subunits. **(B)** The average number of contacts per-unit kDa within the protein-protein interfaces of the complexes studied in this chapter. **(C)** A monomer proximity index, calculated using the square root of the dimer size divided by the linear distance between the monomers that form the biological dimer for each complex. The larger the index value, the more-tightly packed the monomers within the X-ray structure. For all panels, the solid lines shown indicate average values for each metric, while dashed lines indicate ± one standard deviation from the mean. Each value indicated is discussed in detail and defined in the text.

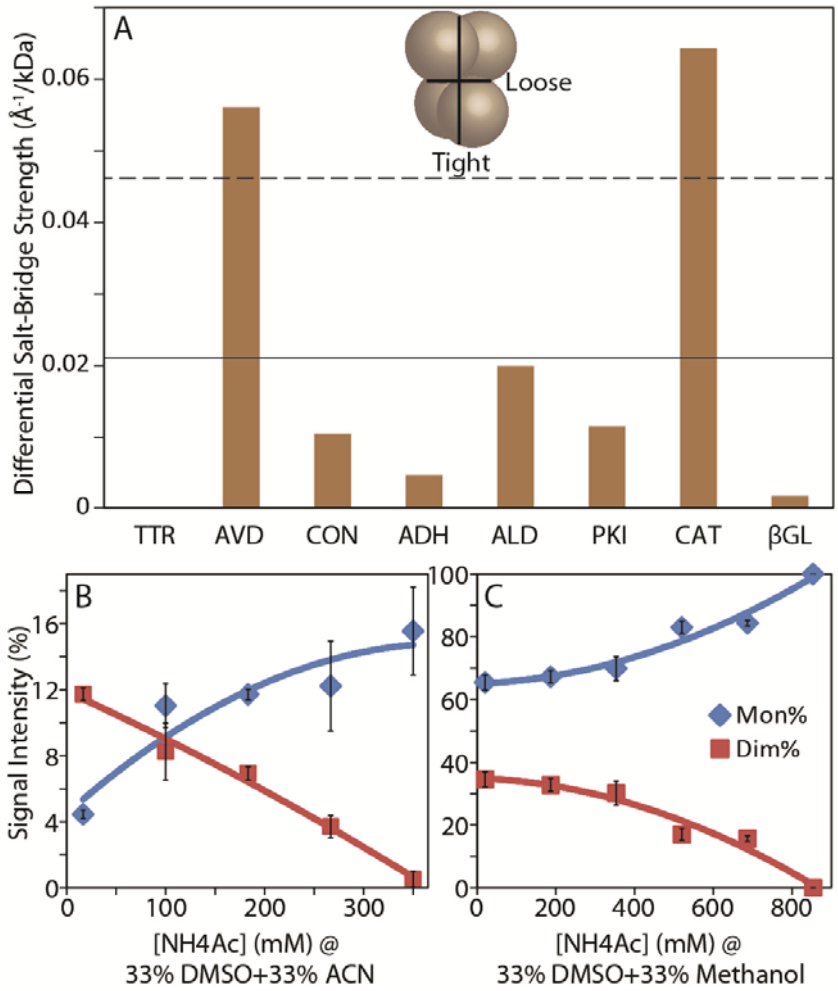

Figure S. 3-4 Correlations between IM-MS disruption experiments using a three-component titration strategy and salt bridge strength calculations from X-ray structure data

(A) A histogram plot showing the calculated strength of the salt-bridge interactions found in the smaller 'loose' interfaces within the X-ray structures of the complexes studied here, over those found in the larger 'tight' interfaces. The solid line shown indicates the mean value for this ratio for all eight tetrameric proteins studied, while dashed line indicates + one standard deviation from the mean. An analysis of this ratio suggests that significant salt-bridge based interface strength asymmetry exists in the AVD and CAT tetramers, but not in the other complexes studied here. **(B)** Monomer and dimer intensity values recorded during the three-component disruption of the AVD tetramer, utilizing NH4Ac, DMSO and acetonitirle (ACN). Organic content in this plot is fixed at 33% DMSO and 33% ACN. **(C)** Monomer and dimer intensity values for the CAT tetramer using a similar screen as in B. Methanol replaces ACN, but total organic content is fixed 33% DMSO and 33% methanol. In both B and C, clear divergent trends are observed between monomer and dimer signal intensity, favoring dimer formation only at the low ionic strength values that favor the retention of inter-protein salt bridge formation.

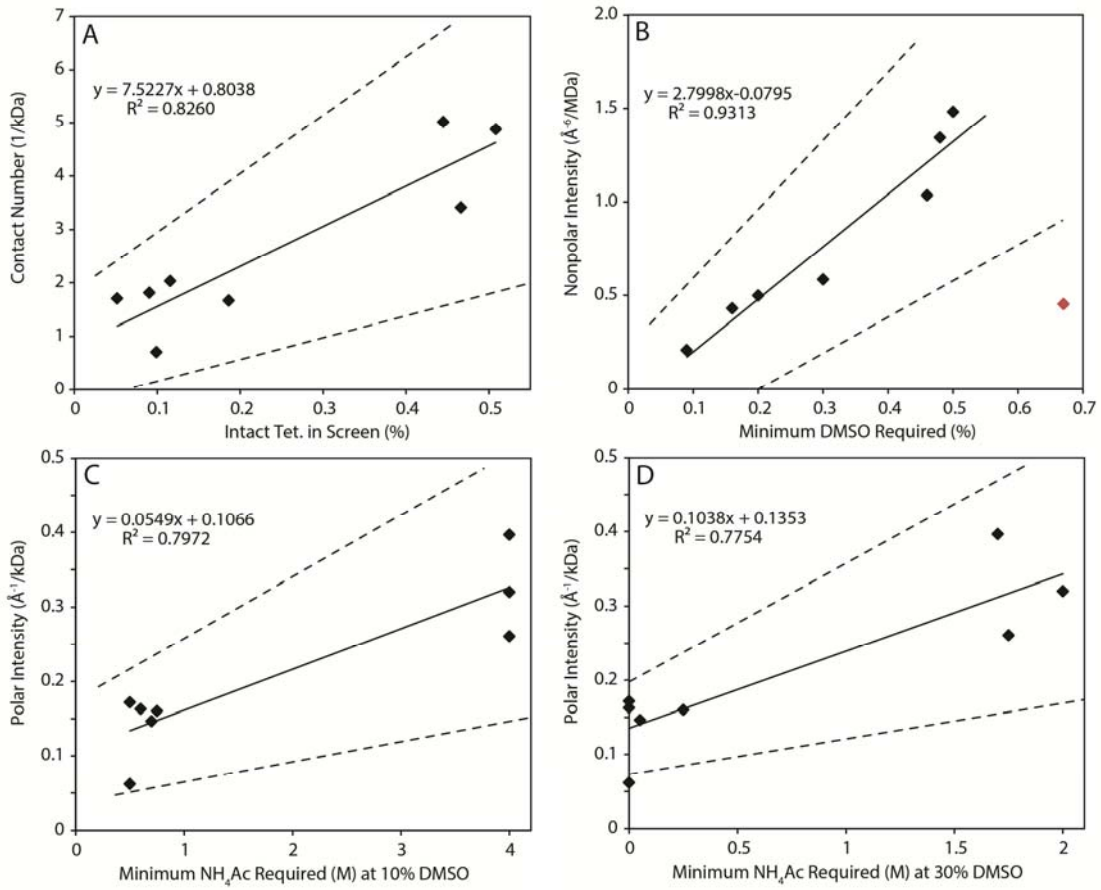

Figure S. 3-5 Correlations between our IM-MS titration experiments (X axis) and X-ray structure data (Y axis), shown as a scatter plot with fitted trend lines. Dashed lines indicate the 95% confidence interval for each fit.

(A) Normalized intact tetramer intensity integrated over all disruption conditions vs. the number of contacts per unit kDa calculated from X-ray data. **(B)** The DMSO % in solution, without added NH_4Ac, required to initiate tetramer disruption (signal drops by 10%) vs. the estimated average strength of hydrophobic contacts within the protein-protein contacts from X-ray ($Å^{-6}$/MDa). Red data point represents PKI, which is excluded from the fit. **(C)** NH_4Ac concentration required for tetramer disruption at fixed 10% and 30% DMSO **(D)** vs. the estimated average strength of the polar interactions ($Å^{-1}$/kDa) within the protein-protein contacts from X-ray. The correlation equations and linear correlation coefficients between IM-MS and X-ray datasets are each shown in the corresponding panel. The error estimates in the text indicates the difference between calculated values from X-ray data and prediction values based on the linear correlations shown.

3.6 References

(1) Alberts, B. *Cell* **1998**, *92*, 291-294.
(2) Pierce, K.; Premont, R.; Lefkowitz, R. *Nature Rev. Mol. Cell Biol.* **2002**, *3*, 639-650.
(3) Wu, C.; Murray, M.; Bernstein, S.; Condron, M.; Bitan, G.; Shea, J.-E.; Bowers, M. *J. Mol. Biol.* **2009**, *387*, 492-501.
(4) Steven, A.; Baumeister, W. *J. Struct. Biol.* **2008**, *163*, 186-195.
(5) Sali, A.; Glaeser, R.; Earnest, T.; Baumeister, W. *Nature* **2003**, *422*, 216-225.
(6) Robinson, C.; Sali, A.; Baumeister, W. *Nature* **2007**, *450*, 973-982.
(7) Dutta, S.; Berman, H. *Structure* **2005**, *13*, 381-388.
(8) Benesch, J.; Robinson, C. *Curr. Opin Struct. Biol* **2006**, *16*, 245-251.
(9) Hernandez, H.; Robinson, C. V. *Nature Protoc.* **2007**, *2*, 715-726.
(10) Parrish, J.; Gulyas, K.; Finley, R. *Curr. Opin. Biotech.* **2006**, *17*, 387-393.
(11) Gingras, A.-C.; Gstaiger, M.; Raught, B.; Aebersold, R. *Nature Rev. Mol. Cell Biol.* **2007**, *8*, 645-654.
(12) Sinz, A. *J. Mass Spectrom.* **2003**, *38*, 1225-1237.
(13) Wales, T.; Engen, J. *Mass Spectrom. Rev.* **2006**, *25*, 158-170.
(14) Konermann, L.; Tong, X.; Pan, Y. *J. Mass Spectrom.* **2008**, *43*, 1021-1036.
(15) Konermann, L.; Stocks, B.; Pan, Y.; Tong, X. *Mass Spectrom. Rev.* **2010**, *29*, 651-667.
(16) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198-207.
(17) Heck, A. *Nature Meth.* **2008**, *5*, 927-933.
(18) Hernandez, H.; Dziembowski, A.; Taverner, T.; Seraphin, B.; Robinson, C. V. *EMBO Rep.* **2006**, *7*, 605-610.
(19) Taverner, T.; Hernandez, H.; Sharon, M.; Ruotolo, B. T.; Matak-Vinkovic, D.; Devos, D.; Russell, R. B.; Robinson, C. V. *Accounts Chem. Res.* **2008**, *41*, 617-627.
(20) Sharon, M.; Robinson, C. *Annual Rev. Biochem.* **2007**, *76*, 167-193.
(21) Leary, J. A.; Schenauer, M. R.; Stefanescu, R.; Andaya, A.; Ruotolo, B. T.; Robinson, C. V.; Thalassinos, K.; Scrivens, J. H.; Sokabe, M.; Hershey, J. W. B. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1699-1706.
(22) Pukala, T. L.; Ruotolo, B. T.; Zhou, M.; Politis, A.; Stefanescu, R.; Leary, J. A.; Robinson, C. V. *Structure* **2009**, *17*, 1235-1243.
(23) Levy, E. D.; Boeri Erba, E.; Robinson, C. V.; Teichmann, S. A. *Nature* **2008**, *453*, 1262-1265.
(24) Zhou, M.; Robinson, C. V. *Trends Biochem.Sci.* **2010**, *35*, 522-529.
(25) Benesch, J. L. P. *J .Am .Soc. Mass Spectrom.* **2009**, *20*, 341-348.
(26) Zhou, M.; Dagan, S.; Wysocki, V. *Angew.Chemie Int. Ed.***2012**, *51*, 4336-4339.
(27) Deng, L.; Broom, A.; Kitova, E.; Richards, M.; Zheng, R.; Shoemaker, G.; Meiering, E.; Klassen, J. *J. Am.Chem. Soc.***2012**, *134*, 16586-16596.
(28) Deng, L.; Kitova, E.; Klassen, J. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 988-996.
(29) Benesch, J.; Aquilina, J.; Ruotolo, B.; Sobott, F.; Robinson, C. *Chem. Biol.* **2006**, *13*, 597-605.

(30) Joanna, F.; Matthew, F. B.; Carol, V. R.; Brandon, T. R. *Chem. Phys. Lett.***2012**, *524*.
(31) Pagel, K.; Hyung, S. J.; Ruotolo, B. T.; Robinson, C. V. *Anal. Chem.***2010**, *82*, 5363-5372.
(32) Zhong, Y.; Hyung, S.-J.; Ruotolo, B. *Expert Rev. Proteomics* **2012**, *9*, 47-58.
(33) Hall, Z.; Politis, A.; Robinson, C. *Structure* **2012**, *20*, 1596-1609.
(34) Zhou, M.; Sandercock, A.; Fraser, C.; Ridlova, G.; Stephens, E.; Schenauer, M.; Yokoi-Fong, T.; Barsky, D.; Leary, J.; Hershey, J.; Doudna, J.; Robinson, C. *Proc. Nat.Acad. Sci. U.S.A.* **2008**, *105*, 18139-18144.
(35) Mason, E. A.; McDaniel, E. W. *Transport Properties of Ions in Gases*; John Wiley & Sons: New York, 1988.
(36) Scarff, C.; Thalassinos, K.; Hilton, G.; Scrivens, J. *Rapid Comm. Mass Spectrom.* **2008**, *22*, 3297-3304.
(37) Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S. J.; Robinson, C. V. *Nature Protoc.***2008**, *3*, 1139-1152.
(38) Zhou, M.; Morgner, N.; Barrera, N.; Politis, A.; Isaacson, S.; Matak-Vinković, D.; Murata, T.; Bernal, R.; Stock, D.; Robinson, C. *Science* **2011**, *334*, 380-385.
(39) van Duijn, E.; Barbu, I.; Barendregt, A.; Jore, M.; Wiedenheft, B.; Lundgren, M.; Westra, E.; Brouns, S. J.; Doudna, J.; van der Oost, J.; Heck, A. *Molec. Cell. Proteomics* **2012**, *11*, 1430-1441.
(40) Baldwin, A.; Lioe, H.; Hilton, G.; Baker, L.; Rubinstein, J.; Kay, L.; Benesch, J. *Structure* **2011**, *19*, 1855-1863.
(41) Politis, A.; Park, A.; Hyung, S.-J.; Barsky, D.; Ruotolo, B.; Robinson, C. *Plos One* **2010**, *10, e12080.*
(42) Bernstein, S.; Dupuis, N.; Lazo, N.; Wyttenbach, T.; Condron, M.; Bitan, G.; Teplow, D.; Shea, J.-E.; Ruotolo, B.; Robinson, C.; Bowers, M. *Nature Chem.* **2009**, *1*, 326-331.
(43) Shoemaker, G.; van Duijn, E.; Crawford, S.; Uetrecht, C.; Baclayon, M.; Roos, W.; Wuite, G.; Estes, M.; Prasad, B. V.; Heck, A. *Molec. Cell. Proteomics* **2010**, *9*, 1742-1751.
(44) Smith, D.; Woods, L.; Radford, S.; Ashcroft, A. *Biophys. J.* **2011**, *101*, 1238-1247.
(45) Giles, K.; Williams, J. P.; Campuzano, I. *Rapid Comm. Mass Spectrom.* **2011**, *25*, 1559-1566.
(46) Zhong, Y.; Hyung, S. J.; Ruotolo, B. T. *Analyst* **2011**, *136*, 3534-3541.
(47) Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. *Anal. Chem.* **2010**, *82*, 9557-9565.
(48) David, E. C.; Martin, F. J. *J. Mass Spectrom.* **1997**, *32*.
(49) Benesch, J.; Ruotolo, B. *Curr. Opin. Sruct.Biol.* **2011**, *21*, 641-649.
(50) Jackson, M.; Mantsch, H. *Biochim. Biophys. Acta* **1991**, *1078*, 231-235.
(51) Han, L.; Ruotolo, B. *Angew. Chem. Int. Ed.* **2013**, 52, 8329–8332.
(52) Wooll, J.; Friesen, R.; White, M.; Watowich, S.; Fox, R.; Lee, J.; Czerwinski, E. *J. Mol. Biol.* **2001**, *312*, 525-540.
(53) Paci, E.; Vendruscolo, M.; Karplus, M. *Biophys. J.* **2002**, *83*, 3032-3038.
(54) Zarrine-Afsar, A.; Wallin, S.; Neculai, A.; Neudecker, P.; Howell, P.; Davidson, A.; Chan, H. *Proc. Nat.Acad. Sci. U.S.A.* **2008**, *105*, 9999-10004.

(55) Wyttenbach, T.; Bowers, M. *The journal of physical chemistry. B* **2011**, *115*, 12266-12275.
(56) Jianwei, L.; John, A. T.; Anne, E. C.; David, E. C. *Int. J. Mass Spectrom.* **1999**, *185-187*.
(57) Pierson, N.; Chen, L.; Valentine, S.; Russell, D.; Clemmer, D. *J. Am. Chem. Soc.* **2011**, *133*, 13810-13813.
(58) Leach, A. R. *Molecular Modeling Principles and Applications; Prentice Hall*: Harlow, England, **2001**.

Chapter 4. The Collisional and Coulombic Unfolding of Gas-Phase Proteins is Highly-Correlated to Their Domain Structure in Solution

4.1 Introduction

Rapidly characterizing the three-dimensional structures of proteins, and their higher-order complexes, is an unquestionably-important goal in the post-genomic era.[1] Large-scale protein structure determination efforts have enabled countless discovories in biochemistry and medicine and, as the component list of the cellular proteome takes shape and definition, the significance of such work has become increasingly evident.[2] In addition, similar endeavors have served to support protein engineering projects that seek to develop optimized enzymes for specific chemical transformations.[3] Both the growing need to structurally characterize small molecule pharmaceuticals bound to their protein targets, as well as the influence of analagous data during the development of biotherapeutics, underscores the impact of protein structure analysis in the context of pharmaceutical discovery and development.[4]

While X-ray and NMR analyses provide invaluable high-resolution protein structure information, mass spectrometry (MS) technologies can be engaged to provide protein structural constraints in a manner orthogonal to these more-established techniques.[5] The

growing suite of MS technologies enables measurements on a range of structural levels, and includes: chemical cross-linking,[6-8] H/D exchange,[9-14] oxidative footprinting,[15-17] ion mobility (IM) separation,[18-23] and MS of intact complexes.[24-26] Many examples exist where MS datasets have provided structural details for protein targets where few existed previously by offering significant advantages in in terms of speed, sample consumption, and complex mixture analysis.[27,28] Many of the contributions of MS-based structural analysis have been related to protein quaternary structure, including models of the eukaryotic exosome,[29] ribosomal initiation factor,[30,31] signalosome,[32] replisome,[33,34] chaperone,[35-39] ATPase,[40] and transcriptional regulation complexes.[41]

Despite successes in defining the overall architecture of multiprotein machines, MS technologies have fewer approaches available to determine the more-local elements of structure within these assemblies. Intra-molecular cross-linking can be used in small systems to constrain protein folding models, but such approaches currently have significant size limitations.[8] H/D exchange and oxidative footprinting data can be interpreted in the context of pre-existing high-resolution structures, but using such data to define unknown structures at a local level is currently a daunting task.[42] IM measurements have been used in conjunction with molecular dynamics for well over a decade to generate high-fidelity structures for peptides and small proteins,[43-45] but the size information contained within the IM measurement often loses its ability to adequately filter models for larger protein systems. As such, there is a clear need to develop MS-based technologies capable of providing local structure information within larger protein monomers.

Gas-phase measurements of global protein structure are now over two decades old,[46] and have revealed that small, single domain protein ions adopt both compact and unfolded conformations in the absence of bulk solvent. These transitions were first recorded in a manner correlated with protein charge state, which is in turn correlated with protein surface area in solution.[47-51] The resulting ions adopt conformers that are compact for low charge states but grow increasingly extended as their overall charge is increased. These experiments revealed a range of Coulombically-unfolded protein conformational families, and have recently led to more extensive studies. For example, IM measurements of ubiquitin ions created from acidified, methanol-containing samples are consistent with the known 'A state' of the protein, a less compact and partially folded form of the biopolymer observed under such conditions in solution.[52-54] Further, IM measurements have identified two distinct families of structures for the intrinsically disordered protein α-Synuclein, a key biomolecule implicated in the etiology of Parkinson's disease.[55] In addition to Coulombic unfolding, compact protein ions can be activated in the gas-phase to produce a range of unfolded conformations. As above, the activation-initiated unfolding of protein ions in the gas phase was first observed in small monomeric systems through collisional[48,51] and thermal[56] means of excitation. Subsequent IM and IM/IM experiments have illustrated the kinetic and energetic stability of protein ion conformers over timescales ranging from seconds to microseconds.[57-59]

Measurements of gas-phase structure have been applied more-recently to intact multiprotein complexes, and have revealed much regarding the structural transitions that occur for these multi-chain protein systems following activation in the absence of bulk solvent. Building upon pioneering work that provided the first entropy estimates for the

transition states associated with protein complex dissociation,[60] and chemical cross-linking experiments that directly related the charge asymmetry found in product ions produced by collision induced dissociation (CID) to protein unfolding,[61] discrete unfolded forms of multiprotein complex ions were captured using IM-MS measurements following collisional activation.[57,62] Our current pool of evidence points to a collisional unfolding mechanism for multi-chain protein complexes that leads to asymmetric structures where a single chain unfolds to a much greater extent than all others within the complex. This asymmetric unfolding of a single chain amongst many is most-likely driven by mobile protons on the surface of the protein,[63] that migrate towards thermally-unfolded regions within the structure.[61-64] As dictated by this mechanism, the general asymmetry of the unfolded multi-chain structure produced following collisional activation depends strongly upon the charge state of the ion selected,[65,66] and the time-scale over which ion activation occurs.[67]

In most of the work discussed above, gas-phase unfolding is employed as a means to study the role of solvent in determining the final native structures adopted by the protein. Recent trends, however, have pointed to the analytical utility of such unfolding data, specifically in its ability to detect subtle shifts in protein structure and stability in a manner correlated with known conformers in solution. Specifically, collision induced unfolding (CIU) data for protein-protein[36,68-70] and protein-ligand[35,71-73] complexes have demonstrated an ability to discover critical stability differences and distinguish solution-phase ligand binding modes by tracking the size of the unfolding protein as a function of the amount of activation energy provided during the experiment. These results are particularly significant given the above discussion, as they link IM-MS data to local

structural elements within large proteins, although thus far such links have largely been generated through a fingerprint-type analysis of the unfolding data collected.[71]

Here, we report on the first use of gas-phase unfolding as a means to determine the number of autonomously-folded domains within monomeric proteins. By selecting a broad range of proteins, containing both different numbers of autonomous domains and a range of local folds, we are able to identify persistent trends in gas-phase unfolding that will likely hold for proteins in general. Collisional and Coulombic unfolding data both provide clear, unmistakable correlations to the known domain structure of the 18 proteins and protein isoforms chosen for our study, ranging from 8 to 78 kDa in intact molecular mass. We include detailed criteria for selecting charge states for CIU analysis, and for defining signal-to-noise levels useful in detecting domains within Coulombic unfolding datasets. We conclude by discussing the implications for the correlations identified herein, both for multi-protein structure determination efforts by MS and for rapid biotherapeutics characterization in the future.

4.2 Experimental

4.2.1 Sample Preparation

Proteins were acquired from the following sources: ubiquitin (bovine, Sigma U6253), linear di-ubiquitin (human, BostonBiochem UC-700-100), linear tri-ubiquitin (human, Enzo life sciences BML-UW0780-0100), linear tetra-ubiquitin (human, Enzo life sciences BML-UW0785-0100), cytochrome C (horse, Sigma C2506), glutathione S-transferase (schistosoma japonicum, provided by the Ragsdale lab at the University of

Michigan, purified and prepared using a previously-published procedure[74]), serum albumin (bovine, Sigma A7906), green fluorescent protein (GFP, Aequorea Victoria, Prospec PRO-687), γ-D crystallin (human, Creative Biomart CRYGD-3634H) and fibronectin III 8-10 N-GST (human, Kerafast EUR116, the GST tag was removed by treatment with a Thrombin CleanCleave Kit from Sigma and purified using a glutathione sepharose 4B column from GE Healthcare). Standards for collision cross-section (CCS) calibration include: a commercially-available peptide mixture (Waters 186002337), cytochrome C, and the protein complexes avidin (Sigma A9275), concanavalin A (Sigma C2010), alcohol dehydrogenase (Sigma A7011), and glutamate dehydrogenase (Sigma, G7882). The peptide mixture (~2 μM for each peptide) was prepared in a water/methanol/ acetic acid (49/49/2) solution. All other proteins were solubilized in 200 mM aqueous ammonium acetate (NH_4Ac) solutions, or kept in their original solution conditions, and flash frozen for storage at -80 ℃. In general, stock samples (≥30 μM), were buffer-exchanged into a 200mM NH_4Ac solution (described above) prior to IM-MS analysis, using a Micro Bio-Spin 6/30 column (Bio-Rad, Hercules, CA). For CIU experiments, all monomeric proteins were diluted into 200 mM NH_4Ac to a final concentration of 5-10 μM before analysis, except glutathione S-transferase, which required treatment with 40% dimethyl sulfoxide and 1 M NH_4Ac to convert the native dimer into its folded monomeric form. For Coulombic unfolding, proteins were denatured in solution using a combination of methanol, acetic acid and, in some cases, dimethyl sulfoxide, optimized for each protein to achieve a maximum range of charge states. Specific denaturation conditions used for each protein are listed in Supplemental Information.

4.2.2 IM-MS Instrument and Methods

IM-MS data were acquired on a Synapt G2 (Waters, Milford, MA) optimized for the transmission of intact protein and protein complexes as described previously,[21] using nano-electrospray ionization (nESI). For CIU experiments, individual protein charge states were first selected by tandem MS using a quadrupole mass filter. The selected ions were then activated in the ion trap region of the instrument using energetic collisions with Ar (3×10^{-2} mbar) prior to IM separation. IM separates proteins according to their orientationally -averaged collision cross section (CCS) by driving ions through a gas-filled chamber with an applied electric field.[21-23,75-78] Ramping the bias voltage value that dictates the static voltage offset between the exit of the quadrupole filter and the inlet of the ion trap (termed 'trap CE' in the instrument control software) allowed for the collection of CIU related unfolding in an energy-resolved manner, revealing gas-phase protein conformations stable on the millisecond timescale, as observed previously.[35,36,68-71,73] In each CIU experiment, collision voltage was ramped in 2V increments until protein backbone fragments were observed. Coulombic unfolding datasets are presented without m/z filtration or any gas phase activation. IM resolution was optimized for all datasets collected by operating the T-wave separator under conditions of high N_2 pressure (3-4 mbar), high wave height (40 V) and medium wave velocity (400-900 m/s).[79]

4.2.3 Data Analysis

Initial IM-MS Data processing was performed using Masslynx v4.1and Driftscope v2.0 (both from Waters, Milford, MA). IM calibration to a He CCS axis was performed based

on a previously-described method, producing highly-linear calibration plots ($R^2 > 0.96$). CIU contour plots charting protein CCS versus collision voltage were generated by Origin software package. The color scale in these plots indicates the signal intensity recorded for each conformer family at the indicated collision energy values, normalized to the highest signal of each collision voltage. Solvent accessible surface areas for all proteins were calculated with GETAREA (www.curie.utmb.edu/getarea.html) using the appropriate Protein Data Bank entries (PDB codes are listed in supplemental information).

4.3 Results and Discussion

Protein coordinate files deposited in the PDB are classified in both the SCOP (Structural Classification of Proteins)[80] and CATH (Class, Architecture, Topology and Homologous superfamily protein structure classification)[81] databases according to their domain structures (Table S.4-1). While both databases assign proteins according to their structures, they do so using different criteria, formalisms, and differing amounts of experimental data.[82-84] For our experiments, we drew from both of these existing databases to assemble a series test proteins covering a broad range of domain structures and local folds in order to test the breadth of the correlations we discovered between solution and gas-phase data.

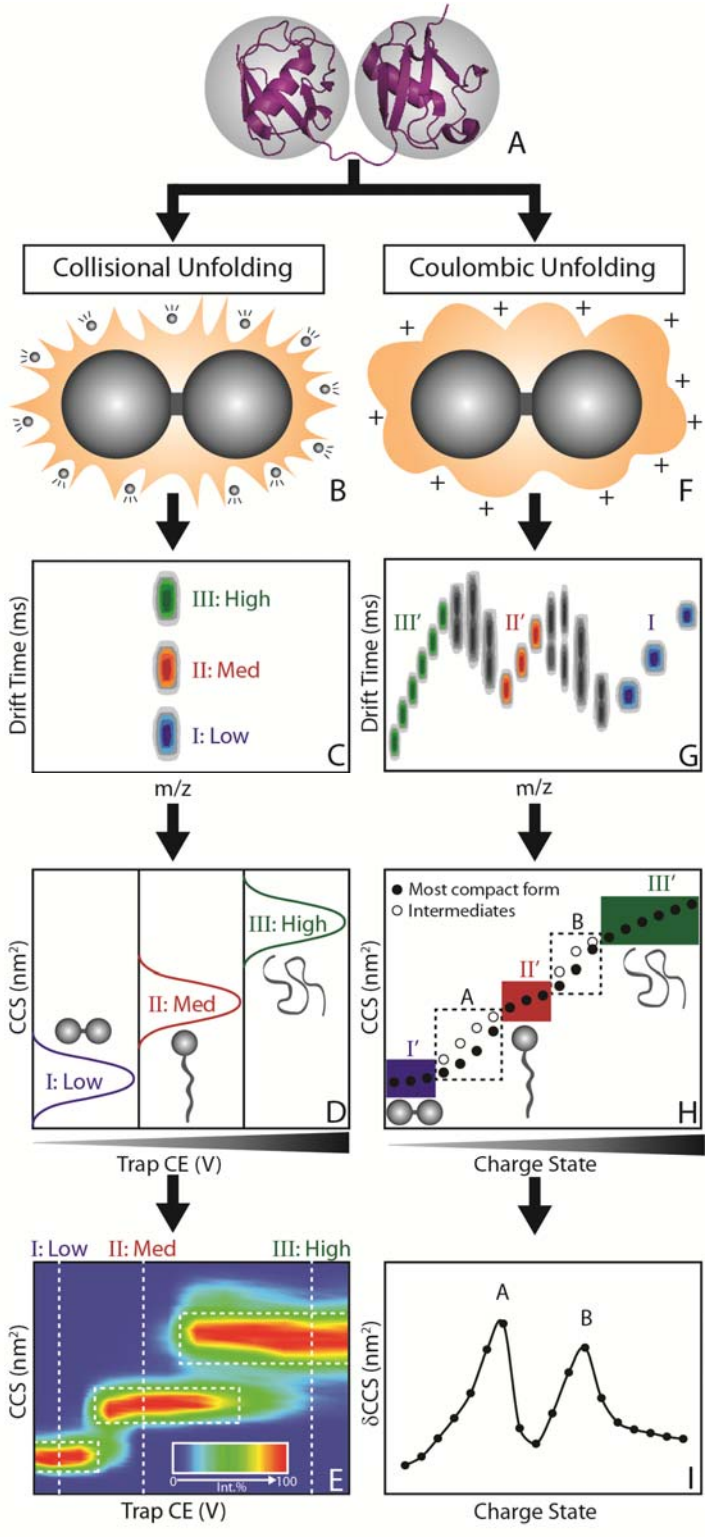

Figure 4-1 A diagram illustrating IM-MS data collection and analysis procedures.
A native two-domain protein **(A**) is subjected to collisional unfolding in the gas-phase **(B**) or Coulombic unfolding through denaturation in solution **(F). (C)** An example overlay of collisional unfolding data collected over a range of collision voltages, resulting in three conformational families I (blue), II (red) and III (green) observed at low, medium and high voltage values respectively. **(D)** The drift time profile is converted into cross-sectional data for each collision voltage. **(E)** Combined data recorded under all collision voltages, where collision cross-section (CCS) is plotted against collision voltage. Colors denote signal intensity as indicated. **(G)** Drift time profile of Coulombically unfolded protein ions. In addition to the three general regions of protein unfolding shown I (blue), II' (red) and III' (green), 2 intermediate stages, marked A and B, are also observed. **(H)** Calibrated data plotting CCS against protein charge state. **(I)** A first derivative plot where δCCS is plotted as a function of charge state, serving to highlight stages A and B.

Figure 4-1 illustrates our general experimental protocol. For example, a monomeric protein with 2 distinct domains (4-1A) can be sampled using nESI in two separate analysis streams. In the first experimental track, protein ions created under native conditions are selected for collisional activation (4-1B). Energy-resolved IM-MS data, acquired over a range of bias voltage values between the exit of the quadrupole mass filter and the ion trap region prior to IM separation, is used to track the unfolding of individual charge states of the protein monomer ion population (4-1C). IM Drift time data is converted to CCS in order to directly correlate the unfolding data to protein sizes and enable comparisons between different proteins and replicate runs on the same protein system (4-1D). These CCS data are then combined into a CIU fingerprint, where the energy required to unfold the protein ions can be readily identified for each conformer family detected (4-1E). Our CIU data analysis involves assessment of the total number of resolved features observed in drift time/CCS space in excess of the initial population detected, and correlates this number with the number of known domains within a given monomer.

The second experimental track involves proteins paritally denatured in solution using a combination of organic and acidified solvents prior to nESI (4-1F). These data are characterized by the extensive charging of the protein ions produced, in a manner correlated with their increased surface area upon unfolding, and the drift time profiles for such ions are recorded for all of the charge states generated (4-1G). As observed previously, protein ions adopt a broad range of unfolded or partially unfolded structures upon charging (Figure S.4-1). The degree of unfolding observed for monomeric protein systems can be divided into three main categories as a function of the charge adopted by

the protein. At low charge states, minimal or no unfolding is observed. Intermediate unfolding is observed as charge is increased; typically accompanied by multiple co-populated protein conformer families and by broadened or multi-modal IM drift time profiles. Ultimately, highly-charged proteins achieve string-like conformations in the absence of solvent. Our analysis tracks all of this data in both IM drift time and CCS space (4-1H), and uses a simple derivative analysis to identify regions within IM data that exhibit the greatest change in CCS (δCCS, 4-1I), which typically overlaps with charge states where the protein ion exists simultaneously in multiple conformations in the gas phase. As above, this enables us to correlate regions of greatest δCCS directly to the number of domains within the protein being analyzed.

4.3.1 Estabilising a General Correlation between Gas-Phase Unfolding and Protein Domain Number

We first examine the unfolding properties of ubiquitin, and various covalently-ligated analogs, in the gas-phase. Monomeric ubiquitin is currently among the most-studied gas-phase proteins.[54,85-92] IM-MS[53,54,92] and tandem IM[87] analysis have both been used to evaluate the influence of charge and energy on this protein system in detail. Recent studies have demonstrated a clear correlation between the folded state of the biopolymer in solution and its resultant gas-phase conformer populations.[54] The native ubiquitin fold is a mixture of α-helix and β-sheet secondary structure, making it an ideal initial target for our approach. A single ubiquitin domain exhibits two conformational states when low charge state protein ions are selected for CIU, and a single δCCS feature in our Coulombic unfolding dataset, both indicative of a single domain structure (Figure 4-2).

As additional ubiquitin domains are added to the C-terminus of the original protein, further CIU and Coulombic unfolding signals are observed, each in precise agreement with the expected response for the known domain structure of the proteins analyzed

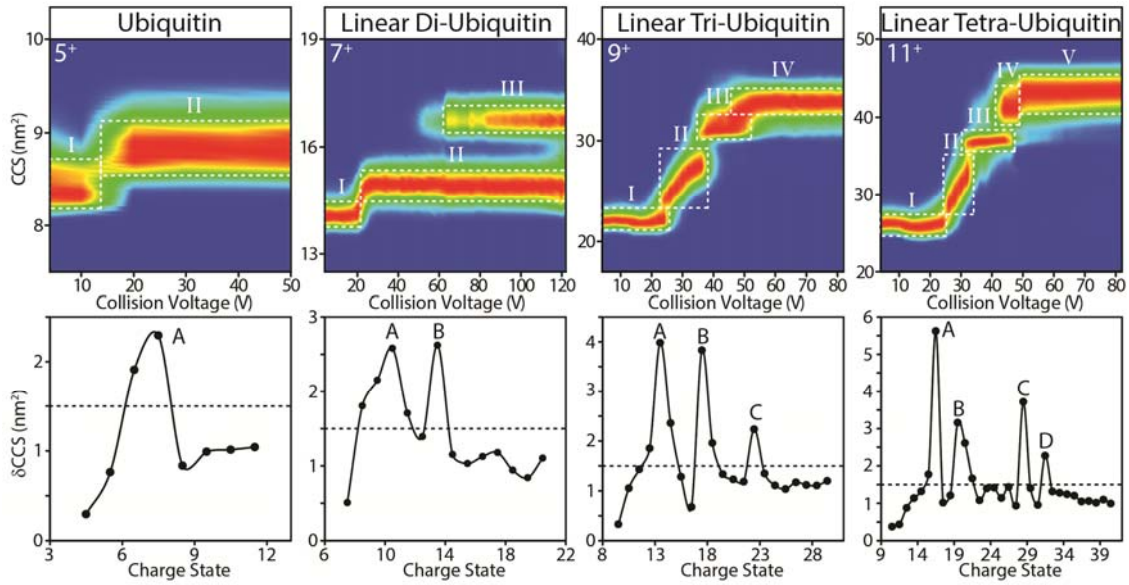

Figure 4-2 Collisional unfolding (top) and Coulombic unfolding (bottom) data of poly-ubiquitin domains linked in a linear fashion.
Charge states used in collisional unfolding experiments are noted on the top left corner. Conformational families detected are noted with numerals I through V. Intermediate unfolding stages identified within Coulombic unfolding experiments are labeled A to D. The dashed line represents an empirically-derived signal cut-off value at δCCS =1.5 nm^2.

Our CIU data on poly-ubiquitin protein constructs reveal some general trends in the domain-related unfolding of gas-phase proteins. For example, while mono- (5^+) and di- (7^+) ubiquitin CIU fingerprints contain clear, step-wise transitions between the different unfolded forms observed, tri- (9^+) and tetra- (11^+) ubiquitin CIU data reveal more-complex transitions between intermediate states. Specifically, conformational families II in both the tri- and tetra-ubiquitin datasets shown in Figure 4-2 appear broadened on the

CCS axis, and possesses a diagonalized shape in CCS/energy space. While high-resolution data is not available for the linear tri- and tetra-ubiquitin constructs studied here,[93,94] and computational tools capable of accurately assessing the structure of the unfolding intermediates probed here are also not yet available, it is likely that these broadened CIU features relate to the more-complex tertiary arrangement of domains in higher-order ubiquitin structures. Despite these complexities, CCS features are resolved in our dataset and the number of these features is clearly correlated with the number of autonomously folded domains within the protein.

Ubiquitin Coulombic unfolding data also reveal trends that can be applied generally to multi-domain proteins studied here by IM-MS. Mono-ubiquitin exhibits maximum δCCS values at ~7 charges, which compares well to previously-reported IM-MS data for this protein.[53] In comparison, di-ubiquitin exhibits maximum δCCS at ~10 charges, and this value increases to ~13 and ~15 charges for tri- and tetra-ubiquitin respectively. This increase correlates closely with the average CCS values for the intact proteins in low charge states and estimates of the same based on a fixed-density approach (linear $R^2 > 0.99$),[21] suggesting that initial domain unfolding is strongly correlated with protein ion surface charge. Furthermore, while di- and tri- ubiquitin require relatively equal amounts of charge to unfold subsequent domains (4 or 5 charges on average), the Coulombic unfolding of tetra-ubiquitin appears less-evenly spaced as a function of charge, requiring between 3 and 9 charges to initiate unfolding of subsequent domains. As above, this more-complicated relationship between charge state and unfolding is likely related to the tertiary arrangement of the domains present in tetra-ubiquitin. Importantly, the additional complexity observed does not prevent us from accurately

recording the number of domains within this protein using Coulombic unfolding measurements. We have assigned an empirical noise value 1.5 nm^2 to all such datasets and, when signals observed above this threshold are collected and counted, a clear, general correlation between unfolding data and known domain structure is observed.

4.3.2 Broadening the Unfolding Dataset using Proteins that are Predominantly α-helical or β-sheet

In order to expand our protein unfolding dataset and evaluate the generality of the correlations observed in Figure 4-2, we applied our IM-MS approach to additional monomeric proteins that A) exhibit a similar number of total domains as the poly-ubiquitin system and B) contain individual domains with significantly different fold types. Data shown in Figure 4-3 is derived from a set of proteins that are constructed from domains which each contain at least 40% α-helix and less than 10% β-sheet. In the case of cytochrome C, as a single domain protein, we observe two conformational families during CIU and one δCCS peak from Coulombic unfolding analysis, in a manner similar to ubiquitin. For glutathione S-transferase, a two-domain monomer in this category, three conformational families and two δCCS peaks are measured during CIU and Coulombic unfolding experiments respectively. We observe the three-domain protein serum albumin to generate four CIU conformational families and three δCCS peaks upon Coulombic unfolding. All of these results conform precisely to the expected number of domains within these proteins, and trends similar to those observed in our poly-ubiquitin analysis are also evident in these data. For example, in a similar fashion to tri- and tetra-ubiquitin, features I and II in serum albumin appear broadened in CCS space. Also,

Coulombic unfolding patterns for serum albumin appear irregularly spaced as a function of charge, when compared to the lower-order domain proteins studied here, in a manner similar to our findings for tetra-ubiquitin. Again, since this is only observed in our data for higher-order domain structures, we attribute this observation to the tertiary arrangement of protein domains within serum albumin, as compared to the more discrete domain structures observed in the high-resolution data for cytochrome c and glutathione S-transferase.

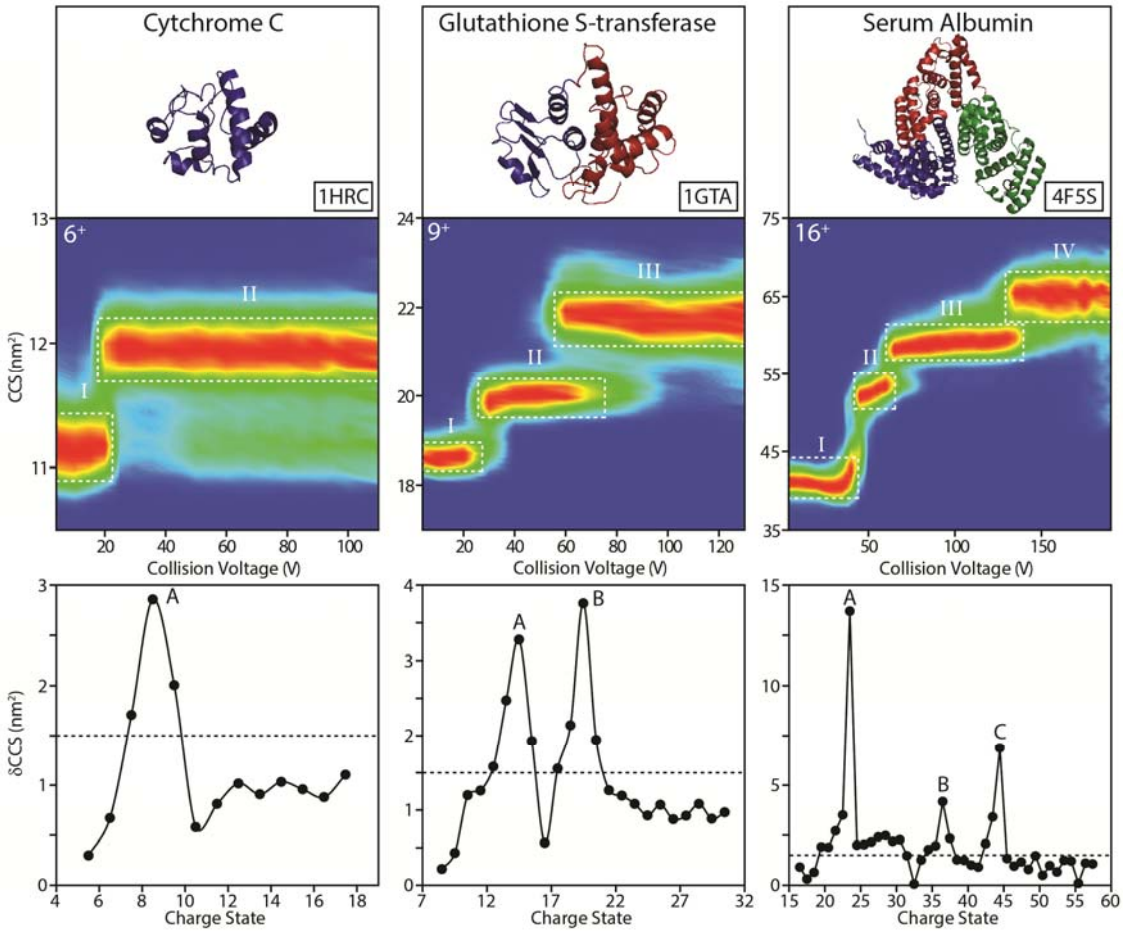

Figure 4-3 X-ray structure (top), collisional unfolding (middle) and Coulombic unfolding (bottom) data for predominantly α-helical proteins.
Including cytochrome C (one domain), glutathione S-transferase (two domains) and serum albumin (three domains). PDB codes for each X-ray dataset is provided in the figure, and is color coded according to domain structure.

While the data shown in Figure 4-3 pertains to primarily α-helix proteins, Figure 4-4 contains data from proteins comprised of domains having more than 40% β-sheet and less than 10% α-helix. We collected CIU and Coulombic unfolding data for GFP, γ-D crystallin and fibronectin, as examples of one, two and three-domain proteins respectively within this protein class. Each protein displays unfolding patterns that correspond precisely to their expected number of domains in solution. In addition, the trends indicated above for poly-ubiquitin and α-helical proteins in Figures 4-2 and 4-3 hold for the β-sheet rich systems shown in Figure 4-4. GFP represents a somewhat unique dataset among those presented here. The data shown in Figure 4-4 for this protein indicates a more-frustrated transition between compact and unfolded forms in both CIU and Coulombic unfolding measurement. Specifically, CIU data show a low intensity feature that appears at 21.5 nm^2 and an acceleration potential of 45V. While this feature is of very low intensity (maximum relative intensity of <10%) and can be easily ignored in the context of our analytical protocol for domain number determination, it does indicate a more-complex gas-phase unfolding transition than observed in other single-domain proteins studied here. Similar trends are detected in our GFP Coulombic unfolding data when the width and asymmetry of the δCCS peak detected is compared with other single domain protein data shown in Figures 4-2 and 4-3, and may reflect the stability and unique architecture of the β-barrel domain within our broader database of proteins. However, as in the case of the CIU data shown in Figure 4-4, these complexities do not hamper our ability to correctly assign the number of domains within GFP based on Coulombic unfolding data.

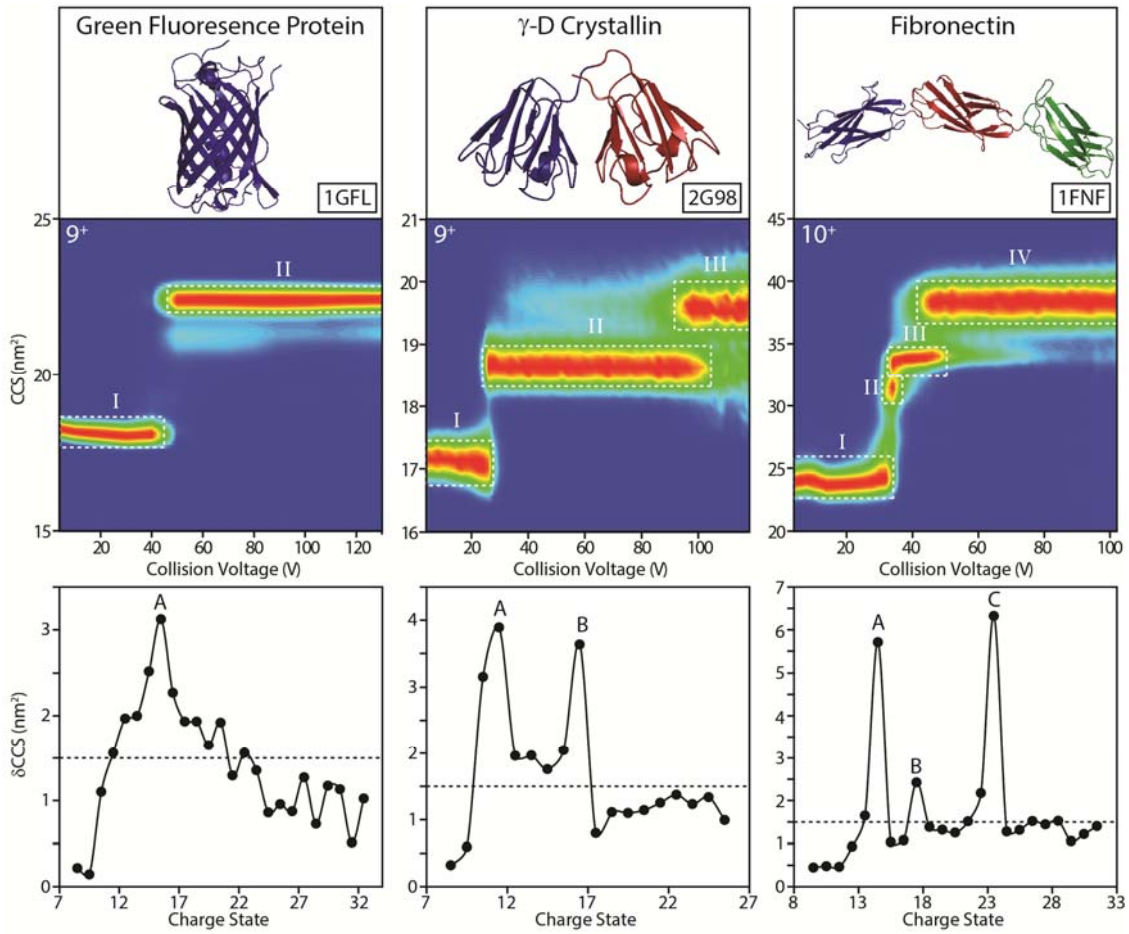

Figure 4-4 X-ray structure (top), collisional unfolding (middle) and Coulombic unfolding (bottom) data predominantly β-sheet proteins.

Including green fluorescence protein (one domain), γ-D crystallin (two domains) and fibronectin (three domains). PDB codes for each X-ray dataset is provided in the figure, and is color coded according to domain structure.

4.3.3 Selecting Optimal Charge States for CIU Protein Domain Analysis

One key element to successfully mapping CIU data onto protein domain structure involves the initial charge state selection. Higher protein charge states, if selected, typically generate a larger number of unfolded transitions that may, or may not, correlate with protein domain structure in a manner dependent upon the overall mass and size of the protein analyzed. As shown on Figure 4-5, all of the charge states identified in this study that produce CIU data in optimal agreement with domain-resolved unfolding patters follow a power-law correlation as a function of protein mass (Figure 4-5A) and solvent accessible surface area (Figure 4-5B). Both experimentally observed trends reported here are mirrored by previously-observed functions that are used to predict average charging of protein ions from native solution in ESI, but are offset to lower charge values.[95] For protein mass-dependent trends (Figure 4-5A), optimal charge states are generally 33% lower than those expected for native proteins analyzed by ESI.[95,96] However, if the solvent accessible surface area is used instead to evaluate this correlation, optimal charge states for CIU domain analysis are approximately 15% lower than the average expected value for proteins electrosprayed from native solutions.[97] Given the data presented here, the correlations shown in Figure 4-5 should provide adequate information to construct protocols that enable the CIU-based domain analysis of proteins having unknown structures.

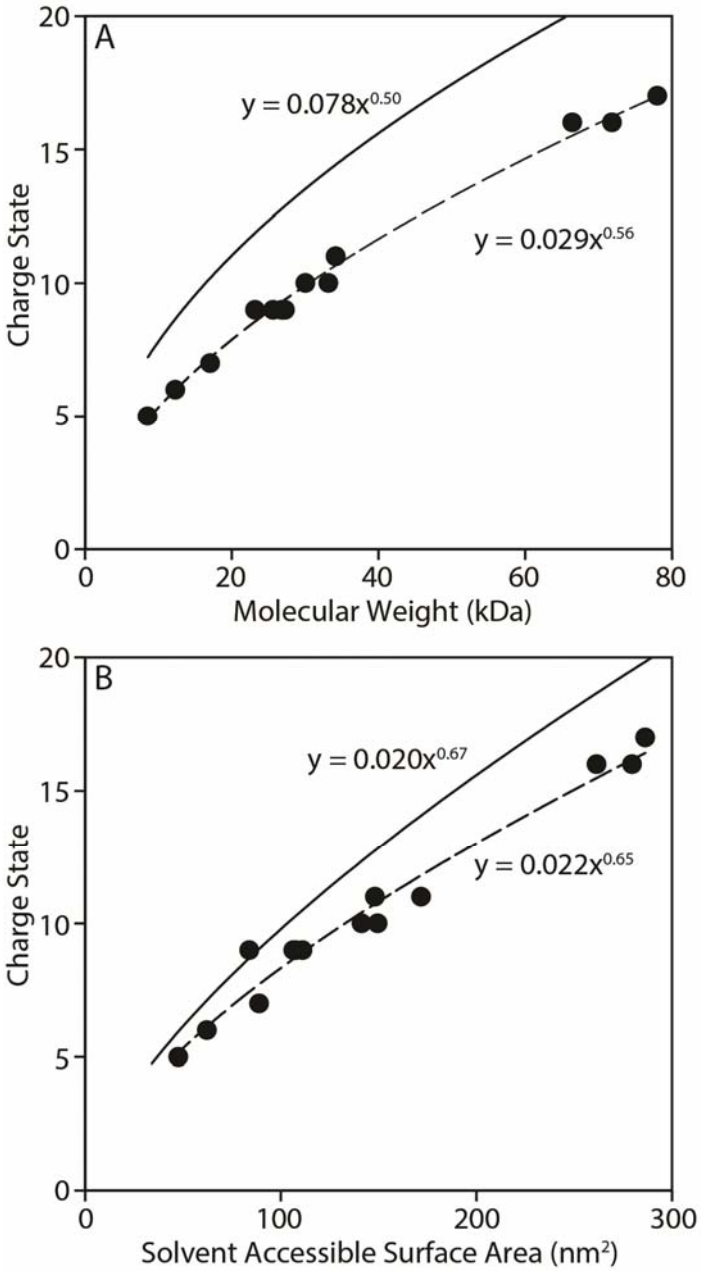

Figure 4-5 Optimal charge states for CIU protein domain analysis
(A) Protein charge states used in CIU experiments plotted against protein mass for thirteen of the systems studied here (--●--). R^2 for the fit shown is 0.99. The solid line indicates expected average charge state from previous experiments that use fixed protein density to estimate surface charging for protein ions produced from ESI under native conditions.[95] **(B)** Protein charge states used in CIU experiments plotted against solvent accessible surface area for the same systems analyzed in A (--●--), as determined from available X-ray structures.[97,110] R^2 for the fit shown is 0.96. The solid line indicates expected average charge states for these protein ion surface areas, generated from ESI under native conditions, derived from previous experimental data. [97]

4.4 Conclusions

In this chapter we have used a group of eighteen proteins to identify and validate a fundamental correlation between gas-phase protein unfolding and protein domain

structure in solution. Using our IM-MS approach, the number of domains within proteins can be identified by recording either the intermediate sizes adopted by in CIU data or by measuring the number of regions within Coulombic unfolding datasets that exhibit a large rate of change. These trends are shown to be independent of protein size (8-78 kDa) and fold (α-helix, β-sheet, and mixed protein folds tested). By selecting charge states that are significantly lower than the average produced by ESI from native solutions, CIU data can be generated that is highly protein domain-correlated. For higher-order domain proteins, we detect slight aberrations in our data that we attribute to tertiary contacts between domains,[98] although we cannot rule out the influence of excess Coulombic or thermal activation in these higher charge state systems, as observed in previous work.[61] Future work in our lab will aim to expand our approach to identify ways to mine this data for more detailed, structurally-important information.

The unfolding observed here demonstrates discrete unfolding stages for each domain, similar to Atomic Force Microscopy (AFM) experiments where proteins tethered to a surface are mechanically unfolded.[99-101] Unlike traditional thermal unfolding or solution-phase unfolding using either guanadinium chloride or urea, in which the homologous domains unfold cooperatively,[102,103] gas-phase unfolding is likely able to make use of subtle structural differences in equivalent domains to initiate the unfolding in only one region of the structure. Once triggered, asymmetric protein unfolding is undoubtedly driven by a mechanism similar to that previously-described for multi-chain non-covalent protein complex ions.[104]

We have also included, primarily in Supplemental Information, a group of examples where the proteins are not linked as a single chain but instead connected elsewhere in the sequence of the previous domain. Unsurprisingly, CIU data for such proteins are often more complex than the simple correlations shown in Figures 4-2 to 4-4 (K48 and K63 linked di-ubiquitin, Figure S.4-2), but since non-linear protein sequences comprise less than 3% of the proteome,[105] and are easily detected by MS,[106] spurious CIU domain assignments in unknowns that possess non-linear sequences are unlikely. Furthermore, our data for cytochrome P450 reductase (CPR), a three-domain protein, produces CIU data in excellent agreement with known domain number, but produces Coulombic unfolding data that is more difficult to interpret, likely due to an inability to identify optimized solution conditions for protein denaturation (Figure S.4-3). Likewise, in the case of the Abelson protein tyrosine kinase (Abl) and transferrin, SCOP and CATH provide different assessments of the total number of autonomous domains contained within the proteins. In the former case, CIU and Coulombic unfolding identify a two domain structure in agreement with CATH. In the latter case, Coulombic data is in agreement with the SCOP classification, and the CIU with CATH, for reasons that are currently unclear (Figure S.4-4). Due to essential differences in methodology, SCOP and CATH domain assignments share classifications in 70% of all databse entries.[84] However, domain assignments that agree with either database are expected to be useful for structural assignments in unknown systems, and our current database of IM-MS data for linear biopolymers provides clear agreements with either one or both databases in all cases. Finally, in the case of billverdin reductase (BVR) (Figure S.4-5), a two domain structure where each domain is spatially-intermingled with the other, the Coulombic

unfolding data is in strong agreement with CATH domain analysis, but CIU produces data that suggests a weak third domain within the system, perhaps due to the 'swapped' nature of the domain structure in the protein.[107,108] As in the case of non-linear protein sequences discussed above, 'swapped' domain architectures are expected to account for an exceedingly small portion of the proteome (less than 0.3% of all complexes), thus minimally impacting the ability of IM-MS to assess the domain structure of unknown proteins.[109] As such, a clear relationship is identified between gas-phase unfolding and known domain architecture, and in future studies, these unfolding data may be used in combination with other biophysical probes to more-completely annotate the autonomously folded regions within currently unknown protein systems.

Acknowledgements to Jessica Rabuck and Matthew Soellner (University of Michigan, Chemistry) for providing Abl protein and for helpful discussions, as well as Ireena Bagai and Stephen Ragsdale (University of Michigan, Biological Chemistry) for providing BVR and CPR. Fundamental gas-phase protein unfolding work in the Ruotolo group is supported by the American Society for Mass Spectrometry (in the form of a Research Award), the Oak Ridge Associated Universities (in the form of a Ralph E. Powe Junior Faculty Enhancement Award), and the National Science Foundation (in the form of a CAREER Award, 1253384).

4.5 Supplemental Information

Table S. 4-1 Detailed information for proteins used in this chapter, including protein names, organisms, masses, PDB codes, secondary structure content (α helix and β sheet percentage) and domain numbers defined by SCOP (red) and CATH (green).

Domain numbers displayed in black indicate unambiguous domain assignments. Linear tri- and tetra- ubiquitin do not have PDB structures. Biliverdin reductase is comprised of 2 mixed domains, meaning each domain consists of different sections from the amino acid chain.

Protein	Organism	Mass (kDa)	PDB code	Secondary Structure	Domain # (SCOP/ CATH)
Ubiquitin	Bovine	8.6	1UBI	25% α+34% β	1
Linear Di-Ubiquitin	Human	17.1	3AXC	22% α+33% β	2
Linear Tri-Ubiquitin	Human	25.7	N/A	Similar to Ubiquitin	3
Linear Tetra-Ubiquitin	Human	34.2	N/A	Similar to Ubiquitin	4
Cytochrome C	Horse	12.4	1HRC	40% α+ 1% β	1
GST Monomer	Schistosoma japonicum	27.3	1GTA	50% α+ 7% β	2
Serum Albumin	Bovine	66.4	4F5S	73% α	3
Green Fluorescent Protein	Aequorea Victoria	26.9	1GFL	10% α+ 49% β	1
Gamma D Crystallin	Human	23.2	2G98	9% α+ 42% β	2
Fibronectin III 8-10	Human	30.1	1FNF	51% β	3
K48 linked Di-Ubiquitin	Human	17.1	1AAR	19% α+ 34% β	2
K63 linked Di-Ubiquitin	Human	17.1	2JF5	23% α+ 32% β	2

NADPH-Cytochrome P450 Reductase	Human	71.8	3QE2	36% α+ 20% β	3
Abelson Protein Tyrosine Kinase	Human	33.2	2GQG	39% α+ 17% β	1/2
Transferrin	Bovine	78.0	1BLF	31% α+ 17% β	2/4
Biliverdin Reductase	Human	34.2	2H63	40% α+ 25% β	2 (mixed)

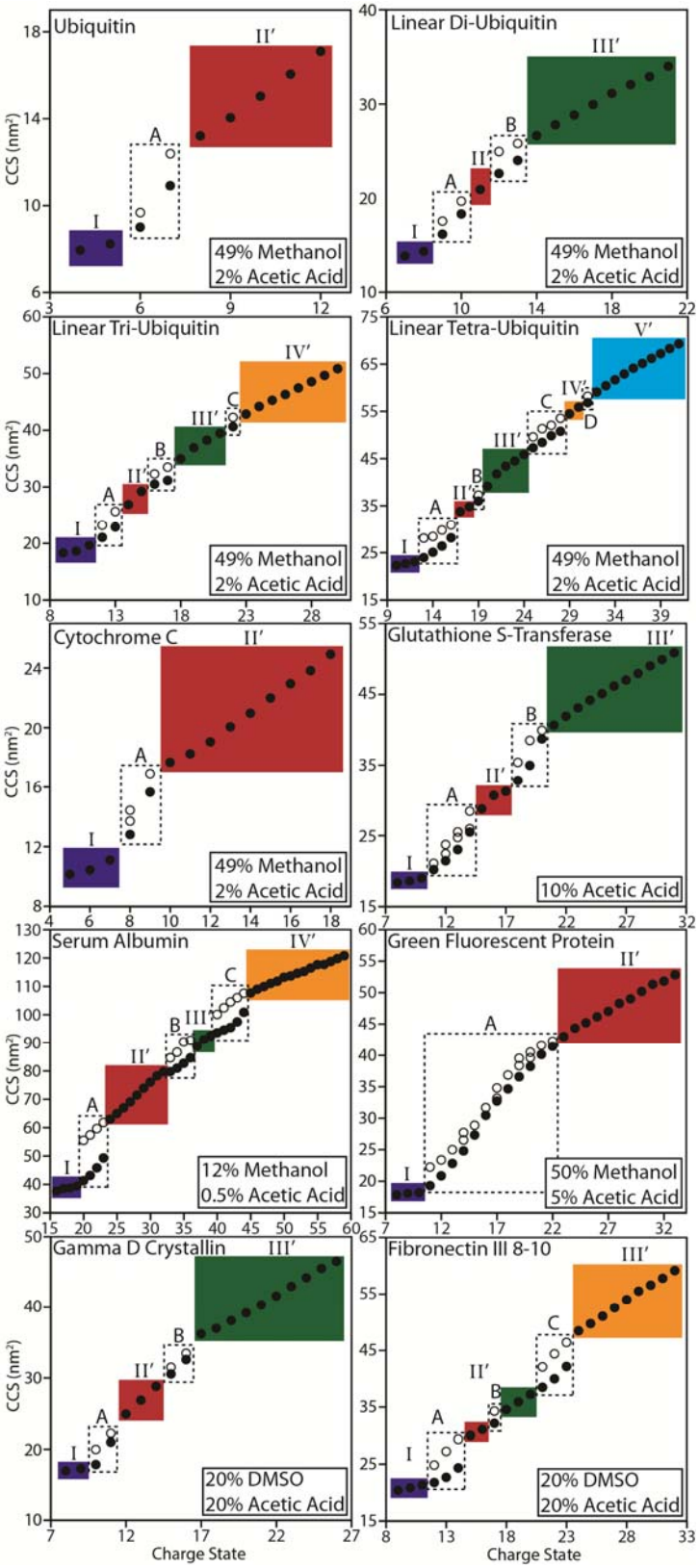

Figure S. 4-1 Coulombic unfolding data for proteins discussed in the main text of this chapter, including 1-4 ubiquitins, 3 primarily α-helix proteins and 3 primarily β-sheet proteins. CCS is plotted against charge state adopted upon solution unfolding. The denaturing solution condition used for each protein is listed in the lower right corner of each plot. Again, the number of known domains in each protein agrees with the number of intermediate stages, labeled A through D, and equals the number of unfolded conformational families identified upon CIU, labeled as I to V'. The most compact structure of each charge state is labeled using closed circles. The intermediate stages having large δCCS values are characterized by multiple unfolded structures that exist simultaneously. These intermediates, rather than the most compact form, are labeled as open circles.

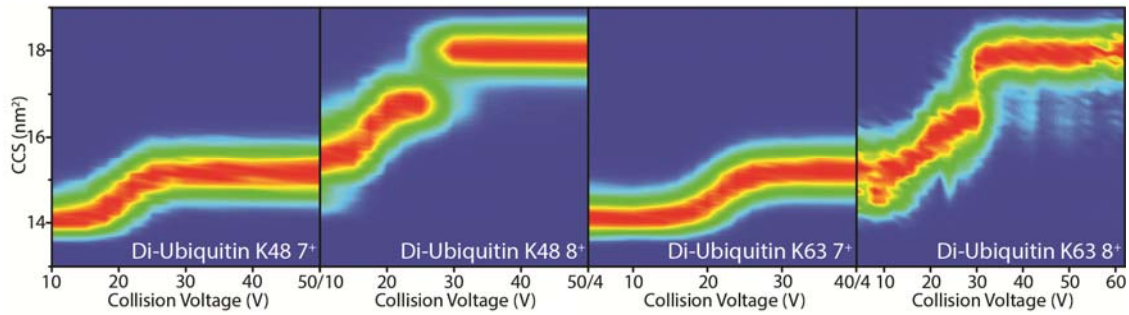

Figure S. 4-2 Collisional unfolding data of non-linear di-ubiquitins, K48 and K63 linked, at charge states 7^+ and 8^+.
No clear indication of a two-domain signature due to the sequence linkages that do not travel directly through the protein termini, although 8^+ displays data with a signature more-closely associated with the expected domain structure of the protein than the 7^+ in this case.

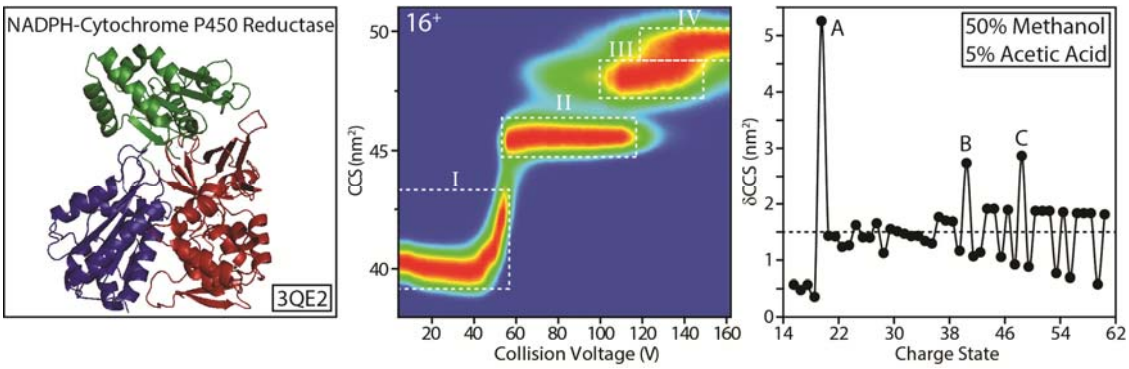

Figure S. 4-3 Crystal structure (left), collisional unfolding (middle) and Coulombic unfolding (right) data for NADPH-Cytochrome P450 Reductase (CPR).
This three-domain protein exhibits a clear three domain CIU pattern. On the other hand, the Coulombic unfolding data is more difficult to interpret. Low intensity features for the final 2 domains, along with a higher degree of noise than expected, frustrate our analysis. It is likely that the solution condition has not been optimized for this protein.

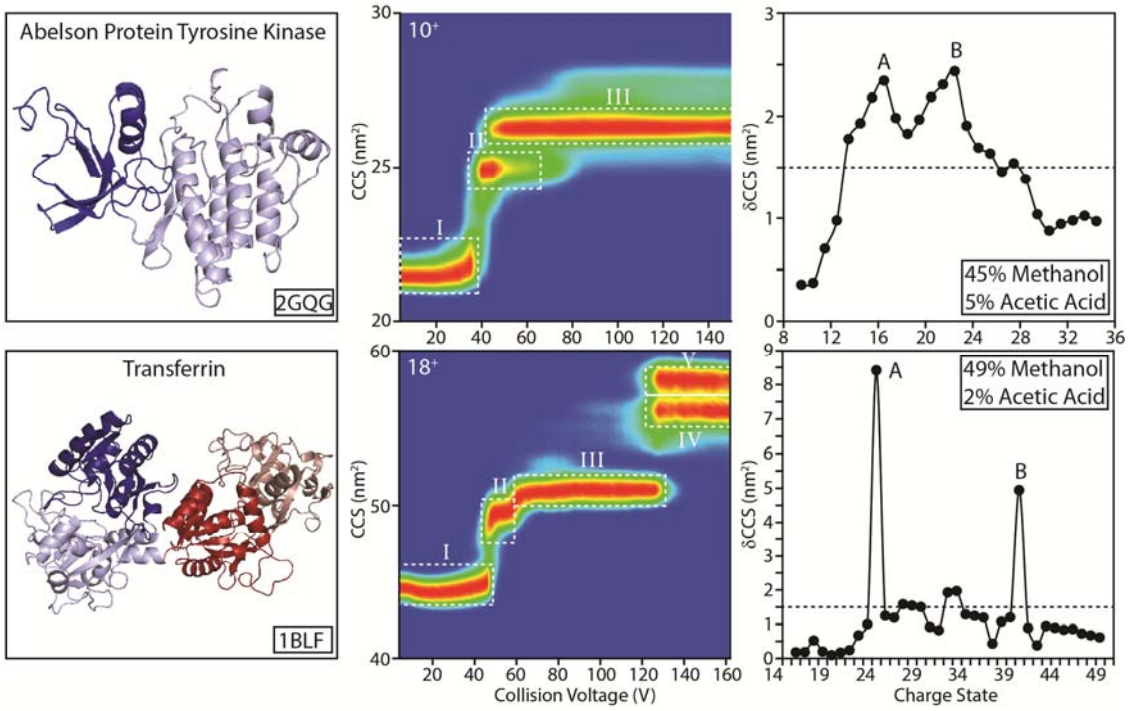

Figure S. 4-4 Crystal structures (left), collisional unfolding (middle) and Coulombic unfolding data (right) for SCOP/CATH inconsistent proteins, Abelson protein tyrosine kinase and Transferrin.

Abelson protein tyrosine kinase is assigned as a two domain protein by CATH (dark blue and light blue), however is considered a one domain protein in SCOP. Both collisional and Coulombic unfolding data indicate a 2-domain protein. We note, however, that the δCCS peaks recorded for this protein during Coulombic unfolding are not as distinct as those found for proteins with unambiguous domain assignments. Transferrin, considered to have 2 distinct large domains with 2 subdomains in each, generates a collisional unfolding pattern that agrees well with CATH (4-domain protein), but a Coulombic unfolding pattern agrees with SCOP (2-domain protein).

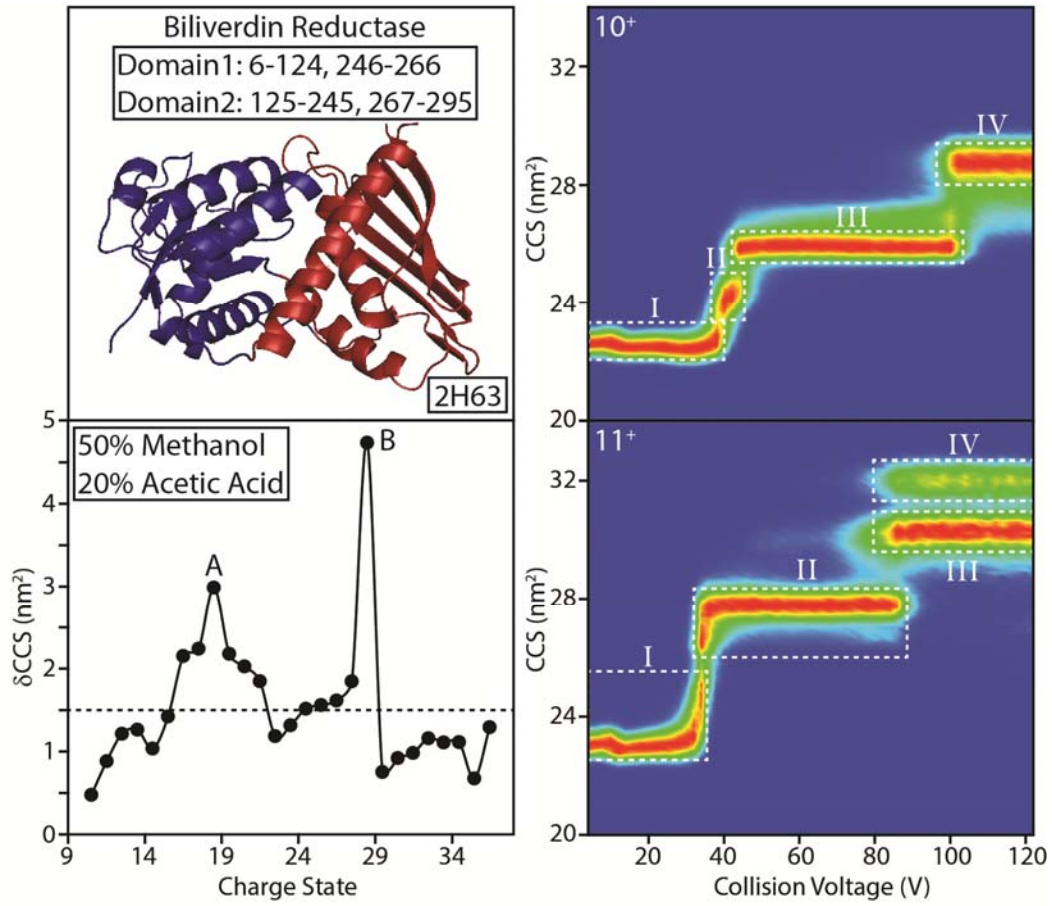

Figure S. 4-5 Crystal structure (top, left), Coulombic unfolding (bottom left) and CIU (right) data of Biliverdin Reductase (BVR).

Each domain of BVR consists of 2 sequence segments, according to CATH. However, these two domains are mixed, meaning that the part of the sequence for domain one overlaps spatially with domain two, and vice-versa. The Coulombic unfolding data indicate a two-domain unfolding pattern. Nevertheless, both 10^+ and 11^+ ions subjected to CIU identify this protein as a having 3-domains, likely due to the inter-mingled nature of the domains in this protein.

4.6 References

(1) Kuhner, S.; van Noort, V.; Betts, M. J.; Leo-Macias, A.; Batisse, C.; Rode, M.; Yamada, T.; Maier, T.; Bader, S.; Beltran-Alvarez, P.; Castano-Diez, D.; Chen, W. H.; Devos, D.; Guell, M.; Norambuena, T.; Racke, I.; Rybin, V.; Schmidt, A.; Yus, E.; Aebersold, R.; Herrmann, R.; Bottcher, B.; Frangakis, A. S.; Russell, R. B.; Serrano, L.; Bork, P.; Gavin, A. C. *Science* **2009**, *326*, 1235.

(2) Robinson, C. V.; Sali, A.; Baumeister, W. *Nature* **2007**, *450*, 973.

(3) Woodley, J. M. *Curr. Opin. Chem. Biol.* **2013**, *17*, 310.

(4) Weston, A. D.; Hood, L. *J. Proteome Res.* **2004**, *3*, 179.

(5) Gingras, A.-C.; Gstaiger, M.; Raught, B.; Aebersold, R. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 645.

(6) Petrotchenko, E. V.; Borchers, C. H. *Mass Spectrom. Rev.* **2010**, *29*, 862.

(7) Sinz, A. *Anal. Bioanal. Chem.* **2010**, *397*, 3433.

(8) Leitner, A.; Walzthoeni, T.; Kahraman, A.; Herzog, F.; Rinner, O.; Beck, M.; Aebersold, R. *Mol. Cell. Proteomics* **2010**, *9*, 1634.

(9) Konermann, L.; Pan, J.; Liu, Y.-H. *Chem. Soc. Rev.* **2011**, *40*, 1224.

(10) Kaltashov, I. A.; Bobst, C. E.; Abzalimov, R. R. *Anal. Chem.* **2009**, *81*, 7892.

(11) Engen, J. R. *Anal. Chem.* **2009**, *81*, 7870.

(12) Wales, T. E.; Engen, J. R. *Mass Spectrom. Rev.* **2006**, *25*, 158.

(13) Englander, S. W. *J. Am. Soc. Mass. Spectrom.* **2006**, *17*, 1481.

(14) Marcsisin, S. R.; Engen, J. R. *Anal. Bioanal. Chem.* **2010**, *397*, 967.

(15) Kiselar, J. G.; Chance, M. R. *J. Mass Spectrom.* **2010**, *45*, 1373.

(16) Konermann, L.; Stocks, B. B.; Pan, Y.; Tong, X. *Mass Spectrom. Rev.* **2010**, *29*, 651.

(17) Wen, J. Z.; Zhang, H.; Gross, M. L.; Blankenship, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6134.

(18) Zhong, Y.; Hyung, S.-J.; Ruotolo, B. T. *Expert Rev. Proteomics* **2012**, *9*, 47.

(19) Uetrecht, C.; Rose, R. J.; van Duijn, E.; Lorenzen, K.; Heck, A. J. R. *Chem. Soc. Rev.* **2010**, *39*, 1633.

(20) Bohrer, B. C.; Mererbloom, S. I.; Koeniger, S. L.; Hilderbrand, A. E.; Clemmer, D. E. In *Annu. Rev. Anal. Chem.* ; Annual Reviews: Palo Alto, 2008; Vol. 1, p 293.

(21) Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V. *Nat. Protoc.* **2008**, *3*, 1139.

(22) McLean, J. A.; Ruotolo, B. T.; Gillig, K. J.; Russell, D. H. *Int. J. Mass spectrom.* **2005**, *240*, 301.

(23) Clemmer, D. E.; Jarrold, M. F. *J. Mass Spectrom.* **1997**, *32*, 577.

(24) Whitelegge, J. *Expert Rev. Proteomics* **2013**, *10*, 127.

(25) Rose, R. J.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. R. *Nat. Methods* **2012**, *9*, 1084.

(26) Heck, A. J. R. *Nat. Methods* **2008**, *5*, 927.

(27) Steven, A. C.; Baumeister, W. *J. Struct. Biol.* **2008**, *163*, 186.

(28) Sali, A.; Glaeser, R.; Earnest, T.; Baumeister, W. *Nature* **2003**, *422*, 216.

(29) Hernandez, H.; Dziembowski, A.; Taverner, T.; Seraphin, B.; Robinson, C. V. *Embo Reports* **2006**, *7*, 605.
(30) Zhou, M.; Sandercock, A. M.; Fraser, C. S.; Ridlova, G.; Stephens, E.; Schenauer, M. R.; Yokoi-Fong, T.; Barsky, D.; Leary, J. A.; Hershey, J. W.; Doudna, J. A.; Robinson, C. V. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18139.
(31) Damoc, E.; Fraser, C. S.; Zhou, M.; Videler, H.; Mayeur, G. L.; Hershey, J. W. B.; Doudna, J. A.; Robinson, C. V.; Leary, J. A. *Mol. Cell. Proteomics* **2007**, *6*, 1135.
(32) Sharon, M.; Mao, H.; Boeri Erba, E.; Stephens, E.; Zheng, N.; Robinson, C. V. *Structure* **2009**, *17*, 31.
(33) Park, A. Y.; Jergic, S.; Politis, A.; Ruotolo, B. T.; Hirshberg, D.; Jessop, L. L.; Beck, J. L.; Barsky, D.; O'Donnell, M.; Dixon, N. E.; Robinson, C. V. *Structure* **2010**, *18*, 285.
(34) McInerney, P.; Johnson, A.; Katz, F.; O'Donnell, M. *Mol. Cell* **2007**, *27*, 527.
(35) Rabuck, J. N.; Hyung, S.-J.; Ko, K. S.; Fox, C. C.; Soellner, M. B.; Ruotolo, B. T. *Anal. Chem.* **2013**, *85*, 6995.
(36) Han, L.; Ruotolo, B. T. *Int. J. Ion Mobil. Spectrom.* **2013**, *16*, 41.
(37) Freeke, J.; Robinson, C. V.; Ruotolo, B. T. *Int. J. Mass spectrom.* **2010**, *298*, 91.
(38) van Duijn, E.; Heck, A. J. R. *Drug Discov. Today Technol.* **2006**, *3*, 21.
(39) van Duijn, E.; Simmons, D. A.; van den Heuvel, R. H. H.; Bakkes, P. J.; van Heerikhuizen, H.; Heeren, R. M. A.; Robinson, C. V.; van der Vies, S. M.; Heck, A. J. R. *J. Am. Chem. Soc.* **2006**, *128*, 4694.
(40) Zhou, M.; Morgner, N.; Barrera, N. P.; Politis, A.; Isaacson, S. C.; Matak-Vinkovic, D.; Murata, T.; Bernal, R. A.; Stock, D.; Robinson, C. V. *Science* **2011**, *334*, 380.
(41) Jeronimo, C.; Forget, D.; Bouchard, A.; Li, Q.; Chua, G.; Poitras, C.; Thérien, C.; Bergeron, D.; Bourassa, S.; Greenblatt, J.; Chabot, B.; Poirier, G. G.; Hughes, T. R.; Blanchette, M.; Price, D. H.; Coulombe, B. *Mol. Cell* **2007**, *27*, 262.
(42) Fitzgerald, M. C.; West, G. M. *J. Am. Soc. Mass. Spectrom.* **2009**, *20*, 1193.
(43) Chen, L.; Shao, Q.; Gao, Y.-Q.; Russell, D. H. *J. Phys. Chem. A* **2011**, *115*, 4427.
(44) Pierson, N. A.; Chen, L. X.; Valentine, S. J.; Russell, D. H.; Clemmer, D. E. *J. Am. Chem. Soc.* **2011**, *133*, 13810.
(45) Fernandez-Lima, F. A.; Wei, H.; Gao, Y. Q.; Russell, D. H. *J. Phys. Chem. A* **2009**, *113*, 8221.
(46) Benesch, J.; Ruotolo, B. *Curr. Opin. Struct. Biol.* **2011**, *21*, 641.
(47) Shelimov, K. B.; Jarrold, M. F. *J. Am. Chem. Soc.* **1996**, *118*, 10313.
(48) Shelimov, K. B.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997**, *119*, 2987.
(49) Fye, J. L.; Woenckhaus, J.; Jarrold, M. F. *J. Am. Chem. Soc.* **1998**, *120*, 1327.
(50) Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1995**, *117*, 10141.
(51) Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997**, *119*, 2240.
(52) Segev, E.; Wyttenbach, T.; Bowers, M. T.; Gerber, R. B. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3077.
(53) Wyttenbach, T.; Bowers, M. T. *J. Phys. Chem. B* **2011**, *115*, 12266.

(54) Shi, H.; Gu, L.; Clemmer, D. E.; Robinson, R. A. S. *J. Phys. Chem. B* **2012**, *117*, 164.
(55) Bernstein, S. L.; Liu, D. F.; Wyttenbach, T.; Bowers, M. T.; Lee, J. C.; Gray, H. B.; Winkler, J. R. *J. Am. Soc. Mass. Spectrom.* **2004**, *15*, 1435.
(56) Li, J. W.; Taraszka, J. A.; Counterman, A. E.; Clemmer, D. E. *Int. J. Mass spectrom.* **1999**, *185*, 37.
(57) Ruotolo, B. T.; Hyung, S. J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V. *Angew. Chem. Int. Ed.* **2007**, *46*, 8001.
(58) Myung, S.; Badman, E. R.; Lee, Y. J.; Clemmer, D. E. *J. Phys. Chem. A* **2002**, *106*, 9976.
(59) Badman, E. R.; Myung, S.; Clemmer, D. E. *J. Am. Soc. Mass. Spectrom.* **2005**, *16*, 1493.
(60) Felitsyn, N.; Kitova, E. N.; Klassen, J. S. *Anal. Chem.* **2001**, *73*, 4647.
(61) Jurchen, J. C.; Williams, E. R. *J. Am. Chem. Soc.* **2003**, *125*, 2817.
(62) Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. *Science* **2005**, *310*, 1658.
(63) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. *J. Mass Spectrom.* **2000**, *35*, 1399.
(64) Hyung, S. J.; Ruotolo, B. T. *Proteomics* **2012**, *12*, 1547.
(65) Hall, Z.; Politis, A.; Bush, M. F.; Smith, L. J.; Robinson, C. V. *J. Am. Chem. Soc.* **2012**, *134*, 3429.
(66) Pagel, K.; Hyung, S. J.; Ruotolo, B. T.; Robinson, C. V. *Anal. Chem.* **2010**, *82*, 5363.
(67) Zhou, M. W.; Huang, C. S.; Wysocki, V. H. *Anal. Chem.* **2012**, *84*, 6016.
(68) Han, L.; Ruotolo, B. *Angew. Chem. Int. Ed.* **2013**, *52*, 8329.
(69) Freeke, J.; Bush, M. F.; Robinson, C. V.; Ruotolo, B. T. *Chem. Phys. Lett.* **2012**, *524*, 1.
(70) Han, L. J.; Hyung, S. J.; Mayers, J. J. S.; Ruotolo, B. T. *J. Am. Chem. Soc.* **2011**, *133*, 11358.
(71) Niu, S.; Rabuck, J. N.; Ruotolo, B. T. *Curr. Opin. Chem. Biol.* **2013**, *17*, 809.
(72) Hopper, J. T. S.; Oldham, N. J. *J. Am. Soc. Mass. Spectrom.* **2009**, *20*, 1851.
(73) Hyung, S.-J.; Robinson, C. V.; Ruotolo, B. T. *Chem. Biol.* **2009**, *16*, 382.
(74) Smith, D.; Johnson, K. *Gene* **1988**, *67*, 31.
(75) Giles, K.; Williams, J. P.; Campuzano, I. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 1559.
(76) Kanu, A. B.; Dwivedi, P.; Tam, M.; Matz, L.; Hill, H. H. *J. Mass Spectrom.* **2008**, *43*, 1.
(77) Creaser, C. S.; Griffiths, J. R.; Bramwell, C. J.; Noreen, S.; Hill, C. A.; Thomas, C. L. P. *Analyst* **2004**, *129*, 984.
(78) Baumbach, J. I.; Eiceman, G. A. *Appl. Spectrosc.* **1999**, *53*, 338A.
(79) Zhong, Y.; Hyung, S.-J.; Ruotolo, B. T. *Analyst* **2011**, *136*, 3534.
(80) Murzin, A.; Brenner, S.; Hubbard, T.; Chothia, C. *J. Mol. Biol.* **1995**, *247*, 536.
(81) Orengo, C.; Michie, A.; Jones, S.; Jones, D.; Swindells, M.; Thornton, J. *Structure* **1997**, *5*, 1093.
(82) Day, R.; Beck, D.; Armen, R.; Daggett, V. *Protein Sci.* **2003**, *12*, 2150.
(83) Hadley, C.; Jones, D. *Structure* **1999**, *7*, 1099.

(84) Csaba, G.; Birzele, F.; Zimmer, R. *BMC Struct. Biol.* **2009**, *9*, 23.
(85) Breuker, K.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18145.
(86) Koeniger, S. L.; Clemmer, D. E. *J. Am. Soc. Mass. Spectrom.* **2007**, *18*, 322.
(87) Koeniger, S. L.; Merenbloom, S. I.; Clemmer, D. E. *J. Phys. Chem. B* **2006**, *110*, 7017.
(88) Wyttenbach, T.; Bowers, M. T. *J. Phys. Chem. B* **2011**, *115*, 12266.
(89) Wright, P. J.; Zhang, J.; Douglas, D. J. *J. Am. Soc. Mass. Spectrom.* **2008**, *19*, 1906.
(90) Skinner, O.; McLafferty, F.; Breuker, K. *J. Am. Soc. Mass. Spectrom.* **2012**, *23*, 1011.
(91) Breuker, K.; Oh, H.; Horn, D. M.; Cerda, B. A.; McLafferty, F. W. *J. Am. Chem. Soc.* **2002**, *124*, 6407.
(92) Purves, R. W.; Barnett, D. A.; Ells, B.; Guevremont, R. *J. Am. Soc. Mass. Spectrom.* **2000**, *11*, 738.
(93) Hospenthal, M.; Freund, S.; Komander, D. *Nat. Struct. Mol. Biol.* **2013**, *20*, 555.
(94) Datta, A.; Hura, G.; Wolberger, C. *J. Mol. Biol.* **2009**, *392*, 1117.
(95) de la Mora, J. F. *Anal. Chim. Acta* **2000**, *406*, 93.
(96) Heck, A. J. R.; van den Heuvel, R. H. H. *Mass Spectrom. Rev.* **2004**, *23*, 368.
(97) Kaltashov, I.; Mohimen, A. *Anal. Chem.* **2005**, *77*, 5370.
(98) Röthlisberger, D.; Honegger, A.; Plückthun, A. *J. Mol. Biol.* **2005**, *347*, 773.
(99) Carrion-Vazquez, M.; Oberhauser, A. F.; Fowler, S. B.; Marszalek, P. E.; Broedel, S. E.; Clarke, J.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*.
(100) Rounsevell, R.; Forman, J.; Clarke, J. *Methods (San Diego, Calif.)* **2004**, *34*, 100.
(101) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J.; Gaub, H. *Science (New York, N.Y.)* **1997**, *276*, 1109.
(102) Strucksberg, K.; Rosenkranz, T.; Fitter, J. *Biochim. Biophys. Acta* **2007**, *1774*, 1591.
(103) Ahmad, B.; Ahmed, M.; Haq, S.; Khan, R. *Biochim. Biophys. Acta* **2005**, *1750*, 93.
(104) Benesch, J. *J. Am. Soc. Mass. Spectrom.* **2009**, *20*, 341.
(105) Jai Kartik, V.; Lavanya, T.; Guruprasad, K. *Int. J. Biol. Macromol.* **2006**, *38*, 174.
(106) Gorman, J.; Wallis, T.; Pitt, J. *Mass Spectrom. Rev.* **2002**, *21*, 183.
(107) Rousseau, F.; Schymkowitz, J.; Itzhaki, L. *Structure (London, England : 1993)* **2003**, *11*, 243.
(108) Byeon, I.-J. L.; Louis, J.; Gronenborn, A. *J. Mol. Biol.* **2004**, *340*, 615.
(109) Liu, Y.; Eisenberg, D. *Protein Sci.* **2002**, *11*, 1285.
(110) Fraczkiewicz, R.; Braun, W. *J. Comput. Chem.* **1998**, *19*, 319.

Chapter 5. Applying IM-MS to Protein Complexes with Unknown Topology and Interface Structure

5.1 Introduction

To validate and extend the methods discussed in chapters 2-4 we have selected two multi-protein systems with unknown structure to study in collaboration with researchers at the University of Michigan: the hetero-dimer of heme oxygenase-2 (HO-2) and NADPH-cytochrome P450 reductase (CPR) in collaboration with the Ragsdale laboratory (Biological Chemistry Department), and a subcomplex from human mitochondrial RNase P (MRPP) in collaboration with the Fierke laboratory (Chemistry Department). Both of these systems have partial high-resolution structure data available, for instance the homotetramer of MRPP2[1,2] and the truncated version of HO-2[3] and CPR[4,5] individually, but lack full sequence structural data that elucidates the arrangement of subunits within the larger complexes formed. Furthermore, the functional forms of both complexes are in the context of the complexes they create, rather than in their monomeric forms. As such, the structures of the intact complexes generated in both of these systems are essential for understanding the function of these complexes. Here, we will apply the IM-MS methods discussed throughout this thesis in an effort elucidate such structures.

5.2 Studying the Heme Oxygenase-2 (HO-2) and NADPH-Cytochrome P450 Reductase (CPR) interaction with IM-MS

5.2.1 Background Information about Heme Degradation

Heme oxygenase (HO) is a membrane-bound microsomal enzyme that catalyzes the degradation of heme to biliverdin, carbon monoxide (CO), and free iron.[6,7] Biliverdin is subsequently converted to bilirubin by a soluble cytosolic enzyme, biliverdin reductase (BVR).[8] In this reaction, HO requires three molecules of oxygen and seven electrons supplied by NADPH-dependent cytochrome P450 reductase (CPR).[6] CPR is a 78 kDa membrane-bound flavoprotein containing one FAD and FMN each.[9] For electron transfer, CPR and its redox partners need to associate with each other.[7,10]

Two principal isoforms of HO have been identified in mammalian cells, HO-1 and HO-2.[11] Both HO-1 and HO-2 exhibit similar catalytic properties; however, they are expressed at different locations in the body and have different modes of regulation. While HO-1 is induced under conditions of heat shock and oxidative stress conditions, and is present in most tissues, HO-2 is constitutively expressed in the brain, testes and carotid bodies.[12] At the sequence level, HO-1 and HO-2 are highly homologous (55% identity, 76% similarity).[13,14] A significant difference in the sequence lies between residues 240-295 (with respect to the HO-2 sequence). Thus far, HO-1 has been the more extensively studied of the two proteins, both structurally and biochemically. This dataset includes X-ray structures of both the rat and human HO-1 in the heme-free[15,16], heme-bound[17,18], verdoheme-bound[19], biliverdin-iron chelate-bound[19], and biliverdin-bound forms[20]. The

crystal structures of rat CPR and rat BVR have been also reported.[4,5,21,22] HO-2 lacks much of the structural data currently available, with high-resolution data available only for truncated human forms of the protein available.[3] The structures of HO-2 complexed to either CPR or BVR are completely unknown.

The full length HO-2 contains 296 amino acids, with only 29-242 presented in the X-ray structure (PDB code: 2QPP), as the C-terminus is relatively unstructured after residue 248. However, this C-terminal region is reported to contain two short amino acid stretches comprising a Cys-Pro dipeptide referred to as the heme regulatory motif (HRM),[23] and is suspected to interact with CPR. Because of the obvious importance of the C-terminus of HO-2 in the context of heme regulation, we chose to work with a near-full length HO-2 in the studies reported here, including residues 1-288. The last 8 amino acids are not included because their inclusion adversely affects the solubility of the protein due to their role in membrane association. Truncated forms of HO-2, including 28-248 and 1-248 are also included in our studies in order to better evaluate the role of the C-terminus in CPR interactions. Also for solubility issues, the CPR construct studied here contains amino acids 69-680, removing the membrane associated N-terminus. Previous studies have shown that HO activity can be reconstructed with the soluble forms of HO and CPR.[24] Below, we employ the approach developed in Chapter 3 of this thesis, automated titration coupled with IM-MS analysis, to quantify and evaluate the nature of the interactions between HO-2 and CPR.

5.2.2 Methods and Materials

Three constructs of HO-2 (1-288, 1-248 and 28-248) and CPR were provided by the Ragsdale lab from the University of Michigan. The proteins exist primarily as monomer when sprayed separately. Heme-containing HO-2 was mixed with CPR in 100mM ammonium acetate (NH_4Ac). The ratio between HO-2 and CPR was adjusted to optimize complex formation, eventually resulting in a 4:1 ratio at 20 μM HO-2 and 5 μM CPR for our experiments. The proteins were then titrated using a combination of DMSO and NH_4Ac, followed by nESI using the nanomate Triversa (Advion, Ithaca NY, USA) for Q-IM-MS analysis on a Synapt G2 (Waters, Milford MA, USA). For detailed accounting of the experimental methods used, see the experimental methods section of Chapter 3 of this thesis.

Structure predictions for the HO-2 C-terminus were performed by the I-TASSER server.[25] This online server allows for the high-quality prediction of 3D protein structure and biological function from amino acid sequence alone. After submitting an amino acid sequence, the server first tries to retrieve template proteins of similar folds from the PDB. Users can also manually submit PDB templates as restraints. After this, continuous fragments excised from the PDB templates are reassembled into full-length models by replica-exchange Monte Carlo simulations and by threading unaligned regions (mainly loops) built by *ab initio* modeling approaches. In cases where no appropriate template is identified from PDB inputs, I-TASSER builds the entire structure using *ab initio* type modeling. As a last step, a simulation is performed comparing the assembled protein and PDB structures in order to refine the global topology. The models generated in the second

simulations are then clustered and the lowest energy structures with optimized potential are selected.

Simulated complex models shown here are generated by the ZDOCK server.[26] ZDOCK searches all possible binding modes in the translational and rotational space between two target proteins. The exhaustiveness of this search is then evaluated using an energy-based scoring function for each contacting model discovered. The scoring function contains terms for the statistical potential, shape complementarity, and electrostatics of the protein complex. Further filtration of models in our experiments was performed manually by examining the heme-orientation, C-terminal contacts in the complex, and collision cross sections (CCS) of the models generated.

5.2.3 Results and Discussion

After mixing the heme-bound HO-2 1-288 with CPR in 100 mM NH_4Ac, a complex between HO-2 and CPR is detected, having a stoichiometry of 1:1. Unlike the protein complexes studied in Chapter 3, HO and CPR are relatively weakly associated, as indicated by the relatively low signal intensity of the dimer relative to total ion current detected in the experiment, with about 18% of our recorded signal resulting from the intact complex. To study the interaction between HO-2 1-288 and CPR, we applied our 2D titration approach, using NH_4Ac and DMSO, to investigate the ability of these two disrupting agents to dissociate the HO-2/CPR dimer formed (Figure 5-1). As shown in Figure 5-1B, the intact complex signal intensity changes significantly as a function of NH_4Ac concentration in solution, indicating the presence of some number of essential

polar contacts that stabilize the assembly. Unlike data shown in Chapter 3, where increasing amounts of DMSO in solution led to with increasing amounts of protein complex disruption, this assembly exhibits a tighter interaction with increasing DMSO up to 40% (by volume). This stabilized complex, however, seems to completely dissociate at DMSO contents above 40%. A close inspection of the IM-MS data reveals that HO-2 unfolds at solution conditions of 40% DMSO (Figure 5-1C) when compared to those with no DMSO added (Figure 5-1A), as indicated by the broadened IM features and higher charge states of the HO-2 monomers detected. This unfolded structure may contribute to the apparently-tightened interaction observed in our dataset. Because of the vulnerability of HO-2 to DMSO based unfolding, hydrophobic interaction between HO-2 and CPR could not been examined through our titration techniques. On the other hand, our NH4Ac based titration data collected at low DMSO concentrations reveals no evidence of unfolding, and the following discussion will focus on our attempts to mine such data to quantify the polar interactions between HO-2 and CPR.

Following our Chapter 3 protocol, complex disruption as a function of NH_4Ac concentration at a fixed DMSO content of 10% (indicated by white dashed line in Figure 5-1B) can be effectively utilized to estimate the number of polar contacts that stabilize protein-protein complexes. As with our previous experiments, this amount of DMSO does not tend to unfold the proteins analyzed, but may disrupt local hydrophobic contacts enough to allow efficient disruption of buried salt bridges, focusing on polar contact only.

Based on our previously-established relationship between NH_4Ac and polar contacts found in X-ray structure of known complexes, derived in Chapter 3, a calibration curve

was constructed (Figure 5-2). By noting the NH_4Ac concentration at which the HO-2/CPR dimer signal drops by 10% relative to its intensity under native conditions (150 mM) and by applying this experimental value to the calibration equation derived in Figure 5-2, the number of polar contacts, including salt-bridges and hydrogen-bonds, between HO-2 1-288 and CPR can be estimated to be 36, with a 95% confidence interval between 13 and 59 contacts.

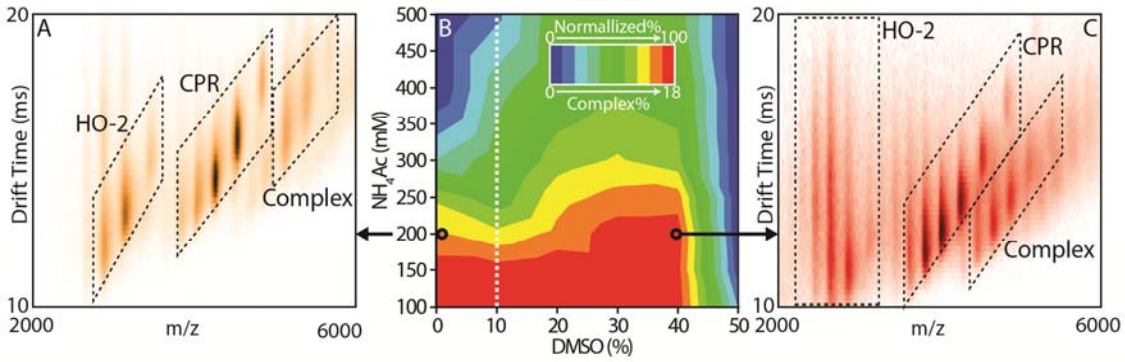

Figure 5-1 2D titration experiments on the HO-2+CPR complex.
(A) IM-MS data collected at 200 mM NH_4Ac without adding DMSO. **(B)** Percent of the signal corresponding to the intact HO-2+CPR complex plotted against DMSO and NH_4Ac percent and concentration respectively, added in solution prior to nESI. **(C)** IM-MS data collected at 200 mM NH_4Ac and 40% DMSO. Broadened features observed at low m/z and high IM drift times indicate the presence of unfolded protein.

Further data were collected on the truncated HO-2 constructs 1-248 and 28-248, mixed with the same CPR protein, in 100 mM NH_4Ac buffer. Both of these truncated HO-2 forms generated 1:1 complexes with CPR, but with significantly lower signal intensities when compared directly to HO-2 1-288 (~5% signal intensity of the complexes recorded, 13% less signal than observed with the longer HO-2 construct). A similar NH_4Ac titration at fixed 10% DMSO as discussed above was performed for the HO-2 1-248/CPR complex and compared with our data for the longer HO-2 construct (Figure 5-3). The

significantly smaller starting ion intensity for the complexes comprised of truncated HO-2, strongly indicates that the C-terminus of HO-2 is important for CPR binding. The N-terminal truncated mutant, HO-2 28-248, forms similar amount of complexes (5%) with HO-2 1-248. However, this double-end truncated construct unfolds significantly at 10% DMSO, making similar IM-MS titration data as acquired for other HO-2 1-288/CPR complexes difficult to obtain. This observation, however, also indicates that the N-terminus may be critically important in HO-2 protein stability overall, and is likely not critical for CPR binding.

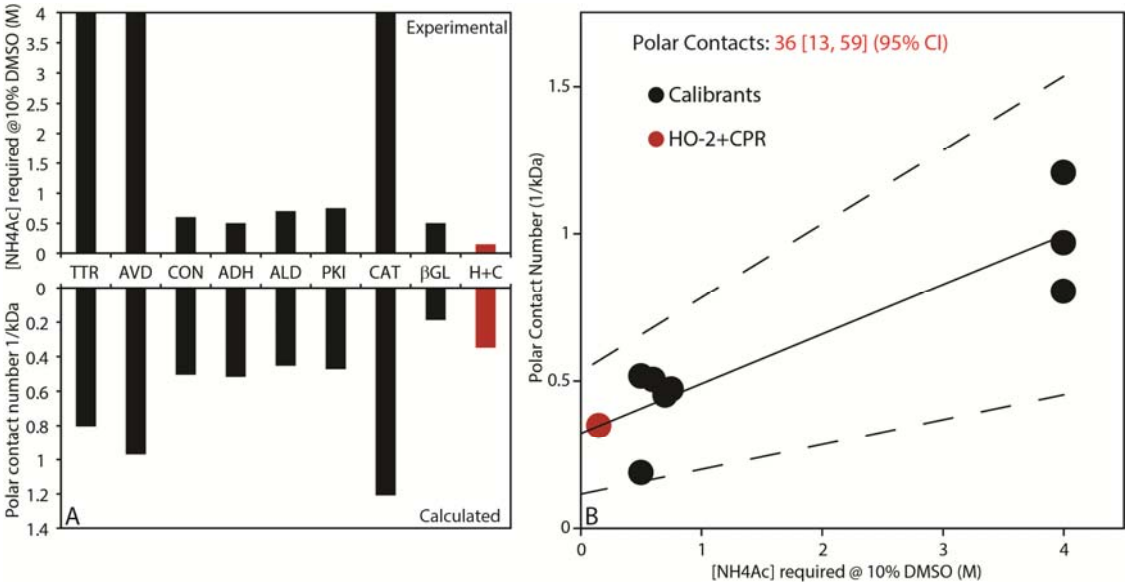

Figure 5-2 NH_4Ac required for dissociation at 10% DMSO versus the number of polar contacts for both calibrants (black) and HO-2 1-288/CPR complex.
(A) Two histogram plots generated for both calibrants and HO-2+CPR, with experimental data on top, and calculated data (either from X-ray or from calibration curve) on bottom. **(B)** Calibration curve for the protein complex disruption experiment, showing both calibrant points (black) and the unknown HO-2/CPR complex. The solid line represents the fitted linear trend and dashed lines report on the upper and lower 95% confidence interval (CI). The fitted linear equation for this experimental and calculated correlation is y=0.168x + 0.324 with an R^2=0.802.

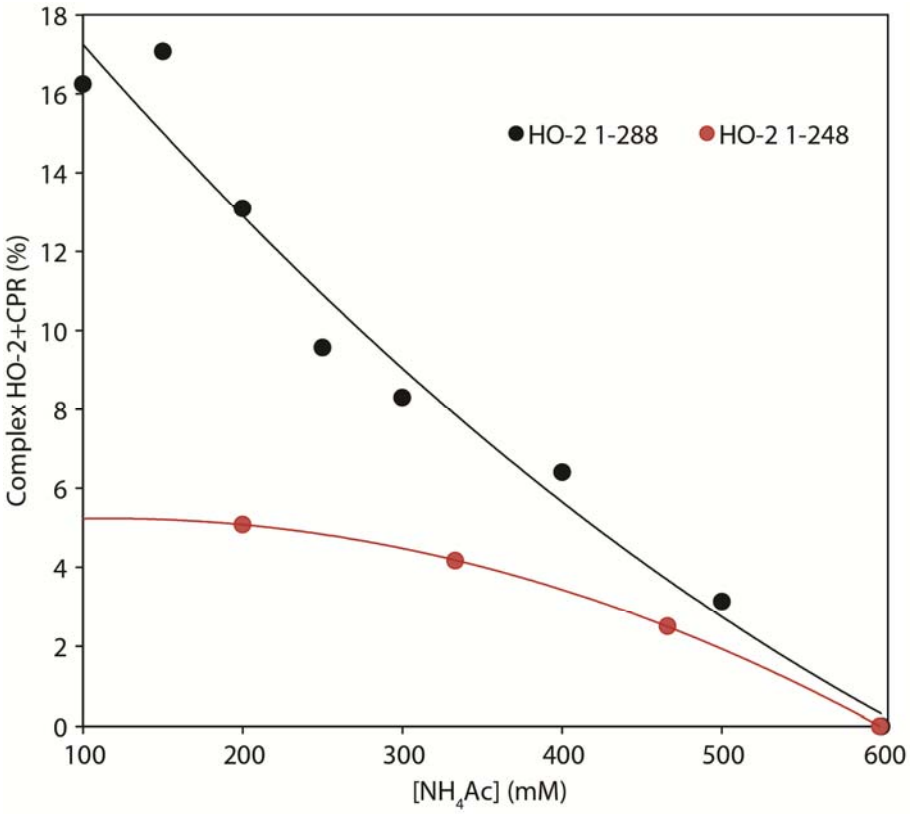

Figure 5-3 Titration experiment increasing the NH_4Ac concentration at fixed DMSO content in solution for HO-2 1-288/CPR (black) and HO-2 1-248/CPR complexes (red).

As the C-terminus, rather than N-terminus, is identified in our data as for a key element in HO-2/CPR complex formation, 3D structure prediction using I-TASSER was conducted to provide a model for residues 243-288, and dock this element to the crystal structure (residues 29-242). While the model generated for the C-terminus is docked in a range of disparate orientations and configurations in our top-scoring models (Figure 5-4A), possibly due to the relatively unstructured and flexible nature of this section of the protein, the average model generated from these top-scoring candidates agrees well with previous biophysical data from the Ragsdale group[23] (Figure 5-4B, red).

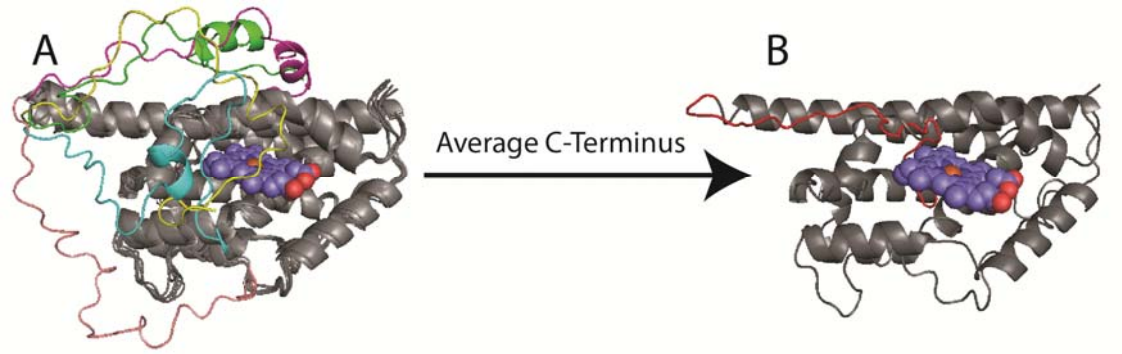

Figure 5-4 HO-2 structures with predicted C-terminal (243-288).
(A) The top 5 models predicted by I-TASSER aligned together, with color coded C-termini from different models and a grey-colored core HO-2 (29-242). Heme is presented with blue and red spheres. **(B)** The average model generated by I-TASSER, with the C-terminus (red) located in the center of the model.

This averaged HO-2 29-288 model is then iterativlely docked onto CPR using a ZDOCK simulation. Figure 5-5 illustrates the docking search process for an optimzed complex model. After exhausting the binding modes between HO-2 29-288 (light green) and CPR (black) (Figure 5-5A), the ZDOCK ranks the complex models based on the ZDOCK scoring function, and filtered by our experimental constraints, which include heme orientation (physical proximity between heme and CPR is necessary for electron and oxygen transfer), HO-2 C-terminal contact with CPR, and CCS values for the intact complex. This final constraint has, in this case, the weakest ability to filter models when compared to the other physical constraints used here, because the CCS (an oreintationally averaged physical size for the complex) provides information only on the center-to-center distance of the interacting proteins and not their relative orientiation. The combination of all restraints leaves us three potential models of similar likelihood and binding orientation (Figure 5-5B).

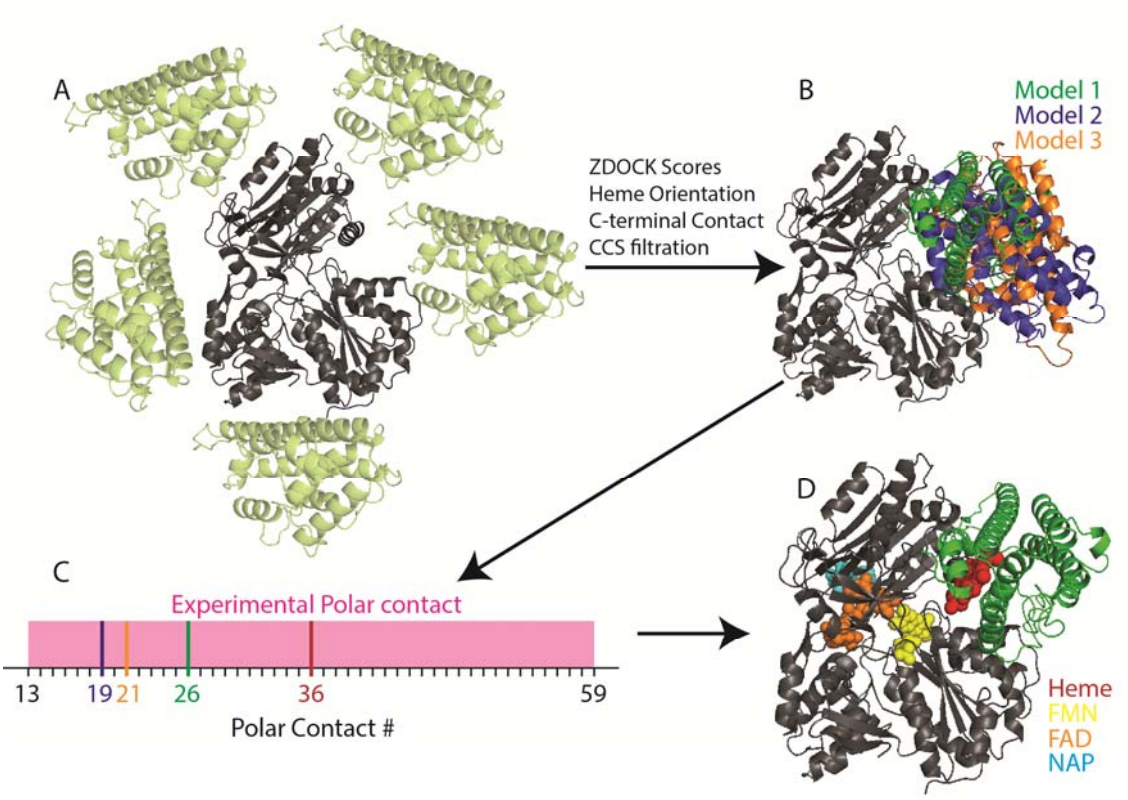

Figure 5-5 Docking simulation for optimized HO-2/CPR complex model using ZDOCK.
(A) HO-2 (light green) docking searches for all possible binding modes with CPR (black). **(B)** The binding modes are evaluated by the ZDOCK energy scoring function and experimental constraints, save IM-MS titration data**. (C)** The number of polar contacts for high-scoring models compared against experimental predictions (pink region). **(D)** The single complex model with the number of polar contacts in closest agreement to experimental prediction.

These optimized models are then filtered against the number of polar contacts present in the HO-2/CPR interface (Figure 5-5C). Note that the number of polar contacts found for all three models are within the 95% confidence interval predicted experimentally, but are in lower number region. This result is anticipated, as our current protein-protein docking protocol does not allow for flexible movement within individual proteins. Model 1 (Figure 5.5D, green model) provides an HO-2/CPR model with polar contacts in closest agreement with our IM-MS titration data, having 26 total polar contacts.

5.3 Topology Construction for Constructing an IM-MS-derived Topology for the Human Mitochondrial Ribonuclease P (RNase P) subcomplex

5.3.1 Background Information on Human Mitochondrial RNase P

Transfer RNA (tRNA) is a central adapter molecule during ribosome-dependent protein synthesis.[27] Biosynthesis of mature functional tRNA requires extensive processing.[27] The early steps for processing immature precursor tRNA (pre-tRNA) include the removal of extraneous sequences at the 5' and 3' termini by ribonucleases.[28] The 5' end maturation of tRNA is catalyzed by ribonuclease P (RNase P), a divalent metal ion-dependent endonuclease responsible for cleavage of the 5' leader from pre-tRNA.[29,30] In the bacterial, archaeal, and the eukaryotic nucleus, the endonuclease ribonuclease P (RNase P) is composed of a catalytic RNA that is assisted by protein subunits.[31-34] Recently, human mitochondrial RNase P protein has gained attention as the catalytic activity of this enzyme derives entirely from proteins.[35,36] This mitochondrial enzyme contains no RNA and consists only of three types of protein subunits mitochondrial RNase P proteins 1, 2, and 3 (MRPP1–3).[37,38] None of the protein-only RNase P shares significant sequence homology with previous identified protein subunits in RNA-based RNase P.[38] Such a dramatic variety in subunit composition makes RNase P, one of the most diverse enzymes in biochemistry, an exciting model system to directly compare RNA and protein based catalysis.[38] Mutations in mtRNase P subunits and mitochondrial transfer RNAs (mt-tRNAs) have been implicated in a number of human diseases.[1,2] Moreover, the majority of mt-tRNAs lack consensus sequences conserved for secondary and tertiary

structures in classic tRNA.[39] Both human mtRNase P and mt-tRNAs are poorly characterized.

Within the human mitochondrial Rnase P protein complex, MRPP2 has been crystallized and charaterized as a homotetramer.[1,2] A homologous structure of MRPP3 has been resolved recently.[40] Due to difficulties inheret in its expression and purification, MRPP1 does not have an experiementally derived high-resolution structure. In addition, the stoichiometry of the subunits in the active MRPP 1:2:3 complex has not yet been established. Here, we describe the application of IM-MS, and the accurate CCS measurment approaches described in Chapter 2 of this thesis, to elucidate the structure and stoichiometry of the MRPP1:2 complex.

5.3.2 Methods and Materials

MRPP1 and 2 proteins were provided by the Fierke lab at the University of Michigan. Protein samples containing MRPP1 and 2 were ionized from 100 mM NH_4Ac buffer solution by nESI. IM-MS data analysis converts the drifttime of each protein into CCS values using calibration curve. For details on the optimal instrument conditions for recording both unknown drift times and calibration curve data, refer to Chapter 2 of this thesis.

CCS values for the subcomplexes and intact assemblies are then utilized as spatial restraints to build a complete 3D topology. CCS constraints are applied through the Integrative Modeling Platform (IMP), a python and C++ based toolbox for complex modeling, with each protein within the assembly represented as a course-grained

sphere.[41] For topology optimization, candidate models are generated by an exhaustive search all of the possible arragments of subunits given connectivity constraints defined by MS experiments. These models are then scored based on CCS, protein density, and symmetry based restraints.[41,42]

5.3.3 Results and Discussions

In addition to the homotetramer of MRPP2, both heterohexamers and heteropentamers comprised of MRPP1 andMRPP2 are observed by IM-MS when the proteins are mixed under native conditions and subjected to nESI (Figure 5-6A). Tandem MS using CID was carried on each population of subcomplexes observed to confirm their composition and stoichiometry (Figure 5-6B and C). This step is essential to precisly assign the intact mass of large protein complexes, as the primary MS data typically contain large amounts of buffer and adducts that skew mas measurments to higher values than expected from the sequences of individual subunits.[43] The stoichiometry of the largest complex between MRPP1 and MRPP2 was determined to be 2:4 from the data collected and summarized in Figure 5-6.

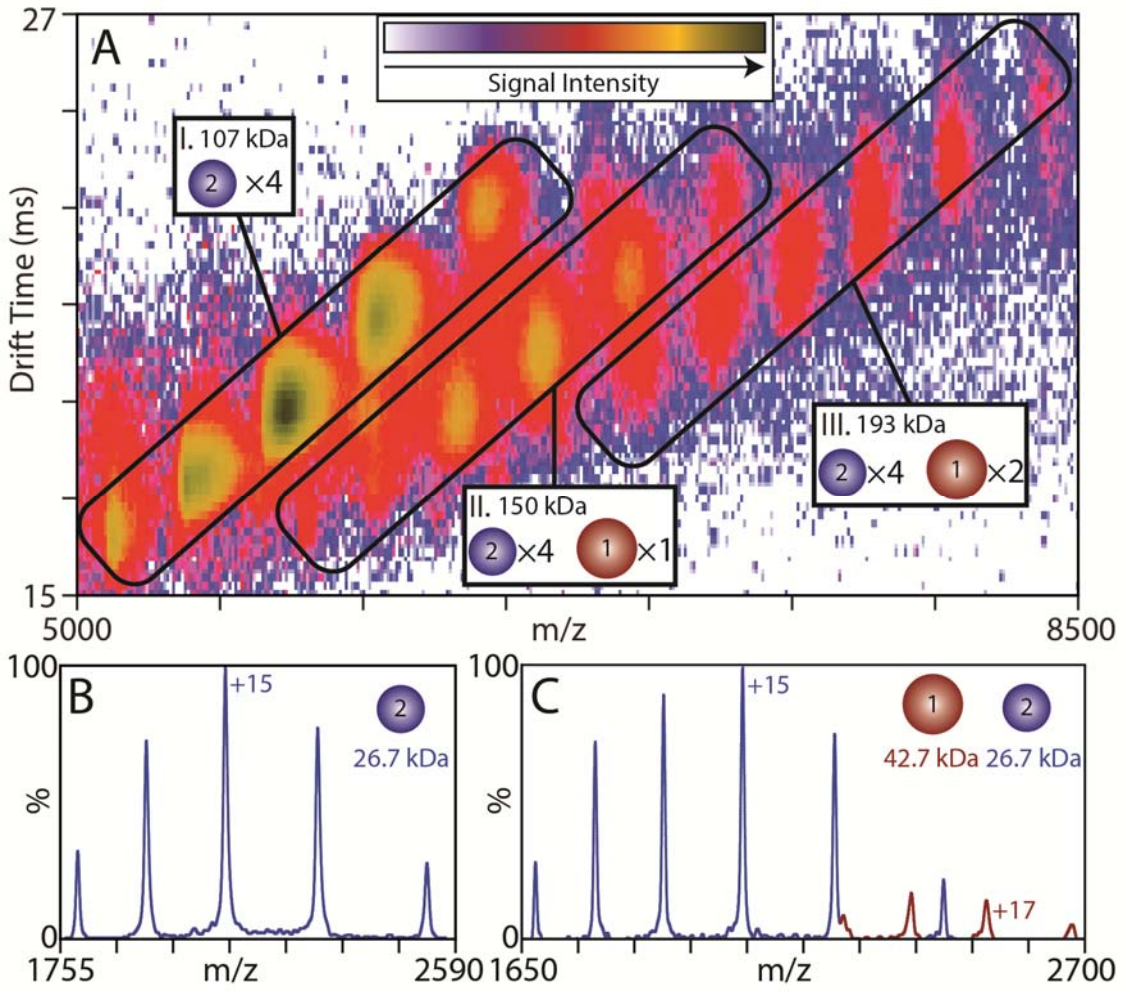

Figure 5-6 Ion mobility-mass spectrometry (IM-MS) analysis of MRPP1·MRPP2 complexes.
(A) Three protein complex stoichiometries are observed, I: the homotetramer of MRPP2; II: a heteropentamer comprised of $(MRPP1)_1(MRPP2)_4$; III: a heterohexamer containing $(MRPP1)_2(MRPP2)_4$; **(B)** Tandem MS analysis for the MRPP2 homotetramer, showing only the monomer region the mass spectrum. **(C)** Tandem MS analysis of the heteropentamer (II) and heterohexamer (III), generating similar tandem MS data in the monomer region, confirming the presence of both proteins in the selected complexes.

CCS values for the MRPP2 homotetramer are in good agreement of the expected CCS values for the complex (PDB code: 2O23). Since its structure is known, the MRPP2 tetramer is set as a fixed core topology modeling shown here. In order to assign

connectivity to the MRPP1:2 complexes observed, detailed MS and IM-MS experiments were carried out. If the MRPP2 tetramer is analyzed in isolation by intact MS, small amounts of both monomer and dimer ion signal can be observed under native conditions (pH 7), strongly indicating a dimer of dimers arrangement. In contrast, isolated MRPP1 does not form any oligomer observable by intact MS. As both heteropentamer and heterohexamer are observed by MS from solutions containing both MPRR1 and MRPP2, it is inferred that, in the heterohexamer, 2 MRPP1 bound to the MRPP2 homotetramer individually. Because of this, in the models generated here, contact between the two copies of MRPP1 is not prohibited, but this connection is also not input into the simulation as a constraint (Figure 5-7A). Symmetry is incorporated into our model as a second class of restraint, by setting the distance between each MRPP1 and the center of MRPP2 homotetramer to equal values. Using such a symmetry requirement for model generation, the top 500 models are selected. These models are then further filtered according to their CCS, with those models within 3% of experimental values collected and clustered for analysis (Figure 5-7B). Although small variations exist within each group, Figure 5-7C illustrates three groups of topologies that agree with all the spatial restraints described above. Of the three classes of structures shown in Figure 5-7, those that belong to the 'Boat II' cluster are most-likely to represent the native structure of the complex. Both the 'Planer' and 'Boat I' clusters have one MRPP2 contact per copy of MRPP2. Based on simple symmetry arguments, these complexes should be able to form homo-octameric complexes, for which we find no experimental evidence. In contrast, those structures that cluster well with the 'Boat II' structure are characterized by two

MRPP2 contacts per copy of MRPP1 in the complex, which predicts a hetero-hexameric assembly structure based on the arguments detailed above.

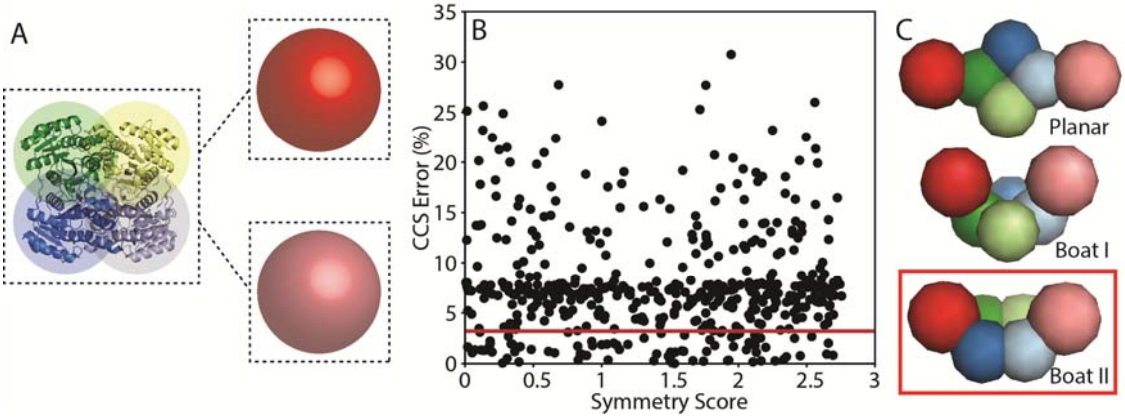

Figure 5-7 Analysis of the modeling output for the MRPP1/2 heterohexamer.
(A) Connectivity is input as the first spatial restraint. **(B)** Symmetry scores, CCS values for the heteropentamer and heterohexamer are used as second level of restraints. The red line indicates 3% CCS error relative to experimental results. **(C)** Generally, three groups of models generate both a reasonable symmetry score and CCS error (under the red line on B): Planar, Boat I and Boat II.

5.4 Conclusions and Future Directions

IM-MS experiments were performed to characterize the interactions within the interface of HO-2/CPR complexes, and this information was used to construct a model of the dimer formed. The number of polar contacts was estimated, and used along with MS data indicating a strong involvement of the C-terminus of the HO-2 protein in the interaction, to generate models. Overall, these data, along with various other restraints, were used to create a putative contact model between HO-2 and CPR. Despite the fact that MS of intact proteins typically provides 'global' constraints for such models, the data here

provides complimentary 'local' information for model generation, and was vital in filtering models generated in rigid docking simulations.

In a second example, IM-MS was applied to determine the stoichiometry and architecture of the MRPP1/2 complex. The stoichiometry of the largest complex between MRPP1 and MRPP2 observed in our experiments was determined to be 2:4 using both intact MS and tandem MS experiments. Potential topologies were determined using spatial restraints derrived from CCS values and symmetry, primarily. The most-likely models generated contain two MRPP2 contacts for every MRPP1 in the complex, thus rationalizing the 2:4 hetero-hexamer stoichiometry measured in our experiments.

While the models described in this chapter are best-described as preliminary, they provide excellent starting points for further experiments that seek to provide additional local constraints on the details of the protein-protein contacts in these systems. For example, detailed protein-protein interaction sites can be mapped by HDX, oxidative labeling and chemical crosslinking experiments combined with MS. Further, integrative modeling combining other low-resolution techniques, such as EM, would be highly beneficial to valdiate some of the models discussed here.

Acknowledgements to our collaborators Prof. Stephen Ragsdale and his group members Ireena Bagai, Angela Fleischhacker and Andrea Morris Spencer for providing HO-2 and CPR proteins, and also to Prof. Carol Fierke and her students Xin Liu and Michael Howard for providing Rnase P protein complexes. I would also like to thank Joseph Eschweiler from Ruotolo group for his contributions in data collection and taking charge of future experiments.

5.5 References

(1) Lustbader, J.; Cirilli, M.; Lin, C.; Xu, H.; Takuma, K.; Wang, N.; Caspersen, C.; Chen, X.; Pollak, S.; Chaney, M.; Trinchese, F.; Liu, S.; Gunn-Moore, F.; Lue, L.-F.; Walker, D.; Kuppusamy, P.; Zewier, Z.; Arancio, O.; Stern, D.; Yan, S.; Wu, H. *Science* **2004**, *304*, 448.

(2) Kissinger, C.; Rejto, P.; Pelletier, L.; Thomson, J.; Showalter, R.; Abreo, M.; Agree, C.; Margosiak, S.; Meng, J.; Aust, R.; Vanderpool, D.; Li, B.; Tempczyk-Russell, A.; Villafranca, J. *J. Mol. Biol.* **2004**, *342*, 943.

(3) Bianchetti, C.; Yi, L.; Ragsdale, S.; Phillips, G. *J. Biol. Chem.* **2007**, *282*, 37624.

(4) Xia, C.; Panda, S.; Marohnic, C.; Martásek, P.; Masters, B.; Kim, J. J. P. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 13486.

(5) Wang, M. *Proc. Natl. Acad. Sci. U.S.A.***1997**, *94*.

(6) Tenhunen, R.; Marver, H.; Schmid, R. *J. Biol. Chem.* **1969**, *244*, 6388.

(7) Yoshida, T.; Noguchi, M.; Kikuchi, G. *J. Biol. Chem.* **1980**, *255*, 4418.

(8) Noguchi, M.; Yoshida, T.; Kikuchi, G. *J. Biol.* **1979**, *86*, 833.

(9) Wang, H.; Lanza, D.; Yost, G. *Arch. Biochem. Biophys.* **1998**, *349*, 329.

(10) Higashimoto, Y.; Sugishima, M.; Sato, H.; Sakamoto, H.; Fukuyama, K.; Palmer, G.; Noguchi, M. *Biochem. Biophys. Res. Commun.* **2008**, *367*, 852.

(11) Maines, M.; Gibbs, P. *Biochem. Biophys. Res. Commun.* **2005**, *338*, 568.

(12) Maines, M. *Annu. Rev. Pharmacol. Toxicol.***1997**, *37*, 517.

(13) Terry, M.; Linley, P.; Kohchi, T. *Biochem. Soc. Trans.* **2002**, *30*, 604.

(14) Watanabe, T.; Kikuchi, M.; Hatakeyama, D.; Shiga, T.; Yamamoto, T.; Aonuma, H.; Takahata, M.; Suzuki, N.; Ito, E. *Dev. Neurobiol* **2007**, *67*, 456.

(15) Sugishima, M.; Sakamoto, H.; Kakuta, Y.; Omata…, Y. *Biochemistry* **2002**.

(16) Zhuang, H.; Kim, Y.-S.; Koehler, R.; Doré, S. *Ann. N. Y. Acad. Sci.* **2003**, *993*, 276.

(17) Sugishima, M.; Omata, Y.; Kakuta, Y.; Sakamoto, H.; Noguchi, M.; Fukuyama, K. *FEBS letters* **2000**, *471*, 61.

(18) Schuller, D.; Wilks, A.; Ortiz de Montellano, P.; Poulos, T. *Nat. Struct. Mol. Biol.* **1999**, *6*, 860.

(19) Lad, L.; Ortiz de Montellano, P.; Poulos, T. *J. Inorg. Biochem.* **2004**, *98*, 1686.

(20) Lad, L.; Friedman, J.; Li, H.; Bhaskar, B.; Ortiz de Montellano, P.; Poulos, T. *Biochemistry* **2004**, *43*, 3793.

(21) Kikuchi, A.; Park, S.; Miyatake, H.; Sun, D.; Sato, M.; Yoshida, T.; Shiro, Y. *Nature Struct. Biol.* **2001**, *8*, 221.

(22) Pereira, P.; Macedo-Ribeiro, S.; Párraga, A.; Pérez-Luque, R.; Cunningham, O.; Darcy, K.; Mantle, T.; Coll, M. *Nature Struct. Biol.* **2001**, *8*, 215.

(23) Yi, L.; Ragsdale, S. *The Journal of biological chemistry* **2007**, *282*, 21056.

(24) Hayashi, S.; Omata, Y.; Sakamoto, H.; Hara, T.; Noguchi, M. *Protein Express. Purif.* **2003**, *29*, 1.

(25) Zhang, Y. *BMC Bioinformatics* **2008**, *9*, 40.

(26) Pierce, B.; Hourai, Y.; Weng, Z. *PloS one* **2011**, *6*.

(27) Phizicky, E. M.; Hopper, A. K. *Genes Dev.* **2010**, *24*.

(28) Hartmann, R.; Gössringer, M.; Späth, B.; Fischer, S.; Marchfelder, A. *Prog. Mol. Biol. Transl. Sci.* **2009**, *85*, 319.
(29) Schön, A. *FEMS Microbiol. Rev.* **1999**, *23*, 391.
(30) Kirsebom, L. *Biochimie* **2007**, *89*, 1183.
(31) Hsieh, J.; Andrews, A.; Fierke, C. *Biopolymers* **2004**, *73*, 79.
(32) Walker, S.; Engelke, D. *Crit. Rev. Biochem. Mol.* **2006**, *41*, 77.
(33) Smith, J.; Hsieh, J.; Fierke, C. *Biopolymers* **2007**, *87*, 329.
(34) Hartmann, E.; Hartmann, R. *Trends Genet.* **2003**, *19*, 561.
(35) Rossmanith, W.; Karwan, R. *Biochem. Biophys. Res. Commun.* **1998**, *247*, 234.
(36) Salavati, R.; Panigrahi, A.; Stuart, K. *Mol. Biochem. Parasitol.* **2001**, *115*, 109.
(37) Walker, S.; Engelke, D. *Cell* **2008**, *135*, 412.
(38) Howard, M.; Liu, X.; Lim, W.; Klemm, B.; Fierke, C.; Koutmos, M.; Engelke, D. *RNA Biol.* **2013**, *10*, 909.
(39) Thomas, D. S.; Stephens, R. M. *Nucleic Acids Res.* **1990**, *18*.
(40) Howard, M.; Lim, W.; Fierke, C.; Koutmos, M. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 16149.
(41) Russel, D.; Lasker, K.; Webb, B.; Velázquez-Muriel, J.; Tjioe, E.; Schneidman-Duhovny, D.; Peterson, B.; Sali, A. *PLoS Biol.* **2012**, *10*.
(42) Hall, Z.; Politis, A.; Robinson, C. *Structure* **2012**, *20*, 1596.
(43) Joanna, F.; Carol, V. R.; Brandon, T. R. *Int. Mass Spectrom.* **2010**, *298*.

Chapter 6. Conclusions and Future Directions

6.1 Conclusions

IM-MS is an emergent technology for structural proteomics, capable of assessing multiprotein topologies from complex mixtures using minute amounts of sample.[1,2] Computational analysis is a key aspect of IM-MS data interpretation, and developments in this area are likely to be rapid over the next five years.[3] The goal of this thesis is to develop protocols that overcome the current challenges in IM-MS experiments aimed at characterizing multiprotein topology and protein interfaces. Like early experiments in high-throughput X-ray structures for proteins, robot-assisted IM-MS is seeking for optimal experimental conditions from a vast array of potential solution states. Detailed subunit sizing and interface properties are revealed by mapping the solution conditions involved. Further, gas-phase and solution-phase unfolding enhance the capabilities for assessing three-dimensional protein architecture, by refining the topology construction to domain level. The number of domains with in single protein subunit could be the key for accurate topology modeling and also assess the intra-chain flexibility. However, challenges remain when apply these protocols to unknown protein systems. The analysis of multiprotein complexes of unknown composition requires integration of intact IM-MS measurements with ‘top-down’ and ‘bottom-up’ style proteomics workflows in order to correctly derive protein connectivity and topology. Two major trends could be involved

in the future application of these protocols: the first is to refine the protocols even more with smaller and simpler systems to accurately assess the basic mechanism behind of gas-phase and solution phase disruption; The second is to take advantages of these protocols for the structural characterization of unknown protein systems, in combination of other techniques, i.e. EM, HDX and crosslinking.

6.2 Future Directions

6.2.1 Refine Protein-Protein Interface Study with Defined Protein Complexes

In chapter 3, we established a strong correlation between the nature/concentration of chemical disrupting agents in solution and the chemical properties of the protein-protein interfaces. However, large error bars, or in other words, less precision is achieved for interface prediction using this correlation, especially when salt bridge interactions are considered. We attribute this lack of precision to the fact that the responses of the protein-interactions.to ammonium acetate and dimethyl sulfoxide are not perfectly orthogonal. In addition, the flexibility of protein structures may also allow long-distance salt-bridge interactions to be accessed during the disrupting process that are not present in the native structure.[4,5] To better understand the solution disruption of multiprotein complexes, and also to push the accuracy/precision of the method presented in chapter 3, experiments should be performed on systems of reduced complexity, constructed from rigid biopolymer systems where models of protein-protein interfaces can be designed in detail. For instance, tailored leucine zipper interfaces[6] and amphipathic helix dimers[7] are two systems that fulfill the above criteria, both having protein structures and interfaces well-

defined, minimally influenced by hydrophobic packing forces. In addition to providing experimental measurements that may lead to high accuracy/precision salt-bridge quantitation in unknown protein interfaces, it is likely that experiments on such small helix-helix interactions will provide important calibration points for the method presented in chapter 3. Such experiments should eventually lead to a broader set of data collected over a very broad range of protein complexes, in an effort to develop an exhaustive protein disruption database covering a greater number of protein structure motifs than presented currently in chapter 3.

6.2.2 Extend the Application on Assemblies with Unknown Architectures, with the Assistance from other Techniques

In parallel with the work described above, other complimentary MS-based tools capable of protein interface foot-printing should be integrated with protein complex disruption experiments. For example, HDX, oxidative labeling, and chemical crosslinking-MS data should be compared with the method presented in chapter 3, evaluating the ability of each to define the regions of protein interaction and the chemical properties of those contacts. By combining all types of MS-related data listed above, higher accuracy protein complex models may be generated.

Examples from chapter 5, including the HO-2+CPR and RNase P systems, should be early systems targeted by the experiments described above. Furthermore, integrated MS analysis can also be applied to more complicated protein assemblies than those shown in chapter 5, including the urease activation complex shown in Figure 6-1. This assembly consists of 18 subunits with 3 folds symmetry, and has been the subject of intact MS

analysis previously[8]. There are two possible mechanisms for the complete assembly of the urease activation complex, and can be differentiated by the topology and connectivity of the intermediate complexes detected during MS analysis. Our preliminary data[9] suggests an activation mechanism that agrees more-closely with the lower path shown in Figure 6-1. Accurate topologies of the intact complex and associated subcomplexes are crucial in understanding the activation mechanism operative in cells.

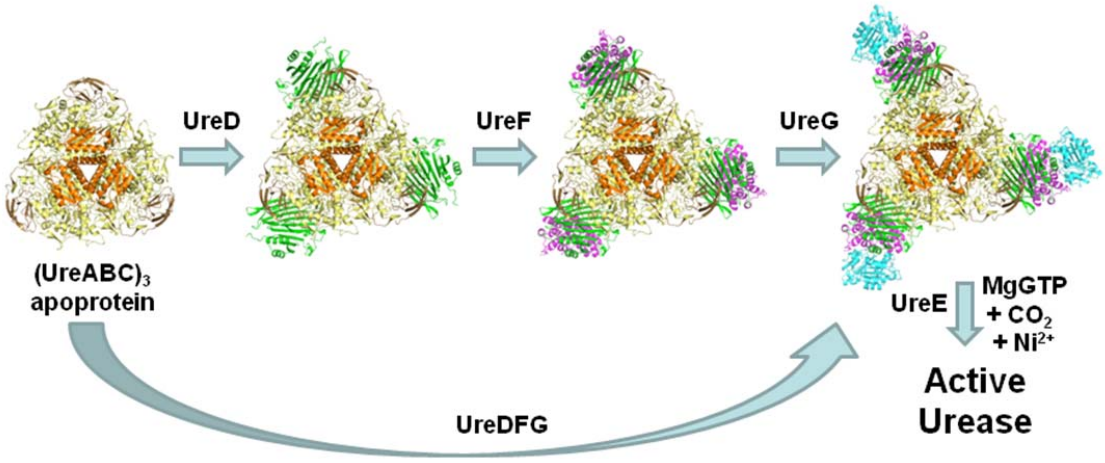

Figure 6-1 Simulated working model of urease activation.
The urease apoprotein (orange UreA, sand UreB, and yellow UreC) sequentially binds UreD (green), UreF (magenta), and UreG (cyan), or binds a complex of these three accessory proteins to form a pre-activation complex. Synthesis of active enzyme requires carbamylation of an active site lysine by CO_2, GTP hydrolysis by UreG, and Ni^{2+} transfer from the UreE metallochaperone. Structures of the docked complexes are derived from computational models. UreD, UreF and UreE structures are adapt from homologous proteins.

On the other hand, integrative structural biology that includes data beyond MS is likely to be main route for determining high-resolution structural models for currently unknown protein complexes. MS, combined with data from X-ray crystallography, NMR spectroscopy, cryo-EM, and SAXS, could generate a comprehensive analyses of large, dynamic, macromolecular machines that incorporates many levels of structural

information.[10-12] Such "hybrid approaches", will undoubtedly require further developments in bio-informatics and software, which will likely be a focus of many active research groups in the near future. For the scope of the examples shown in this thesis, the RNase P complex discussed in chapter 5 would be an excellent target for such experiments, especially those that combine EM and IM-MS data.

6.2.3 Antibody Disulfide Mapping

Antibodies are broadly-important in biotechnology, disease, and biophysics[13,14]. Pharmaceutical companies are interested in developing methodologies to quickly determine if an antibody produced during scaled-up procedures is identical to those generated in control experiments, especially in terms of their disulfide connectivity[15-19]. Many years ago, MS-based protocols were developed for disulfide mapping that involve proteolysis followed by tandem MS[20]. While powerful, this approach is slow and labor-intensive. Preliminary data for IgG1, shown in Figure 6-2, indicates that different CIU patterns likely develop for intact heavy chains in different disulfide bound states. Further experiments should seek to validate the differences shown in Figure 6-2 for systems where the disulfide patterns are known precisely, and evaluate this approach for systems containing many antibody isomeric forms.

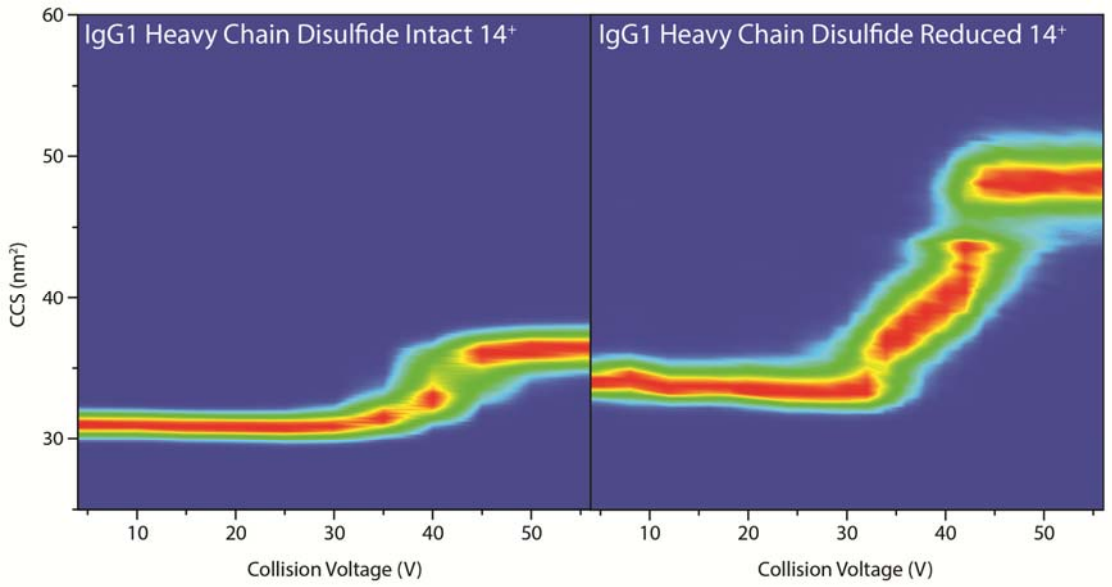

Figure 6-2 Collisional unfolding data of IgG heavy chain, with intact disulfide bonds or reduced disulfide bonds.

These two datasets display significantly different structural features derived from samples with or without intact intra-molecular disulfide bonds. The disulfide-reduced heavy chain has a low energy structure that is larger than disulfide intact, and extends into significantly more unfolded intermediates and final structures than the disulfide intact version. As these datasets are acquired from identical charge states, and under the same instrument conditions, this difference is best interpreted as being related to presence or absence of disulfide bonds.

6.3 References

(1) Ruotolo, B.; Benesch, J.; Sandercock, A.; Hyung, S. J.; Robinson, C. *Nat. Protocols* **2008**, *3*, 1139.
(2) van Duijn, E.; Barbu, I.; Barendregt, A.; Jore, M.; Wiedenheft, B.; Lundgren, M.; Westra, E.; Brouns, S. J.; Doudna, J.; van der Oost, J.; Heck, A. *Mol. Cell. Proteomics* **2012**, *11*, 1430.
(3) Politis, A.; Park, A.; Hyung, S.-J.; Barsky, D.; Ruotolo, B.; Robinson, C. *PloS one* **2010**, *5*.
(4) Paci, E.; Vendruscolo, M.; Karplus, M. *Biophys. J.* **2002**, *83*, 3032.
(5) Zarrine-Afsar, A.; Wallin, S.; Neculai, A.; Neudecker, P.; Howell, P.; Davidson, A.; Chan, H. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 9999.
(6) Landschulz, W.; Johnson, P.; McKnight, S. *Science* **1988**, *240*, 1759.
(7) Lin, W.-H.; Hung, C.-H.; Hsu, C.-I.; Lin, J.-Y. *J. Biol. Chem.* **1997**, *272*, 20044.
(8) Marsh, Joseph A.; Hernández, H.; Hall, Z.; Ahnert, Sebastian E.; Perica, T.; Robinson, Carol V.; Teichmann, Sarah A. *Cell* **2013**, *153*, 461.
(9) Farrugia, M.; Han, L.; Zhong, Y.; Boer, J.; Ruotolo, B.; Hausinger, R. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1328.
(10) Steven, A.; Baumeister, W. *J. Struct. Biol.* **2008**, *163*, 186.
(11) Robinson, C.; Sali, A.; Baumeister, W. *Nature* **2007**, *450*, 973.
(12) Alber, F.; Dokudovskaya, S.; Veenhoff, L.; Zhang, W.; Kipper, J.; Devos, D.; Suprapto, A.; Karni-Schmidt, O.; Williams, R.; Chait, B.; Rout, M.; Sali, A. *Nature* **2007**, *450*, 683.
(13) Sliwkowski, M. X.; Mellman, I. *Science* **2013**, *341*, 1192.
(14) Scott, A. M.; Wolchok, J. D.; Old, L. J. *Nat. Rev. Cancer* **2012**, *12*, 278.
(15) Wypych, J.; Li, M.; Guo, A.; Zhang, Z. Q.; Martinez, T.; Allen, M. J.; Fodor, S.; Kelner, D. N.; Flynn, G. C.; Liu, Y. Q. D.; Bondarenko, P. V.; Ricci, M. S.; Dillon, T. M.; Balland, A. *J. Biol. Chem.* **2008**, *283*, 16194.
(16) Bagal, D.; Valliere-Douglass, J. F.; Balland, A.; Schnier, P. D. *Anal. Chem.* **2010**, *82*, 6751.
(17) Jones, L.; Zhang, H.; Cui, W.; Kumar, S.; Sperry, J.; Carroll, J.; Gross, M. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 835.
(18) Dillon, T. M.; Ricci, M. S.; Vezina, C.; Flynn, G. C.; Liu, Y. D.; Rehder, D. S.; Plant, M.; Henkle, B.; Li, Y.; Deechongkit, S.; Varnum, B.; Wypych, J.; Balland, A.; Bondarenko, P. V. *J. Biol. Chem.* **2008**, *283*, 16206.
(19) Allen, M. J.; Guo, A.; Martinez, T.; Han, M.; Flynn, G. C.; Wypych, J.; Liu, Y. D.; Shen, W. D.; Dillon, T. M.; Vezina, C.; Balland, A. *Biochemistry* **2009**, *48*, 3755.
(20) Gorman, J. J.; Wallis, T. P.; Pitt, J. J. *Mass Spectrom. Rev.* **2002**, *21*, 183.