Development of Collision Induced Unfolding for the Characterization of Large Multiprotein Complexes

by

Chunyi Zhao

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2021

Doctoral Committee:

Professor Brandon T. Ruotolo, Chair
Professor Phillip C. Andrews
Professor Kristina I. Håkansson
Professor David H. Sherman

Chunyi Zhao

cyzhao@umich.edu

ORCID iD: 0000-0003-4651-064X

DEDICATION

I dedicate this dissertation to my parents: Jingmei Yan and Wenmin Zhao. Without their unconditional love and support, I would not have accomplished this work.

ACKNOWLEDGEMENT

First and foremost, I would like to thank my advisor Professor Brandon Ruotolo for his guidance and support through the past five years. As an excellent role model and a scientist, he has inspired me with his passion for science and great work ethic. His full support in my research and my career choice made my doctoral experience enjoyable and allowed me to explore more possibilities.

I would also like to thank my committee members Professor David Sherman, Professor Kristina Håkansson and Professor Phillip Andrews. You all have been very supportive and encouraging during meetings and discussions. I also appreciate how you have challenged me to help me improve my understanding of the research and to spark creative ideas. I wouldn't have accomplished all these projects without the collaboration and help from you.

I've also had the privilege of working with an incredible group of colleagues in the Ruotolo lab. Special thanks to Dr. Joseph Eschweiler, who trained me on the instruments and guided me into the structural biology field. I'd also like to thank Dr. Sugyan Dixit and Dr. Dan Polasky for having the patience to answer my endless research questions and for influencing me with your enthusiasm and passion for science, especially Sugyan who also offered a lot of emotional support. I am also thankful for Dr. Varun Gadkari, who is always super supportive and perceptive both inside and outside of the lab. I had a good time working with Dr. Yuwei Tian and Dr. Sarah Fantin and enjoyed our girls time together. My desk neighbor, Kristine Parson, has

iv

and support, both personal and scientific. Your patience and kindness give me all the strength to overcome the lows. I am fortunate to have you.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

Figure 3-1. (a) Schematics of individual steps in tirandamycin oxidative tailoring. Complete oxidative modifications are highlighted in red. TamI (a P450 enzyme) mediates the initial hydroxylation of TirC at C10, to form TirE, after which TamL oxidizes TirE to the ketone (TirD). Subsequent epoxidation (C11-C12 alkene) and C18 hydroxylation, both catalyzed by TamI, complete the cascade. (b) MS Spectra of TamI and TamL. (c) Isolated 13+ TamL-TirC

LIST OF APPENDICES

ABSTRACT

Protein complexes carry out numerous critical functions in cells and as such represent a key class of drug targets associated with myriad human diseases. Methods for rapid evaluation of protein complex structure and stability are, therefore, extremely important in ongoing efforts to discover new pharmaceuticals. While technologies such as ion mobility spectrometry-mass spectrometry (IM-MS) and collision induced unfolding (CIU) have been established as useful techniques for the rapid analysis of protein quaternary structure and stability using small amounts of unpurified sample, there are still significant challenges associated with the interpretation of such data, especially in the context of large multi-domain protein targets. In this dissertation, we develop CIU based approaches to target such multi-domain proteins in order to leverage IM-MS for future efforts focused on stability assessments, probing protein-substrate interactions, and next-generation protein engineering efforts.

In Chapter 2, we study the CIU of HSA dimer ions as a model multiprotein complex unfolding. Through the novel combination of domain-specific chemical probes, domain-deleted protein constructs, CIU combined with electron capture dissociation (ECD), and steered MD, we were able to demonstrate, for the first time, a detailed mechanism for multiprotein complex CIU. Specifically, our data indicates that a single monomer within the complex is responsible for the CIU transitions observed for the dimer, and that the CIU observed is domain correlated.

The remaining chapters contained within this thesis are concerned with developing IM-MS and CIU methods for the analysis of biosynthetic enzymes, a class of multi-protein complexes

carrying out the synthesis of a host of natural products. In Chapter 3, we apply IM-MS and CIU

to probe the transient physical association between co-dependent enzymes TamI (an iterative

cytochrome P450 monooxygenases) and TamL (a flavin adenine dinucleotide-dependent

oxidase), involved in late-stage oxidation of tirandamycin antibiotics. Our results demonstrate

that TamI and TamL form a biocatalytically competent heterodimeric complex *in vitro*, and we

utilize IM-MS to measure binding affinities between a range of tirandamycin antibiotics with

TamI. Furthermore, we employ domain-specific chemical probes and CIU to mechanistically

reveal that the loop region, which is the "lid" of the TamI substrate pocket unfolds at lower

activation energies, while the heme binding pocket unfolds at higher CIU energies.

In chapter 4, we turn our attention to type I polyketide synthases (PKS), which form an

enzymatic assembly line for the production of polyketide natural products using a series of

modules that include keto synthase (KS) and acyltransferase (AT) domains. we demonstrate the

current limits of quantitative CIU technology by probing the stability of a ~280kDa PKS dimer

protein complex, detecting evidence of stability shifts associated with substrate binding that

account for <0.1% of the mass for the intact assembly. In Chapter 5, We use IM-MS to detect

different conformational states of a 207kDa di-domain KS-AT dimer and are able to capture

stability differences between these different conformations using CIU. Furthermore, through

tracking these forms as a function of time, we elucidate a detailed disassembly pathway for KS-

AT dimers for the first time.

To conclude, this dissertation has focused on new developments in technology targeting the

structure of gas-phase multiprotein complexes and its application in drug discovery. Significant

progress has been made in advancing CIU technology for its application in structural biology.

Chapter 1 Introduction

Dynamic functions of multiprotein complexes impact almost every aspect of biochemistry. Despite this, our ability to assess the structures of such macromolecules in a high-throughput mode lags significantly behind similarly rapid assays of protein function. Nano electrospray ion mobility-mass spectrometry (nESI-IM-MS) is becoming an increasingly important tool for characterizing intact multiprotein complexes in terms of their size, stoichiometry and composition. However, there are still many challenges associated with applying IM-MS toward the characterization of protein structure and stability.

1.1   Protein Complexes and Biochemistry

Proteins are the main actors in cellular biochemistry, carrying out numerous critical functions specified by the information encoded within genes. To understand protein function, the determination the architectures of their protein complexes is an important step. While the prevalence of protein complexes in cells is clear, recent studies have shown that the typical number of cellular complexes has been previously underestimated due the lack of technology sufficient to track transient assemblies.[1] Broad efforts have been made to analyze the structures of large protein complexes.[2,3] A number of innovations have catalyzed these efforts including those associated with biochemistry, molecular biology, computational biology, nuclear magnetic resonance (NMR), as well as advances in light and electron microscopy (EM) instrumentation.[1]

In addition to the protein-protein interactions discussed above, other protein based biomolecular interactions can occur with nucleic acids, ligands, cofactors, or metal ions. In order to understand

protein function, it is crucial to investigate these interactions. Protein-ligand interactions are some of the most well-studied phenomena in biochemistry.[4] Specifically, protein-ligand binding can be used to search for biomarkers, determine disease progression, and elucidate drug actions.[5] A significant portion of the work in these fields focuses on the measurement of interaction strengths between proteins and small molecules using dissociation constant values (k_D) as a quantitative measure of the binding event. Technologies to measure k_D values are generally time-consuming and require purified high-concentration samples.[6,7] Recently, many gas-phase technologies have been developed to detect protein-ligand interactions and simultaneously study the impact of such ligands on protein structure, and these techniques will be discussed both in the sections below, and in subsequent chapters.

One of the classes of multiprotein complexes discussed at length in this dissertation is polyketide synthases (PKSs), which form an enzymatic assembly line with a modular architecture and are responsible for the synthesis of polyketide natural products, accounting for the core structures or complete chemical entities associated with many clinically approved therapeutic agents.[8] Type I PKS modules utilize continuous modules to process, elongate and terminate the polyketide chain.[9] Studying the interactions between such PKS modules and the protein subunits, as well as the structural changes upon substrate/ligand binding, are essential for the development of new polyketide-type drugs.

1.2 Approaches for Multi-Protein Complexes Structure Characterization

Many structural biology technologies are well-established for biomolecular structure characterization at different levels of resolution.[10] Lower resolution analyses of protein structure often reveals information including the relative positions and interactions of the macromolecules

involved in complex formation. At higher resolution, the atomic details of the architecture involved in such assemblies are often revealed, including the relative orientations and folds of the interacting components. High-Resolution Protein Structure Determination

X-ray crystallography (XRD) nuclear magnetic resonance spectroscopy (NMR), and cryo-electron microscopy (cryoEM) represent the most widely-used technologies currently available for high-resolution protein structure characterization.[11–20] There have been many remarkable accomplishments in the area of XRD technology. For example, information such as the amplitudes and even the phases of structure factors within a crystal sample are now measured and used when applying this method in a modern context to solve protein structures. XRD data together with molecular dynamics based optimization are used to achieve an atomic structure of the biological units that comprise the crystal lattice. Regardless, there remain many challenges surrounding the ability of XRD to broadly capture targets through the proteome in physiological relevant states.

On the other hand, NMR techniques are used frequently to determine the dynamics of proteins in solution in contrast to XRD and cryoEM. NMR uses an orthogonal pulsed magnetic field to measure the magnitude of nuclear magnetic moments of the excited nuclei after aligning those nuclei in a strong external magnetic field, and then uses defects in the observed frequencies of their nuclear spins to deduce atomic geometries.[21] Despite the clear utility of NMR in protein structural biology, a major limitation in the broad application of this technology remains its size limitations (50-100kDa, generally) associated with the proteins and complexes that produce high-quality measurements.

Finally, cryo-EM uses a beam of electrons aimed at a flash-frozen sample in vitreous ice to generate atomic-resolution structure data. Two-dimensional images are first created by surveying the EM data acquired across a wide range of particles, that are grouped into class averages. Eventually three-dimensional shape and the symmetry of an assembly are revealed by assembling the class-average structural analysis into a three-dimensional model, or through the use of electron tomography with leverages angle resolved EM data. While all of these methods are well-established structural biology tools and have been highly successful in many cases, large amounts of high-stability, high-purity protein are typically required in order to derive atomic structures.[22]

1.2.1 Lower-Resolution Protein Structure Analysis

Other than atomic-resolution approaches, a number of lower-resolution ways to characterize the structure of multi-protein assemblies have also been widely used, such as electron microscopy (EM)[2] and small angle X-ray scattering (SAXS)[23] that can be used obtain structure data such as complex shape and topology, along with information on subunit secondary and tertiary structure. As a specific example of such an approach, SAXS allows structure measurement for protein conformations and orientations in solution. SAXS is competent for assemblies with size 50-250 kDa, and can be used when relatively small amounts of protein are available.[24] However, most lower resolution technologies also have their own limitations. As a result, there is an urgent need to develop new approaches to generate subunit stoichiometry, composition, and shape information for heterogeneous macromolecular complexes of biological importance.

1.2.2 Companion Technologies for MS studies of Multi-Protein Complexes

One addition to the series of biophysical tools discussed in the section above is mass spectrometry (MS), which has been established as a key technology for protein structure characterization due to recent improvements in resolution, speed, accuracy and sensitivity. In contrast with cryoEM, NMR, and XRD, which are all difficult to automate and operate in a high-throughput mode, MS has the capability to achieve high-throughput measurements. Furthermore, in situations where high-resolution structures are difficult to obtain due to low amounts or purities of the protein sample targeted, MS can contribute significantly since only a small amount of relatively unpurified protein sample is needed for accurate measurements. Moreover, the integration of novel chemical labeling has improved the capacity of MS to obtain higher order protein structural information,[25] with the most common classes of chemical labeling forms used for MS based protein structure analysis being chemical cross-linking (CXL)[26,27], hydrogen/deuterium exchange (HDX)[28], and oxidative footprinting (OFP)[29,30].

1.3   Ion Mobility – Mass Spectrometry (IM-MS) Instrumentation

Though the origins of ion mobility spectrometry can be traced back to the beginning of the 20th century, more recent coupling of the technique to MS, creating a tandem tool termed Ion Mobility Mass Spectrometry (IM-MS) which is often coupled to soft ionization techniques such as electrospray ionization (ESI) and Matrix-assisted laser desorption ionization (MALDI), has gained importance as a useful approach for the analysis of bio- macromolecules. By using IM-MS to acquire information related with the mobility, collision cross-section (CCS) and ultimately the conformation of biomolecules in the gas phase, it is able to assess the conformational

dynamics[31], unfolding pathways[32], ligand-induced conformational changes[12], aggregation

intermediates, and quaternary structures[16].
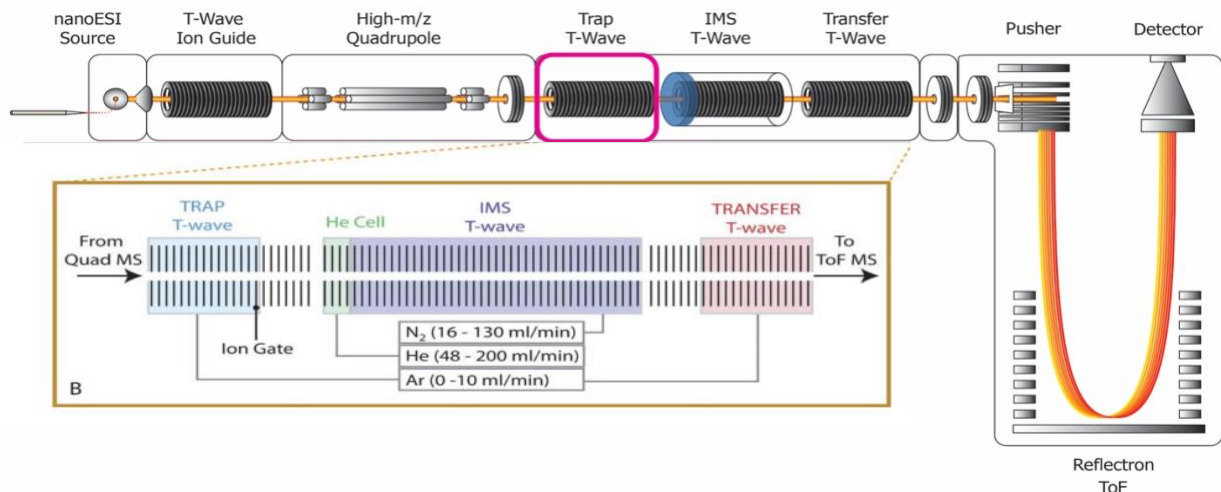
Figure 1-1. Schematic diagram of the commercially available Waters Synapt G2 HDMS nanoelectrospray-quadrupole-ion mobility-time-of-flight mass spectrometer. The general schematic of the instrument indicates five main instrument regions: ion generation (using a nESI ion source), ion guide (using a T-wave SRIG) ion selection (using a modified quadrupole mass analyzer capable of selecting ions up to 32,000 m/z), ion mobility separation (using a tri-wave region described in the inset), and ion mass analysis (using a time-of-flight mass analyzer capable of ~40,000 mass resolving power).

Once ions are ionized and desolvated by the nESI source, they are introduced into a T-wave ion

guide and then guided into quadrupole mass spectrometer working as a mass filter. Selected ions

are further transferred and trapped within a T-wave ion trap prior to the gate in front of the ion

mobility cell. After trapped for a predefined period, ions are released into the mobility cell and

separated based on their mobility as they are transmitted through the ion guide filled with inert

gas under the influence of a weak electric field. The exiting ions are then transferred to a T-wave

transfer ion guide and focused into the ToF mass spectrometer for m/z analysis. After a gate

pulse, 200 mass spectra are recorded by the mass analyzer. After one round of 200 pushes, the

gate pulse process repeats and a new group of 200 spectra acquired. As a result, the overall

mobility spectrum range is defined by $200 \times t_{pp}$, where t_{pp} is the record length of the ToF experiment. Therefore, the arrival time of an ion can be calculated as bin number$\times t_{pp}$.

Several commercial IM-MS instruments are available using different combinations of IM and MS technologies. Data shown in Chapters 3-5 in this thesis was acquired on the Synapt G2 instrument described above[33,34], whereas data shown in Chapter 2 was acquired on both the Synapt G2 and a modified Agilent 6560 drift tube IM-MS instrument. This instrument has been described previously in the literature, including the modifications unique to the platform used in this work.[35,36] Briefly, the drift tube utilized within the 6560 uses a linear field gradient in order to carry out IM separation, in contrast to the TWAVE device used in the Synapt G2. In addition, the Agilent 6560 used here has been modified to include altered ion optics that allow for collision induced unfolding (CIU) experiments of large ions, as well as electron capture dissociation (ECD).[36–38]

1.3.1 Protein Ion Generation by Electrospray Ionization

The advent of ESI as a method for generating macro-sized biomolecular ions created a revolution in MS, enabling vast new insights into cellular biochemistry.[39] More broadly, ESI has evolved into a broadly-used analytical tool, capable of generating molecular ions from wide range of targets, from small molecules to megadalton-scale multi-protein complexes.[40] The basic operating principles of ESI are straight forward: sample is loaded into a capillary tube coated with conductive metal as depicted in Figure 3, and is introduced to a mass spectrometer driven by a flow induced both by applied packing pressure supplied by pneumatic or a syringe-driven pump, and an applied electrical potential (0.8-2.5kV). A Taylor cone[41] is formed at the tip of the capillary and charged droplets are consequently produced and drawn into a proximal vacuum

interface. The size of the ESI droplets reduces as a result of solvent evaporation, which continues until the Columbic repulsion between the increasingly crowded charges within the droplets becomes sufficiently strong such that the surface tension cannot retain droplet integrity. At this point, termed the 'Rayleigh limit'[42], an asymmetric droplet fission event occurs, creating a number of smaller charged droplets alongside the original droplet having shed sufficient charge to retain integrity. This process of evaporation followed by a Rayleigh fission event cycles until desolvated multiply charged protein ions are formed via the charged residue model[43], as depicted in Figure 1-2.
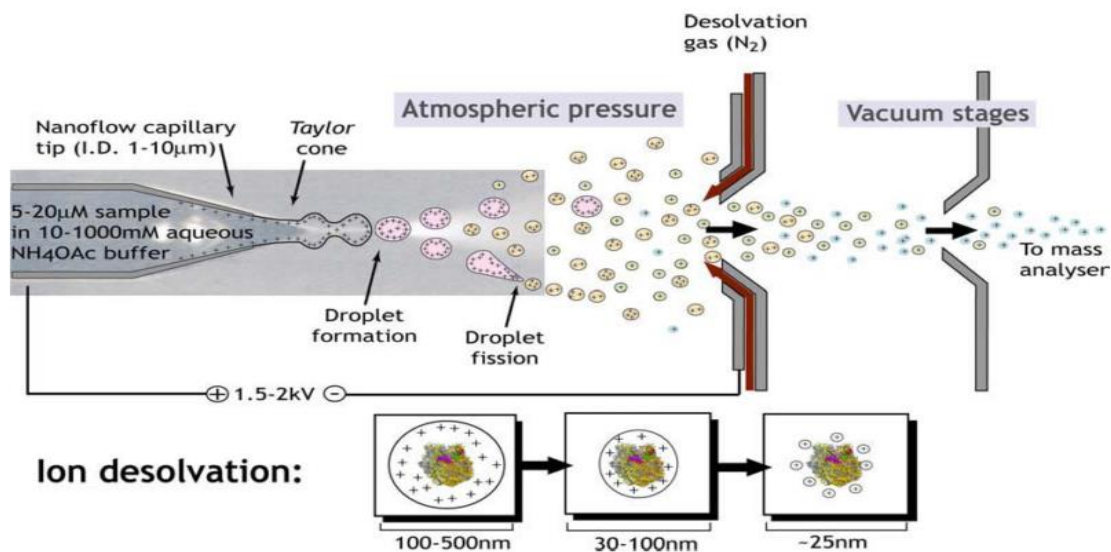
Figure 1-2. Schematic depiction of the positive ion mode nESI process.

To enhance ionization efficiency, nESI, a miniaturized version of ESI, was introduced[44,45]. Through decreasing the flow rate from the microliter per minute range to those in the nanoliter per minute range, the nESI emitter produces droplets of reduced size compared with conventional ESI, which greatly facilitates the desolvation process (Figure 1-2). In specific, the diameters of initial droplets created by conventional ESI are of the order of several micrometers, while those formed via nESI are estimated to be on the order of 150-200nm.[46–48] As a result,

nESI is estimated to exhibit 2000 times greater ionization efficiencies when compared with conventional ESI, defined as the number of analyte ions recorded at the detector divided by the number of analyte molecules sprayed in ESI-MS.

The final stage of ESI ion formation process is still an active area of research. Currently there are many proposed models, two of which have the most evidence: The charge residue model (CRM)[48,49] as discussed above, postulates that evaporation and Coulombic fission occur until a droplet containing a single residual analyte ion remains. Complete evaporation of the solvent within this droplet eventually yields a "naked" analyte ion, called a charged residue, by which the overall charge state of the observed ions is proportional to the surface area of the protein in solution. CRM is the most widely accepted model associated with the formation of multiprotein complex ions. Alternatively, the ion evaporation model (IEM)[50] argues that prior to compete desolvation of the droplet, the repulsion forces that exist between the charged analyte ions and the other charges in the droplet is strong enough to overcome solvation forces so that analyte ion is ejected from the droplet into gas phase. Neither the CRM nor the IEM can account for all the experimental observations made from ESI-MS. In most cases, both mechanisms operate together to produce ions, in that IEM rids charged ions from droplets prior to CRM ion formation. In addition to the CRM and IEM processes described above, chain ejection model (CEM) has recently been proposed to explain the observation of high charge state ions generated from denaturing solutions [43,51]. The CEM mechanism describes a process by which a disordered biopolymer chain can be partially ejected from a droplet, and how subsequent proton transfer events to the unfolded structure can lead to the final ejection of the extended chain.

1.3.2   High Mass Protein Ion Transmission, Selection, and Detection

The majority of structural MS experiments are performed using hybrid mass spectrometers, combining a quadrupole mass filter with an orthogonal time-of-flight (ToF) mass analyzer. In ToF mass analyzers, ions are accelerated by a known electric field. A ToF instrument typically consists of a series of lenses which serve focus the primary ion beam. These ions then enter a pusher region, where ions experience a time-varying field which accelerates the ions in packets into an evacuated flight tube of known length. Ions are accelerated such that they achieve identical kinetic energies, which then adopt different drift velocities and thus drift times in a field free drift tube in a manner directly correlated with their m/z ratio . ToF analyzers positioned orthogonally to the original ion beam axis, as is the case in Q-ToF instruments commonly used for structural MS, act to segment the continuous ion beak emanating from the nESI source into packets for ToF analysis, leaving the majority of ions un analyzed by the ToF. In order to enhance the duty cycle of orthogonal ToF arrangements, an RF-only ion trap can be used to confine ions prior to the pusher and then released stored ions in a pulse synchronized with that of the pusher.

All ions must traverse the same fixed distance in the flight tube to reach the detector. Because of their different m/z values, ions possess different velocities after acceleration by a field that imparts identical kinetic energies to all ions. The time each ion takes to reach the detector is related to its m/z by:

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

Where t is the flight time, d is the fixed distance that ion travels through the drift tube and U is the electric potential applied in the pusher. Most commercially available ToF instruments are equipped with one or more ion mirrors commonly referred to as 'reflectrons'. When ions are ejected from the pusher into the flight tube, two ions with the same m/z theoretically have the same velocity. In fact, this is not always the case when differences in starting positions within the pusher, and different initial kinetic energies in the direction of flight, are considered. In order to compensate for the different kinetic energies that ions of same m/z can possess in practical ToF analyzers, reflectrons are thus introduced.[52] In order to normalize such spreads in ion kinetic energies, the reflectron uses a series of ring electrodes to create an electric field, which can redirect the ions to a subsequent reflectron or the final detector. Ions that possess a greater kinetic energy in the direction flight are able to penetrate deeper into the reflections field to achieve longer flight distances when compared with ions of lower kinetic energies. With appropriate configuration, ions with equal m/z but different starting kinetic energies will adopt corrected flight times, and thus produce mass spectra with higher resolving power.[52]
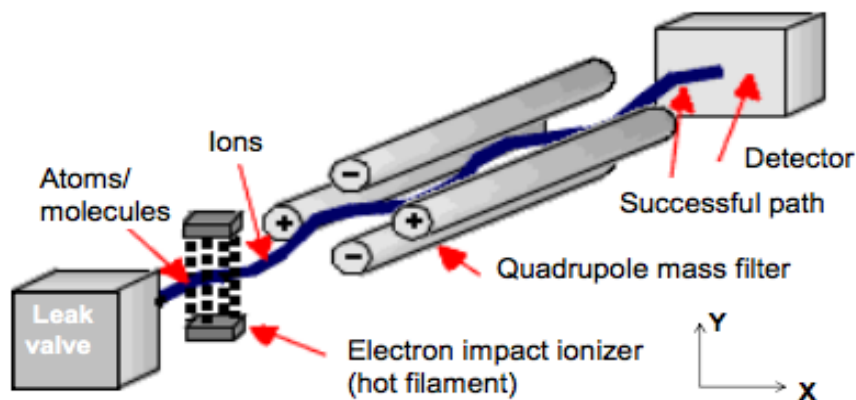
Figure 1-3. Schematic of a quadrupole mass filter.

Another mass analyzer commonly used in structural MS is the quadrupole. A quadrupole mass filter consists of two perpendicularly mounted pairs of parallel conductive rods as shown in Figure 1-3.[53] Two opposite rods are paired together with DC and RF voltages applied to each rod-pair. Two opposite rods in X-direction have an applied potential as $(U+V\cos(\omega t))$ and the other two rods in Y-direction have a potential of $-(U+V\cos(\omega t))$, where U refers to a DC voltage with the same magnitude for both rod-pairs and $V\cos(\omega t)$ stands for a RF voltage to oscillate ions at the same frequency. By applying both voltages, the trajectory of ions can be tuned to transmit ions in a flight path centered between the four rods.

There are two modes of operating the quadrupole mass filter. One is narrow-band mode, for a given DC and RF voltages, only ions of a certain m/z ratio can pass through the quadrupole, with all other ions deflected from their original path. In the X-direction, positively charged ions with low m/z are redirected by RF voltages more significantly than larger ions and encounter one of the rods and are neutralized. Therefore, the X-direction functions as a high-pass mass filter. On the other hand, in the Y-direction negative DC voltages attract positively charged analyte ions while RF voltages help to stabilize the trajectory of ions in the center of the device. Thus Y-direction works as a low-pass mass filter. Alternatively, in RF-only mode, where the DC potential is zero, the quadrupole can also act as a broad band mass filter. In this situation, ions having a wide range of m/z are stabilized so as to pass through the mass filter, typically spanning from 0.8 to 4-5 times the m/z of the optimally stabilized ion in the RF field.

Quadrupoles operate as continuous mass filters, making this technique ideal to combine with continuous ion sources, like ESI. Quadrupoles have also been combined successfully with both ion mobility[54] and time-of-flight[55,56] analyzers to provide effective tandem MS instrument platforms for complex mixture analysis, trace detection, and structural biology. To achieve

enhanced transmission efficiency for large protein complexes, quadrupole mass filters within the Synapt G2 IM-MS system operate at a reduced RF frequency, making them capable of selecting ions up to 32,000 m/z, in contrast to the typical range for such devices (<4,000 m/z).[34]

1.3.3 IM Separation

Ion Mobility (IM) separates ions based on their ability to traverse a chamber filled with inert neutral gas under the influence of an electric field.[57,58] When combined with mass spectrometry (MS), the resulting multi-dimensional IM-MS technique can provide a sensitive and efficient means of analyzing complex mixtures ranging from crude oil to cellular extracts.[59–61] By applying the IM-MS approach to challenges in structural biology and ligand screening, this technology can be used tocollect useful structure and stability information for proteins that are difficult to analyze by other approaches.[12,62,63]
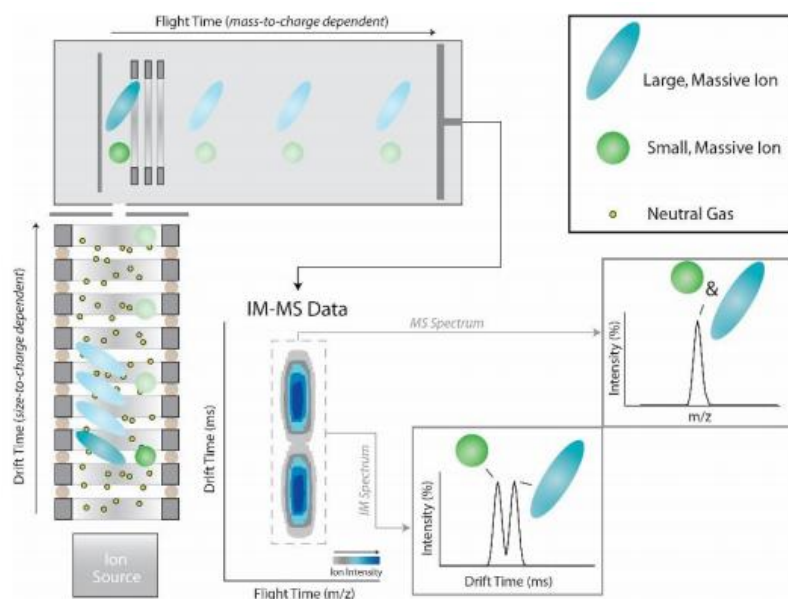
Figure 1-4. Ion mobility-mass spectrometry data acquisition and basic principles. Ions are generated at the ion source (lower left), and are directed to an ion guide filled with neutral gas molecules and driven by an electric field. The ions migrate through this region according to their size-to-charge ratio. The resulting data contains information of 3-dimension, including ion intensity, size, and mass information.

Multiple IM analyzers have been coupled with MS, including drift tubes, trapped IM (TIMS), and differential mobility analyzers.[64] Most of the work discussed in this thesis utilizes traveling wave (T-wave) based IM analysis.[33,65] T-wave IM devices are comprised of gas-filled SRI[65,66] with opposite phases of RF potential applied to adjacent ring electrodes to confine analyte ions radially (Figure 1-5). Meanwhile, a pulsed DC voltage is superimposed onto the confining RF voltage and stepped through the device in order to propel ions against the resistive force provided by the background gas. This moving electric field is referred as 'travelling wave' as ions can 'surf' on the wave in order to reduce their transit time if the field intensity is high or the ion experiences minimal resistance from the background gas (e.g. has a high ion mobility or a low CCS). During mobility separations, ions experience a large number of collisions with the background gas and roll over the top of traveling waves. The larger the ion, the more collisions experienced, thus the mobility of the ions is inversely correlated with the number of roll over events, and the longer transit times through the T-wave device that results. Through the appropriate choice of travelling wave velocity, pulse height and gas pressure, T-wave IM separators can work effectively in the context of a tandem mass spectrometer.
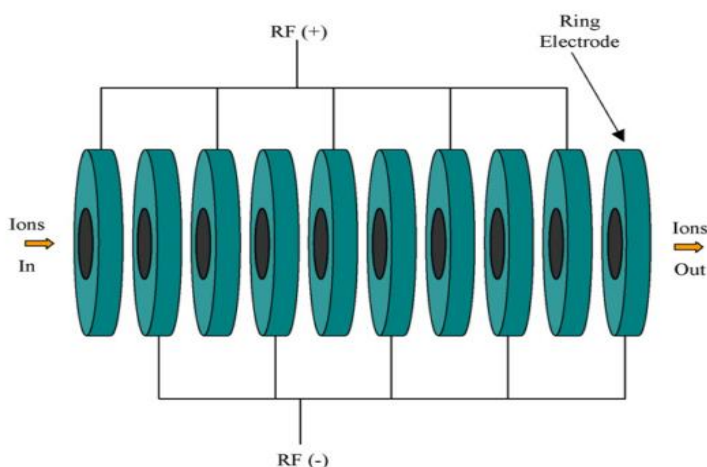
Figure 1-5. A stacked ring ion guide (SRIG).

1.4   IM-MS Methodologies for the Analysis of Multi-Protein Complexes Structure

1.4.1   Collision Induced Unfolding for Measuring the Stability of Proteins and Complexes

As the size and complexity of biomolecules increases, the CCS information generated from IM measurements may not be sufficient on their own to differentiate protein structures. Collision induced unfolding (CIU) functions as a supplement to IM experiments and acts to collisionally activate ions in the gas phase in order to populate unfolded states for subsequent IM analysis. CIU experiments are typically performed in a trapping region prior to an IM analyzer, and can be performed in a quadrupole-selected or 'all ion' mode.

Early CIU studies were focused on the analysis of gas-phase protein structures and the characterization of the folding funnel adopted by proteins in a solvent free environment.[67] Subsequent CIU measurements shifted to focus on analytical objectives associated with the differentiation of subtly different protein tertiary structures .[68] Subsequent CIU experiments probed changes in protein complexes triggered by ligand binding and protein-protein interactions, and highlighted the ability of gas-phase unfolding to detect subtle differences in

protein stability.[12,69–72] More recent studies using CIU have quantified the level of cooperative stabilization which occurs within protein complexes upon ligand binding [73], revealed a domain-correlated mechanism of gas-phase unfolding[74], assessed stability changes within complexes when bound to a wide range of anions and cations[75,76], elucidated the selectivity of membrane protein-ligand complexes[77], and uncovered details of disulfide bond structure within intact antibodies[78].

Simultaneously with the methodological advancements discussed above, software for the processing of CIU data has also been developed in order to enable the quantitative analysis of CIU data with increased throughput and accuracy and advance the ability of CIU to differentiate protein structures.[77,79,80] For example, CIUSuite 2 is able to annotate CIU features, extract CIU50 stability values, detect CIU features, and thus enables the fully automated analysis of protein stabilities.[80]

1.4.2   The Collision Induced Dissociation (CID) Mechanism of Multi-Protein Complexes

Protein topology analysis routinely requires information regarding the connectivity and composition associated with the target assembly. Tandem MS technologies, typically relying upon collision induced dissociation (CID), are generally utilized as part of a broader strategy to reveal the organizational principles of multiprotein complexes.[81] Despite the prevalence of CID, many activation tools have been applied to protein complex ions, including: electron transfer dissociation (ETD)/electron capture dissociation (ECD), surfaced induced dissociation (SID), and photo-dissociation (PD).[82,83] CID of protein complexes can be conducted on a range of instrument platforms, including Q-ToF and orbitrap instruments modified to carry out native MS. The CID experiments discussed in this thesis occur in the ion trap T-wave region prior to the IM

separator on a Synapt G2 instrument, where ions are accelerated in an electric field and collide with neutral gas at relatively high kinetic energies. After a series of energy transfer events between the background gas and the precursor ion, protein complexes dissociate into product ions.

The most commonly observed CID dissociation pathway for multi-protein complexes produces product ions that are assymetric in terms of their mass and charge relative to the precursor ions involved. Typically, selected precursor complex ions decay to produce product ions that include both highly charged single subunits and charge stripped oligomers that have lost the subunits that form a separate product ion population following collisional activation. .[84,85] This asymmetric charge partitioning CID mechanism has been observed across a wide range of protein complexes, including the cytochrome C dimer where two different monomer CSD are observed[86], and the GroEL tetradecamer, where the precursor ions dissociate into monomer and 13mer, while the monomer carries away 58.5% of the parent ions charge.[87] In addition to the asymmetric charge partitioning mechanism described above, other CID dissociation pathways have also been observed, often driven by an altered precursor ion charge state.[88,89]

Although the mechanistic underpinnings of protein complex CID had been proposed prior to the initial IM-MS measurements on protein complexes, the combination of CIU and CID is proved to be instrumental in revealing the role of protein unfolding in the context of multiprotein CID. For example, CIU and CID techniques were combined to study tetrameric transthyretin (TTR) assemblies, and demonstrate that such ions can populate a range partially folded intermediate states, indicating that one or more subunits can experience unfolding during the dissociation process.[69] Another more recent study focusing on human serum albumin (HSA) monomers revealed a domain-correlated CIU, further indicating that protein unfolding with large system

17

proceeds in a manner correlated with folded subunits, even if those units are tethered by covalent bonds.[90]

1.5   References

(1)   Sali, A.; Glaeser, R.; Earnest, T.; Baumeister, W. *Nature* **2003,** *422*, 216.
(2)   Baumeister, W.; Steven, A. C. *Trends in Biochemical Sciences* **2000,**
(3)   Saibil, H. R.; Orlova, E. V. *Nat. Struct. Biol.* **2002,**
(4)   Du, X.; Li, Y.; Xia, Y. L.; Ai, S. M.; Liang, J.; Sang, P.; Ji, X. L.; Liu, S. Q. *International Journal of Molecular Sciences* **2016,**
(5)   Geoghegan, K. F.; Kelly, M. A. *Mass Spectrom. Rev.* **2005,**
(6)   Saggio, I.; Gloaguen, I.; Poiana, G.; Laufer, R. *EMBO J.* **1995,**
(7)   Kim, P.; Zhao, J.; Lu, P.; Zhao, Z. *Nucleic Acids Res.* **2017,**
(8)   Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R. H.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. *Nature* **2014,** *510*, 512.
(9)   Whicher, J. R.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R. H.; Håkansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. *Nature* **2014,** *510*, 560.
(10)   Steven, A. C.; Baumeister, W. **2008,** *163*, 186.
(11)   Clemmer, D. E.; Jarrold, M. F. *Journal of Mass Spectrometry* **1997,**
(12)   Hyung, S. J.; Robinson, C. V.; Ruotolo, B. T. *Chem. Biol.* **2009,** *16*, 382.
(13)   Loo, J. A.; Berhane, B.; Kaddis, C. S.; Wooding, K. M.; Xie, Y.; Kaufman, S. L.; Chernushevich, I. V. *J. Am. Soc. Mass Spectrom.* **2005,**
(14)   Politis, A.; Park, A. Y.; Hyung, S. J.; Barsky, D.; Ruotolo, B. T.; Robinson, C. V. *PLoS One* **2010,** *5*,
(15)   Pukala, T. L.; Ruotolo, B. T.; Zhou, M.; Politis, A.; Stefanescu, R.; Leary, J. A.; Robinson, C. V. *Structure* **2009,** *17*, 1235.
(16)   Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. *Science (80-.).* **2005,** *310*, 1658.
(17)   Uetrecht, C.; Rose, R. J.; van Duijn, E.; Lorenzen, K.; Heck, A. J. R. *Chem. Soc. Rev.* **2010,** *39*, 1633.
(18)   Van Duijn, E.; Barendregt, A.; Synowsky, S.; Versluis, C.; Heck, A. J. R. *J. Am. Chem. Soc.* **2009,**
(19)   Wang, S. C.; Politis, A.; Di Bartolo, N.; Bavro, V. N.; Tucker, S. J.; Booth, P. J.; Barrera, N. P.; Robinson, C. V. *J. Am. Chem. Soc.* **2010,**
(20)   Wyttenbach, T.; Bowers, M. T. *Annu. Rev. Phys. Chem.* **2007,**
(21)   Rabi, I. I.; Zacharias, J. R.; Millman, S.; Kusch, P. *Physical Review* **1938,**
(22)   Robinson, C. V.; Sali, A.; Baumeister, W. *Nature* **2007,**
(23)   Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A. *Quarterly Reviews of Biophysics* **2007,**
(24)   Förster, F.; Webb, B.; Krukenberg, K. A.; Tsuruta, H.; Agard, D. A.; Sali, A. *J. Mol. Biol.* **2008,**
(25)   Hyung, S. J.; Ruotolo, B. T. *Proteomics* **2012,**

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

(27) Jin Lee, Y. *Molecular BioSystems* **2008,**

(28) Konermann, L.; Pan, J.; Liu, Y. H. *Chem. Soc. Rev.* **2011,**

(29) Kiselar, J. G.; Chance, M. R. *J. Mass Spectrom.* **2010,**

(30) Konermann, L.; Stocks, B. B.; Pan, Y.; Tong, X. *Mass Spectrometry Reviews* **2010,**

(31) Jenner, M.; Ellis, J.; Huang, W. C.; Lloyd Raven, E.; Roberts, G. C. K.; Oldham, N. J. *Angew. Chemie - Int. Ed.* **2011,** *50*, 8291.

(32) Shi, H.; Pierson, N. A.; Valentine, S. J.; Clemmer, D. E. *J. Phys. Chem. B* **2012,** *116*, 3344.

(33) Giles, K.; Williams, J. P.; Campuzano, I. *Rapid Commun. Mass Spectrom.* **2011,** *25*, 1559.

(34) Zhong, Y.; Hyung, S.-J.; Ruotolo, B. T. *Analyst* **2011,** *136*, 3534.

(35) Vallejo, D. D.; Polasky, D. A.; Kurulugama, R. T.; Eschweiler, J. D.; Fjeldsted, J. C.; Ruotolo, B. T. *Anal. Chem.* **2019,** *91*, 8137.

(36) Gadkari, V. V.; Rojas Ramirez, C.; Vallejo, D. D.; Kurulugama, R. T.; Fjeldsted, J. C.; Ruotolo, B. T. *Submitt. Rev.* **2020,**

(37) Fort, K. L.; Cramer, C. N.; Voinov, V. G.; Vasil'Ev, Y. V.; Lopez, N. I.; Beckman, J. S.; Heck, A. J. R. *J. Proteome Res.* **2018,** *17*, 926.

(38) Williams, J. P.; Morrison, L. J.; Brown, J. M.; Beckman, J. S.; Voinov, V. G.; Lermyte, F. *Anal. Chem.* **2020,**

(39) Fenn, J. B.; Mann, M.; Meng, C. K. A. I.; Wong, S. F.; Whitehouse, C. M. *246*,

(40) Snijder, J.; Rose, R. J.; Veesler, D.; Johnson, J. E.; Heck, A. J. R. *Angew. Chemie - Int. Ed.* **2013,** *52*, 4020.

(41) *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.* **1964,**

(42) Rayleigh, Lord. *London, Edinburgh, Dublin Philos. Mag. J. Sci.* **1882,**

(43) Konermann, L.; Rodriguez, A. D.; Liu, J. *Anal. Chem.* **2012,**

(44) Wilm, M. S.; Mann, M. *Int. J. Mass Spectrom. Ion Process.* **1994,**

(45) Wilm, M.; Mann, M. *Anal. Chem.* **1996,** *68*, 1.

(46) Juraschek, R.; Dülcks, T.; Karas, M. *J. Am. Soc. Mass Spectrom.* **1999,**

(47) Karas, R. M.; Bahr, · U; Dülcks, · T. *Fresenius J Anal Chem* **2000,**

(48) Dole, M.; Mack, L. L.; Hines, R. L.; Chemistry, D. O.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. *J. Chem. Phys.* **1968,**

(49) Mack, L. L.; Kralik, P.; Rheude, A.; Dole, M. *J. Chem. Phys.* **1970,**

(50) Iribarne, J. V.; Thomson, B. A. *J. Chem. Phys.* **1976,**

(51) Ahadi, E.; Konermann, L. *J. Am. Chem. Soc.* **2011,**

(52) Mamyrin, B. A.; Karataev, V. I.; Shmikk, D. V; Zagulin, V. A. *Sov. Phys. - JETP* **1973,** *37*, 45.

(53) Finnigan, R. E. *Anal. Chem.* **1994,** *66*, 969A.

(54) Lawrence, A. H.; James Barbour, R.; Sutcliffe, R. *Anal. Chem.* **1991,**

(55) Morris, H. R.; Paxton, T.; Dell, A.; Langhorne, J.; Berg, M.; Bordoli, R. S.; Hoyes, J.; Bateman, R. H. *Rapid Commun. Mass Spectrom.* **1996,**

(56) Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. *Rapid Commun. Mass Spectrom.* **1997,**

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

(58) Smith, D. P.; Knapman, T. W.; Campuzano, L.; Malham, R. W.; Berryman, J. T.; Radford, S. E.; Ashcroft, A. E. *Eur. J. Mass Spectrom.* **2009,** *15*, 113.

(59) Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997,** *119*, 2240.

(60) Valentine, S. J.; Liu, X.; Plasencia, M. D.; Hilderbrand, A. E.; Kurulugama, R. T.; Koeniger, S. L.; Clemmer, D. E. *Expert Rev. Proteomics* **2005,** *2*, 553.

(61) Fernandez-Lima, F. A.; Becker, C.; McKenna, A. M.; Rodgers, R. P.; Marshall, A. G.; Russell, D. H. *Anal. Chem.* **2009,** *81*, 9941.

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

(63) Freeke, J.; Bush, M. F.; Robinson, C. V.; Ruotolo, B. T. *Chem. Phys. Lett.* **2012,** *524*, 1.

(64) May, J. C.; McLean, J. A. *Analytical Chemistry* **2015,**

(65) Thalassinos, K.; Slade, S. E.; Jennings, K. R.; Scrivens, J. H.; Giles, K.; Wildgoose, J.; Hoyes, J.; Bateman, R. H.; Bowers, M. T. *Int. J. Mass Spectrom.* **2004,**

(66) Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H. *Rapid Commun. Mass Spectrom.* **2004,** *18*, 2401.

(67) Shelimov, K. B.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997,**

(68) Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997,**

(69) Ruotolo, B. T.; Hyung, S.-J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V. *Angew. Chemie* **2007,** *119*, 8147.

(70) Hopper, J. T. S.; Oldham, N. J. *J. Am. Soc. Mass Spectrom.* **2009,** *20*, 1851.

(71) Han, L.; Hyung, S. J.; Mayers, J. J. S.; Ruotolo, B. T. *J. Am. Chem. Soc.* **2011,**

(72) Zhou, M.; Dagan, S.; Wysocki, V. H. *Analyst* **2013,**

(73) Niu, S.; Ruotolo, B. T. *Protein Sci.* **2015,** *24*, 1272.

(74) Zhong, Y.; Han, L.; Ruotolo, B. T. **2014,** 9209.

(75) Han, L.; Hyung, S. J.; Ruotolo, B. T. *Angew. Chemie - Int. Ed.* **2012,** *51*, 5692.

(76) Han, L.; Hyung, S. J.; Mayers, J. J. S.; Ruotolo, B. T. *J. Am. Chem. Soc.* **2011,** *133*, 11358.

(77) Laganowsky, A.; Reading, E.; Allison, T. M.; Ulmschneider, M. B.; Degiacomi, M. T.; Baldwin, A. J.; Robinson, C. V. *Nature* **2014,**

(78) Tian, Y.; Han, L.; Buckner, A. C.; Ruotolo, B. T. **2015,**

(79) Eschweiler, J. D.; Rabuck-Gibbons, J. N.; Tian, Y.; Ruotolo, B. T. *Anal. Chem.* **2015,** *87*, 11516.

(80) Polasky, D. A.; Dixit, S. M.; Fantin, S. M.; Ruotolo, B. T. *Anal. Chem.* **2019,** *91*, 3147.

(81) Dixit, S. M.; Polasky, D. A.; Ruotolo, B. T. *Current Opinion in Chemical Biology* **2018,**

(82) Wysocki, V. H.; Joyce, K. E.; Jones, C. M.; Beardsley, R. L. *J. Am. Soc. Mass Spectrom.* **2008,**

(83) Kempkes, L. J. M.; Martens, J.; Berden, G.; Oomens, J. *J. Mass Spectrom.* **2018,**

(84) Schwartz, B. L.; Bruce, J. E.; Anderson, G. A.; Steven, A.; Rockwood, A. L.; Smith, R. D. **1995,**

(85) Jurchen, J. C.; Williams, E. R. *J. Am. Chem. Soc.* **2003,** *125*, 2817.

(86) Jurchen, J. C.; Garcia, D. E.; Williams, E. R. **2003,**

(87) Sobott, F.; Robinson, C. V. **2004,** *236*, 25.

(88) Jones, C. M.; Beardsley, R. L.; Galhena, A. S.; Dagan, S.; Cheng, G.; Wysocki, V. H. *J. Am. Chem. Soc.* **2006,**

(89) Van Den Heuvel, R. H. H.; Van Duijn, E.; Mazon, H.; Synowsky, S. A.; Lorenzen, K.; Versluis, C.; Brouns, S. J. J.; Langridge, D.; Van Der Oost, J.; Hoyes, J.; Heck, A. J. R. *Anal. Chem.* **2006,**

(90) Eschweiler, J. D.; Martini, R. M.; Ruotolo, B. T. *J. Am. Chem. Soc.* **2017,** *139*, 534.

(91)   Eschweiler, J. D.; Martini, R. M.; Ruotolo, B. T. **2017,**
(92)   Zhang, Y.; Cremer, P. S. **2009,**
(93)   Susa, A. C.; Xia, Z.; Williams, E. R. *Anal. Chem.* **2017,** *89*, 3116.

Chapter 2 The Gas-Phase Unfolding of Protein Complexes Follows a Largely Domain-Correlated Mechanism

2.1   Introduction:

In order to understand protein function, the structures of large multi-protein complexes must be determined. Tools like X-ray crystallography (XRD), nuclear magnetic resonance spectroscopy (NMR) and cryo-electron microscopy are well-established structural biology tools for high-resolution protein structure characterization, but typically require large amounts of purified protein present in a structural homogeneous state to be effective. In addition, a number of lower-resolution ways to characterize the structure of multi-protein assemblies have also been widely used, such as small angle X-ray scattering (SAXS). However, these technologies also have their own limitations, leaving a large section of the proteome relatively unknown from a structural biology perspective.

Recently developments in native ion mobility-mass spectrometry (IM-MS) have been developed in an effort to fill the gaps associated with protein structure determination discussed above, and has proved successful in efforts to map the overall quaternary structures, stoichiometries, and topologies for a range of assemblies that have resisted prior structural biology campaigns. However, IM-MS techniques provide only coarse-grained models of multi-protein complexes, and struggle to determine the local details of protein structure. Indeed, recent computational efforts have indicated that protein complex structures coarse-grained at the level of individual

proteins can contain significant errors which are largely ameliorated by moving to domain-level coarse graining.[1] Therefore, there is a clear need to develop MS-based technologies capable of providing rapid and robust information on the local structures of subunits within larger multi-protein systems.

In an effort to obtain such domain-level information for the subunits within protein assemblies, our group has been developing collision induced unfolding (CIU) methods.[2,3] In the past, CIU has been used to explore the domain-specific unfolding of gas-phase serum albumins[4] as well as assess protein stability shifts triggered by ligand binding within monomeric proteins.[3,5] All of these studies indicate potential applications for CIU as a tool to study large protein structure of unknown folds, as well as local stability measurements of domains within protein complexes. The ability to access such information from multiprotein complexes relies on a deep understanding of the CIU mechanism for gas-phase protein ions, which remains lacking Multiple previous reports have sought to develop a detailed mechanism of gas-phase protein unfolding. For example, a CIU survey of monomeric proteins revealed domain correlated unfolding events across a range of sequences and protein sizes. Furthermore, by deploying domain-specific chemical probes and domain-deleted constructs, a detailed CIU mechanism of the three-domain model protein human serum albumin (HSA) was constructed whereby domain 2 unfolds upon collisional activation, followed by domain 3 and finally domain 1 at highest energies.[4] Despite these and other advances in our mechanistic understanding of CIU, such details are limited, especially in the context of multiprotein complexes.

Prior work has explicitly linked the CIU and CID of multiprotein complexes.[6,7] Such links allows us to draw hypotheses directly from the extensive library CID data that has been reported for protein assemblies.[8–11] Primary observations of protein complex CID detected product ions

for protein complexes that included highly charged monomers and lowly-charged oligomers stripped of individual monomeric subunits.[12] This asymmetric product ion distribution was further studied so as to link the unfolding of monomeric units within the complex to the eventual product ion distribution, invoking CID intermediate states that included a single unfolded protein bound to the remaining, nominally folded, protein complex.[13,14] IM-MS data for the TRAP 11-mer and the transthyretin tetramer (TTR) linked explicitly CIU and CID processes, and while these studies could not rule out the presence of protein complexes where several protein subunits have undergone CIU, the data shown were consistent with protein complex CIU states where the conformational changes within a single monomer produced upon collisional heating gave rise to the gas-phase unfolding observed.[6]

In this study, we use a series of chemical probes and domain-deleted constructs to study the CIU of HSA dimers as a model for multiprotein CIU more generally. Our data enables us to provide detailed annotations CIU features produced from a protein assembly for the first time, and enables us to build a detailed mechanism for HSA dimer CIU. Furthermore, we integrate electron capture dissociation and steered molecular dynamics (MD) simulations to confirm our mechanism and provide additional details. Our data identify a single chain unfolding process of HSA dimers, consistent with the asymmetric charge partitioning mechanism of protein complex CID. In addition, we observe strong evidence of a largely domain-correlated unfolding process for the unraveling HSA monomer, detecting clear unfolding events for domains 1 and 3, as well as a two-stage unfolding event for domain 2, for low charge state dimer ions.

2.2 Experimental Methods

Sample Preparation. Wild type (WT) human serum albumin were purchased from Sigma-Aldrich as lyophilized powders at purities greater than 97%. The lyophilized powders were then diluted to 100 μM in 200 mM ammonium acetate and aliquoted and stored at −80 °C for future use. 8- Anilino-1-naphthalenesulfonic acid (ANS) ammonium salt hydrate (97%), indomethacin (IDM, 99%), and hemin (HMN, 98%) were also purchased from Sigma-Aldrich. Recombinant albumin domains 1, 2, and 3 as well as the domain 1−2 fusion protein (HSA D12) were purchased from Albumin Biosciences (Huntsville, AL) as lyophilized powders. Recombinant albumin domains were diluted to 180 μM in 200 mM ammonium acetate and stored at −80 °C prior to IM-MS analysis.

IM-MS Data Collection and Analysis

Protein samples were buffer exchanged into 200mM ammonium acetate buffer using Micro Bio-Spin 30 columns (Bio-Rad, Hercules, CA). Sample aliquots (~5μL) were analyzed by IM-MS on a quadrupole-ion mobility- time-of-flight mass spectrometer (Q-IM-ToF MS) instrument (Synapt G2 HDMS, Waters, Milford, MA) as described previously.[15] Protein ions were generated using a nESI source in the positive mode. Capillary voltages of 1.4−1.6 kV were applied, with the sampling and extraction cones set to 10 V and 5 V, respectively. The source backing pressure was adjusted to ~8 mbar. The IM T-wave ion guide was operated a 4 mbar with wave height and wave velocity values of 15 V and 150 m/s, respectively. Mass spectra and drift time distributions were obtained for the ions at multiple trap collision energies in steps of 2 V from 10 to 200 V. All collision cross section (CCS) values were measured, which relate IM drift times directly to ion size and shape.

Extracted CIU data were analyzed using CIUSuite 2.[16] CIU fingerprints were plotted as a 2D contour plot, in which ion intensities were normalized to a maximum value of 1 at each collision voltage and smoothed using a Savitzky-Golay filter. Feature detection was applied within the CIU data sets to describe CIU transitions. RMSD plots and values were generated using the RMSD comparison module of CIUSuite2.

Chemical Probe CID Analysis

HSA constructs were incubated with chemical probes at a 1:1 ratio using 10 μM protein solutions and a DMSO content of less than 1%. Signals for singly bound and apo protein dimer ions were quadrupole isolated and the intensity of both forms were tracked and normalized to plot the CID curve associated with each probe-bound complex.

Sample Preparation & IM-MS Measurements

Lyopholized HSA D12 was reconstituted into 100 mM NH4Ac and then it was buffer exchanged using Bio-Rad BioSpin 6 columns (6 kDa MWCO) before analysis. For CIU-ECD experiments, a modified Agilent 6560 drift tube ion mobility-quadrupole time-of-flight (DTIM-Q-TOF) (Santa Clara, CA) was used.[17,18] Using static nanoelectrospray ionization (nESI), ions were generated from a 10uM solution in 200mM NH_4Ac. Gold-coated borosilicated nESI emitters were used with a capillary voltage of 1400 V. Sulfur hexafluoride (SF_6) was used as drying gas with a flow of 2 L/min at 25 °C. The front funnel and trap funnel operated at 4−4.50 tor, while the drift tube with <18.5V/cm was operated at 3.95 torr under high purity nitrogen. Mass calibration in the extended mass range and sensitivity mode was performed using Agilent tune mix and the Agilent Jet Stream ESI source.

CIU-ECD Experiments

The Agilent 6560 instrument modifications include the addition of a new capillary exit (CE) lens immediately after the ion transfer capillary.[17,18] This enables a controlled voltage change between CE and fragmentor lens (F). The largest difference (also referred to as collision voltage) achieved is 450V if F = 400V. This large accelerating field allows the ions to undergo higher energy collisions for CIU and CID experiments. SF6 was used as an electron scavenger to prevent arcing between CE/F lenses, as well as to enable larger center-of-mass frame energy collisions with analyte ions. Another modification to the Agilent 6560 is the addition of an electromagnetostatic linear electron capture dissociation cell (e-MSion, Corvallis OR) positioned before a shortened collision cell.[18–20] This allows for post-IM electron capture dissociation (ECD) experiments on IM-separated analyte ions. Lastly, the instrument mass range was extended to 20,000 m/z and the top and bottom slit potentials before the TOF entrance were optimized for transmission of large ions (sensitivity tune mode).

CIU experiments of D12 were acquired by increasing CE voltage relative to the F voltage in 10V steps with an acquisition time of 20s per step. Arrival Time Distributions (ATDs) for each activation step and each charge state was obtained using the MIDAC Agilent extractor in CIUSuite2.[16] CIU fingerprints were obtained by plotting ATDs as a function of CE voltage in CIUSuite2. IM-MS data was analyzed using Agilent IM-MS Browser v10 as well as Agilent Navigator v8. Using the e-MSion ECD cell, fragmentation of D12 was performed following quadrupole selection of the 13^+ charge state. The ECD filament was operated at 1.2 µA to allow analyte ion transmission as well as electron emission and ECD. [18,20] ECD fragmentation data was accumulated for 15 min for selected voltage values in a manner correlated with CIU data.

CIU-ECD Data Analysis

Data were extracted in .csv format using Agilent MassHunter IM-MS Browser 10.0. Next, this

data was imported into mMass version 5.5.0 for further mass calibration and manual peak

picking.[21] Mass calibration was optimized in order to decrease false positives when searching for

internal fragments (*vide infra*). For fragments where the monoisotopic peak is not discernible,

the average peptide fragment mass was obtained. Next, using the averagine method,

monoisotopic masses were calculated using a custom Excel spreadsheet.[22]For annotating

terminal and internal fragments, custom python scripts were used. We developed an iterative

approach to peptide fragment ion identification, where MS data were fed through our protocol,

all identified signals removed, and the remaining signals were allowed to move through the

identification procedure for subsequent passes. For HSA D12, in the first pass only the *c*, *z*, *x*,

and *y* fragment ion types were considered in our identification procedure, which have been

observed in ECD fragmentation spectra previously.[23,24] In subsequent passes, water and

ammonium losses were considered. The maximum peptide fragment ion charge state considered

for all passes through our identification procedure was 6^+ and the mass tolerance was 10 ppm.

In order to identify internal fragments, we used a more stringent approach than what is described

above for terminal fragment ion identification. Our internal fragment ion identification procedure

also involved multiple passes through our protocol, but we focused on keeping the available

search space narrow for each pass. Fragment ion spectra were segmented into regions of m/z

space and the charge state was determined for all fragment ion signals detected, allowing us to

populate a list of identified fragment ion masses. This mass list was further limited to all those

that shared the same charge state, and these mass values were searched against the available

internal fragment ion pool iteratively until high-quality identification was achieved. The only ion

types considered for these internal fragment searches were *c-z* and *c•-z/z•-c* as previously

reported for internal fragments produced by z• radical cascades.[25] Another characteristic of ECD is its ability to reduce disulfide bonds.[24,26,27,28] HSA contains 17 disulfide bonds along with two reduced cysteines.[29] Our internal fragment ion search results assumes that ECD may act to break only a few such disulfides, but this number was limited in order to narrow the overall search space.[27] We also considered the possibility that internal fragment ions can be produced with all of their disulfide bridges intact. In other cases where a disulfide bond has been broken, certain neutral mass losses need to be considered when analyzing the resulting internal fragment ion pool. Our internal fragment identification pipeline produces candidate mass values using the following workflow: First the number of cysteines was determined for each internal sequence possible of a specific length range and charge. Next, based on the sequence for HSA found in the Uniprot database (Q56G89_HUMAN), the location of the 11 disulfides in domains 1 and 2 of HSA were determined. For each internal fragment, a disulfide pattern analysis was performed: If both cystines involve in a bringing pair were not within the same internal fragment, they were considered to be in their reduced forms or in a known modified state. The known cysteine chemical modifications associated with disulfide bond reduction have been previously reported in the literature.[27,30]

Following theoretical ion matching performed using deconvoluted masses using a mass measurement error tolerance of 1 ppm, matches were scored based on how well the theoretical isotopic envelope matched in the recorded signal profile from the MS data. A score of 100 was assigned to all isotopologues that were precisely matched to theoretical isotopic envelopes. Only matches with less than 1ppm error and a combined score (weighted average of m/z and intensity scores) of 70 were considered in our analysis.

Molecular Dynamics Simulations

Molecular dynamics (MD) simulations were performed using CHARMM on a workstation with 48 CPU core 2.10 Ghz Intel Xeon processor. The CHARMM36 force field was employed for all MD simulations.[31] To generate models of HSA D12, CHARMM-GUI[32] was used to import the crystal structure of HSA from the PDB (4K2C). Missing N-terminal residues were added, while domain 3 residues were deleted. The output structure from CHARMM-GUI was then used as the starting point for downstream MD simulations.

A charge placement algorithm was developed so that all possible charge isomers that act to produce the experimentally observed protein charge state could be generated for our MD workflow. Positive charges were placed on arginine and lysine, while negative charges were placed on glutamic acid and aspartic acid. All other residues were kept neutral. Our charge placement algorithm generates every possible charge isomer by iteratively neutralizing acidic and basic residues so that the final net charge is equivalent to the experimentally observed charge state. Since the exact location of charges within protein ions are unknown, we employed reasonable distance restrictions to mimic Coulombic repulsion, where charges of the same polarity were spaced by least 3 residues apart in our final models, and additional steps were undertaken to distribute the charges evenly throughout the protein sequence. Charge isomers used for MD simulations were randomly sampled and included both basic (all acidic residues are neutralized) and zwitterionic (both negative and positive charges present) protein ion charge isomers.

Following MD simulations, IMPACT[33] was used to determine the theoretical CCS values of all models generated. Since a large number of the models generated during our MD runs represent protein ions that have undergone significant unfolding, we took additional steps in order to ensure the accuracy of our model CCS values. For example, to obtain more accurate theoretical

CCSs, 2000 structures were randomly sampled from the extracted structures, and Collidoscope[34] was used to calculate theoretical CCSs using the trajectory method (TM). The correlation between IMPACT projected area (PA) and Collidoscope TM was plotted, and a simple regression was used to build a calibration function capable of converting the rapid IMPACT PA CCS values into to approximated TM CCS for all the structures.

Following charge placement, models of HSA D12 were subjected to energy minimization, which included 8 of a steepest descent scheme incorporating the adopted basis Newton-Raphson (ABNR) method. The energy minimized domain-deleted HSA construct models were then equilibrated *in vacuo* using gradual heating from 100K to 300K over 200ps, with a simulation time step of 2 fs. The final equilibrated structure was then used as a starting structure for replica exchange MD (REMD) simulations[35]. Our REMD runs utlized 20 replicas ranging from 300K to 800K over 20 ns simulation timescale. The exchange frequency was set to 1ps. Structures were then extracted from all the REMD trajectory files. These structures were then filtered based on the experimental CCS recorded for HSA D12in order to obtain compact models.

In order to simulate the CIU experiment, we constructed a steered MD (SMD) approach for HSA D12, where the lowest energy compact model was selected as a starting structure. Prior to performing SMD, energy minimization of these models was performed as outlined above. The energy minimized compact HSA D12 models were subjected to simulated heating *in vacuo* while fixing certain residues to allow unfolding in regions that were experimentally relevant. Heating was performed by gradually ramping the temperature from 300K to 950K over 3ns of simulation time. Structures from the trajectory files were extracted and theoretical CCS were calculated followed by filtering based on experimental CCS values. CCS values recorded for HSA D12 monomer dissociated from HSA D12 dimer were also used for filtering model structures.

For both the compact HSA D12 and unfolded HSA D12 modeling results, clustering was performed to better identify structural preferences. We employed a Kabsch[36] algorithm where the pairwise RMSD of all structures within the available pool were computed. To determine the number of structural ensembles, an f-test was performed. With the determined number of clusters, output from our Kabsch analysis was used to sort structures into clusters. HSA D12 dimer models were generated using the docking algorithm in PyRosetta.[37] The lowest energy model of appropriate CCS from each structural ensemble was selected from the filtered models of SMD for docking with compact HAS D12 models to create putative structures for dimer CIU intermediates. For each docking scenario, 10,000 trajectories were initiated, and their CCS values computed, in order to generate final models.

2.3 Results and Discussion

Comparing HSA dimer and monomer CIU

Previous studies have indicated that the CIU of low charge state multi-domain proteins is correlated with domain level structure[15]. Previous work from our group used HSA as a model system system to further study the domain-correlated unfolding of gas-phase protein ions[4]. Each HSA monomer is a single polypeptide chain consisting three homologous domains (D1, D2, and D3), each bearing high levels of sequence (70-90%) and structural identity. To elucidate the level of domain correlation that exists in the CIU of multiprotein complexes, the work reported here focuses on non-specific HSA dimers in an effort to build off of these previous CIU measurements.

Figure 2-1A shows the CIU fingerprint of the 15+ HSA monomer produced under native conditions, which is consistent with previous work.[4] As expected, HSA monomer undergoes CIU

transitions upon collisional activation, resulting in a total of N transitions and adopts N+1

features, where N is the number of domains in the native structure. A comparison with CIU data

acquired for HSA dimers, including the 21^+ (Figure 2-1B), 22^+ (Figure 2-1C), and 23^+ (Figure 2-

1D), reveals both similarities and differences when compared to HSA monomer data. HSA dimer

CIU reveals 5 total features across all three charge states probed, and the centroid drift times of

these features, along with their CCS values, appear similarly consistent. For example, the RMSD

of the centroid time among each features is calculated as within 5%, indicating negligible

differences across all CIU of different charge states. We note that the HSA dimer and monomer

CIU data exhibits broad similarities in terms of the shapes of the features recorded, with each

structural change in the protein reported in a relatively sharp transition between CIU features.

We have further analyzed the ΔCCS values that characterize the CIU transitions captured for

HSA monomers and oligomers, and find that RMSD of the ΔCCS values ranges from 0.3% to

9%, further suggesting that HSA monomers and dimers are undergoing CIU in a similar

manner.[4]

Myriad clear differences exist, however, between HSA monomer and dimer CIU datasets.

Beyond the absolute differences in CCS, CIU50s and drift times for the ions involved, the total

number of features detected over the acceleration voltage range probed here is different between

the two ion classes, with dimers producing an additional CIU feature when compared to HSA

dimers. To explain this key difference, we are left with three potential explanations that comport

with previous work in this area: 1) that the CIU observed is monomer driven and domain

correlated, but the extra charge pool present within the HSA dimer relative to what is available in

the monomer at low charge states is causing sub-domain CIU events to occur, 2) that the extra

CIU features are reporting on unfolding events occurring in both HSA monomers within the

complex, which may or may not be domain correlated, or 3) that the dimer CIU observed is driven by a single monomer unfolding event, but is no longer correlated with the native domain structure of HSA. In an effort to parse these scenarios, we designed a set of experiments aimed at probing the level of domain correlation and monomer involvement in the CIU processes observed for HSA dimers.
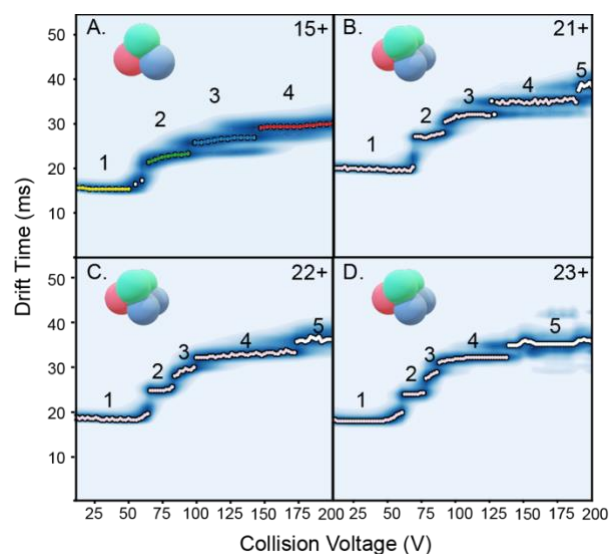
Figure 2-1. CIU fingerprints with annotated features for human serum albumin monomer of 15+ (A), human serum albumin dimer 21[+] (B), 22[+] (C), and 23[+] (D). Four features are detected for HSA monomer ions and five features are detected for dimer ions across different charge states.

Domain-specific chemical probes indicate HSA homodimer CIU is domain correlated

In order to evaluate the level of domain correlation associated with HSA dimer CIU, we deployed a series of HSA domain-specific chemical probes as CID markers to track the unfolding process of different domains. We used a similar strategy go study HSA monomer CIU previously, revealing probe CID events correlated at or near the onset voltage values of the CIU features observed.[4] Furthermore, these results indicted domain 2 (D2) to be the least stable and responsible for the lowest energy CIU event observed in HSA monomers, followed by domain 3 (D3) and domain 1 (D1) For this work, we chose hemin as the marker for D1 due to its high

binding affinity and specificity for HSA D1.[38,39] Indomethacin[40] was used as the marker for D2 and ANS[41] was used for tracking D3 based on similar criteria, although ANS can also access a low-affinity D2 binding site.

Upon mixing probe molecules in solution with HSA prior to nESI-IM-MS, we primarily observe 1:1 and 2:1 HSA-ligand complexes, and select the later for CID analysis. Figure 2-2 shows such CID results for 22^+ HSA dimer-ligand complex ions. As expected, hemin dissociation (Figure 2-2A) occurs between the transition from CIU feature 4 to CIU feature 5 of the HSA homodimer CIU fingerprint, which indicates the feature 5 is associated with D1 unfolding. In contrast, we observe indomethacin ejection (Figure 2-2B) during the initial phase of the HSA dimer CIU process at relative low collision voltages, implicating D2 in the generation of CIU features 2 and 3. Surprisingly, we observe a bimodal CID curve for ANS ejection (Figure 2-2C) with a first transition that falls between feature 1 and feature 2 and a second transition that mainly occurs at the end of feature 4. As mentioned above, ANS exhibits weak affinity for D2, and we interpret this first ejection event as evidence of ANS binding in this alternate site, with the majority of ANS attached to D3 which we previously used to assign the HSA monomer CIU feature 3 primarily to D3 unfolding.[41] Therefore, based on the data show in Figure 2-2, we were able to confirm the domain-correlated nature of HSA dimer CIU, and our analysis yielded likely identifications for CIU features 4 and 5, which appear to be associated with D3 and D1 respectively. Putatively, features 2 and 3 appeared to be associated with D2 based on the data shown in Figure 2-2, and we sought to confirm this assignment using a series of additional experiments described below.
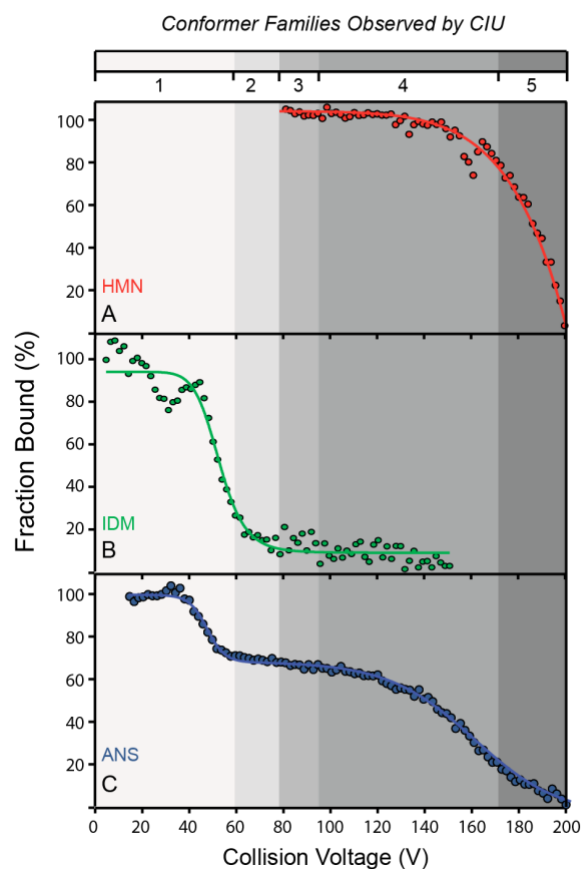
Figure 2-2. Correlation between HSA domain-specific chemical probes CID curves with 22+ HSA dimer CIU fingerprints. CID breakdown curves are fitted with sigmoid functions. Data sets are shown for hemin bound to HSA domain 1 (A), indomethacin bound to HSA domain 2 (B), and ANS bound to HSA domain 3 (C). All plots are overlaid on top of a grey scale indicating the voltage ranges where HSA dimer CIU features are observed.

Domain deleted constructs further illuminate the mechanism of HSA dimer CIU

To further probe the level of domain correlation observed in HSA homodimer CIU, we performed CID chemical probe experiments using a domain-deleted HSA construct comprised of only D1 and D2 (D12). The CIU fingerprint of the 13+ D12 monomer Contains three features, one fewer than is observed in full length HSA as observed previously (Figure I-1), consistent with our previously reported domain-correlated CIU mechanism for HSA monomers. Similarly, the CIU fingerprint we record for the 17+ D12 homodimer contains four features, 1 fewer than detected for the full-length HSA dimer. The number of features detected for domain deleted

36

HSA dimer constructs was consistent across a range of D12 dimer charge states. We note that the remaining features within the D12 dimer fingerprint bear a strong similarity to those detected for the full-length dimer, especially the stabilities and relative CCSs of features 2 and 3 in both datasets. Also, we note that the total number of features detected in our D12 dimer CIU dataset remains one greater than that observed for the D12 monomer, indicating that the additional feature we detect in full length HSA dimers is likely not due to D3. To investigate the domain-correlated CIU of D12 dimers further, we again deployed our domain-specific ligand CID assay, the results from which are shown in Figures 2-3B and C. As expected, bound hemin is ejected from D12 dimer ions in a manner correlated with the appearance of feature 4, suggesting this feature is related to D1 unfolding (Figure 2-3B). In contrast, Indomethacin is ejected from the complex primarily across features 2 and 3 (Figure 2-3C). Taken together, we surmise that both features 2 and 3 observed in across full length and domain-deleted HSA dimers are the result of D2 unfolding.

In addition to our experiments focusing on D12 homodimers, we also conducted a series of CIU experiments targeting D12-D123 heterodimer, where we incubated D12 with full length HSA (D123) in order to generate a noncovalent heterodimer comprised of 5 total domains (see Supporting Information Figure I-2). CIU fingerprints collected for ions across the 19^+, 20^+, and 21^+ charge states (Figure I-3), reveal highly correlated fingerprints to those captured for D12 dimers. For example, D12-D123 CIU data for 21+ shares a quantitatively similar fingerprint to those recorded for the 17^+ D12 homodimer ions, as determined through RMSD values being 9.02% obtained by a subtractive analysis of the two fingerprints obtained (Figure I-4). This analysis then reveals that the CIU of D12-D123 heterodimer ions is driven by the unfolding of D12 alone, with D123 operating as a structurally invariant observer to the process, further

supporting a model where CIU of multiprotein complexes more generally are best characterized

as monomeric unfolding events tethered to a larger, relatively static oligomer.
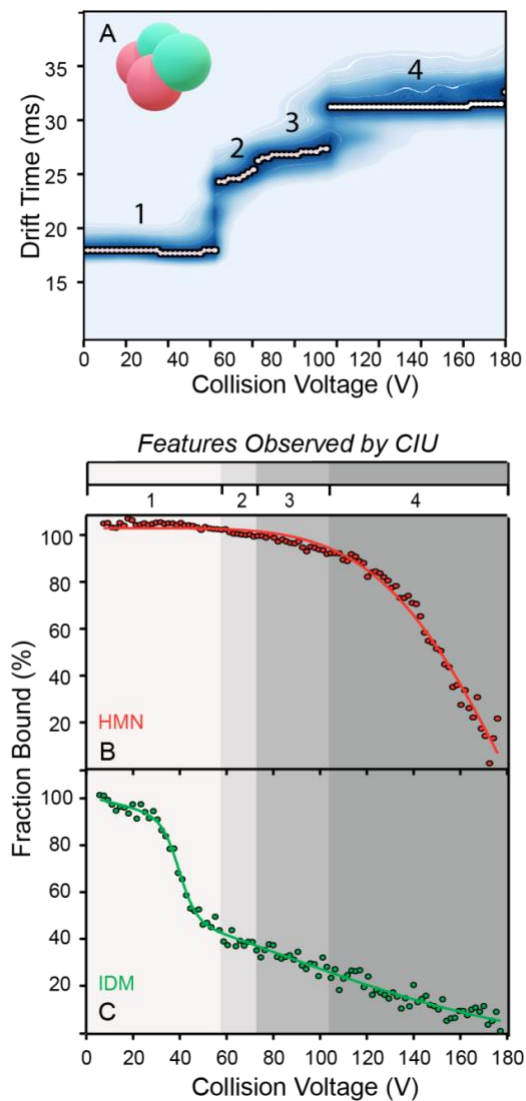
Figure 2-3. Correlation of HSA domain-specific chemical probes dissociation curve with D12 dimer CIU. (A) CIU fingerprints with annotated features for HSA D12 dimer construct. Illustration of CID breakdown curves for D12-ligand complexes. Data sets are shown for domain-specific chemical probe hemin bound with domain 1 (B) and indomethacin bound with domain 2 (C). The color scale indicates voltage ranges where different CIU features of D12 dimer are observed.

CIU-ECD of HSA D12

Prior reports have indicated that ECD fragment ion production can be linked directly to protein structure, whereby electron capture and bond cleavage is preferential at unstructured sites within a given protein sequence.[42–46] As such, we sought to combine ECD fragmentation with CIU in order to directly annotate the features observed within the latter datatype. When transferred to our modified drift tube IM-MS instrument equipped with ECD, monomers, we observe four total CIU features for 13+ D12 monomers (Figure 2-4A). Based on the CIU D12 monomer fingerprint obtained in these preliminary experiments, we then proceeded to collect ECD fragmentation data for D12 parent ions following quadrupole section of 13^+ ions at the following levels of CIU activation: 140, 265, 330, and 480V (Figure 2-4A, B).

For ECD data collected at a CIU acceleration of 140 V, the most intense signals recorded correspond to a z-ions series originating from the c-terminus of D12 (for detailed ID information, see Figure I-5). We further observed that some z-ions were identified along with their y-ions counterparts (Figure I-5), acting as further confirmation of our identifications.[23] At values above 1900 m/z we observed a cluster of signals corresponding to peptide fragments having between 3 and 5 charges that did not correspond to terminal fragments of D12 (gray shaded area in Figure 2-4B). ECD data collected at 265V of CIU acceleration voltage revealed a dramatically altered fragmentation pattern for D12. For example, the c-terminal ions observed were now limited to $z_8^+ - z_{11}^+$, the most intense ion signals in the series in our 140V dataset. In addition, a greater number of fragment ion signals were detected above 1900 m/z and with greater intensity relative both the terminal fragments observed under these conditions and the singles in this region present at 140V. When we increased the CIU acceleration voltage to a value of 300V, we noted that the previously observed z-ion series was completely absent, and the signals in the region above 1900 m/z continued to grow in intensity and complexity. When the CIU acceleration

39

voltage is increased to 400V, most of the fragment ions observed in ECD acquired at lower voltages have been replaced by new signals that do not match expected terminal fragment masses across expected fragment ion types. (See Supp I-6)

Using the approach outlined in the methods section above, we attempted to identify a filtered set of the ECD fragments observed at *m/z* above 1900, focusing on just those ions that exhibited intensity profiles that correlated with the appearance of CIU features 2, 3 and 4.Out of the 23 signals observed having charge states of either 4+ or 5+, our approach was only able to confidently identify two such ECD fragments, exhibiting a score > 70 (see supp Table I-7). One corresponded to an internal fragment within the D1 sequence (Figure 2-4B, red square) and the other to a D2 internal fragment (Figure 2-4B, green rhombus) as indicated in Figure 2-4C. Of the two, the D2 fragment appeared to be more strongly correlated with the onset of CIU, as this signal is not detected at voltages insufficient to initiate D12 unfolding. As such, this fragment sequence was subsequently used a restraint in our steered MD simulations.

Steered MD Simulations Reveal Putative Structures for D12 CIU Intermediates

Following the computational generation of compact starting structures for D12 in a range of charge isomeric states, SMD simulations (Figure 2-4D) were performed using two restraint sets: 1) all of D2 residues were allowed to undergo CIU; 2) only those residues identified by ECD analysis, limited to residues 271-385 in D12 sequence, were allowed to unfold. Broadly, all SMD simulations, produced D12 monomer models that agreed with experimental CCS values obtained from ejected monomer ions generated from D12 dimer CID (Figure I-6).

Following clustering, SMD D12 models exhibiting lowest energy were docked to compact models produced for HSA D12, and quantitatively evaluated. For those models produced using

our first set of restraints, which allowed all D2 residues to undergo CIU, model CCSs spanned a wide range and exhibited a multimodal distribution that encompassed experimental CCS values recorded for CIU features 2, 3, and 4 (Figure 2-4E), clearly illustrating that unfolding across both D12 subunits need not be invoked in order to produce CCS values large enough to encompass those observed experimentally. For those SMD models produced using our second set of restraints, which further limited CIU to just those residues identified in ECD experiments as linked to initial D12 CIU, we record a far narrower CCS distribution. Furthermore, this model set gives rise to a clearly bimodal distribution of theoretical CCSs that align well with the experimental CCS values obtained for CIU features 2 and 3 (Figure 2-4F). This remarkable level of correlation between experiment and theory, achieved by restraining models using two intendent pieces of experimental information, lends credence to an HSA dimer CIU model whereby D2 undergoes two stages of unfolding, followed by nominally single stage unfolding events for D3 and D1.
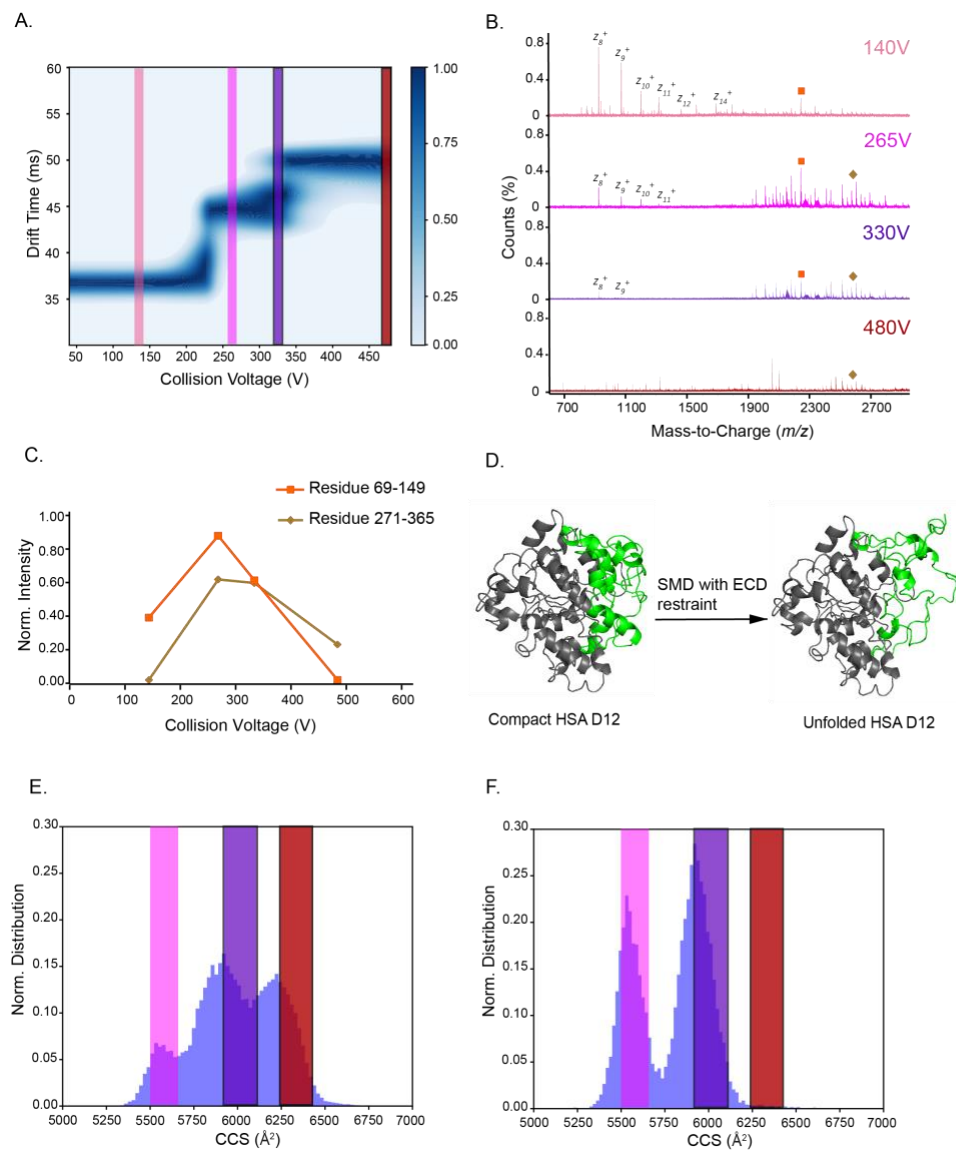
Figure 2-4. CIU-ECD of HSA D12. **(A)** A CIU fingerprint of the 13+ D12 monomer acquired on drift tube IM-MS is shown. D12 monomer CIU data exhibits four features. For each feature ECD fragmentation data was obtained, as indicated by the color-coded regions of drift time/collision voltage space (dark grey = 140V, green = 265 V, gold = 330 V, pink = 480 V). **(B)** ECD fragmentation is displayed as acquired from each acceleration voltage range as indicated in A. ECD fragments that originate from the c-terminus of D12 are labeled in red, while the only those internal fragments that we were able to identify with high confidence are marked with either a red square or a green rhombus that match in the data shown in panel C. **(C)** The normalized ion intensities recorded for ECD internal fragments identified in B are plotted against CIU collision voltage values. The plots are further annotated with the sequences associated with the identified fragments. **(D)** Models generated by MD simulations. The left shows a compact model of HSA D12 that agrees with native MS data. The right model is an unfolded model of HSA D12 using ECD fragments observed in (C) as a restraint for SMD. **(E)** Distribution of CCS of all HSA D12 dimer generated by docking compact models with D2 unfolded models. **(F)** Distribution of CCS of all HSA D12 dimer generated by docking compact models with unfolded models with the ECD restraint. The colored boxes in (E) and (F) indicates the CCS ranges which were also used in (A).

42

2.4   Conclusions

Here, we describe the most comprehensive mechanistic analysis of multiprotein CIU to date. By comparing the CIU fingerprints obtained for HSA monomer and homodimers, we find striking similarities between the features observed. These observations also extended to a domain deleted D12 construct and to SMD results, all of which strongly indicate that the CIU features observed for HSA dimer ions are primarily associated with a single HSA monomer unfolding event. Through an analysis of chemical probe CID results and CIU of domain-deleted HSA constructs, we confirm that the CIU of HSA dimers are domain-correlated, although serve to generate sub-domain CIU events, giving rise to a total of n+1 total transitions, where n is the number of domains in the HSA construct. Lastly, by performing MD simulations for HSA D12 dimer using ECD fragments as additional experimental restraints, we assign the sub-domain CIU events to D2, the least stable of the three HSA domains.

The MS approaches used here, and the conclusions reached associated with HSA dimer CIU, likely extend to the gas-phase unfolding of multiprotein complexes in general. Our work is consistent with asymmetric charge partitioning mechanisms of multi-protein complex CID that invoke the unfolding of monomers prior to ejection as product ions. The combination of CIU and ECD can clearly be expanded to include other proteins and complexes for annotation of gas-phase unfolding data in order to enable targeted uses of CIU experiments that will dramatically expand the utility of the technique for protein engineering. Finally, give the size and complexity of HSA monomers, it is likely that our general observations regarding domain-correlated, monomer driven CIU, likely expand to other multiprotein complex ions, laying a foundation for extracting domain-level information from multiprotein CIU. Clearly, more data and examples

will need to be acquired in order to validate CIU methods for enabling the construction of next-generation domain-level IM-MS models of protein quaternary structure.

2.5   References

(1)   Eschweiler, J. D.; Frank, A. T.; Ruotolo, B. T. *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 1991.
(2)   Benesch, J. L. P. *J. Am. Soc. Mass Spectrom.* **2009,** *20*, 341.
(3)   Hyung, S. J.; Robinson, C. V.; Ruotolo, B. T. *Chem. Biol.* **2009,** *16*, 382.
(4)   Eschweiler, J. D.; Martini, R. M.; Ruotolo, B. T. *J. Am. Chem. Soc.* **2017,** *139*, 534.
(5)   Hopper, J. T. S.; Oldham, N. J. *J. Am. Soc. Mass Spectrom.* **2009,** *20*, 1851.
(6)   Ruotolo, B. T.; Hyung, S.-J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V. *Angew. Chemie* **2007,** *119*, 8147.
(7)   Bornschein, R. E.; Ruotolo, B. T. *Analyst* **2015,**
(8)   Schwartz, B. L.; Bruce, J. E.; Anderson, G. A.; Steven, A.; Rockwood, A. L.; Smith, R. D. **1995,**
(9)   Loo, J. A. *J. Mass Spectrom.* **1995,**
(10)   Felitsyn, N.; Kitova, E. N.; Klassen, J. S. *Anal. Chem.* **2001,** *73*, 4647.
(11)   Benesch, J. L. P.; Aquilina, J. A.; Ruotolo, B. T.; Sobott, F.; Robinson, C. V. *Chem. Biol.* **2006,** *13*, 597.
(12)   Light-Wahl, K. J.; Schwartz, B. L.; Smith, R. D. *J. Am. Chem. Soc.* **1994,** *116*, 5271.
(13)   Jurchen, J. C.; Garcia, D. E.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **2004,** *15*, 1408.
(14)   Jurchen, J. C.; Williams, E. R. *J. Am. Chem. Soc.* **2003,** *125*, 2817.
(15)   Zhong, Y.; Han, L.; Ruotolo, B. T. **2014,** 9209.
(16)   Polasky, D. A.; Dixit, S. M.; Fantin, S. M.; Ruotolo, B. T. *Anal. Chem.* **2019,** *91*, 3147.
(17)   Vallejo, D. D.; Polasky, D. A.; Kurulugama, R. T.; Eschweiler, J. D.; Fjeldsted, J. C.; Ruotolo, B. T. *Anal. Chem.* **2019,** *91*, 8137.
(18)   Gadkari, V. V.; Rojas Ramirez, C.; Vallejo, D. D.; Kurulugama, R. T.; Fjeldsted, J. C.; Ruotolo, B. T. *Submitt. Rev.* **2020,**
(19)   Fort, K. L.; Cramer, C. N.; Voinov, V. G.; Vasil'Ev, Y. V.; Lopez, N. I.; Beckman, J. S.; Heck, A. J. R. *J. Proteome Res.* **2018,** *17*, 926.
(20)   Williams, J. P.; Morrison, L. J.; Brown, J. M.; Beckman, J. S.; Voinov, V. G.; Lermyte, F. *Anal. Chem.* **2020,**
(21)   Strohalm, M.; Kavan, D.; Novák, P.; Volný, M.; Havlíček, V. *Anal. Chem.* **2010,**
(22)   Senko, M. W.; Beu, S. C.; Mclafferty, F. W. *J Am Soc MAss Spectrom* **1995,** *6*, 229.
(23)   Zhurov, K. O.; Fornelli, L.; Wodrich, M. D.; Laskay, Ü. A.; Tsybin, Y. O. *Chem. Soc. Rev.* **2013,** *42*, 5014.
(24)   Lermyte, F.; Valkenborg, D.; Loo, J. A.; Sobott, F. *Mass Spectrom. Rev.* **2018,** 1.
(25)   Xiaojuan Li, Cheng Lin, Liang Han, Catherine E. Costello, and P. B. O. *J. Am. Soc. Mass Spectrom.* **2010,** *21*, 646.
(26)   Mormann, M.; Eble, J.; Schwöppe, C.; Mesters, R. M.; Berdel, W. E.; Peter-Katalinić, J.; Pohlentz, G. *Anal. Bioanal. Chem.* **2008,** *392*, 831.
(27)   Ganisl, B.; Breuker, K. *ChemistryOpen* **2012,** *1*, 260.
(28)   Lee, Y.; Oh, H. Bin. *Mass Spectrom. Lett.* **2010,** *1*, 5.

(29) Sugio, S.; Kashima, A.; Mochizuki, S.; Noda, M.; Kobayashi, K. *Protein Eng.* **1999,** *12,* 439.

(30) Lee, M.; Lee, Y.; Kang, M.; Park, H.; Seong, Y.; Sung, B. J.; Moon, B.; Oh, H. Bin. *J. Mass Spectrom.* **2011,** *46,* 830.

(31) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009,**

(32) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. *J. Comput. Chem.* **2008,**

(33) Marklund, E. G.; Degiacomi, M. T.; Robinson, C. V.; Baldwin, A. J.; Benesch, J. L. P. *Structure* **2015,** *23,* 791.

(34) Ewing, S. A.; Donor, M. T.; Wilson, J. W.; Prell, J. S. *J. Am. Soc. Mass Spectrom.* **2017,**

(35) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999,**

(36) Kabsch, W. *Acta Crystallogr. Sect. A* **1976,**

(37) Chaudhury, S.; Lyskov, S.; Gray, J. J. *Bioinformatics* **2010,**

(38) Buchholz, T. J.; Geders, T. W.; Bartley, F. E.; Reynolds, K. A.; Smith, J. L.; Sherman, D. H. *ACS Chem. Biol.* **2009,** *4,* 41.

(39) Zunszain, P. A.; Ghuman, J.; Komatsu, T.; Tsuchida, E.; Curry, S. *BMC Struct. Biol.* **2003,** *3,* 1.

(40) Metallo, S. J. *Curr. Opin. Chem. Biol.* **2010,** *14,* 481.

(41) Bagatolli, L. A. *J. Fluoresc.* **1996,** *6,* 33.

(42) Breuker, K.; Oh, H. Bin; Horn, D. M.; Cerda, B. A.; McLafferty, F. W. *J. Am. Chem. Soc.* **2002,**

(43) Horn, D. M.; Breuker, K.; Frank, A. J.; McLafferty, F. W. *J. Am. Chem. Soc.* **2001,**

(44) Lermyte, F.; Sobott, F. *Proteomics* **2015,**

(45) Zhang, H.; Cui, W.; Wen, J.; Blankenship, R. E.; Gross, M. L. *Anal. Chem.* **2011,**

(46) Zhou, M.; Liu, W.; Shaw, J. B. *Anal. Chem.* **2020,** *92,* 1788.

Chapter 3 Ion Mobility-Mass Spectrometry and Electron Microscopy Reveal Protein-Protein and Protein-Substrate Interactions between co-Dependent Oxidative Enzymes in the Tirandamycin Biosynthesis Pathway

Kinshuk Raj Srivastava[2#*], Chunyi Zhao[1*], Yogan Khatri[2], Min Su[2], Rachel Martini, Rosa Vasquez Espinoza[2], Brandon T. Ruotolo[1], and David H. Sherman[2]

*These authors contributed equally to this work

Abstract

Biosynthetic enzymes have evolved to engage in transient interactions with partner proteins to catalyze diverse reactions on complex substrates. To ensure efficient performance, these multi-component enzymes co-localize and are presumed to form dynamic complexes. In this report, we probe the transient physical association between co-dependent enzymes TamI (an iterative cytochrome P450 monooxygenases) and TamL (a flavin adenine dinucleotide-dependent oxidase), involved in late-stage oxidation of tirandamycin antibiotics.[1] We employed an integrated experimental approach that combines native ion mobility mass spectrometry (IM-MS), negative stain Electron Microscopy (EM), and biochemical studies to characterize the TamI-TamL complex, probe its topology, and biocatalytic activity. Co-expressed sample containing TamI and TamL proteins, as well as an engineered fusion construct were shown to catalyse the conversion of TirC to TirA & TirB following multi-step oxidative cascade. To characterize the unusual reactivity of TamI, we employed collision-induced dissociation (CID) and collision-induced unfolding (CIU) to probe mechanistically the tirandamycin bound and free form of TamI. The results provided new information about the changes in stability of TamI in response to substrate and heme co-factor binding. Upon activation, we observed that at lower energy, the

loop region, which is the "lid" of the TamI substrate pocket unfolds while loss of heme from the binding pocket occurs at higher energies. Finally, our results demonstrate that TamI and TamL interact and form a biocatalytically competent heterodimeric complex *in vivo*.

3.1 Introduction

Studies of natural product biosynthetic pathways have provided important insights into the assembly and tailoring of bioactive secondary metabolites. Many of these molecules are clinically approved pharmaceuticals and continue to inspire the search for new medicines and other high-value chemicals.[2–5] Biosynthetic pathways typically involve sequential catalytic events, and often their function is coordinated by transient, noncovalent physical interactions. However, the temporal and spatial formation of supramolecular complexes of biosynthetic enzymes is underexplored due to a lack of robust analytical methods for their detection and analysis. It is generally assumed that multi-enzyme complexes sequester reactive biosynthetic intermediates and facilitate their efficient translocation among active sites, which maximizes the catalytic performance of the pathway. [6–13] Owing to their dynamic and transient nature, the protein-protein interactions among individual biosynthetic enzymes remain largely uncharacterized. Research over the past three decades shows a rapidly increasing interest in identifying and characterizing diverse biosynthetic pathways from microbes, plants and animals. Investigating transient protein-protein and protein-substrate interactions in multi-component enzyme systems will advance our understanding of the forces that drive substrate positioning and shuttling of intermediates between enzyme active sites. This knowledge will help enable development of rational approaches to efficiently exploit biosynthetic enzymes to engineer new reaction cascades for assembly of diverse pharmacophores and other high-value chemicals.[8,14–22]

The dynamic and transient nature of protein-substrate and protein-protein interactions within biosynthetic pathways presents a significant challenge for their detection and characterization by high-resolution structural biology tools.[23] The importance of natural product biosynthesis has spawned many approaches aimed at ameliorating such challenges. For example, mass spectrometry (MS) based structural biology probes have emerged as a family of technologies capable of capturing transient protein-substrate and protein-protein interactions.[24] Briefly, ion mobility (IM) separates ions through a fixed-length drift region filled with inert gas, under the force of a weak electric field. The time an ion spends in this region, referred to as drift time (t_D) is effected by charge and the number of ion collisions with the background gas. These drift times can be converted to rotationally averaged collision cross sections, a measure of the effective surface area of the ion.[25,26] Since native structure and non-covalent interactions are preserved, gas phase activation approaches such as collision induced dissociation (CID), and collision induced unfolding (CIU) can be utilized to study the binding strength of protein-protein and protein-substrate complexes, and monitor the changes in structural folds and stability.[27,28] Specifically, CIU experiments have been employed to explore domain-specific unfolding of gas-phase serum albumins[29] as well as assess protein stability shift triggered by ligand binding[28,30]. This gas phase stability assay is analogous to solution phase unfolding measurements and has proven useful to analyze transient, low abundance species that otherwise elude characterization by bulk solution methods. Despite the potential insights that IM-MS may provide on the function and structure of biosynthetic enzymes, relatively few reports have applied this technology to these types of protein complexes.[31]

Previously, we have reported the *in vitro* functional characterization of two oxidative enzymes comprised of a biosynthetic cytochrome P450 (TamI) and a flavin adenine dinucleotide-

dependent oxidase (TamL). They act co-dependently to catalyse late-stage formation of the C10

keto, C11/C12 epoxide, and C18 hydroxyl functionality on the tirandamycin C (TirC) bicyclic

ketal moiety to yield potent antibiotics TirA and TirB (Figure 3-1a). The resulting structural

variations were found to be a vital contributor to maximizing Tir antibiotic potency against

vancomycin-resistant *Enterococcus faecalis*. The multifunctional TamI cytochrome P450

enzyme catalyzes initial C-H hydroxylation of TirC at C10, to form TirE. Subsequently, TamL

catalyzes oxidation of the C10 hydroxyl to carbonyl to yield TirD. Further epoxidation and

hydroxylation catalyzed by TamI complete the oxidative cascade to yield TirB. Although

mechanistic details about Tir biosynthesis are beginning to emerge[32], little is known about the

basis for efficient exchange of intermediates between the active-site pockets of TamI and TamL.

We hypothesized that these two proteins form a complex to accelerate the processing of

intermediates of the oxidative cascade to generate the highly tailored tirandamycin TirA and

TirB antibiotics.

In this report, we employed a range of analytical methods, including IM-MS, Electron

Microscopy (EM), and in vitro enzymatic assays to probe the functional protein-substrate and

protein-protein interactions adopted by TamI and TamL. We measured the affinity of substrate

binding to both enzymes, and further elucidated the effects of TirC and heme cofactor binding on

the structure of TamI. Our findings demonstrated that TamI exists as a monomer and TamL as a

dimer. Further, we measured the gas-phase ligand ejection of both TirC and heme from the TamI

monomer. By mapping these observations onto the CIU data collected for this CYP450, we were

able to link specific unfolding events associated with both the Tir substrate and heme binding

pockets. Thus, the CIU of the mobile loop lid region of TamI unfolds at the lowest energy,

followed by a loss of structure in the heme binding pocket at higher energies. In addition, we

directly observed heterodimer and heterotrimer complexes between TamI and TamL. Furthermore, single particle negative stain EM experiments revealed a stable heterodimeric complex and overall topology of the TamI:TamL species.

3.2   Results

Assessing the oligomeric distributions of TamI and TamL under native conditions

Previous work demonstrated that TamI and TamL act co-dependently through the repeated exchange of TirD and TirE intermediates (Figure 3-1a), which catalyzes oxidation of the initial TamI substrate, TirC, to generate TirA and TirB.[1] Although the functional cooperation between TamI and TamL was evident, direct support for a stable TamI-TamL protein complexes remained to be established. When we examined TamI and TamL individually using IM-MS, we observe that both proteins exist as a mixture of monomers and dimers under native conditions (Figure 3-1). For TamI, we observed that the monomer dominates the oligomer distribution, showing a 7-fold preference over the dimer. In contrast, TamL primarily adopts a dimeric state, which is 4-fold more abundant than the TamL monomers detected in our IM-MS data. All TamI and TamL oligomer assignments were confirmed using tandem MS analysis.

Substrate and heme binding to apo TamI

Following the analysis of TamI and TamL native oligomers, we next assessed the substrate binding affinity of TamI using native IM-MS. Our MS data confirmed prior measurements of the binding ability between both TirC and TirD with TamI (Figure II-1)[1]. By measuring the relative intensities of the two bound and apo forms of TamI over a range of compound concentrations in our native IM-MS data, we determined a K_d value of 2.7±0.4μM for TamI−TirC binding, and

29.1±4.3µM for TirD–TamI complexes. Both values agree well with previous measurements.[1]

To further investigate the binding of heme and TirC to apo TamI, CID experiments were

conducted by isolating TirC bound TamI at 13+ charge state (Figure 3-1C). The data revealed

that the TirC intermediate is ejected from the protein at lower activation energies than the heme

co-factor. This result aligns well with previously reported x-ray analysis of TamI binding pockets

for heme and TirC, which indicates coordinate versus non-covalent bonding for heme and TirC,

respectively.
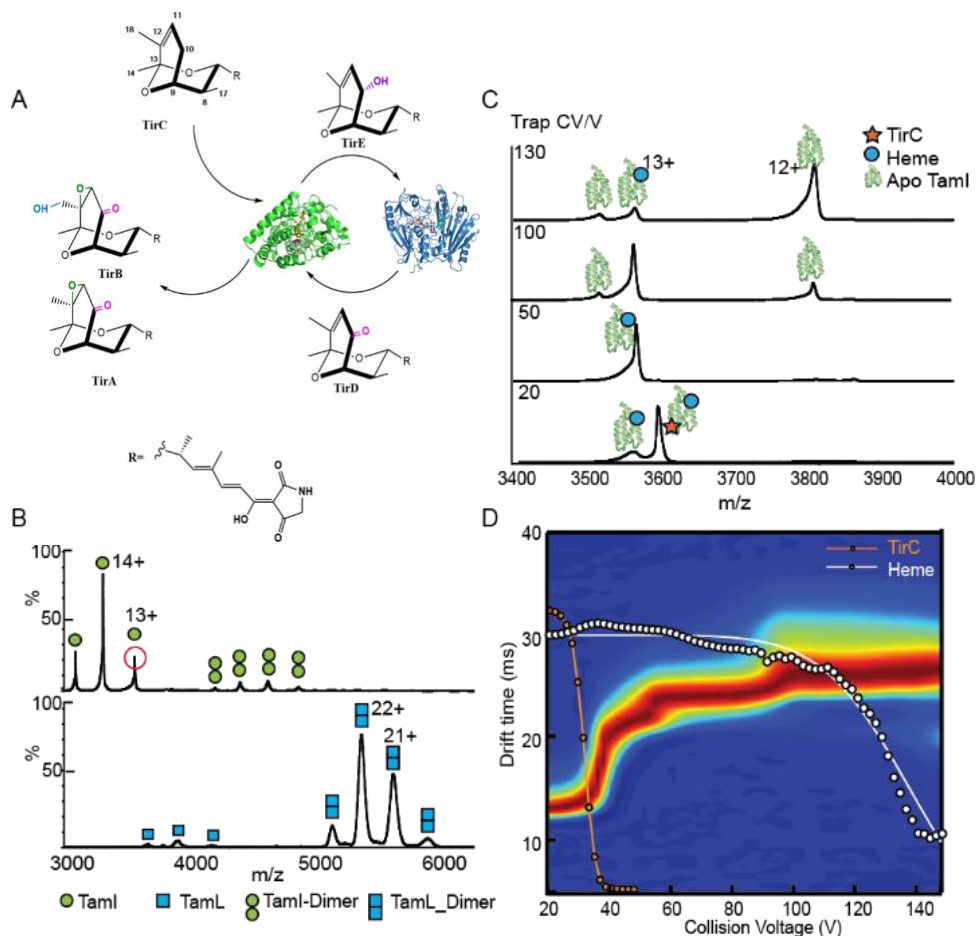
Figure 3-1. (a) Schematics of individual steps in tirandamycin oxidative tailoring. Complete oxidative modifications are highlighted in red. TamI (a P450 enzyme) mediates the initial hydroxylation of TirC at C10, to form TirE, after which TamL oxidizes TirE to the ketone (TirD). Subsequent epoxidation (C11-C12 alkene) and C18 hydroxylation,

both catalyzed by TamI, complete the cascade. (b) MS Spectra of TamI and TamL. (c) Isolated 13+ TamL-TirC MS spectra at different activation energy. (d) CIU & CID correlation map of TamI-TirC at 13+.

Correlations between collision induced unfolding and collision induced dissociattion data allows for the annotation of the TamI P450 unfolding pathway

Previous reports have revealed the ability of CIU (Collision Induced Unfolding) to detect minor differences in protein stability due to changes in structure at the domain level.[29] In order to reveal the domain correlated CIU mechanism of TamI, we sought to map ligand ejection data from CID (Collision Induced Dissociation) with different structural features observed by CIU for the same ions. Upon slowly increasing the activation energy experienced by 13+ TamI ions (Figure 3C), TirC ejection was observed, followed by heme ejection at higher energies. This data enabled correlation of the first CIU transition observed to the region surrounding the TamI substrate binding domain, and the last transition to the heme binding pocket. Similar observations were recorded for TirD−TamI complexes and across all charge states observed for both protein-substrate complexes. Since TirD and TirC have significantly different binding affinities for TamI, these data enabled us to associate directly the ligand ejection to structural changes in the enzyme during CIU, rather than as a product of the interaction strength between the ligands and protein.

CIU Reveals Shifts in TamI stability upon substrate binding

Having structurally-annotated the TamI CIU fingerprint data, we then investigated TamI stability shifts triggered by substrate binding.[33] Thus, we recorded CIU50 values (defined as the voltage at which 50% of a relatively compact state of the protein transitions to a more extended state) to assess overall TamI stability.[34–36] Figure 3-2, displays CIU fingerprints and measured CIU50 values for both holo- and TirC bound 12+ TamI protein ions. The CIU50 for holo-TamI was

found to be 47.8±0.2V, and 49.8±0.1V for TamI−TirC complexes, an increase of 4.2% over

heme-bound TamI. These differences were also visualized through difference analysis of the

CIU data, which revealed a large RMSD value (9.05%) for comparisons between holo- and TirC

bound TamI. These results were consistent across a range of TamI-ligand complex charge states

(Figure II-3). Both CIU50 values and total fingerprint RMSD values confirmed that TirC binding

increases the stability of TamI. Subtle changes in TamI structure induced by substrate binding,

and observed through superimposing x-ray crystal structures of substrate-free and substrate-

bound forms of TamI (Figure 3-2D) likely reflect the observed stability shift in the CIU data.
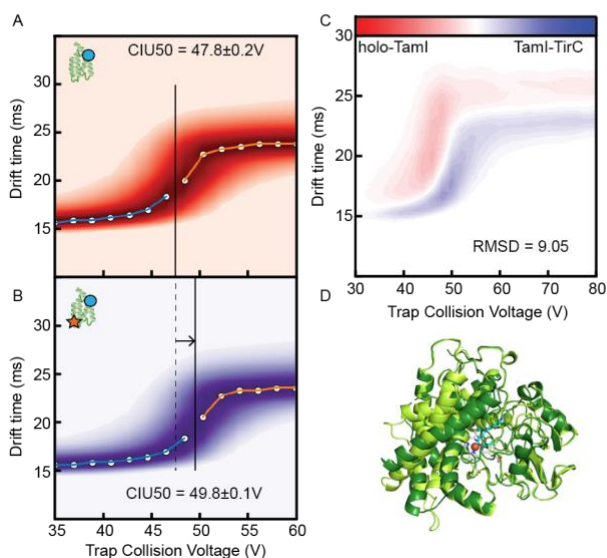
Figure 3-2. (a) Holo-TamI unfolding fingerprint. (b) TamI-TirC unfolding fingerprint. (c) CIU comparison plot. (d) Superimposed crystal structure of substrate free and substrate bound form of TamI (PDBID 6XA2)

Evidence for TamI−TamL Heterocomplex Formation

We tried to reconstitute TamI:TamL complex *in vitro* by incubating individually purified protein

in varied stoichiometry followed by gel filtration chromatography, but we couldn't isolate the

heterodimer which could be due to low affinity of transient protein:protein complex. Next, we

incubated TamI and TamL protein with a heterobifunctional crosslinker (LC-SMCC) and subjected the sample to native IM-MS analysis. Our results (Figure II-5) revealed that a combination of chemical cross-linking and native IM-MS enabled capture of the TamI:TamL complex for structural analysis. In addition to the TamI:TamL heterodimer, we also observed TamL homodimer and homotrimer, consistent with previous reports of TamL self-oligomerization. In an effort to further characterize the hetero-oligomerization of TamI and TamL, we co-expressed these enzymes in *E. coli*, and subsequently purified and characterized them using IM-MS and Electron Microscopy (EM). For the co-expressed and purified sample we will use terminology "*in vivo* co-expressed complex" while the complex assembled with individually purified protein in vitro we define them as "*in vitro* reconstituted complex" in the manuscript. The data showed that the association between TamI and TamL is sufficient when these enzymes are co-expressed and we could purify stable complexes by Ni^{2+}-NTA and size exclusion chromatography. The native IM-MS spectra (Figure 3-3A) acquired for these samples displayed signals corresponding to TamL monomers, homodimers, and homotrimers, as well as TamI monomers. In addition, heterocomplexes corresponding to the TamI-TamL heterodimer, and TamI-$TamL_2$ heterotrimer were detected. As discussed above, each oligomeric state identified above is corroborated by tandem MS data (Figure II-4). Collision cross sections (CCSs), protein size values determined by IM-MS are consistent with predicted CCS values for TamI (monomer, -0.28% deviation) and TamL (deviations of -3.91% and -0.63% for monomer and dimer, respectively) based on CCS values predicted from the corresponding protein crystallography data (TableII-1). Moreover, the measured CCSs indicated a 7% decrease in size for the heterodimer compared to the TamL homodimer. Notably, since the MS spectrum reflects the authentic population of *in vivo* co-expressed protein complexes, it reveals that $TamL_2$

homodimer dominates the oligomer population with the relative intensity over 3-folder greater than the corresponding TamI:TamL heterodimer.
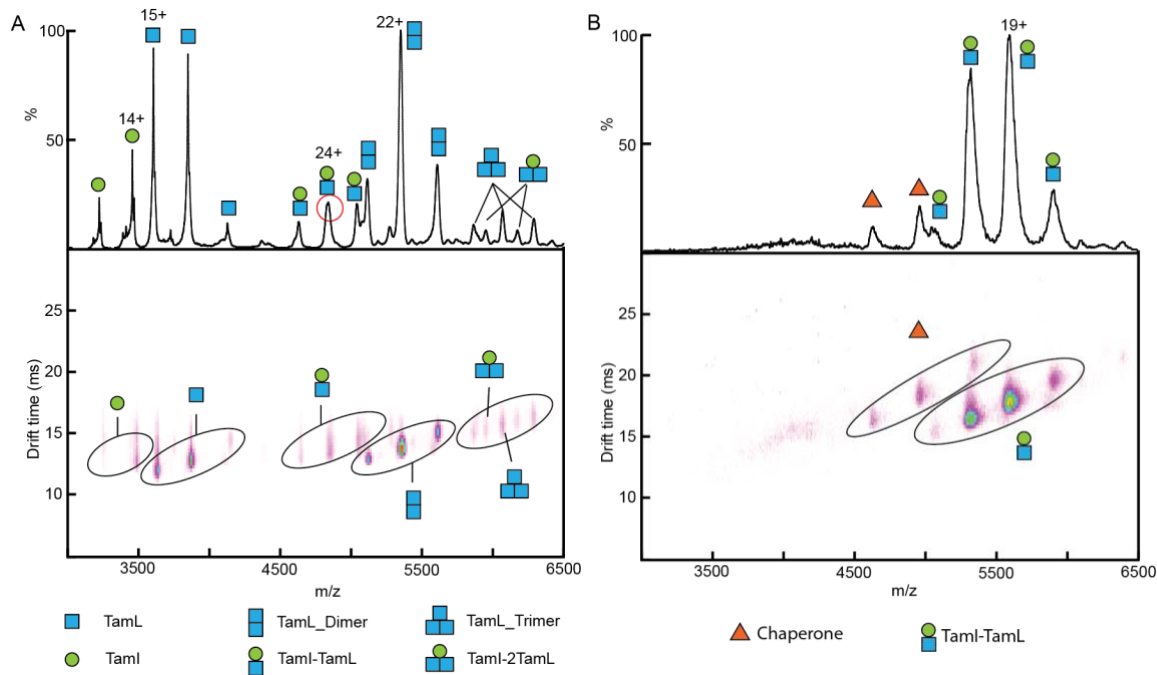
Figure 3-3. (a) Mass and Ion Mobility Spectra of TamI-TamL in vivo derived co-expression sample. (b) Mass and Ion Mobility Spectra of in vitro derived TamI-TamL sample.

Next, we performed single particle negative staining EM in an effort to visualize the TamI:TamL heterodimer complex and topology. Cell lysate from *E. coli* co-produced TamI/TamL were purified by size exclusion chromatography (Figure II-4, panel D). The fractions corresponding to dimers were pooled for EM imaging. The 2D class averages of the co-produced TamI/TamL samples exhibited well-defined classes (Figure 3-4, middle panel left two columns), along with a set of that display less-defined features an overall similar topology (Figure 3-4, middle panel right three columns). Since the gel filtration purified dimer fractions of co-expressed sample is a mixture of homo- and hetero-dimer, we expect that the panel of 2D class averages will have representative 2D classes corresponding to both homo- and hetero-dimer. Therefore, we sought to separate the heterodimer classes by excluding the TamL$_2$ homodimer. We rendered a simulated electron density

map from the reported TamL$_2$ homodimeric x-ray structure (UCSF Chimera), and subsequently generated the projections (EMAN2 e2project3d.py) to compare with the experimental 2D class averages. These projection images clearly resemble a subset of the 2D averages (Figure 3-4, middle panel left column), indicative of the TamL$_2$ homodimer population in the co-expressed sample.
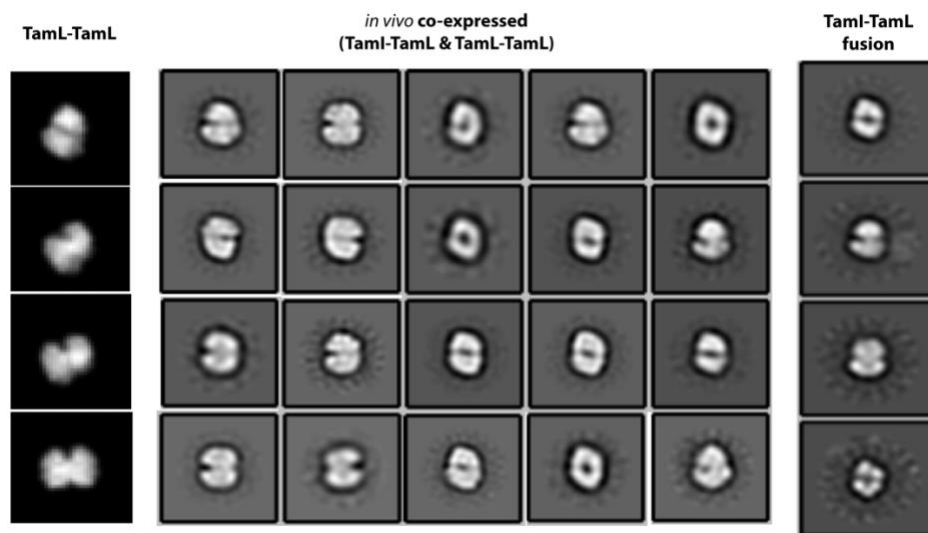
Figure 3-4. Representative 2D class averages of purified a) in vivo co-expressed sample, b) TamI-TamL fusion construct. The simulated projection map of TamL homodimer from their crystal structure.

Investigation of fusion construct further supports TamI:TamL complex formation

To further probe structural features of the TamI:TamL heterodimer, we generated a fusion construct using a 23-amino acid flexible linker between the two proteins. Comparison of the co-expressed and purified dimeric fractions with the TamI:TamL fusion protein showed that they possessed the same retention time in size exclusion chromatography (Figure 3-3 and Figure II-6). The molecular weight was consistent with the predicted molecular mass calculated from the amino acid sequences. MS analysis on the other hand (Figure 3-3B) revealed that the fusion construct generated a single TamI-TamL heterodimer whereas the TamI:TamL produced through *in vivo* co-expression was comprised of six different oligomeric states. Moreover, the TamI-TamL

heterodimeric fusion protein, and the *in vivo* co-expressed heterodimer showed highly similar CCS (4% difference). In addition, we measured the FWHM (Full Width at Half Maximum) of mobility distribution and observed clear differences among heterodimer samples generated by *in vivo* co-expression, cross-linking, and the fusion construct (Table II-2). The TamI:TamL fusion construct sample exhibited the highest resolution, and observed the lowest resolution *in vivo* co-expressed samples. To further develop our understanding of the fusion protein structure, we performed EM imaging experiments. Interestingly, we observed the globular shape of particles in EM images (Figure 3-4) which supports our hypothesis that TamI and TamL interact and form an enzyme-enzyme complex to facilitate the multi-step oxidative cascade for the conversion of TirC to TirA and TirB. These 2D class averages likely resemble a subset of structures from the co-expressed sample (Figure 3-4, middle panel right 2 columns).

Coupled TamI:TamL activity

We tested the catalytic activity of *in vivo* assembled TamI:TamL complex and the corresponding recombinant TamI-linker-TamL fusion protein (Figure 3-5). We observed that both complexes are catalytically active and are able convert TirC to TirA along with smaller amounts of TirB, which is the terminal product of the oxidative cascade. Results presented here along with the previous studies demonstrate that the complex, non-covalent or covalent fusion construct, between TamI and TamL, can efficiently catalyze the consecutive steps of an unusual multi-step oxidative cascade for post-assembly-line tailoring of bioactive tirandamycins.
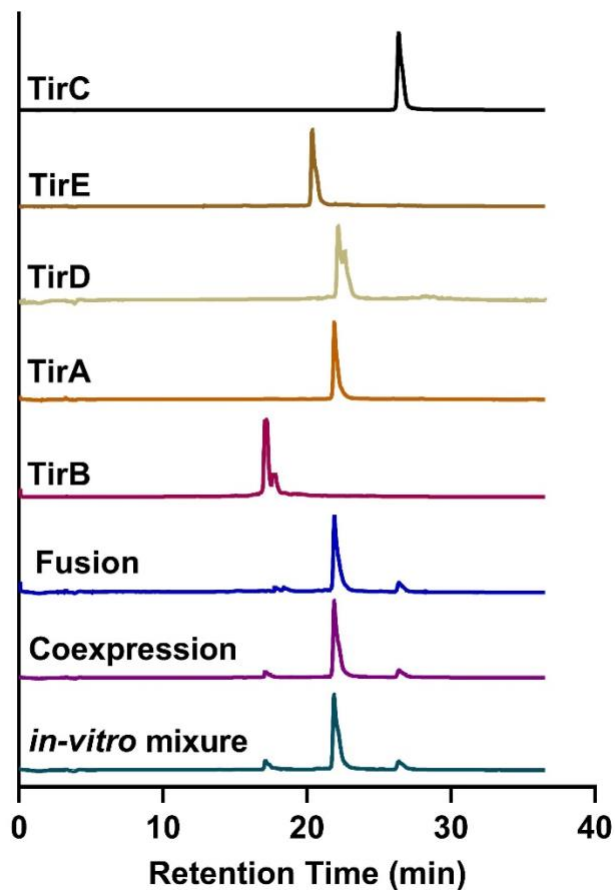
Figure 3-5. In vitro reconstitution of TamI and TamL mediated oxidation steps by TamI:TamL complexes that include recombinant fusion protein, purified in vivo expressed complex along with the *in vitro* assembled complex of TamI and TamL. Peaks were identified by MS and comparison to authentic standards (TirC, TirE, TirD, TirA, and TirB). Traces indicate absorbance at 354 nm.

3.3 Discussion

Natural product pathways are typically comprised of multi-enzyme systems that include a complex network of interactions to mediate processing of substrates along the path to a fully formed metabolite. Important examples have been demonstrated functionally and also at a structural level for HMG synthase and its cognate acyl carrier protein in the curacin pathway, and for the post-PKS co-dependent tailoring enzyme complexes in gilvocarcin and mithramycin.[37] Based on growing evidence about weak enzyme-enzyme complexes, it is apparent that enzyme assemblies are a cardinal feature of metabolism across all domain of life,

and occurs in central metabolisms as well as specialized metabolisms of plants and fungi. Currently, the biochemical relevance of pairwise interactions between sequential enzymes and significant clustering of enzymes in different biosynthetic pathways remains underexplored. It is commonly believed that the formation of dynamic enzyme-enzyme complexes promotes substrate channelling, which is a facilitated transfer of intermediate metabolites from one enzyme to another enzyme without equilibration of intermediates to bulk aqueous solvents. This phenomenon is presumed to help protect labile or toxic intermediates from being released. This also has implications for regulating metabolic flux at a branch point in the biosynthetic pathway. Thus, channelled intermediates avoid being utilized by competing enzymes at biosynthetic branch points. As proposed previously, two enzymes (TamI and TamL) are efficiently catalysing four steps of oxidative cascade for production of bioactive tirandamycins, we hypothesized that these enzyme might be interacting to facilitate the exchange of intermediates in their active pockets to avoid the loss of intermediates through diffusion in bulk solvent. To test our hypothesis, we gathered evidences by combination of molecular biology, IM-MS and EM techniques and our results demonstrate that TamI and TamL forms complex *in vivo*. Furthermore, measured CCSs indicated a 7% decrease in size for the heterodimer (TamI:TamL) when compared to the TamL homodimer (PDB: 2Y08, 2Y3R) suggesting that a structural rearrangement occurs upon heterodimer formation. Observation of globular topology and catalytic activity of fusion construct provides further support that TamI and TamL interacts to effectively catalyze the conversion of TirC to bioactive tirandamycines.

Tirandamycin congeners (Tir A, TirB, TirC, TirD, TirE) vary from one another based on the extent of oxidative tailoring in the bicyclic ketal moiety, which is key for their biological activity. The iterative TamI oxidative C-H functionalization cascade and the co-dependency of

TamI and TamL for the TirE → TirD oxidation led us to hypothesize that these two enzymes might interact to accelerate the processing of intermediates to yield final products TirA and TirB. Using an integrative approach that combined IM-MS and negative staining EM, we demonstrated that TamI and TamL forms a weak, transient protein-protein complex to facilitate the multi-step oxidative tailoring process. We captured the TamI:TamL complex through protein production in *E. coli*. The engineered TamI-linker-TamL fusion construct was also observed to have a similar size and shape compared to the native dimers. In the EM images of TamI-linker-TamL fusion construct, the two domain corresponding to the enzymes are in close proximity, which could be due to interactions across the interfaces of TamI and TamL. Further, we tested the catalytic activity of *in vivo* co-expressed TamI:TamL native complex, and the engineered TamI-linker-TamL fusion construct. Both complexes are functionally active and able to catalyze all steps of the oxidative cascade for converting TirC to TirA and TirB. These observations are consistent with previously reported *in vitro* reconstitution studies. Taken together, the results strongly support our hypothesis that transient physical association between TamI and TamL is an important feature of the tirandamycin biosynthetic pathway. Perhaps, due to the significant contribution of oxidative tailoring to antibiotic potency and the inability of TamI to efficiently catalyze TirE to TirD conversion drove recruitment of the flavoenzyme to evolve TamI/TamL co-dependent oxidative cascade.

Utilizing this state-of-art technique IM-MS, we measured the binding affinity of the tirandamycin substrates to TamI P450, and obtained detailed information about the changes in protein conformation and stability upon substrate binding. Our binding affinity measurements indicate that TamI strongly binds TirC and TirD, and these results further support the proposed role of TamI in TirC→TirE and TirD→TirA conversion. We were unable to detect the complex

of TamI with TirE and TirA, which could be due to the less binding affinity of TamI for these tirandamycins which are the products of TamI. This observation is further supported by our previous reported UV spectroscopy based measurement of K_m values, where K_m for Tir A and TirE is ~10 times higher than the TirC and TirD.[1]

IM-MS enables probing of free energy differences of stabilization of protein structure due to substrate and co-factor binding.[38–41] Here, we exploited the unique attribute of IM-MS techniques and examined the effect of tirandamycin substrate and P450 enzyme heme co-factor binding to the TamI conformational energy landscape and overall structural stability. Our data indicates that at lower collision voltages, TirC dissociates from the protein, suggesting that the loop region, which is the "lid" of TamI substrate pocket unfolds first. While the loss of heme co-factor was only observed at rather higher activation voltages, indicating the remote helical bundle comprising the heme binding pocket undergoes unfolding at later stage with higher applied activation voltages. Thus, these data highlight the protective effect of the heme cofactor on the stability of the TamI P450 structure. We envisage that these results could have broad implications on the stabilizing effect of substrate and heme cofactor for other P450 enzymes and is consistent with previous observations.[42–44] There are selected studies that investigate the folding/unfolding energy landscape of heme-proteins such as cytochrome C, hemoglobin, and myoglobin, which provide insight about the effect of substrate and heme cofactor on hemoproteins structure. However, there is a paucity of information on the stabilization of cytochrome P450 enzymes by heme cofactor despite their significant role in metabolism of pharmaceuticals and xenobiotics. Unlike another hemoproteins, the heme in cytochrome P450s is deeply buried in the hydrophobic pocket. Usually, cytochrome P450s have hydrophobic active site cavity for substrate binding, and heme prosthetic group is buried deep inside and coordinated

with a highly conserved cysteine.[43–46] IM-MS enabled us to develop qualitative picture of unfolding landscape of a P450 enzyme for the first time. The results presented here further supports the structural insight that heme cofactor is buried deep insight, therefore, higher energy required for unfolding of the heme binding domains and breaking of Cys-heme bond to observe ejection of heme.

3.4   Conclusion and outlook

In the present study, we captured and structurally characterized the transient complex between co-dependent oxidative enzymes, TamI and TamL, which catalyze late-stage multi-step oxidative tailoring of tirandamycin natural products. Our study highlights the functional significance of enzyme-enzyme complexes during the biosynthesis to facilitated the transfer of metabolic intermediates from one enzyme to another. Further, we created the fusion construct by linking TamI and TamL *via* a non-interacting flexible linker and investigated the structure and catalytic activities. Results demonstrate that fusion construct adopt globular structure and is catalytically active. This results further provide supportive evidence of TamI ad TamL complex formations during biosynthesis of bioactive tirandamycins. Additionally, detailed IMMS based characterization of multi-functional enzyme, TamI, provide evidence of substrate binding assisted stabilization of TamI structure. IM-MS further helps to elucidate the conformational energy landscape of TamI, a multifunctional cytochrome P450 enzyme in which substrate binding domain unfolds at lower energy and heme binding domain unfolding happens at higher energy indicating the protective role of heme prosthetic group in stabilization of P450 enzyme's structure. The insight gained here will be useful for engineering TamI, and possibly other P450 enzymes. Additionally, the approach of combining IMMS with negative staining EM appears to

be a very impactful analytical approach to detect and further characterize the stoichiometry and topology of transient protein-protein complexes in diverse biosynthetic/metabolic pathways.

3.5 Material & Methods

Materials & General Experimental Procedure

Unless otherwise noted, chemicals, reagents and solvents were purchased from EMD Millipore, Sigma-Aldrich, and Thermo Fisher Scientific. Agarose for gel electrophoresis was purchased from BioExpress (VWR). Kanamycin sulfate, isopropyl-β-D-thiogalactopyranoside (IPTG), and dithiothreitol (DTT) were obtained from Gold Biotechnology. Tris-HCl, ampicillin disodium salt and NADP$^+$ were purchased from Amresco. Chloramphenicol was obtained from Roche. Thiamine and NADPH were purchased from Chem-Impex. δ-aminolevulinic acid was purchased from Oakwood Chemical. Lysozyme was purchased from RPI. Imidazole was purchased from AK Scientific. Amicon Ultra centrifugal filters used for protein concentration were from EMD Millipore. PD-10 columns were purchased from GE Healthcare. Glucose-6-phosphate was from Biosynth, and glucose-6-phosphate dehydrogenase (yeast) was from Alfa Aesar. Succinimidyl-4(N-maleimidomethyl) cyclohexane (LC-SMCC) was purchased through Thermo Scientific.

Deionized water was obtained from a Milli-Q system (EMD Millipore) using Q-Gard 2/Quantum Ex Ultrapure organex cartridges. Media components for E. coli growth were purchased from EMD Millipore, Sigma-Aldrich, and Thermo Fisher Scientific unless otherwise specified. Glycerol was purchased from BDH (VWR). LB broth (Miller) and LB agar (Miller) were obtained in pre-made granulated form from EMD Millipore. TB broth was made from individually purchased components and consisted of 4% (v/v) glycerol. Media and solutions were autoclaved or sterile filtered before use. For bacterial culture, all media solutions were

autoclaved and buffers, and other solutions, pH was monitored using a VWR sympHony SB70P

pH meter calibrated according to the manufacturer's specifications. When specified, room

temperature (rt) was ~22-23 °C.

DNA oligonucleotides for cloning and mutagenesis were purchased from Integrated DNA

Technologies. PCR was performed using a Bio-Rad iCycler thermal cycler system. Restriction

endonucleases and other associated molecular biology reagents were purchased from New

England Biolabs. Invitrogen PCR cleanup/gel extraction and plasmid miniprep kits were

purchased from Thermo Fisher Scientific. All DNA manipulations were accomplished following

the manufacturer's protocols. DNA concentrations and protein purity indexes were measured

using a NanoDrop ND-1000 spectrophotometer. The DNA sequencing was performed at the

University of Michigan DNA Sequencing Core. E. coli DH5α was used for plasmid preparation

and maintenance while E. coli BL21(DE3) was used for protein overexpression. Chemically

competent E. coli cells were prepared using the method of Inoue[47]. Optical density (OD_{600}) was

measured using an Eppendorf BioPhotometer.

Protein purifications were performed manually using prepacked Ni^{2+}-NTA columns with GE

peristaltic pump P1 and automatically with GE columns and AKTA Pure FPLC machine, IM-MS

studies were performed on (please write details of instruments). Reversed-phase high-performance

liquid chromatography (RP-HPLC) purification was performed using Luna C18 columns with the

following specifications: dimensions, 150 x 4.6 mm; particle size, 5 μm; pore size, 100 Å. and a

solvent system of acetonitrile and H_2O supplemented with 0.1% formic acid. Electron microscopy

experiments were performed in the Electron Microscopy facility at the Life Science Institute, the

University of Michigan using an FEI morgangi and Tecnai T12 electron microscope equipped with

a LaB6 filament.

Cloning and preparation of recombinant enzymes

For expression of TamL we used previously prepared PSJ2-TamL vector[1]. The vector pSJ2 is an in-house derivative of pET-21a containing a coding region for an N-terminal His$_8$-tag, and insertion of our gene cassettes creates an in-frame fusion to this tag. For TamI, In summary, genes corresponding to TamI with append NdeI/EcoRI site at the 5'- and 3'-termini, respectively was synthesized from Genscript. The gene construct were subsequently digested with NdeI and EcoRI and ligated to the in-house derivative of pet28b overexpression vector containing N-terminal His8-tag. To create the TamI-linker-TamL construct, we used previously prepared plasmid pet28b-TamI-RhFRED as a templet where TamI is cloned in between NdeI/EcoRI sites and RhFRED is cloned between EcoRI/HindIII sites of the pet28b vector(ref). Using previously prepared plasmid pSJ2-TamL as a template, we PCR amplified the linker-TamL sequence under standard condition using forward primer 5'-atta*GAATTC*GGTGGCGGTGGCAGTGGTGGCAGTGGTGGCGGT GGCAGTGGTGGCGGTGGCAGTGGTGGCGGTGGCAGTATGAAGCACATCGATTCCGTG-3' (the italic bases represent EcoRI sites and the underline bases represent the nucleotides corresponding to 23-residue long linker G4SGGS(G4S)$_3$), and reverse primer 5'-GGTGCTCGAGTGCGGCCGC*AAGCTT*TCAGGCCGGCGGAACCCGCACGGAGAGGGCGTGCC-3' (italic bases represent HindIII sites). The amplified DNA fragments were isolated and purified prior to restriction digestion with EcoRI and HindIII. The resulting cDNA containing linker-TamL sequence with stop codon before the HindIII sites were digested with ECoRI/HindIII enzyme pairs and further ligated into EcoRI/HindIII digested pet28b_TamI-RhFRED construct (to remove the RhFRED) yielding the recombinant plasmid pet28b_TamI-linker-TamL for overexpression of N-His$_8$-tagged TamI-linker-TamL fusion protein. The authenticity of all

constructs was verified by DNA sequencing, and the sequencing results verified that the entire open reading frame in each case were free of unwanted mutations. These constructs were then transformed into *E. coli* BL21(DE3) overexpression host, and the resulting proteins were purified using nickel affinity chromatography as described below.

Overexpression and purification of TamI and TamI-linker-TamL recombinant proteins

The construct pet28b_TamI and pet28b_TamI-linker-TamL were used to transform E. coli BL21(DE3) for protein overexpression. The individual colonies were selected for overnight growth (37 °C) in 10 mL of LB containing kanamycin (50 μg/mL). 6 x 1 L of TB (3 L Erlenmeyer flasks) supplemented with kanamycin (50 μg/mL), and glycerol (4% v/v) were inoculated with the 10 mL overnight seed cultures and incubated at 37 °C (200 rpm). When the OD reached 0.6-1.0, the 1 L cultures were allowed to cool in an ice-water bath (10-20 min) before IPTG (0.2 mM) and δ-aminolevulinic acid (1 mM) were added to induce protein expression and to allow for facile production of the heme cofactor in *E. coli*, respectively. The cultures were allowed to grow at 18 °C for 18-20 h before the cells were harvested and stored at -80 °C until used for protein purification. All following steps were performed at 4 °C. The cells were thawed and resuspended in 30 mL of lysis buffer (50 mM Tris-HCl (pH = 7.4 at room temperature (rt)), 50 mM NaCl, 10% (v/v) glycerol, 1 mM PMSF, 0.5 mg/mL lysozyme, 2 mM MgCl$_2$, 5-10 U/mL Benzonase nuclease) per 1 L of original overexpression culture (i.e., 2-3 mL per 1 g of cells). The cell suspension was incubated on a nutating shaker for 1 h prior to sonication using a Model 705 Sonic Dismembrator (Thermo Fisher Scientific) and centrifugation at 50,000 x g for 30 min to remove cellular debris. To the resulting clarified lysate were added 4 M solutions of NaCl and imidazole such that the final concentrations of each were 300 mM and 10 mM, respectively. The lysate was filtered through a syringe-operated 0.45 μm filter (Corning)

before loading onto a prepacked column containing 7mL of Ni-NTA resin (Qiagen) using peristaltic pumps (GE Healthcare). The loaded material was washed with 5 CV of wash buffer (50 mM Tris-HCl (pH = 7.4 at rt), 300 mM NaCl, 10 mM imidazole, 10% (v/v) glycerol) followed by 10 CV of 5% elution buffer (50 mM Tris-HCl (pH = 7.4 at rt), 300 mM NaCl, 300 mM imidazole, 10% (v/v) glycerol) to remove additional protein contaminants. TamI and Tam-linker-TamL fusion proteins were eluted with a linear gradient of 10-50% elution buffer over 30 CV. Fractions containing pure material as assessed by SDS- PAGE and by monitoring absorbance (A_{420}/A_{280}) were pooled and concentrated using 30 kD MWCO centrifugal filters. Concentrated protein was desalted by loading onto PD-10 columns followed by through dialysis in storage buffer (50 mM Tris (pH = 7.4), 1 mM EDTA, 0.2 mM DTT, 10% (v/v) glycerol). Aliquots of purified protein were flash frozen in liquid N_2 and stored at -80 °C until needed for biochemical experiments. Following this protocol, the yield of functional P450 (as assessed by obtaining CO difference spectra according to the established protocol. TamI-linker-TamL protein were further subjected to size exclusion purification via an ÄKTA FPLC system (GE Healthcare) with Superdex 200 Increase 10/300 GL column in storage buffer (50 mM Tris (pH = 7.4), 1 mM EDTA, 0.2 mM DTT, 5% (v/v) glycerol) for further use in Mass and EM experiments.

To remove His-tag N-terminal His-tag, fractions, the protein samples after Ni^{2+}-NTA purifications were pooled and subsequently incubated with purified His-TEV protease to cleave off the N-terminal His-tag. A ratio of 1: 0.035for protein: His-TEV were incubated to ensure cleavage of His-tag and SDS page gel indicates the definite shift in the sizee of His_8-tagged TamI to TEV-digested TamI. Protein samples were further dialyzed in storage buffer overnight. Un-cleaved TamI was removed by applying the digest a Ni^{2+}-NTA column equilibrated in loading buffer, by collecting the flow through un-cleaved TamI would stay on the column while

leaving the cleaved product to flow through. Subsequent imidazole elution results in elution of His-TEV protease and His-tag. Flow-through fractions were pooled and concentrated using an Amicon Millipore concentrator with a cut-off membrane of 30 kD, and the concentrated protein solutions were then flash frozen in Liq. N_2 and stored in -80°C.

Overexpression and purification of recombinant TamL proteins

Constructs bearing TamL was purified following previously reported protocol.[1] In summary, construct bearing TamL were previously cloned into the pSJ2 overexpression vector and transformed into the *E. coli* BL21(DE3) CodonPlus RP overexpression host. A single transformant was used to begin a small seed culture that was passaged into 1 L of LB amended with 100 μg/mL of ampicillin and 40 μg/mL of chloramphenicol. This culture was grown with shaking at 37°C to an OD_{600} of 0.8, induced with 0.4 mM IPTG and grown overnight with shaking at 15°C. Cells were recovered by centrifugation and subsequently resuspended in 40 mL of lysis buffer (20 mM NaH_2PO_4, 500 mM NaCl, pH 7.8, 1 mg/mL BSA). Lysis was achieved by sonication for 4 min at 50% power using 1⁄2 inch disruptor in 2-second bursts. The lysate was clarified by centrifugation at 60,000 RPM for 20 min at 4°C. The clarified lysate was slowly loaded, using peristaltic pump, into pre-packed columns containing 7mL Ni^{2+}-NTA resin. The resin was then washed with a step gradient of imidazole in phosphate buffer (20 mM NaH_2PO_4, 500 mM NaCl, pH 6.0), with the His-tagged proteins being eluted at 100 mM imidazole. The protein was exchanged into storage buffer (20 mM NaH_2PO_4, 100 mM NaCl, pH 8.0) by dialysis with three-time exchange of buffers and subsequently protein solutions were concentrated on a Millipore Amicon Ultra centrifugal concentrator (30 kD MWCO), and flash frozen in Liq N_2 for flash freezing and storage in -80 °C freezing. For the use in EM imaging, the purified protein

solutions were further subjected to another round of purification in size exclusion buffer ((50mM Phosphate buffer, 150mM NaCl, 2% glycerol, 1mM DTT pH 7.4) using with Superdex 200 Increase 10/300 GL column via an ÄKTA FPLC system (GE Healthcare). To obtain the monomer fraction for crosslinking experiments, the cells were resuspended in lysis buffer containing 1mM DTT. Following above mentioned protocol we purified the TamL in similar condition with additional 1mM DTT in each buffer. Purified fraction of TamL were the subsequently dialyzed into storage buffer contains 1mM DTT. Subsequently, the TamL were subjected to size exclusion chromatographic purification as described above in size exclusion buffer (50mM Phosphate buffer, 150mM NaCl, 2% glycerol, 1mM DTT pH 7.4).

Chemical crosslinking of TamI and TamL

Chemical crosslinking were performed with heterobifunctional crosslinker LC-SMCC which contains N-hydroxysuccinimide (NHS) ester and maleimide groups that allow covalent conjugation of amine- and sulfhydryl-containing group of proteins. We planned to crosslink TamI N-terminus primary amino group with NHS ester of LC-SMCC and free sulfhydryl group of TamL monomer. The crosslinking experiments were conducted in two step in conjugation buffer Phosphate buffered saline (20mM sodium phosphate, 150mM NaCl, pH 7.2) at 4°C for 2 hours. In first step, TamI (0.1mM) were incubated with 10 fold molar excess of LC-SMCC crosslinker in conjugation buffer for 2 hours at 4°C. To remove excess of LC-SMCC, the reaction mixture were desalted twice with PD10 desalting column using conjugation buffer. Subsequently, we mixed the TamL monomer fractions with desalted TamI protein mixture in equimolar molar ratio in conjugation buffer and incubated the mixture for 2 hours at 4°C followed by further desalting by PD10 desalting column before subjecting to Mass based analysis.

Co-expression and purification of TamI and TamL proteins

The co-expression of both proteins were performed by considering the compatibilities of plasmids pet28b_TamI and pSJ2-TamL into BL21(DE3) strains as well as the efficiency of each plasmid transformation. The antibiotic selection marker for pet28b_TamI is Kanamycin while for pSJ2-TamL plasmid the selection marker is ampicillin. We co-transformed *E. coli* BL21(DE3) with pet28b_TamI and pSJ2_TamL plasmid by heat-shock and plated them on a solid agar plates. The co-transformed bacteria grow under double antibiotic selection (ampicillin 100 μg/ml, kanamycin 50 μg/ml) since the pet28b-TamI vector contains an antibiotic resistance cassette against kanamycin, and the pSJ2-TamL against ampicillin. We observed low transformation efficiency, we selected colonies from plates with both antibiotics, and colonies were grown over-night in 5 ml LB medium (+antibiotics) at 37 °C. 6 x 1 L of TB (2 L Erlenmeyer flasks) supplemented with kanamycin (50 μg/mL), ampicillin (100 μg/mL), and glycerol (4% v/v) were inoculated with the 10 mL overnight seed cultures and incubated at 37 °C (200 rpm). When the OD reached 0.6-1.0, the 1 L cultures were allowed to cool in an ice-water bath (10-20 min) before IPTG (0.4 mM) and δ-aminolevulinic acid (1 mM) were added to induce protein expression and to allow for facile production of the heme cofactor in *E. coli*, respectively. The cultures were allowed to grow at 18 °C for 18-20 h before the cells were harvested and stored at -80 °C until used for protein purification. All subsequent steps of purification were performed at 4 °C following the same protocol as we discussed above for TamI expression. When required, the protein aliquots after Ni-NTA purification were further subjected to another round of purification using with Superdex 200 Increase 10/300 GL column via an ÄKTA FPLC system (GE Healthcare).The purity of aliquots were then accessed for SDS PAGE gel and Mass-based characterization of the existence of both

TamI and TamL gene in the cell lysate. Importantly, the TamI:TamL complex co-eluted as judged by SDS-PAGE gel and IM-MS data of fraction pooled from size exclusion chromatography.

Enzyme assays

Isolation of TirA, TirB, TirC and TirD and preparation of TirE by preparative-scale enzymatic conversion have been described previously. The *in vitro* enzymatic conversions of tirandamycins were performed in a total volume of 250 μl of conversion buffer (50 mM NaH_2PO_4 , pH 7.4, 1 mM EDTA, 0.2 mM dithioerythritol, 10% glycerol) containing 500 μM substrates, 2 μM TamL and 2 μM TamI, or ~ 2 μM TamI-linker-TamL fusion protein, partnered by 4 mM spinach ferredoxin, 2 μM ferredoxin-$NADP^+$ reductase,1mM $MgCl_2$, 5 mM glucose-6-phosphate, and 1 U/mL glucose-6- phosphate dehydrogenase. in storage buffer (50 mM NaH_2PO_4 (pH = 7.4), 1 mM EDTA, 0.2 mM DTT, 10% (v/v) glycerol). Reactions were incubated at 30 °C (100 rpm) for 3h before to quenching followed by extraction, using 2×200 ml of $CHCl_3$. The resulting organic extract was dried, redissolved in 120 ml of methanol and analysed by addition of 100 μL of methanol with HPLC. All reactions were performed and analyzed in triplicate. HPLC analysis was performed on a Schimadzu HPLC system with UV detection at 354 nm using a Phenomenex Luna C18 column with the following specifications: dimensions, 250 x 4.6 mm; particle size, 5 μm; pore size, 100 Å. For analysis of reactions HPLC conditions we used mobile phase (A = deionized water + 0.1% formic acid, B = acetonitrile + 0.1% formic acid); and the gradient 15% B for 1 min, 15% to 100% B over 35 min, 100% B for 1 min.

Ion Mobility – Mass Spectrometry

Protein samples were buffer exchanged into 200mM ammonium acetate buffer using Micro Bio-Spin 30 columns (Bio-Rad, Hercules, CA). Sample aliquots (~5μL) were analyzed by IM-MS on a quadrupole-ion mobility- time-of-flight mass spectrometer (Q-IM-ToF MS) instrument (Synapt G2 HDMS, Waters, Milford, MA). Protein ions were generated using a nESI source operated in the positive polarity with applied capillary voltages of 1.4−1.6 kV. For improved sample ion desolvation, sampling and extraction cones were operated at 50 and 5 V, respectively, and the source backing pressure was adjusted to ~8mbar. The travelling wave ion mobility (TWIM) cell was operated at 4 mbar with wave height and wave velocity values of 15 V and 150 m/s, respectively. All CCSs were calculated based on a previously described calibration framework, using class specific calibrant ions, including transthyretin, concanavalin A, alcohol dehydogenase, and glutamate dehydrogenase.[48] which relate IM drift times directly to ion size and shape, were calibrated using ions of known CCS as described previously and detailed in Table II-1. Raw data was analysed in MassLynx (Waters, Milford, MA). Raw data extractions were performed using TWIMExtract[49], and CIU analyses were completed using CIUSuite2[33].

Binding Affinity Measurement

In order to measure the binding affinity of TamI for TirC, 20uM of TamI was titrated with varying amounts of TirC, resulting in final samples with protein:ligand ratios ranging from 0.6 to 3, which were than incubated on ice for 10min. The ratio of TirC bound TamI over free TirC was plotted against the concentration of TirC bound TamI, giving a Kd of 2.7±0.4. To measure Kd of TirD bound TamI, we incubated TamI with TirC and TirD at 1:1 ratio separately, by comparing the relative intensity of TirC bound TamI with TirD bound TamI, we estimate the Kd for TamI and TirD binding as 29.1±4.3uM.

Correlation of TamI CIU and CID

TamI was incubated with TirC at 1:2 ratio. Both CIU and CID data were collected for 12+, 13+ and 14+ ions by applying collision voltage in the Trap cell region of the instrument prior to IM analysis. The collision voltage was increased in a stepwise manner from 20-200V in 2V increments. The charge states of interest were analysed individually by quadrupole selecting narrow m/z windows correlating to ligand bound TamI. As the selected substrate bound TamI ions unfolded, the ligand was first ejected, resulting in the appearance of unbound TamI at a lower m/z. CID ligand ejection curves were plotted by tracking the relative intensity of ligand bound TamI states over all other observed TamI ion populations. We correlated TirC/heme dissociation curves with CIU fingerprints recorded for all three charge states.

Stability Shift Analysis

TamI was incubated with TirC/TirD at 1:1 ratio for 10min. CIU data were collected for both substrate free and substrate bound form of TamI. Extracted CIU data were analyzed using a home-built software package termed CIUSuite 2, a custom software developed for analysing and presenting CIU data. CIU fingerprints were plotted as a 2D contour plot, in which ion intensities were normalized to a maximum value of 1 at each collision voltage and smoothed using a Savitzky-Golay filter. RMSD plots and values were generated using RMSD comparison module, with the ability to compare CIU fingerprints of TamI with and without substrate loaded. CIU50 values, defined as the voltage at which 50% of a relatively compact state of the protein transitions to a more extended state[33], were also measured for TamI with and without substrate, enabling to measure the stability shift upon substrate loading.

Electron Microscopic Imaging

Sample quality and homogeneity were evaluated by negative staining using Morgagni (Thermo Fisher Scientific) electron microscope whereas the final data collections were performed on Tecnai 12 (Thermo Fisher Scietific) electron microscope. Proteins were purified freshly using size exclusion chromatography and samples were prepared immediately for EM experiments using a conventional negative staining protocol. Briefly, eluted fractions were diluted 250-fold into EM buffer (20mM Tris, 50mM NaCl, 2% glycerol, pH 7.4) just before coating on to the grid and 3.5 μL of the sample was adsorbed to a glow-discharged carbon-coated copper grid (Electron Microscopy Sciences, Hatfield, PA, USA) and stained with 0.75% uranyl formate (Polysciences Inc, Warrington, PA, USA) solution. Samples were imaged at room temperature with a Tecnai 12 electron microscope equipped with a LaB_6 filament and operated at an acceleration voltage of 120 kV. Images were recorded on an Ultral4000 (Gatan Inc.) CCD camera using low-dose procedures at a nominal magnification of ×35,238 and a defocus value of -1.0 μm. The images were binned (2×2 pixels) to obtain a pixel size of 4.2 Å on the specimen level for the subsequent data processing. A total of 8857 particles from 30 images were automatically excised for co-expressed sample using e2boxer.py, whereas 36563 particles from 75 images were excised for TamI-TamL fusion protein sample. 2D classification were carried out using Relion software.

Acknowledgement

3.6   References

(1)   Carlson, J. C.; Li, S.; Gunatilleke, S. S.; Anzai, Y.; Burr, D. A.; Podust, L. M.; Sherman, D. H. *Nat. Chem.* **2011,** *3*, 628.
(2)   Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J. *Nature Reviews Drug Discovery* **2015,**
(3)   Newman, D. J.; Cragg, G. M. *Journal of Natural Products* **2016,**
(4)   Shen, B. *Cell* **2015,**

(5)   Karageorgis, G.; Foley, D. J.; Laraia, L.; Waldmann, H. *Nat. Chem.* **2020,**

(6)   Sweetlove, L. J.; Fernie, A. R. *Nature Communications* **2018,**

(7)   Zhang, Y.; Fernie, A. R. *Antioxid. Redox Signal.* **2020,**

(8)   Bassard, J.-E.; Møller, B. L.; Laursen, T. *Curr. Mol. Biol. Reports* **2017,**

(9)   Laursen, T.; Borch, J.; Knudsen, C.; Bavishi, K.; Torta, F.; Martens, H. J.; Silvestro, D.; Hatzakis, N. S.; Wenk, M. R.; Dafforn, T. R.; Olsen, C. E.; Motawia, M. S.; Hamberger, B.; Møller, B. L.; Bassard, J. E. *Science (80-.).* **2016,**

(10)   Jørgensen, K.; Rasmussen, A. V.; Morant, M.; Nielsen, A. H.; Bjarnholt, N.; Zagrobelny, M.; Bak, S.; Møller, B. L. *Current Opinion in Plant Biology* **2005,**

(11)   Møller, B. L. *Science (80-.).* **2010,**

(12)   Zhang, Y.; Fernie, A. R. *Plant Commun.* **2020,**

(13)   Bassard, J. E.; Halkier, B. A. *Phytochemistry Reviews* **2018,**

(14)   Singleton, C.; Howard, T. P.; Smirnoff, N. *Journal of Experimental Botany* **2014,**

(15)   Smirnoff, N. *Plant Physiol.* **2019,**

(16)   Pouvreau, B.; Vanhercke, T.; Singh, S. *Plant Science* **2018,**

(17)   You, C.; Zhang, Y. H. P. *ACS Synth. Biol.* **2013,**

(18)   Han, P.; Zhou, X.; You, C. *Front. Bioeng. Biotechnol.* **2020,**

(19)   Klein, W. P.; Thomsen, R. P.; Turner, K. B.; Walper, S. A.; Vranish, J.; Kjems, J.; Ancona, M. G.; Medintz, I. L. *ACS Nano* **2019,**

(20)   Behrendorff, J. B. Y. H.; Borràs-Gas, G.; Pribil, M. *Trends in Biotechnology* **2020,**

(21)   Zhang, Y. H. P. *Biotechnology Advances* **2011,**

(22)   Vázquez-González, M.; Wang, C.; Willner, I. *Nature Catalysis* **2020,**

(23)   Ward, A. B.; Sali, A.; Wilson, I. A. *Science (80-.).* **2013,** *339*, 913.

(24)   Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E. *Nat. Chem.* **2014,** *6*, 281.

(25)   Ellis, H. W.; Pai, P. Y.; McDaniel, E. W.; Mason, E. A.; Viehland, L. A. *At. Data Nucl. Data Tables* **1976,** *17*, 178.

(26)   Cawley, A. T.; Flenker, U. *J. Mass Spectrom.* **2008,** *43*, 854.

(27)   Benesch, J. L. P. *J. Am. Soc. Mass Spectrom.* **2009,** *20*, 341.

(28)   Hyung, S. J.; Robinson, C. V.; Ruotolo, B. T. *Chem. Biol.* **2009,** *16*, 382.

(29)   Eschweiler, J. D.; Martini, R. M.; Ruotolo, B. T. *J. Am. Chem. Soc.* **2017,** *139*, 534.

(30)   Hopper, J. T. S.; Oldham, N. J. *J. Am. Soc. Mass Spectrom.* **2009,** *20*, 1851.

(31)   Chan, D. S. H.; Hess, J.; Shaw, E.; Spry, C.; Starley, R.; Dagostin, C.; Dias, M. V. B.; Kale, R.; Mendes, V.; Blundell, T. L.; Coyne, A. G.; Abell, C. *Biochem. J.* **2019,**

(32)   Newmister, S. A.; Srivastava, K. R.; Espinoza, R. V.; Caddell Haatveit, K.; Khatri, Y.; Martini, R. M.; Garcia-Borràs, M.; Podust, L. M.; Houk, K. N.; Sherman, D. H. *ACS Catal.* **2020,**

(33)   Polasky, D. A.; Dixit, S. M.; Fantin, S. M.; Ruotolo, B. T. *Anal. Chem.* **2019,** *91*, 3147.

(34)   Zhao, Y.; Singh, A.; Li, L.; Linhardt, R. J.; Xu, Y.; Liu, J.; Woods, R. J.; Amster, I. J. *Analyst* **2015,** *140*, 6980.

(35)   Nyon, M. P.; Prentice, T.; Day, J.; Kirkpatrick, J.; Sivalingam, G. N.; Levy, G.; Haq, I.; Irving, J. A.; Lomas, D. A.; Christodoulou, J.; Gooptu, B.; Thalassinos, K. *Protein Sci.* **2015,** *24*, 1301.

(36)   Niu, S.; Ruotolo, B. T. *Protein Sci.* **2015,** *24*, 1272.

(37)   Gober, R.; Wheeler, R.; Rohr, J. *Medchemcomm* **2019,**

(38)   Eschweiler, J. D.; Kerr, R.; Rabuck-Gibbons, J.; Ruotolo, B. T. *Annual Review of Analytical Chemistry* **2017,**

(39)   Beveridge, R.; Migas, L. G.; Payne, K. A. P.; Scrutton, N. S.; Leys, D.; Barran, P. E. *Nat. Commun.* **2016,**

(40)   Schenk, E. R.; Almeida, R.; Miksovska, J.; Ridgeway, M. E.; Park, M. A.; Fernandez-Lima, F. *J. Am. Soc. Mass Spectrom.* **2015,** *26*, 555.

(41)   Zhang, X.; Quinn, K.; Cruickshank-Quinn, C.; Reisdorph, R.; Reisdorph, N. *Current Opinion in Chemical Biology* **2018,**

(42)   Correia, M. A.; Sinclair, P. R.; De Matteis, F. *Drug Metabolism Reviews* **2011,**

(43)   Rudolf, J. D.; Chang, C. Y.; Ma, M.; Shen, B. *Natural Product Reports* **2017,**

(44)   Zhang, X.; Li, S. *Natural Product Reports* **2017,**

(45)   Urban, P.; Lautier, T.; Pompon, D.; Truan, G. *International Journal of Molecular Sciences* **2018,**

(46)   Guengerich, F. P.; Waterman, M. R.; Egli, M. *Trends in Pharmacological Sciences* **2016,**

(47)   Inoue, H.; Nojima, H.; Okayama, H. *Gene* **1990,**

(48)   Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. *Anal. Chem.* **2010,** *82*, 9557.

(49)   Haynes, S. E.; Polasky, D. A.; Dixit, S. M.; Majmudar, J. D.; Neeson, K.; Ruotolo, B. T.; Martin, B. R. *Anal. Chem.* **2017,** *89*, 5669.

Chapter 4 Gas-Phase Unfolding Reveals Stability Shifts Associated with Substrate Binding in Modular Polyketide Synthases

4.1   Introduction

Protein–protein and protein–substrate interactions are centrally important to many biological processes and have thus been investigated extensively using a variety of techniques.[1] Protein-substrate interactions are especially essential to the biosynthesis of natural products; however, due to their dynamic and transient nature, interactions among biosynthetic pathway enzymes remain mostly uncharacterized. A thorough understanding of these transient protein-protein and protein-substrate interactions is critical for understanding the biosynthesis of natural products, as these actions involve an intricate interplay of substrate binding and shuttling process among a series of binding pockets arrayed across large, multi-domain enzymes. Protein stability is considered a key biophysical attribute that dictates much of protein-ligand binding dynamics [2]. Shifts in protein stability likely occur in biosynthetic enzymes during substrate binding, but due to the size of the proteins involved, the biophysical shifts involved during the complex relay mechanisms involved in natural product biosynthesis remain largely unknown.[3]

Mass Spectrometry (MS) has been widely used to study the interactions between proteins and small molecules. Earlier studies have demonstrated the capability of MS to probe the noncovalent interactions responsible for molecular recognition within a wide range of biopolymers.[4,5] More recent studies have applied native MS technology in order to directly detect the influence of ligand binding on the stability of intact multiprotein complexes.[6,7] By coupling

ion mobility separation in tandem with mass spectrometry (IM-MS), the gas-phase structures of protein complexes can be rapidly assessed. IM is able to measure ion-neutral collision cross sections (CCSs), which can be used to provide useful structure information and aid in the construction of protein complex models[,8]. Collision induced unfolding (CIU) builds upon standard IM measurements by providing both the means to resolve protein conformers that present unresolvable differences in ground state CCS, as well as an opportunity to measure protein stabilities from small amounts of unpurified samples By manipulating the acceleration voltage that ions experience upon entering the ion trap prior to the IM separator, protein ions can undergo stepwise collisional activation leading to CIU, a process that can be tracked with IM and used to study the details of protein-ligand interactions.[9]

Prior CIU studies have focused on both building a complete mechanistic picture of gas-phase protein unfolding and deploying CIU technology to detect subtle changes in protein structure.[10].[11] Specifically, CIU has been used to differentiate biosimilar antibodies, assess the specificity of lipid binding within membrane protein complexes, and distinguish conformationally-selective kinase inhibitor binding modes. Despite these many successes, CIU has been primarily been used to analyze single domain proteins less than 70 KDa. Furthermore, the stability differences reported within protein CIU data, often reported as CIU50 values that report the acceleration voltage required to convert 50% of the preceding conformer into the following intermediate state, typically report on significant structural changes within smaller protein systems, leading to relatively large CIU50 differences. Clearly, pushing CIU studies toward protein complexes comprised of large, multi-domain subunits face many analytical challenges. Chief among these overcoming the loss in sensitivity associated with CIU data

acquired for larger protein systems, where substrate binding may only account for a small percentage of the overall mass of the assembly.

Polyketide synthases (PKSs) are a class of such multi-domain, multienzymes, that use their modular architecture to produce polyketide natural products that form the basis for nearly one-third of pharmaceuticals.[12] Type I PKS modules utilize discrete modules to sequentially elongate, process and terminate a polyketide chain.[2] Figure 4-1A provides an illustration of the catalytic cycle for PikAIV responsible for production of the pikromycin and Figure 4-1B represents the hemoenzymatic synthesis of tylactone via JuvEIV and JuvEV. Each module that extends the sequence of an intermediate contains three critical domains: acyl carrier protein (ACP), keto synthase (KS) and acyltransferase (AT) domains. An AT acts to load the ACP with a building block from a specific acyl-coenzyme A (acyl-CoA), while the KS accepts the upstream polyketide intermediates and catalyzes a decarboxylative Claisen reaction resulting in chain elongation. Next, the ACP shuttles the elongating polyketide intermediates among domains, transferring them to respective catalytic domains for loading, extension and keto-group processing or modification. Additionally, a ketoreductase (KR) domain is responsible to reduce a β-keto group to a β-hydroxyl. At the end of many PKS pathways, the completed polyketide product is offloaded by a terminal thioesterase (TE) domain. In order to study the full-length type I PKS module, previous work applied cryo-electron microscopy (cryo-EM) to visualize the structure of a pikromycin PKS module 5 (PikAIII) and its structural rearrangements during the catalytic cycle when loaded with corresponding substrates.[13,14] In addition to PikAIII, pikromycin PKS module 6 (PikAIV) and juvenimicin PKS module 7 (JuvEV) are also of high interest because they represent terminal modules in their respective biosynthetic systems . Currently, limited structural information is available for PikAIV and JuvEV, leaving the

mechanistic details surrounding their functions relatively opaque. For example, JuvEV presents a similar multi-domain organization to PikAIII but can accommodate a larger polyketide and selects for a malonyl rather than a methylmalonyl extender unit.[15] As such, JuvEV represents a high-value target for future cryoEM investigations[13,14].

Here, we discuss PikAIV and JuvEV IM-MS and CIU data, where we probe the shifts in stability that occur upon loading both polyketide substrates and methylmalonyl (MM)/malonyl groups within the enzyme active sites. We describe the development of CIU methods capable of extracting information from these large, multi-domain, multi-protein complexes and observe clear evidence of stability shifts associated with substrate binding events that account for <0.1% of the mass for the intact assembly. This work represents the first use of CIU to measure stability differences in substrate-bound multi-domain, multimeric protein complexes in this size range.
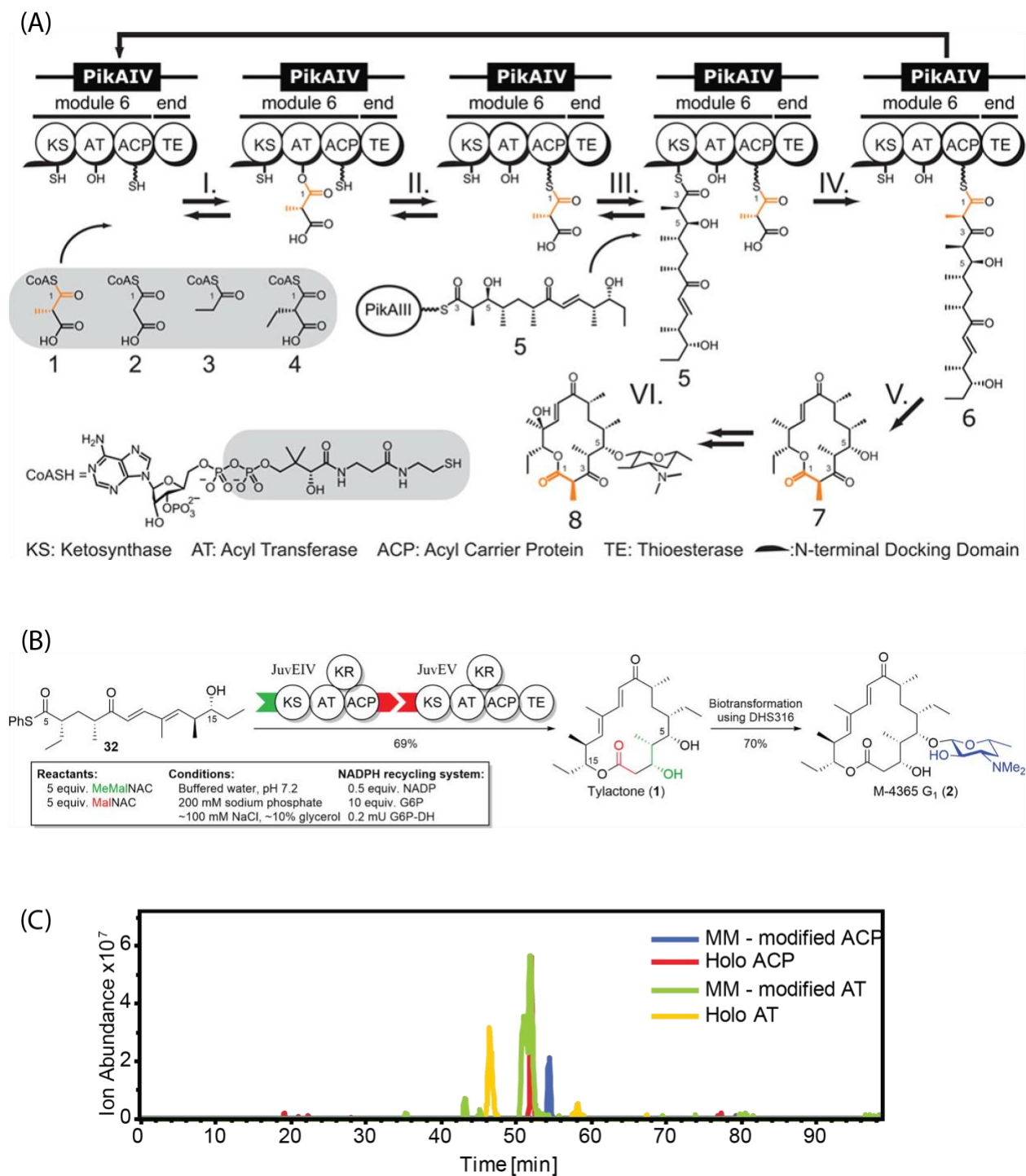
Figure 4-1. (A) Catalytic Cycle for PikAIV.[16] (B) Chemoenzymatic synthesis of tylactone. (1) and glycosylation to (2) via JuvEIV and JuvEV.[15] (C) Extracted ion chromatograms (EICs) for unloaded and MM-loaded ACP and AT domains. We were able to verify MM loaded to PikAIV through the observation of peptide elution time differences associated with holo and MM modified forms of both the ACP and AT domains.

4.2   Materials and Methods

Sample preparation

Full length, PikAIV and JuvEV proteins comprised of four/ five domains each accordingly, were purified by size exclusion chromatography as described previously[13] before buffer exchange into 500 mM ammonium acetate. To prepare methylmalonyl (MM)/malonyl-loaded proteins, 20 uM holo–PikAIV/holo–JuvEV was incubated with 500uM MM-*N*-acetyl cysteamine (NAC)/malonyl-NAC for 30 min at room temperature and confirmed the presence of MM/malonyl-acyl carrier protein (ACP) by bottom-up LC-FT-ICR MS. To prepare hexaketide loaded PikAIV, 20uM holo-PikAIV was incubated with 500uM hexaketide for 30 min at room temperature, as described previously.[13]

IM-MS data collection

Protein samples were buffer exchanged into 500mM ammonium acetate buffer using Micro Bio-Spin 30 columns (Bio-Rad, Hercules, CA). Sample aliquots (~5μL) were analyzed by IM-MS on a quadrupole-ion mobility-time-of-flight mass spectrometer (Q-IM-ToF MS) instrument (Synapt G2 HDMS, Waters, Milford, MA). Protein ions were generated using a nESI source operated in positive mode with an applied capillary voltage of 1.4−1.6 kV. For improved sample ion desolvation, sampling and extraction cones were operated at 50 and 10 V, respectively, and the source backing pressure was adjusted to ~8mbar. The travelling wave ion mobility (TWIM) cell was operated at 4 mbar with wave height and wave velocity values of 15 V and 150 m/s, respectively. Raw data was analyzed in MassLynx (Waters, Milford, MA). Raw data extractions were performed using TWIMExtract[17].

CIU Data analysis

CIU data were collected for holo, substrate free and substrate bound form of PikAIV/JuvEV. Extracted CIU data were analyzed using a home-built software package termed CIUSuite2[18], a custom software package developed for analyzing and presenting CIU data. CIU fingerprints were plotted as a 2D contour plot, in which ion intensities were normalized to a maximum value of 1 at each collision voltage and smoothed using a Savitzky-Golay filter. RMSD plots and values were generated using the RMSD comparison module, with the ability to compare CIU fingerprints of protein with and without substrate loaded. CIU50 values, defined as the voltage at which 50% of a relatively compact state of the protein transitions to a more extended state[18], were also measured for holo–PikAIV, MM–PikAIV, hexaketide–PikAIV, holo–JuvEV, and malonyl–JuvEV , enabling stability shift measurements associated with substrate loading.

4.3   Results and Discussion

CIU Tracks Shifts in PikAIV Substrate Binding-associated Stability Shifts

To both probe the extent to which CIU-based stability shifts can be detected in large multi-domain protein systems, as well as provide quantitative measurements of protein stability shifts triggered by substrate loading in such biosynthetic enzymes, we collected CIU fingerprints for the 38$^+$ PikAIV dimer occupying the holo state, where the ACP bearing a phosphopantetheine (Ppant) arm, the MM-loaded state, the hexaketide-loaded state, and the MM–hexaketide-loaded state. We began by preparing each of these PikAIV states and verifying their status using an LC-MS/MS methodology described previously. By tracking CID product ion populations, we were able to verify Ppant, MM and hexaketide attachment to peptide fragments produced following gas-phase activation (Figure 4-1C). Furthermore, we observed significant elution time

differences between the holo and MM PikAIV forms of both the ACP and AT domain. Taken together, these data validate to the status of the PikAIV forms studied here.

CIU data collected for 36^+, 37^+ and 38^+ PikAIV dimers exhibit two main features with one clear transition, over the acceleration voltage range probed here. We note that this CIU does not appear to be correlated with the number of expected native domains within the PikAIV construct probed here, which includes four autonomously folded regions per monomer as described above, and surmise that the overall molecular mass of the oligomer studied here, ~280 kDa, may preclude such an assessment in the energy range accessible on current instrumentation. In order to quantify any shifts in PikAIV stability upon substrate attachment, we recorded CIU50 values as an indication of protein stability across data acquired for four different ligand-attached versions of the PikAIV protein. The CIU50 values captured for holo–PikAIV, MM–PikAIV, hexaketide–PikAIV, and MM–hexaketide–PikAIV were 45.3±0V, 48.2±0.8V, 52±0.8V and 64.5±0.1V respectively. These data indicate that both MM and hexaketide attachment is capable of increasing the stability of holo-PikAIV, with MM and hexaketide loaded together maximizing the stability increase for holo–PikAIV by 42%. Our CIU data are the product of multiple measurements, resulting in consistent CIU50 values that exhibit with negligible standard deviation values (Figure 4-2E).
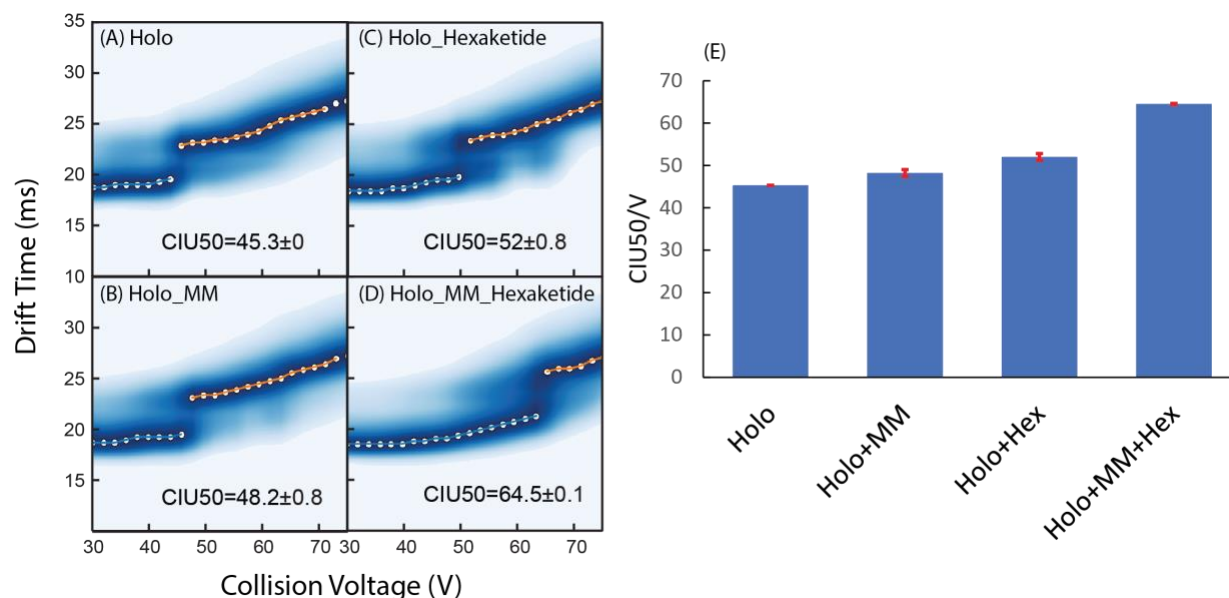
Figure 4-2. CIU fingerprints of holo–PikAIV (A), MM–PikAIV (B), hexaketide–PikAIV (C), and MM–hexaketide–PikAIV (D) of 38[+] with CIU50 values quantified. (E) Histogram charting the CIU50 values of all four samples with error bars measured.

CIU Reveals Shifts in JuvEV stability upon Substrate Binding

CIU fingerprints collected for 42-45[+] JuvEV dimers in both holo (unbound) and malonyl bound states indicate increased stabilty for the biosynthetic enzyme upon malonyl binding. Specifically, CIU fingerprints and measured CIU50 values for 44[+] JuvEV dimer ions are displayed in Figure 4-3A (holo) and Figure 4-3B (malonyl loaded). We observe three highly diagonalized CIU features with two unfolding transitions in this data. The relatively undefined nature of the transitions observed for JuvEV clearly complicate the analysis of such CIU data, but despite the shape of the features and transitions observed we found JuvEV CIU data to be highly reproducible, thus allowing us to confidently extract CIU50 values. We quantified the CIU50 for the first structural transition of the holo-JuvEV dimer as 116.6 ± 0.4V and 126.2 ± 1.4V for Malonyl–JuvEV. In addition, we record CIU50 2 as 155.6 ± 0.6V for holo–JuvEV and 167.2 ± 1.8V for Malonyl–JuvEV accordingly. Both CIU50 1 and CIU50 2 values show that

85

JuvEV with malonyl loaded is significantly stabilized. Furthermore, we computed fingerprint-wide RMSD values of 4% for replicate CIU data recorded for holo-JuvEV (Figure 4-3C), directly indicating the excellent reproducibility of CIU data we record for this complex. When simlar RMSD values are computed for datasets including both holo-JuvEV and Malonyl-JuvEV CIU fingerprints, the result is a value of 16% (Figure 4-3D) which suggests a significant increase in JuvEV stability upon malonyl loading. Error bars for the CIU50 values reported in Figure 4-3E further illustrate the reproducibility of our measurements. Taken together, the data shown in Figures 4-2 and 4-3 strongly supports the ability of CIU to quantifiably and reproducibly differentiate stability differences in large multiprotein complexes upon substrate binding.
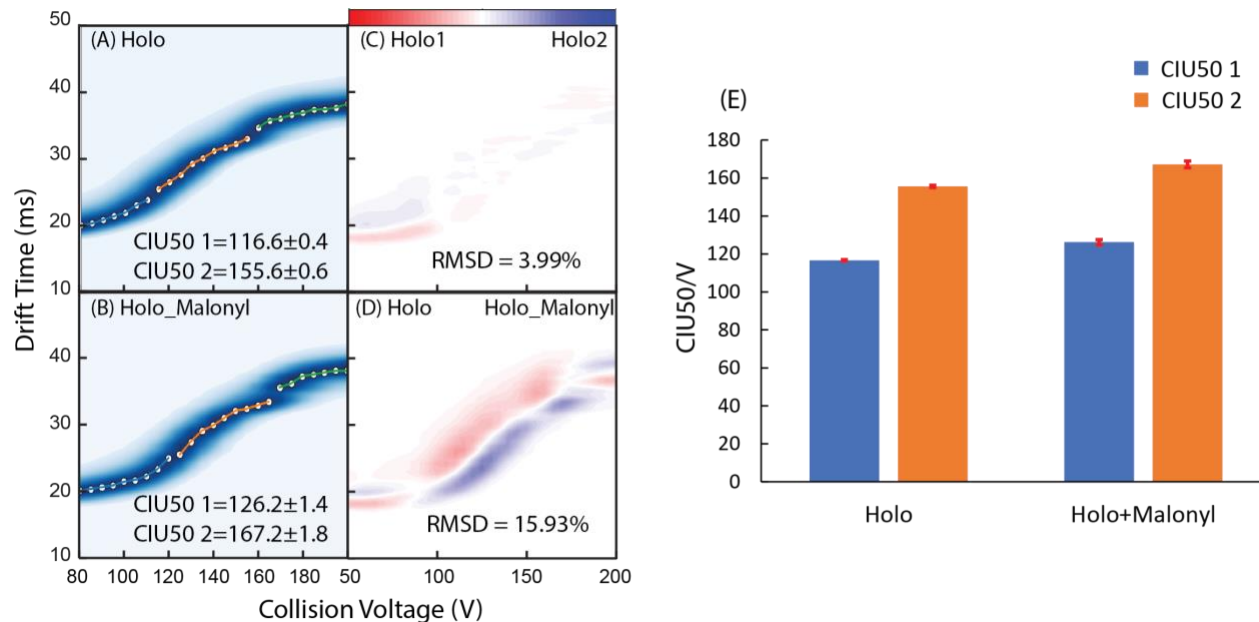
Figure 4-3. Quantifying the reproducibility and differentiating power for the CIU of multiprotein complexes. CIU fingerprints of holo-JuvEV (A) and malonyl loaded (B) with CIU50 values quantified. (C) Difference plot for holo-JuvEV replicates. (D) Difference plot comparing holo-JuvEV (red) and malonyl loaded JuvEV (blue). (E) Histogram charting the CIU50 1 (blue) and CIU50 2 (orange) values of both JuvEV samples with error bars indicated.

4.4   Conclusions

Here, we have endeavored to demonstrate the general ability of CIU to assess stability shift in large protein complexes comprised of multi-domain subunits involved in natural product biosynthesis. Our CID experiments using LC-Fourier-transform ion cyclotron resonance (FT-ICR) served to validate the modification of MM on both the ACP and AT domains, allowing us to link our CIU results to specific loading states of the enzymes studied. We used IM-MS to detect stability shifts for PikAIV–substrate complexes, including Ppant, hexaketide, and MM loading events for the first time. In addition, we have also been able to detect stability shifts upon binding a malonyl substrate to JuvEV, which has a multi-domain organization similar to PikAIII but lacks detailed structure data currently. As the small molecules tethered to the proteins studied in this report represent <1% of the total molecular mass of the systems interrogated, the data presented in this report provides the clearest example of the sensitivity limits of CIU technology. Our data also deepen our understanding of PikAIV and JuvEV function. We interpret the stability shifts reported here as indicative of a conformational change in the ~280kDa PikAIV and JuvEV dimers upon substrate loading, which may link to prior structural data collected for related pikromycin biosynthetic enzymes.[15,16,19] Future work will seek to marshal CIU data in combination with other structural biology probes in order to further elucidate the mechanistic relationship that exists between structure changes and function in these enzymatic complexes.

4.5 References

(1)   Rizwan, M.; Tse, J. W.; Nori, A.; Leong, K. W.; Yim, E. K. F. in *Principles of Regenerative Medicine* **2019,**

(2)   Deller, M. C.; Kong, L.; Rupp, B. *Acta Crystallogr. Sect. Struct. Biol. Commun.* **2016,** *72*, 72.

(3)   Li, C.; Zhang, R.; Wang, J.; Wilson, L. M.; Yan, Y. *Trends in Biotechnology* **2020,**

(4)   Ganem, B.; Li, Y. T.; Henion, J. D. *Journal of the American Chemical Society* **1991,**

(5)   Hu, P.; Loo, J. A. *J. Mass Spectrom.* **1995,**

(6)   McCammon, M. G.; Scott, D. J.; Keetch, C. A.; Greene, L. H.; Purkey, H. E.; Petrassi, H. M.; Kelly, J. W.; Robinson, C. V. *Structure* **2002,**

(7)   Kitova, E. N.; Seo, M.; Roy, P. N.; Klassen, J. S. *J. Am. Chem. Soc.* **2008,**

(8)   Hyung, S. J.; Robinson, C. V.; Ruotolo, B. T. *Chem. Biol.* **2009,** *16*, 382.

(9)   Niu, S.; Ruotolo, B. T. *Protein Sci.* **2015,**

(10)   Shelimov, K. B.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997,**

(11)   Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997,**

(12)   Whicher, J. R.; Smaga, S. S.; Hansen, D. A.; Brown, W. C.; Gerwick, W. H.; Sherman, D. H.; Smith, J. L.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R. H.; Håkansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G.; Trindade-Silva, A. E.; Lim-Fong, G. E.; Sharp, K. H.; Haygood, M. G.; Sashital, D. G.; Butcher, S. E.; Kittendorf, J. D.; Sherman, D. H.; Hansen, D. A.; Rath, C. M.; Eisman, E. B.; Narayan, A. R. H.; Kittendorf, J. D.; Mortison, J. D.; Yoon, Y. J.; Sherman, D. H.; Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R. H.; Håkansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G.; Buchholz, T. J.; Geders, T. W.; Bartley, F. E.; Reynolds, K. A.; Smith, J. L.; Sherman, D. H. *Nature* **2014,** *510*, 512.

(13)   Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R. H.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. *Nature* **2014,** *510*, 512.

(14)   Whicher, J. R.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R. H.; Håkansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. *Nature* **2014,** *510*, 560.

(15)   Lowell, A. N.; Demars, M. D.; Slocum, S. T.; Yu, F.; Anand, K.; Chemler, J. A.; Korakavi, N.; Priessnitz, J. K.; Park, S. R.; Koch, A. A.; Schultz, P. J.; Sherman, D. H. *J. Am. Chem. Soc.* **2017,** *139*, 7913.

(16)   Bonnett, S. A.; Rath, C. M.; Shareef, A. R.; Joels, J. R.; Chemler, J. A.; Hkansson, K.; Reynolds, K.; Sherman, D. H. *Chem. Biol.* **2011,** *18*, 1075.

(17)   Haynes, S. E.; Polasky, D. A.; Dixit, S. M.; Majmudar, J. D.; Neeson, K.; Ruotolo, B. T.; Martin, B. R. *Anal. Chem.* **2017,** *89*, 5669.

(18)   Polasky, D. A.; Dixit, S. M.; Fantin, S. M.; Ruotolo, B. T. *Anal. Chem.* **2019,** *91*, 3147.

(19)   Buchholz, T. J.; Geders, T. W.; Bartley, F. E.; Reynolds, K. A.; Smith, J. L.; Sherman, D. H. *ACS Chem. Biol.* **2009,** *4*, 41.

Chapter 5 Time-Resolved Ion Mobility-Mass Spectrometry Reveals Structural Transitions in the Disassembly of Modular Polyketide Syntheses

5.1 Introduction

Polyketide natural products account for either the core structures or complete chemical entities for many biologically important and clinically approved therapeutic agents. Polyketides are synthesized through the action of type I polyketide synthases (PKSs), which constitute a class of enzymatic assembly lines that exhibit modular architectures.[1] Accordingly, it is important to elucidate the architecture and dynamics PKS modules in order to enable rational bioengineering efforts aimed at creating high-value chemicals and novel drugs. However, PKS structural information is currently limited, thus obtaining information on their modular configuration, dynamism, and mechanisms of assembly is crucial for the success of future protein engineering efforts.

Type I PKS modules utilize functionally distinct domains to sequentially elongate, process and terminate polyketide chains. The most relevant modules for our studies include: the acyl carrier protein (ACP), keto synthase (KS) and acyltransferase (AT) domains. In addition to these domains, modules may also contain elements that successively modify the keto group to a hydroxyl (keto reductase (KR)), a double bond (dehydratase (DH)), or a single bond (enoyl reductase (ER)). At the end of many PKS modules, a terminal thioesterase (TE) domain offloads the final polyketide product. Previous studies have reported crystal structures for various excised PKS catalytic domains, KS-AT di-domains and docking domains[2–16], with some models based on

porcine fatty acid syntase[17]. Cryo-electron microscopy (cryo-EM) has been used to directly visualize the structure of a pikromycin PKS module 5 (PikAIII) and its structural rearrangements during the catalytic cycle when loaded with corresponding substrates. In general, however, only limited structural information is available for this class of large macromolecular complexes.

Currently, many technologies are well-established for biomolecular structure characterization at different levels of resolution.[18] For high-resolution protein structure determination, the atomic details of protein architecture are depicted. X-ray crystallography (XRD), nuclear magnetic resonance spectroscopy (NMR), and cryoEM offer routes to such data that have served to illuminate much of cellular biochemistry.[19–28] While these well-established structural biology tools are highly successful in many cases, large amounts of high-stability, high-purity protein are still typically required in order to derive atomic structure information .[29]

In order to overcome such technical challenges, native mass spectrometry (MS) has been developed to identifying transient, low abundance complexes. Native MS technologies work to preserve native protein structures in solution prior to ionization and introduction to vacuum, and such data can readily report on protein complex stoichiometry and interaction strengths between subunits.[30] When coupled with ion mobility (IM), a gas phase technique which separates ions based on their size, shape, and charge, the combined IM-MS technology can provide restraints sufficient to build 3D models of multiprotein topology.[31] In such models, coarse graining at the level of individual subunits is commonly practiced, which can lead to high fidelity protein quaternary structures for assemblies built from small, globular subunits, but can lead to unphysical models when larger, multi-domain proteins are involved. To overcome these challenges, gas-phase protein unfolding can be employed in order to resolve closely-related protein tertiary structures and assay the domain structures of multi-domain proteins. [32] Such

collision induced unfolding (CIU) experiments are realized through the stepwise heating of isolated gas-phase protein ions, where IM-MS is employed to record the changes in protein structure engendered during the collisional heating process. CIU fingerprinting data can reveal disulfide variants within biotherapeutic proteins[33] and detect subtle differences in protein stability associated with local structure changes too subtle to detect by IM separation alone.[20,34–37 38] Altogether, native MS, ion mobility, and CIU anchor a set of useful technologies that populate a continually expanding toolbox of structural MS technologies that have dramatically enhanced our understanding of protein complex architecture and function.

In this report, we study the 207kDa KS-AT di-domain protein dimer as a model of the broader PKS enzyme class. In our studies, we were able to observe three conformational states for KS-AT monomers and two conformational states for KS-AT dimers. A comparison between our measured collision cross section (CCS) values extracted from our dimer IM measurements with estimated CCS values derived from previously reported KS-AT dimer structures and models indicate that we observe evidence oligomers that occupy both the "AT-out" and "AT down" configurations. We further noted that these forms evolved as a function of time, and therefore performed a series of time-resolved IM-MS measurements to study the structural evolution of KS-AT oligomers, which eventually leads to the complete decay of KS-AT dimer states. CIU data acquired at time points that favor each KS-AT dimer forms suggest that the more compact state is ~ 30% more stable than its more elongated state. By integrating our time-resolved IM-MS and CIU data, we propose a detailed disassembly pathway for KS-AT dimers that elucidates this process for the first time.

5.2 Experimental Section

Sample Preparation

KS-AT di-domain dimer used in our study is excised from full length PikAIII. Protein sample was expressed and purified as described earlier[39] before buffer exchanged into 500mM ammonium acetate using Micro Bio-Spin 30 columns (Bio-Rad, Hercules, CA).

IM-MS Data Collection and CCS Analysis

Sample aliquots (~5μL) were analyzed by IM-MS on a quadrupole-ion mobility- time-of-flight mass spectrometer (Q-IM-ToF MS) instrument (Synapt G2 HDMS, Waters, Milford, MA) as described previously.[40] Protein ions were generated using a nESI source in the positive mode. Capillary voltages of 1.6−1.7 kV were applied, with the sampling and extraction cones set to 50 V and 10 V, respectively. The source backing pressure was adjusted to ~8 mbar. The IM T-wave ion guide was operated a 4 mbar with wave height and wave velocity values of 15 V and 150 m/s, respectively.

All experimental collision cross section (CCS) values were calibrated using ions of known CCS as described previously in order to relate IM drift times directly to ion size and shape.[41] Theoretical CCS ranges for "AT-out" and "AT- down" KS-AT homodimer as well as the native monomer were also calculated from the published PDB structure using the trajectory method estimation (TJM) value reported by IMPACT[42] as the upper bound and 90% of the IMPACT TJM projected as the lower bound accounting for known gas-phase compaction of protein ions.[43,44]

Time Resolved IM-MS Experiments

KS-AT samples stored in buffers that lack glycerol and NaCl leads to disassembly of the native dimer. Immediately following buffer exchange into 500mM ammonium acetate, protein sample aliquots were injected for IM-MS analysis in order to record a time 0 dataset. Additional aliquots were then tested every 40min for a total of 12 time points. KS-AT monomer peak intensity information was collected by integrating charge states ranging from 21^+ to 25^+, and different conformational states were extracted based on the drift time differences observed in order to determine the relative intensity for each conformation. Similarly, KS-AT dimer peak intensity was collected by combining the signal intensities for dimer charge states ranging from 28^+ to 33^+, followed by the drift time resolved extraction of extended and compact dimer signal intensity information.

CIU Data Analysis

In order to collect CIU data for isolated extended monomer and dimer states, time 0 data was used for all aliquots tested directly following buffer exchanged into 500mM ammonium acetate, and the charge states that only presenting the extended state at time 0 were selected for CIU comparisons. Conversely, CIU data for compacted monomer and dimer states were extracted at following extended incubation times in 500 mM ammonium acetate, and charge states were further selected and narrowed down. CIU data were analyzed using CIUSuite 2[45]. CIU fingerprints were plotted as a 2D contour plot, in which ion intensities were normalized to a maximum value of 1 at each collision voltage and smoothed using a Savitzky-Golay filter. Feature detection was applied within the data sets to describe CIU transitions.

5.3   Result and Discussion

Different conformational states observed by IM-MS

To probe the structure and dynamics of PKS enzymes, we began our studies by focusing on the KS-AT di-domain homodimer excised from the pikromycin PKS module 5 (PikAIII) since these two critical domains form the centerpiece of all PKS systems. Previous reports have found that the KS-AT homodimer housed within the full length PKS module adopts a "AT-down" conformation, creating an arch-shaped symmetric dimer with a single active site, as shown in Figure 5-1A.[46] Other prior work was able to dissect a PKS module into individual active KS-AT and ACP components which could be reconstituted in order to perform all of the normal catalytic functions of the original homodimeric module.[47] Following these efforts, the high-resolution structure of the 194-kDa homodimeric fragment containing the KS–AT di-domain of module 5 within 6-deoxyerythronolide B synthase (DEBS), which is also part of the PKS family, was solved to a resolution of 2.7Å. This di-domain contained the full-length KS and AT modules as well as three flanking peptide linkers, each with a unique secondary structure. This DEBS KS-AT di-domain adopts an "AT-out" conformation as shown in Figure 5-1B.

In order to probe the conformational landscape of KS-AT homodimers, we began by collecting IM-MS for our KS-AT di-domain samples extracted from PikAIII, and the resulting data is shown in Figure 5-1C. We observe signals corresponding to both KS-AT dimers and monomers across a range of charge states with both states exhibiting similar ion intensity. IM data reveals evidence for at least three conformational families for KS-AT monomers and two different conformational families for KS-AT homodimers. In order assign the KS-AT homodimer conformational states observed in our experiments, we recorded CCS values for each form and compared these experimental values with those estimated for models drawn from the "AT-out" and "AT-down" KS-AT homodimer forms discussed above. As expected, our results indicate agreement with both the "AT-out" and "AT-down" models for the extended and compact KS-AT

dimer CCS values recorded respectively across a range of charge states. This strongly suggests that the "AT-down" and "AT-out" forms represent conformational states that exist in equilibrium within a native KS-AT ensemble. We also extracted KS-AT monomer models from existing dimer constructs, and note that intermediate monomer CCS values agrees well with he expected CCS range for the native KS-AT monomer.

A

B

C

| Relative Intensity (%) | 22+ 21+ | 29+ 28+ |

Drift time (ms)

m/z

—Extended Monomer   —Extended Dimer
—Intermediate Monomer   —Compact Dimer
—Compact Monomer

D

CCS (Å²)

Charge State

☐ 'AT Out' Model Range
☐ 'AT Down' Model Range
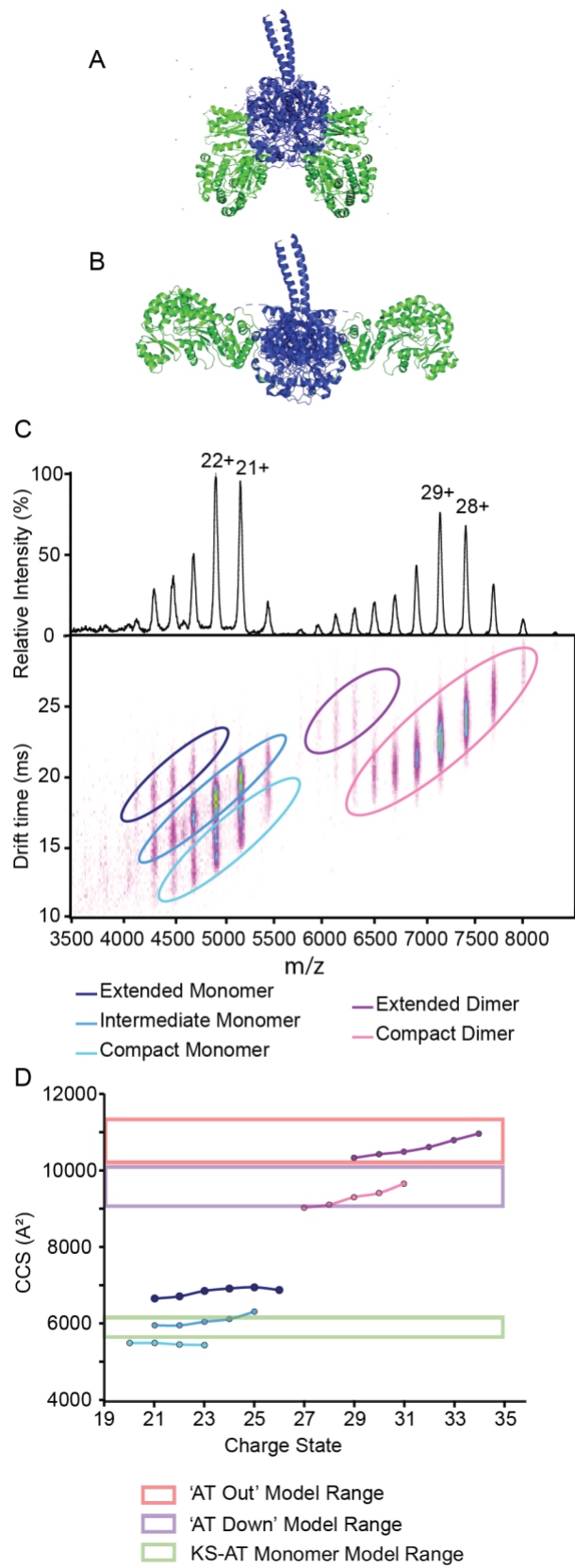☐ KS-AT Monomer Model Range

Figure 5-1. (A) High-resolution structure of the of KS-AT dimer in the "AT-down" configuration generated from a previously reported full length model[48]. (B) High-resolution structure of the of KS-AT dimer in the "AT-out" configuration based on the X-ray crystal structure of the KS-AT didomain from DEBS module5 (PDB 2HG4)[49]. (C) IM-MS data acquired for KS-AT monomer and dimer at 80min time point. IM data suggests the presence of three conformations for the KS-AT monomer including extended (dark blue), intermediate (blue) and compact states (light blue), and two conformations for KS-AT dimers including extended (purple) and compact states (pink). (D) CCS measurements of the three monomer and two dimer conformational families detected for KS-AT oligomers observed across different charge states; estimated CCS values for the "AT-out" and "AT-down" KS-AT dimer model ranges, as well as native monomer model range based on reported high-resolution structure information.

Time-Resolved IM-MS Tracks KS-AT Conformational Changes

After we established a correlation between the KS-AT conformations observed in our experiments and those reported previously, we aimed to further explore the evolution of this conformational ensemble over time. As indicated above, KS-AT samples stored in buffers that lack NaCl and glycerol are inherently unstable, leading ultimately to the disassembly of the native dimer. In order to capture this disassembly process, we initiated a series of time-resolved IM-MS measurements to track the relative intensity of various conformational families we detect at time 0 over an eight-hour period. Figure 5-2A shows drift time data acquired for the 22^+ KS-AT monomers over 4 different time points. Over the course of three hours, monomer conformations shift from an ensemble that favors the intermediate state to one that favors compact CCS values, with the extended state representing a minor yet decaying feature of the conformational profile overall. In parallel, we mapped the shifts in drift time data recorded for KS-AT dimers, and a similar analysis conducted over 4 time points over the same 3 hour timespan discussed above, indicates a conformational ensemble balanced between the 'AT-down' and 'AT-out' forms at time 0, which shifts to favor the more compact 'AT-down' form over time.

When we extend our analysis to the full 440 min timespan probed in our experiment, the observations made above are confirmed (Figure 5-2C), with the KS-AT dimer decaying

dramatically in intensity to favor the 'AT-down' state, and monomers strongly favoring their most compact configuration. When integrated, our data points to both compact monomer and dimer KS-AT forms as those that are thermodynamically favored under our destabilizing experimental conditions.
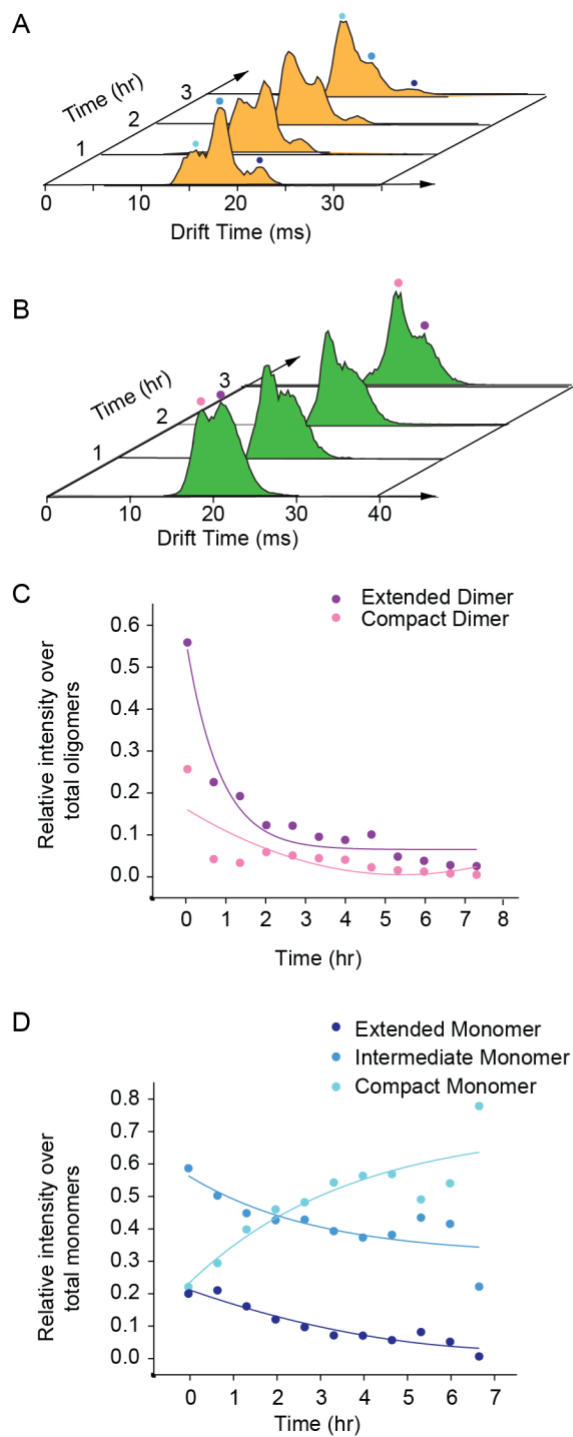
Figure 5-2. (A) Drift time data for 22+ KS-AT monomer ions plotted at 4 different time points, with compact (cyan) intermediate (light blue) and extended (dark blue) states shown. (B) Drift time data for 33^+ KS-AT dimer with compact (light pink) and extended (purple) states labelled. (C) Time resolved experiment recorded across 440min, tracking the relative intensity of the extended dimer (purple) and compact dimer (pink). (D) Relative intensities of extended (dark blue), intermediate (blue), and compact (light blue) monomer forms as a function of time.

Gas-Phase Unfolding Reveals Stability Differences in KS-AT Conformations

Having assessed the time-dependent conformational shifts of KS-AT oligomers under destabilizing buffer conditions, we sought to independently assess the stabilities of the conformational states observed in our IM-MS data. To this end, we deployed CIU measurements on carefully selected charge states and time points in order to assess each state in relative isolation. Figure 5-3A and Figuer 5-3B display such a CIU experiment performed for both the extended and compact 29+ KS-AT dimer. Across both datasets, we observe three different CIU features. We note that the remarkable similarity in the CIU fingerprints for these two forms, with a lack of a clear transition between what we have assigned as the 'AT-down' and 'AT-out' in analyses above, further confirms that the ground state conformational ensembles observed in our IM-MS are not the product of gas-phase activation. Upon comparison, the most significant difference revealed in the CIU fingerprints recorded for the two KS-AT conformational families relates to the CIU50 stability value recorded. Specifically, the CIU50 we observe for the extended dimer is 66.3±1.2V and 83.8±2.1V for its compact state. This result indicates that the compact KS-AT dimer state is 30% more stable than the extended dimer form, an observation which is consistent our time-resolved IM-MS data that also indicated the greater stability of the compact conformer family.
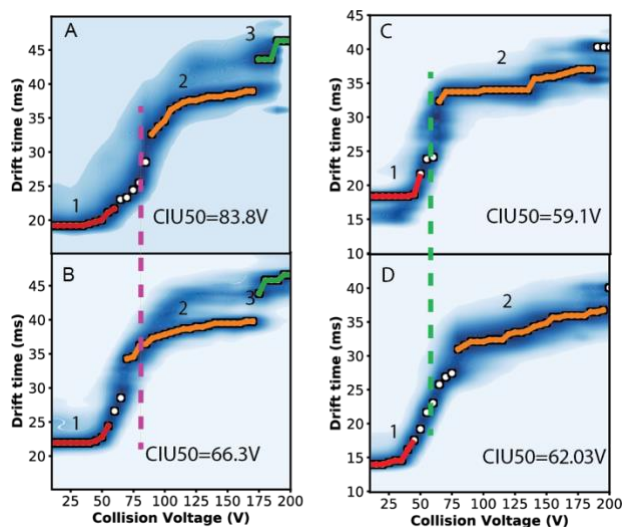
Figure 5-3. CIU fingerprints of the compact dimer (A) and the extended dimer (B) with features annotated. CIU50 values were measured for both dimer states, with the compact dimer being 83.8±2.1V and the extended dimer being 66.3±1.2V.

We conducted a similar set of CIU experiment for KS-AT monomers. Since the resolved structural states we observe are more difficult to isolate in the same manner we describe for our dimer CIU analysis, we were only able to successfully capture CIU data for the extended compact monomer states. As shown in Figure 5-3C and Figure 5-3D, two CIU features with a single sharp transition to significantly more extended state are observed for both monomer states probed here, with the compact monomer exhibiting greater stability than the extended state. As discussed above in reference to our dimer CIU data, our monomer stability assessment similarly comports with our time-resolved IM-MS data, where we observed the most compact monomer conformation to be the most stable.

A Model of KS-AT Disassembly Based on IM-MS Data

Taken together, our data allows us to propose a mechanism for the disassembly of the KS-AT di-domain dimer. Our data indicates that the KS-AT extended dimer decays to a more stable compact form, which eventually decays into monomeric units. Meanwhile, the KS-AT monomer

101

population exists as an equilibrium ensemble of at least three conformational families, which eventually decays to the most compact state observed. Due to its agreement with model CCS values extracted from KS-AT dimer structure data, we propose that this form is generated directly from dimer dissociation events, but is rapidly shuttled to other forms.
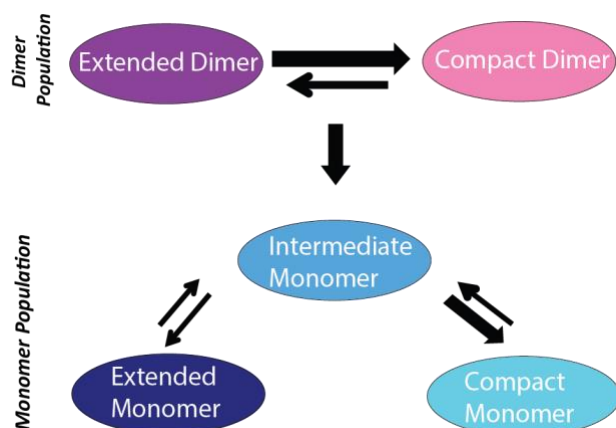

Figure 5-4. Disassembly pathway of KS-AT system including the two dimer conformations and three monomer conformations we observe experimentally.

5.4 Conclusions

Here, we present the first detailed analysis of PKS oligomer disassembly. Our work focuses on the centrally important KS-AT di-domain dimer, where we deploy time-resolved IM-MS, CCS, and CIU analyses in order to propose a comprehensive model of KS-AT dimer disassembly. Our CCS analysis reveals a that both KS-AT monomers and dimers exist as ensembles of transient conformational families, with the dimer occupying both of the previously reported quaternary structures previously reported for this system. Time-resolved IM-MS data indicate that KS-AT dimers and monomers evolve structurally as a function of time, with both forms decaying ultimately to compact states. In addition, we utilize CIU to explore the stabilities of individual conformational states isolated for both KS-AT complexes and subunits, revealing that the

preference for compact forms detected in our time resolved data appears to be linked to the greater thermal stability of these conformers. Overall, our IM-MS data is able to track KS-AT structures that are likely too transient to be captured for high-resolution structure analysis, highlighting the continuing importance of integrated structural biology campaigns for uncovering the role of dynamism in protein function.

Although the experimental data presented in this study provides an important contribution to the paucity of structural information surrounding PKS proteins, the disassembly pathway established here remains a putative insight. Future efforts in molecular modeling, using accumulated CCS data from our study, are needed in order to fully map the pathway outlined here. Overall, native MS, IM, and CIU have a bright future in PKS structural biology. We envision both dedicated IM-MS data collection campaigns aimed at assessing the role of PKS modules in protein complex formation and disassembly, as well as using IM-MS data to screen conditions for those most appropriate for downstream high-resolution structure analysis.

5.5 References

(1)   Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2012,** *75*, 311.
(2)   Broadhurst, R. W.; Nietlispach, D.; Wheatcroft, M. P.; Leadlay, P. F.; Weissman, K. J. *Chem. Biol.* **2003,**
(3)   Buchholz, T. J.; Geders, T. W.; Bartley, F. E.; Reynolds, K. A.; Smith, J. L.; Sherman, D. H. *ACS Chem. Biol.* **2009,** *4*, 41.
(4)   Gehret, J. J.; Gu, L.; Gerwick, W. H.; Wipf, P.; Sherman, D. H.; Smith, J. L. *J. Biol. Chem.* **2011,**
(5)   Scaglione, J. B.; Akey, D. L.; Sullivan, R.; Kittendorf, J. D.; Rath, C. M.; Kim, E. S.; Smith, J. L.; Sherman, D. H. *Angew. Chemie - Int. Ed.* **2010,**
(6)   Kosol, S.; Gallo, A.; Griffiths, D.; Valentic, T. R.; Masschelein, J.; Jenner, M.; de los Santos, E. L. C.; Manzi, L.; Sydor, P. K.; Rea, D.; Zhou, S.; Fülöp, V.; Oldham, N. J.; Tsai, S. C.; Challis, G. L.; Lewandowski, J. R. *Nat. Chem.* **2019,**
(7)   Tsai, S. C.; Lu, H.; Cane, D. E.; Khosla, C.; Stroud, R. M. *Biochemistry* **2002,**
(8)   Bonnett, S. A.; Whicher, J. R.; Papireddy, K.; Florova, G.; Smith, J. L.; Reynolds, K. A. *Chem. Biol.* **2013,**
(9)   Tang, Y.; Chen, A. Y.; Kim, C. Y.; Cane, D. E.; Khosla, C. *Chem. Biol.* **2007,** *14*, 931.

(10) Tang, Y.; Kim, C.-Y.; Mathews, I. I.; Cane, D. E.; Khosla, C. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103*, 11124.

(11) Zheng, J.; Keatinge-Clay, A. T. *J. Mol. Biol.* **2011,**

(12) Zheng, J.; Taylor, C. A.; Piasecki, S. K.; Keatinge-Clay, A. T. *Structure* **2010,**

(13) Keatinge-Clay, A. T. *Chem. Biol.* **2007,**

(14) Keatinge-Clay, A. *J. Mol. Biol.* **2008,**

(15) Akey, D. L.; Razelun, J. R.; Tehranisa, J.; Sherman, D. H.; Gerwick, W. H.; Smith, J. L. *Structure* **2010,**

(16) Zheng, J.; Gay, D. C.; Demeler, B.; White, M. A.; Keatinge-Clay, A. T. *Nat. Chem. Biol.* **2012,**

(17) Maier, T.; Leibundgut, M.; Ban, N. *Science (80-.).* **2008,**

(18) Steven, A. C.; Baumeister, W. **2008,** *163*, 186.

(19) Clemmer, D. E.; Jarrold, M. F. *Journal of Mass Spectrometry* **1997,**

(20) Hyung, S. J.; Robinson, C. V.; Ruotolo, B. T. *Chem. Biol.* **2009,** *16*, 382.

(21) Loo, J. A.; Berhane, B.; Kaddis, C. S.; Wooding, K. M.; Xie, Y.; Kaufman, S. L.; Chernushevich, I. V. *J. Am. Soc. Mass Spectrom.* **2005,**

(22) Politis, A.; Park, A. Y.; Hyung, S. J.; Barsky, D.; Ruotolo, B. T.; Robinson, C. V. *PLoS One* **2010,** *5*,

(23) Pukala, T. L.; Ruotolo, B. T.; Zhou, M.; Politis, A.; Stefanescu, R.; Leary, J. A.; Robinson, C. V. *Structure* **2009,** *17*, 1235.

(24) Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. *Science (80-.).* **2005,** *310*, 1658.

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

(26) Van Duijn, E.; Barendregt, A.; Synowsky, S.; Versluis, C.; Heck, A. J. R. *J. Am. Chem. Soc.* **2009,**

(27) Wang, S. C.; Politis, A.; Di Bartolo, N.; Bavro, V. N.; Tucker, S. J.; Booth, P. J.; Barrera, N. P.; Robinson, C. V. *J. Am. Chem. Soc.* **2010,**

(28) Wyttenbach, T.; Bowers, M. T. *Annu. Rev. Phys. Chem.* **2007,**

(29) Robinson, C. V.; Sali, A.; Baumeister, W. *Nature* **2007,**

(30) Benesch, J. L. P.; Ruotolo, B. T. *Current Opinion in Structural Biology* **2011,**

(31) Jenner, M.; Ellis, J.; Huang, W. C.; Lloyd Raven, E.; Roberts, G. C. K.; Oldham, N. J. *Angew. Chemie - Int. Ed.* **2011,** *50*, 8291.

(32) Mayer, P. M.; Blanchet, V.; Joblin, C. *J. Chem. Phys.* **2011,**

(33) Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997,**

(34) Ruotolo, B. T.; Hyung, S. J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V. *Angew. Chemie - Int. Ed.* **2007,** *46*, 8001.

(35) Hopper, J. T. S.; Oldham, N. J. *J. Am. Soc. Mass Spectrom.* **2009,** *20*, 1851.

(36) Han, L.; Hyung, S. J.; Mayers, J. J. S.; Ruotolo, B. T. *J. Am. Chem. Soc.* **2011,**

(37) Zhou, M.; Dagan, S.; Wysocki, V. H. *Analyst* **2013,**

(38) Niu, S.; Ruotolo, B. T. *Protein Sci.* **2015,** *24*, 1272.

(39) Caldara-Festin, G.; Jackson, D. R.; Barajas, J. F.; Valentic, T. R.; Patel, A. B.; Aguilar, S.; Nguyen, M. C.; Vo, M.; Khanna, A.; Sasaki, E.; Liu, H. W.; Tsai, S. C.; Smith, J. L. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112*, E6844.

(40) Zhong, Y.; Han, L.; Ruotolo, B. T. **2014,** 9209.

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

(42) Marklund, E. G.; Degiacomi, M. T.; Robinson, C. V.; Baldwin, A. J.; Benesch, J. L. P. *Structure* **2015,** *23*, 791.

(43) Bornschein, R. E.; Ruotolo, B. T. *Analyst* **2015,**

(44) Bornschein, R. E.; Niu, S.; Eschweiler, J.; Ruotolo, B. T. *J. Am. Soc. Mass Spectrom.* **2016,**

(45) Polasky, D. A.; Dixit, S. M.; Fantin, S. M.; Ruotolo, B. T. *Anal. Chem.* **2019,** *91*, 3147.

(46) Whicher, J. R.; Smaga, S. S.; Hansen, D. A.; Brown, W. C.; Gerwick, W. H.; Sherman, D. H.; Smith, J. L.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R. H.; Håkansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G.; Trindade-Silva, A. E.; Lim-Fong, G. E.; Sharp, K. H.; Haygood, M. G.; Sashital, D. G.; Butcher, S. E.; Kittendorf, J. D.; Sherman, D. H.; Hansen, D. A.; Rath, C. M.; Eisman, E. B.; Narayan, A. R. H.; Kittendorf, J. D.; Mortison, J. D.; Yoon, Y. J.; Sherman, D. H.; Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R. H.; Håkansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G.; Buchholz, T. J.; Geders, T. W.; Bartley, F. E.; Reynolds, K. A.; Smith, J. L.; Sherman, D. H. *Nature* **2014,** *510*, 512.

(47) Chen, A. Y.; Schnarr, N. A.; Kim, C. Y.; Cane, D. E.; Khosla, C. *J. Am. Chem. Soc.* **2006,** *128*, 3067.

(48) Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R. H.; Hakansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. *Nature* **2014,** *510*, 512.

(49) Tang, Y.; Kim, C. Y.; Mathews, I. I.; Cane, D. E.; Khosla, C. *Proc. Natl. Acad. Sci. U. S. A.* **2006,**

Chapter 6 Conclusions and Future Directions

6.1   Conclusions and Future Directions of CIU to Explore Multiprotein Complexes Gas-
       Phase Unfolding Mechanism

Our studies in gas-phase protein unfolding have been focused on combining other analytical
tools to complement IM-MS in order to understand protein unfolding process at domain level.
More importantly, we attempt to advance CIU technology into a more robust tool to bring the
measurement toward mainstream of structural biology tools. In chapter 2, we use HSA dimer
ions as a model multiprotein complex to study the unfolding mechanism. Previous study found
that the ejection of small molecules during the CIU process was highly correlated with specific
structural transitions, which allowed us to develop a model of albumin unfolding based on CID
of small molecules of known binding locations.[91] With similar approach combined utilization of
domain-specific chemical probes and domain-deleted protein constructs, we were able to assign
CIU features to the unfolding of different domains of HSA dimer. Our study further indicates
that a single monomer within the complex is responsible for the CIU transitions observed for the
dimer.

In chapter 3, we demonstrate the capability of CIU to elucidate the domain-specific gas-phase
unfolding mechanism of TamI by correlating CID and CIU data for this biosynthetic enzyme, in
preparation for future protein engineering efforts that leverage CIU data extensively.
Specifically, we reveal that the loop region, serving as the "lid" of the TamI substrate pocket

unfolds at lower activation energies, while the heme binding pocket unfolds at higher energies. For the first time, we provided insight about the conformational energy landscape of a cytochrome P450 enzyme, which can be used to engineer stable TamI, and possibly other P450 enzymes, for diverse biocatalytic and synthetic biology applications.

To fully understand multiprotein complexes unfolding mechanism, more detailed case studies of proteins with different tertiary structures must be invested. In addition, it's extremely difficult to obtain domain level information while domain-level local structure information for large multiprotein system is needed to reduce the errors compared to subunit level models. Another challenge is to find domain-specific ligand in order to correlate the ligand dissociation curve with protein unfolding. Lastly, although we were able to prove the HSA dimer unfolding follows the single chain unfolding theory, it's still not fully understood about how the presence or absence of charge may affect the information content of the experiment.

Further directions to fully understand multiprotein unfolding include combining IM-MS with novel analytical tools to unveil more information content from experiments. In Chapter 2, with the help from steered MD simulation using CIU-ECD fragment data as restraints, models were generated to rule out the possibilities of unfolding mechanisms. Structural modeling has proved to be a powerful tool when using data generated from IM-MS experiments as restraints. With the recent development in coarse-grained models building, CIU has the great potential to expand its application to broader and more complicated multiprotein complex system to achieve more thorough understanding of multiprotein unfolding mechanism.

6.2 Conclusions and Future Directions of CIU for Protein Engineering Efforts

Our work concerned with biosynthetic enzymes focus on pushing IM-MS and CIU methods to probe larger multiprotein complexes without sensitivity loss in order to unveil transient protein-protein and protein-substrate interactions. In chapter 3, we apply IM-MS and CIU to demonstrate that TamI and TamL interact and form a biocatalytically competent heterodimeric complex *in vivo*, and the binding affinities between a range of tirandamycin antibiotics with both wild type and fusion protein constructs were measured utilizing IM-MS. In chapter 4 and 5, we focus on type I PKS system. We elucidate a detailed disassembly pathway for KS-AT dimers for the first time. In Chapter 5, we studied the stability of a ~280kDa PKS dimer protein complex in order to demonstrate the current limits of quantitative CIU technology. Stability shifts associated with substrate binding that account for <0.1% of the mass for the intact assembly were detected. It's the first use of CIU to measure the stability differences in substrate-bound multiprotein complexes at this level.

Using IM-MS for biosynthetic enzymes studies presented in this dissertation contributes to the deeper understanding of protein-substrate interaction in order to shed light on structural information. However, to unveil more detail information of molecular architecture, there are still gaps need to be filled using other analytical tools. In addition to CIU, CID and ECD are also considered as powerful approaches for energy activation to obtain more information. In Chapter 2, ECD was coupled with CIU in order to get fragmentation information to add restraints to structural modeling. In the Appendix, we also presented the utilization of CID as another approach to capture protein structural difference. Furthermore, with the use of smaller emitter needles, CIU technology can be applied to broader sample pool with more complicated buffer conditions beyond ammonium acetate.

The insights gained here proved its great usage for protein engineering, especially for P450 enzymes and PKSs, having the potential to improve enzyme properties, create novel activities and alter enzyme substrate selectivity. Additionally, our work illustrated that combining IMMS with cryo-EM is a very impactful analytical approach to detect and further characterize the stoichiometry and topology of transient protein-protein, protein-substrate complexes in complicated biosynthetic pathways, where IMMS is built as a screening paradigm for CryoEM sample preparation by developing predictive signatures for the construction of high-resolution structure (Figure 6-1).
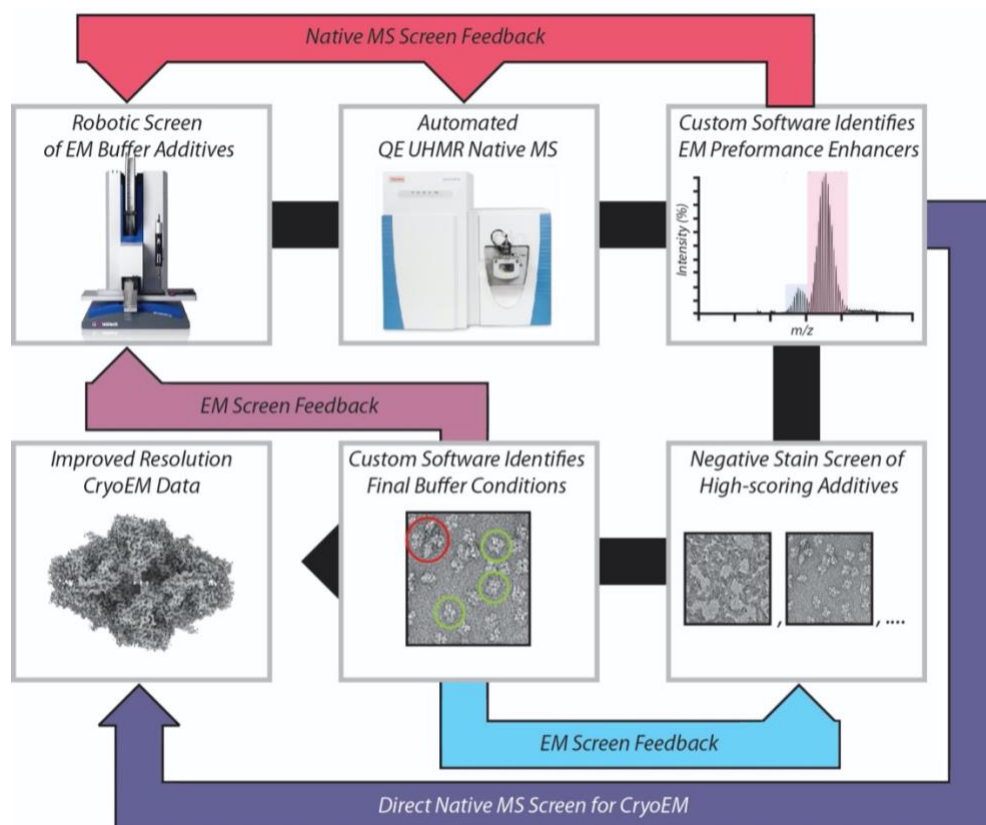

Figure 6-1. Illustration of the high-throughput MS screening for the high-resolution CryoEM images production.

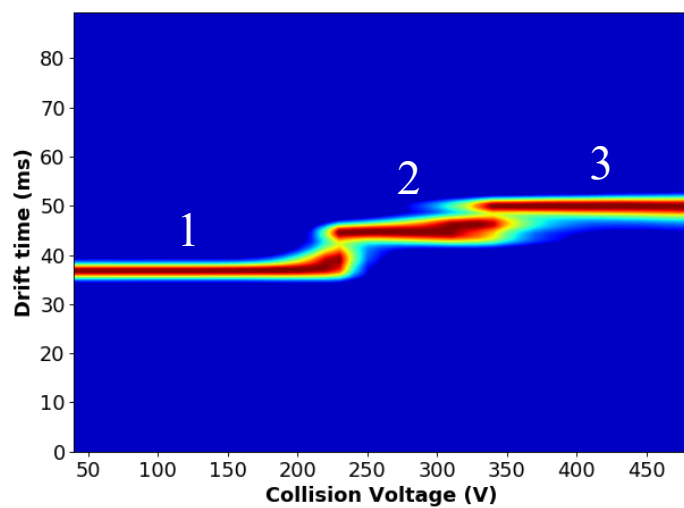# Appendix I. Supporting Information for Chapter 2

Figure I-2. The CIU fingerprint of D12 monomer of 13+ with three features observed.
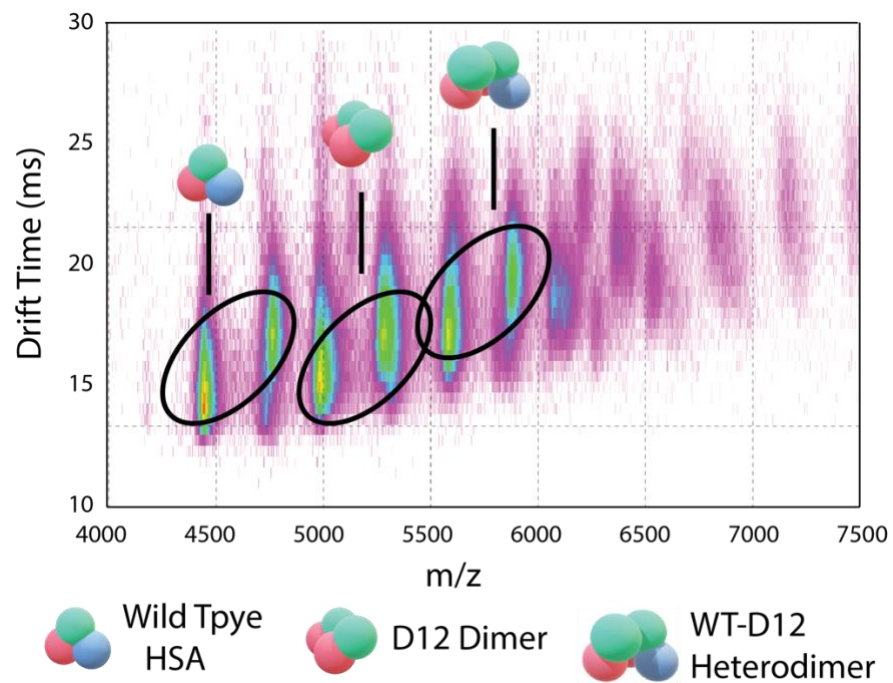
Figure I-3. Mobility data for sample mixture of D12 dimer and wildtype HSA, showing the presence of wild type HSA monomer, D12 dimer, and D12-D123 heterodimer formation.
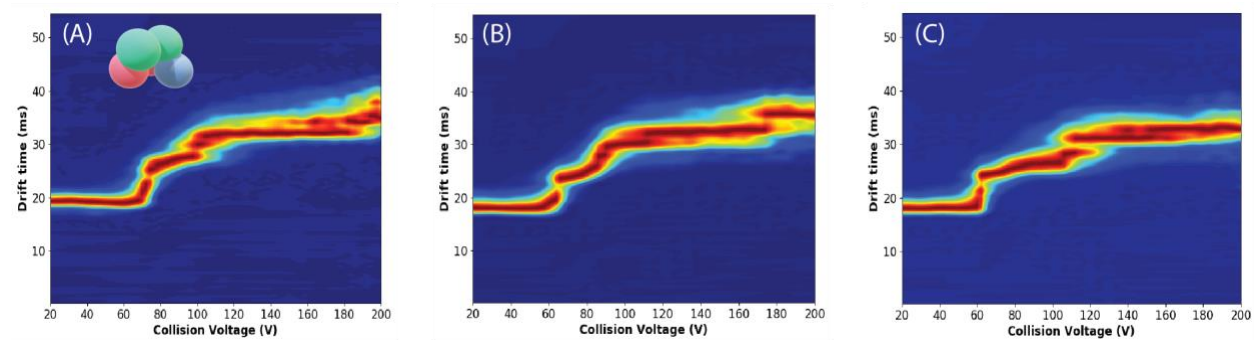

Figure I-4. CIU fingerprints for D12-D123 HSA heterodimer of 19+ (A), 20+ (B), and 21+ (C).
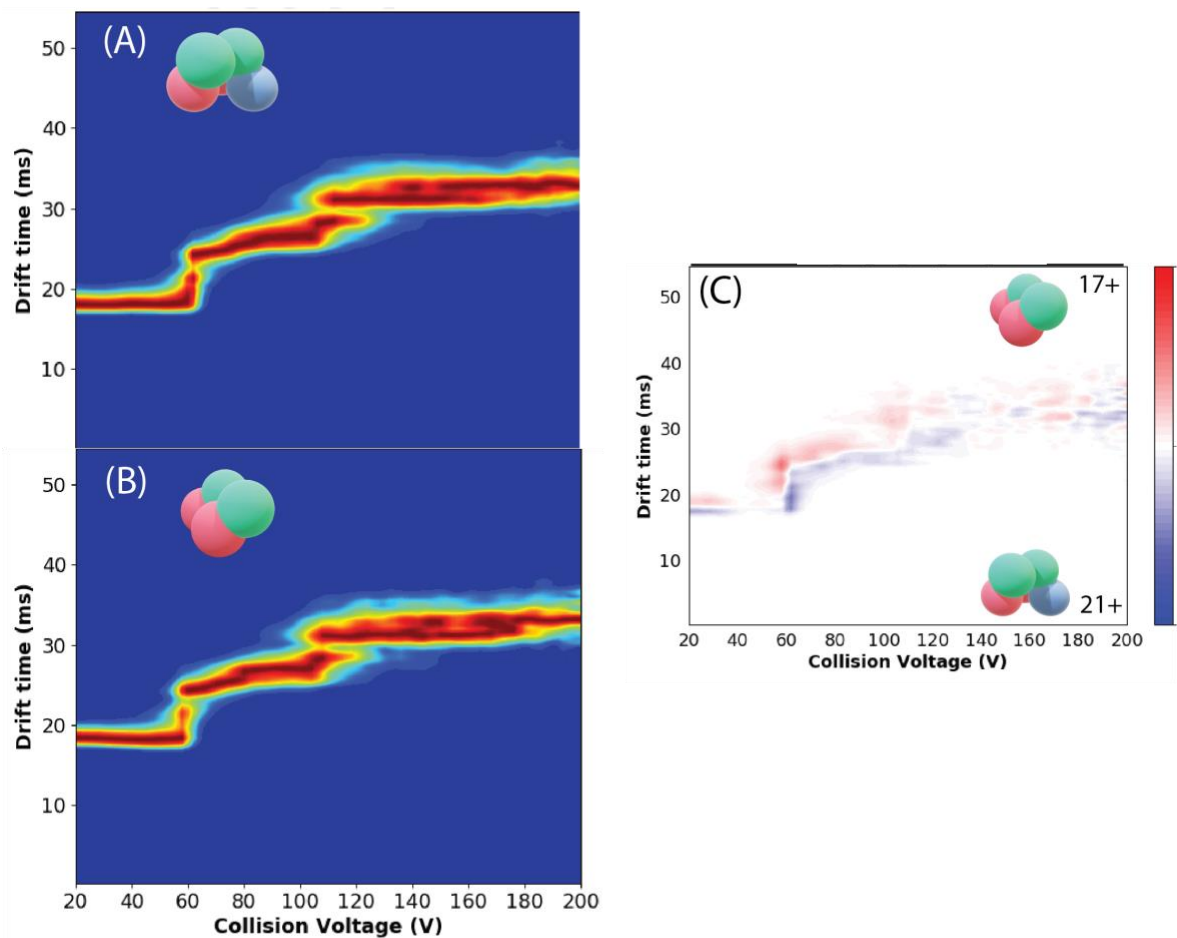
Figure I-5. (A) CIU fingerprints of HSA D12-D123 heterodimer of 21+. (B) CIU fingerprints of D12 dimer of 17+. (C) CIU difference plot comparing HSA D12-D123 of 21+ (blue) and D12 dimer of 17+ (red)

(A)

DAHKSEVAHRFKDLGEENFKALVLIAFAQYLQQCPFEDHVKLVNEVTEFAKTCVADES

AENCDKSLHTLFGDKLCTVATLRETYGEMADCCAKQEPERNECFLQHKDDNPNLPRLV

RPEVDVMCTAFHDNEETFLKKYLYEIARRHPYFYAPELLFFAKRYKAAFTECCQAADK

AACLLPKLDELRDEGKASSAKQGLKCASLQKFGERAFKAWAVARLSQRFPKAEFAEVS

KLVTDLTKVHTECCHGDLLECADDRADLAKYICENQDSISSKLKECCEKPLLEKSHCIAE

VENDEMPADLPSLAADFVGSKDVCKNYAEAKDVFLGMFLYEYARRHPDYSVVLLLRL

AKTYETTLEKCCAAADPHECYAKVFDEFKPLVEEPQ

(B)

Table I-1. (A) HSA D12 sequence. The construct encompasses residues 1-385 of HSA. (B) Terminal fragments matches with sequence, m/z value, charge, error tolerance and ion type.

| Sequence	*m/z*	*z*	Error (ppm)	ion type
YAKVFDEFKPLVEEPQ	961.987	2	-7.385	(z)16
AKVFDEFKPLVEEPQ	880.455	2	2.787	(z)15
AKVFDEFKPLVEEPQ	1759.903	1	4.877	(z)15
KVFDEFKPLVEEPQ	844.9364	2	0	(z)14
KVFDEFKPLVEEPQ	1688.866	1	0	(z)14
FDEFKPLVEEPQ	1461.702	1	-3.888	(z)12
DEFKPLVEEPQ	1314.634	1	-6.085	(z)11
EFKPLVEEPQ	1199.607	1	3.886	(z)10
EFKPLVEEPQ	1215.626	1	-1.352	(y)10
FKPLVEEPQ	1070.564	1	0.2848	(z)9
FKPLVEEPQ	1086.583	1	0	(y)9
KPLVEEPQ	923.496	1	-1.959	(z)8
KPLVEEPQ	939.515	1	0.9254	(y)8
PLVEEPQ	811.420	1	-2.126	(y)7
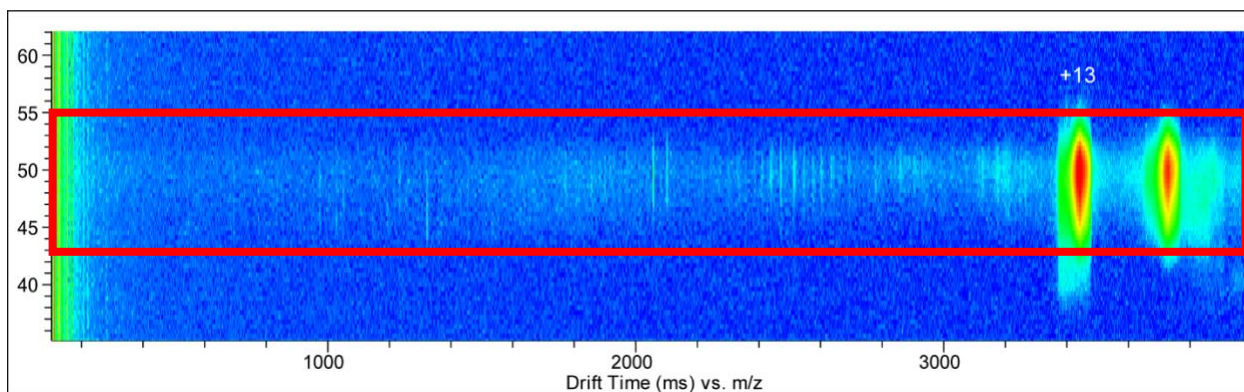
Figure I-6. (A) IM-ECD-MS spectrum of HSA D1213+ at collision voltage of 400V. The new signals that appeared at this voltage do not come from in-source CID. If that was the case the CID fragment would have been separated in the IM dimension. However, in the spectrum above the fragments occupy the same drift bins as the precursor ion.

Table I-2. Internal fragment matches from CIU-ECD experiments of HSA D1213+. The following information for each internal fragment identified is shown: m/z values, charge state, error tolerance, the number of disulfides bonds (ss_count), and the modification for the reduced cysteines present (mods). The mod 'chhsshl' represent a loss of -CH_2SSH (79 Da) from a cysteine residue.

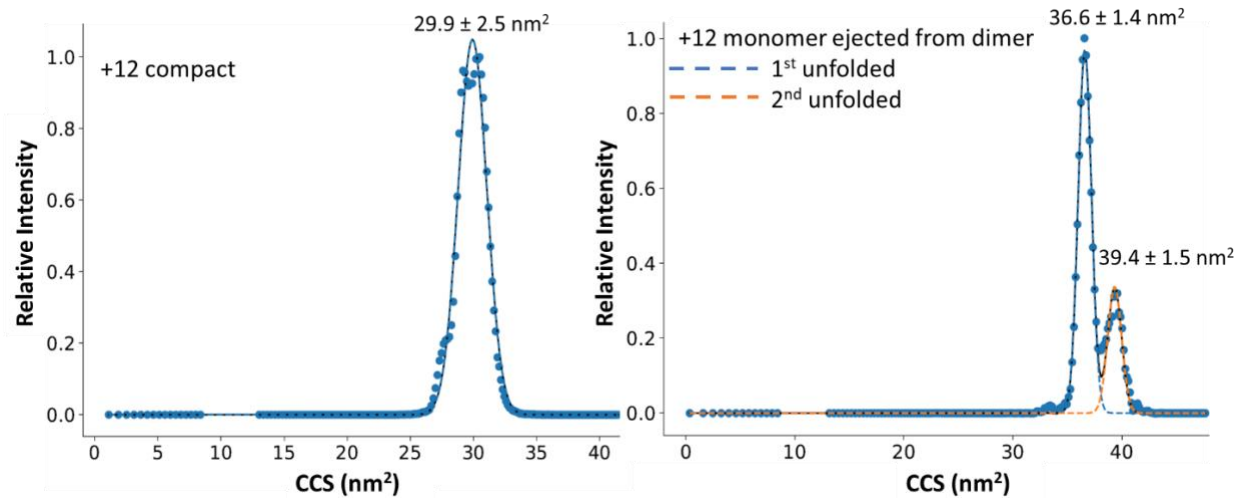| z	*m/z*	ion_type	mods	ss_count	error	isomz_score	isoint_score	fragment_score
4	2601.255	*c*271-*z*14	('chhsshl', 'chhsshl')	2	-0.35814	100	67	72
4	2241.295	*c*69-*y*230	('chhsshl',)	3	-0.48644	100	75	78

Figure I-6. CCS distributions of HSA D12 monomer (a) from native MS and (b) ejected from HSA D12 dimer.

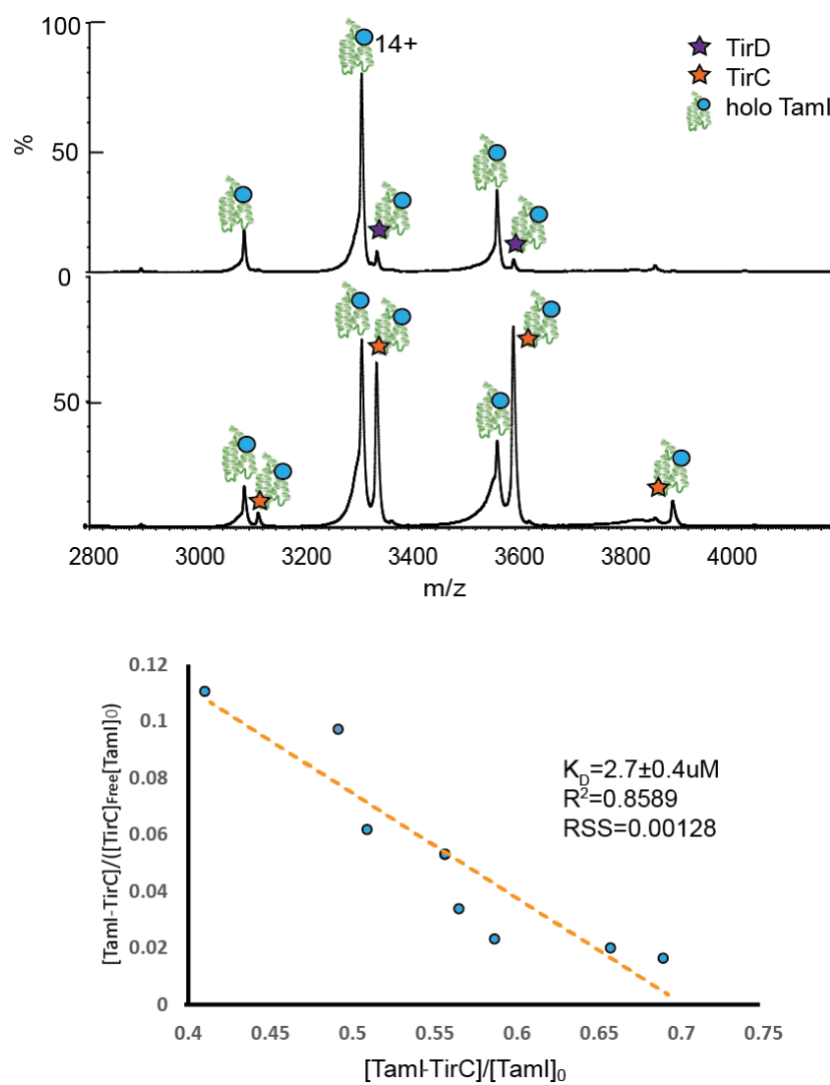# Appendix II. Supporting Information for Chapter 3

Figure II-1. Substrate binding affinity measurement of TamI using native IM-MS.
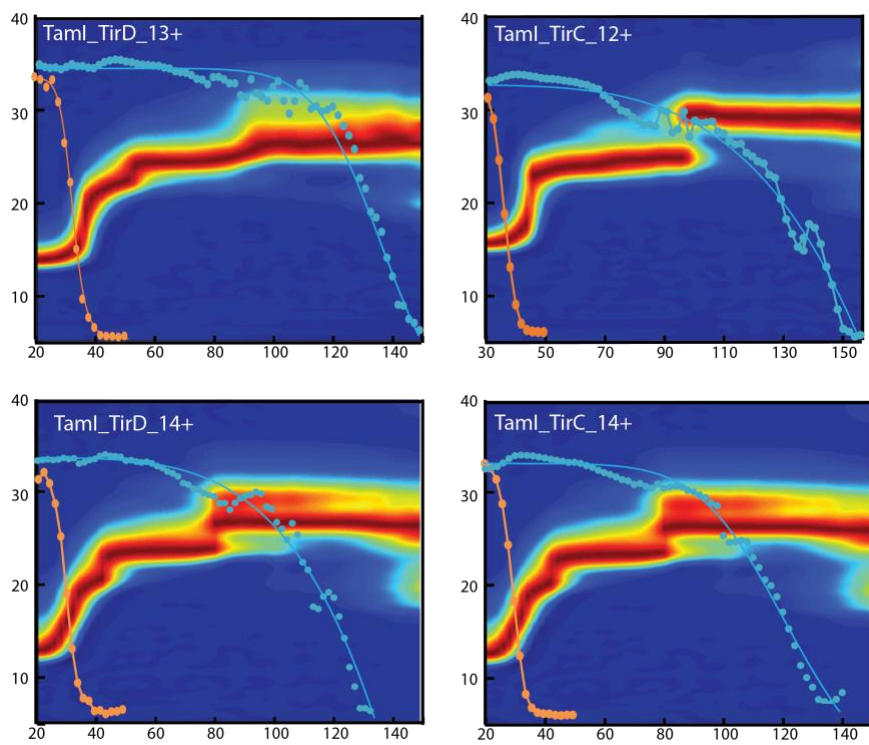
Figure II-2. Correlation of TirC and heme dissociation with TamI unfolding at charge states of 12+ and 14+. Correlation of TirD and heme dissociation with TamI unfolding at charge states of 13+ and 14+.
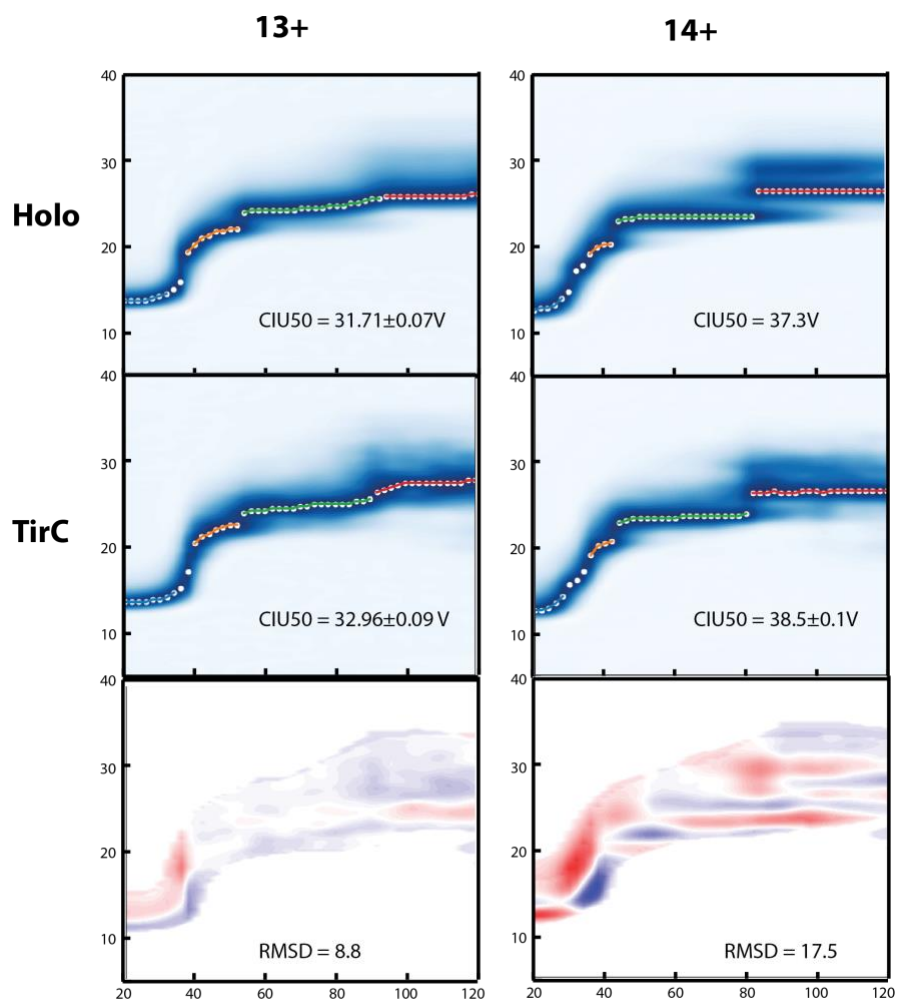
Figure II-3. TirC binding stabilization of TamI unfolding at charge states of 13+ and 14+
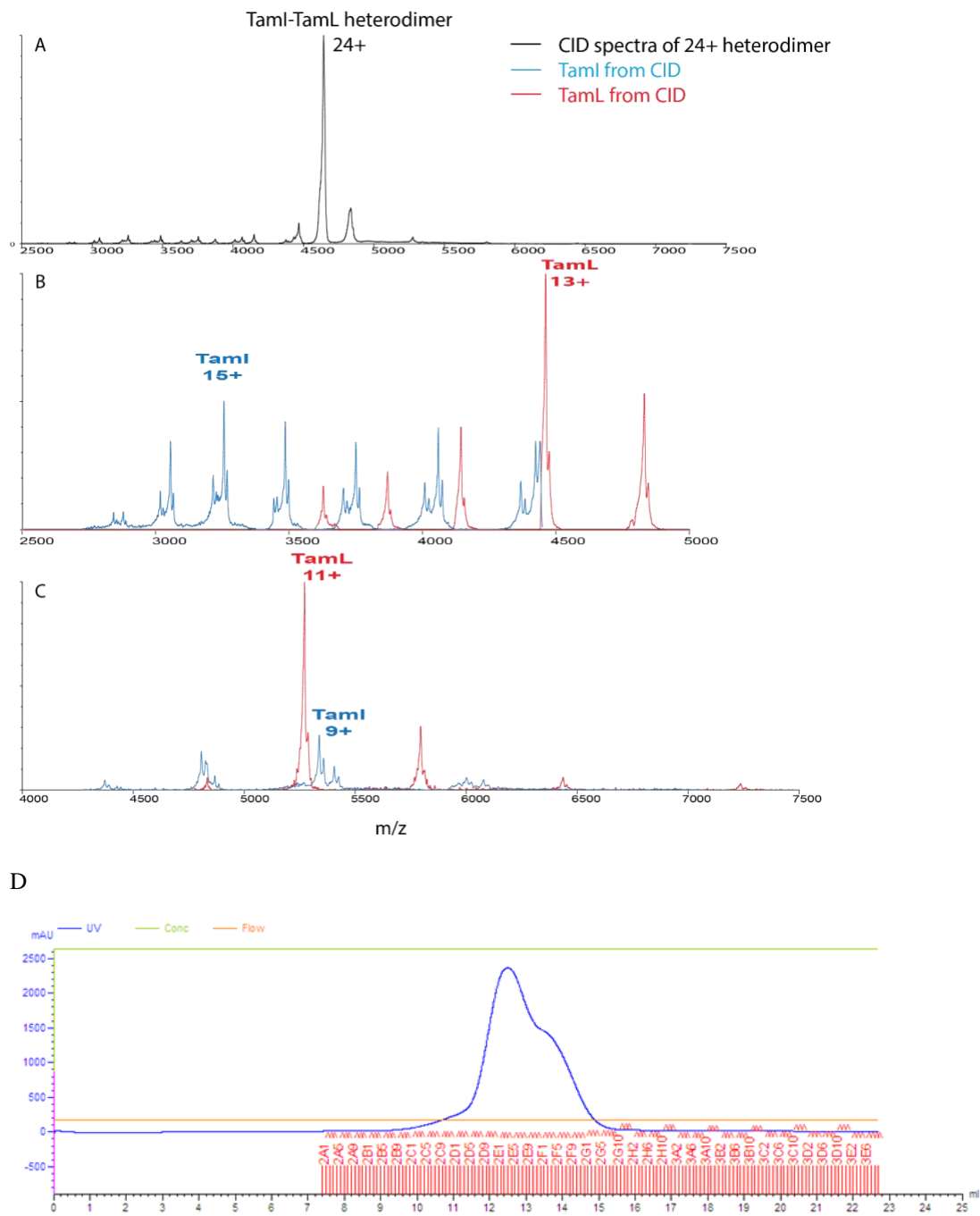
Figure II-4. Panel A, B, and C present the Collision Induced Dissociation (CID) of coexpressed TamI:TamL to confirm the evidence of heterodimer. Panel D, represent the chromatograms of co-expressed TamI:TamL when purified through size exclusion chromatographic techniques.

Table II-1. CCS measurement by Ion Mobility Mass Spectrometry compared with reported CryoEM. TamL dimer has two conformations: elongated dimer is more consistent with CryoEM, the compact dimer might be the artifact of gas phase compaction.

| Protein	z	Average CCS	Least Square Analysis CCS Error (+/-)	PA*1.15	TJM	AverageCCS (all charge states)	error%
TamI	13	3177.9	55.9				
	14	3182.2	39.0	2950	3230	3220.9	-0.28%
	15	3302.7	42.4				
TamL_Dimer compact	21	5603.8	79.7				
	22	5493.5	72.4	5467	6183	5507.7	-10.92%
	23	5425.9	75.2				
TamL_Dimer elongated	23	5983.7	76.5				
	24	6194.8	74.2				
	25	6188.0	73.6	5467	6183	6144.1	-0.63%
	26	6209.9	75.6				
TamL_Monomer	14	3592.3	51.2				
	15	3543.1	73.1	3383	3733	3586.9	-3.91%
	16	3572.4	64.0				
	17	3639.8	66.2				
Fusion-Heterodimer	18	5489.5	71.8				
	19	5447.9	77.7			5468.7	
Coexpressed-Heterodimer	23	5686.9	92.1	5357	5811	5741.9	-1.19%
	24	5796.9	128.2	5332	5814	5741.9	-1.23%
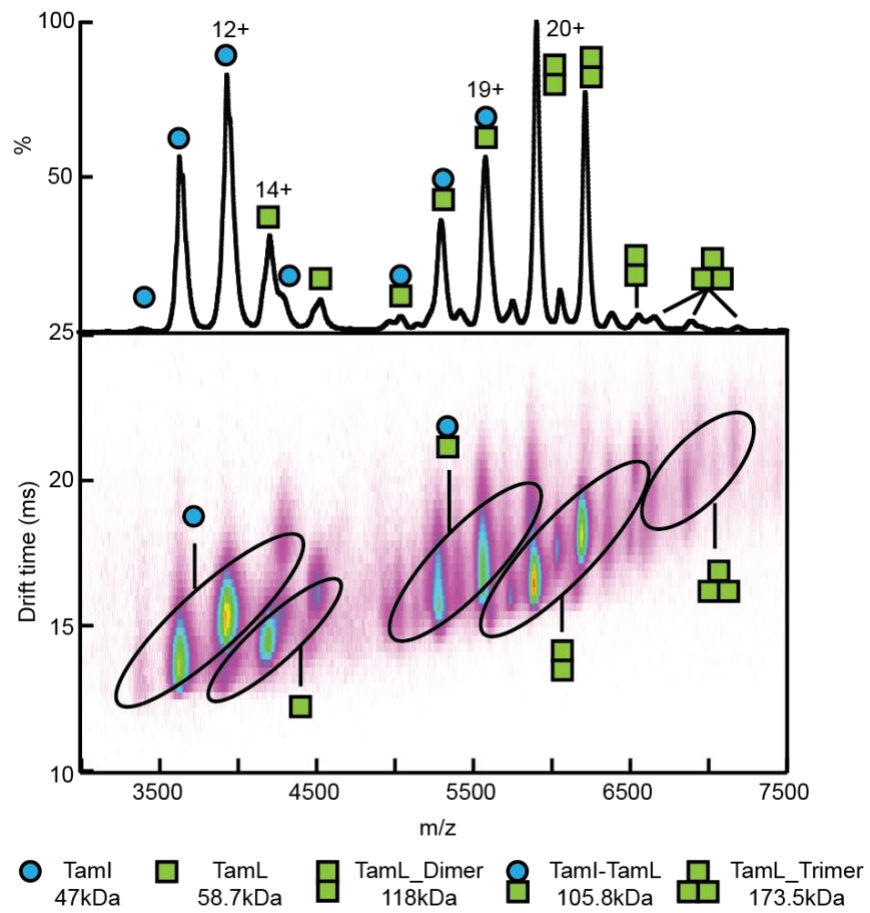
Figure II-5. Mass and Ion mobility spectra of TamI-TamL crosslinking sample.

Table II-2. FWHM data observed for heterodimer generated by in vivo co-expression, cross-linking, and the fusion constructs.

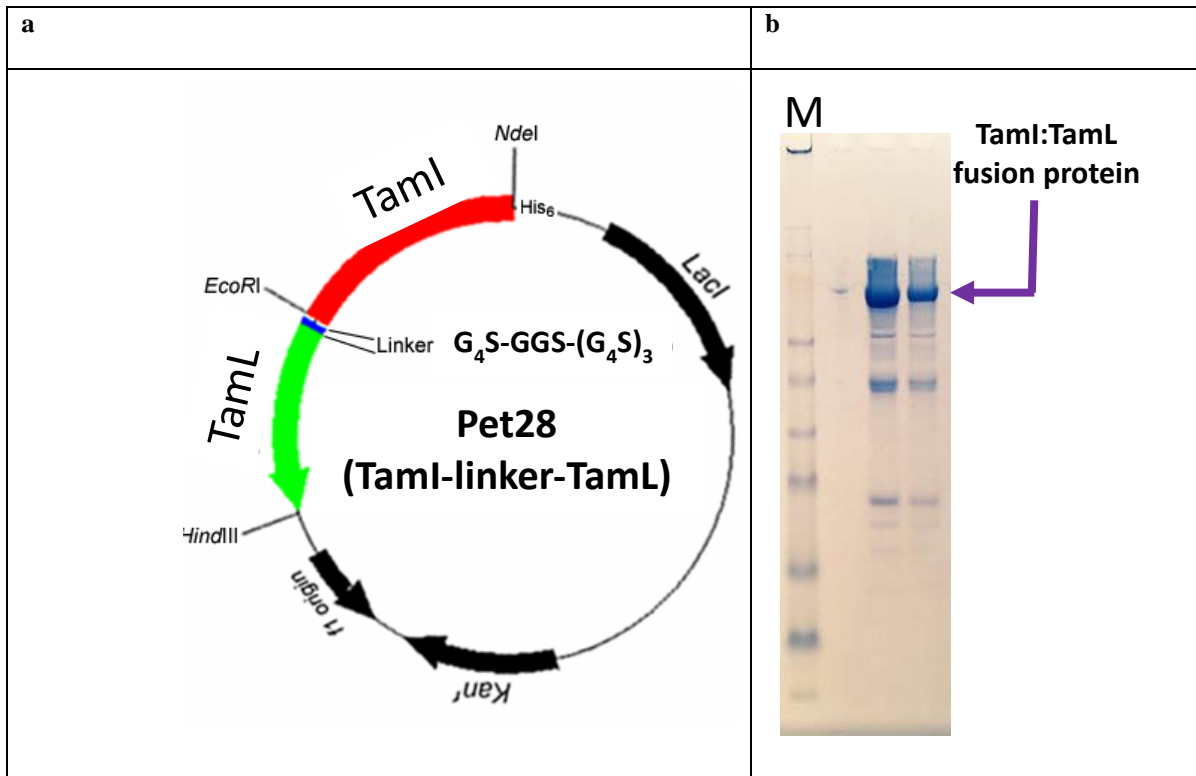|	fusion				cross-linked			co-expressed	
	18+	19+	20+		19+	20+		23+	24+
FWHM	1.82	1.57	1.34		2.45	2.29		2.98	3.74
x	18.49	16.96	15.78		16.79	15.94		13.88	13.81
FWHM/x	0.10	0.09	0.08		0.15	0.14		0.21	0.27
average	0.09				0.14			0.24	

Figure II-6. TamI:TamL fusion construct design and purification. a) plasmid map, b) SDS page gel electrophoresis image where Lane 1 from left is marker lane and lane 2, 3, 4 display TamI:TamL fusion protein purified and loaded at varied concentration.
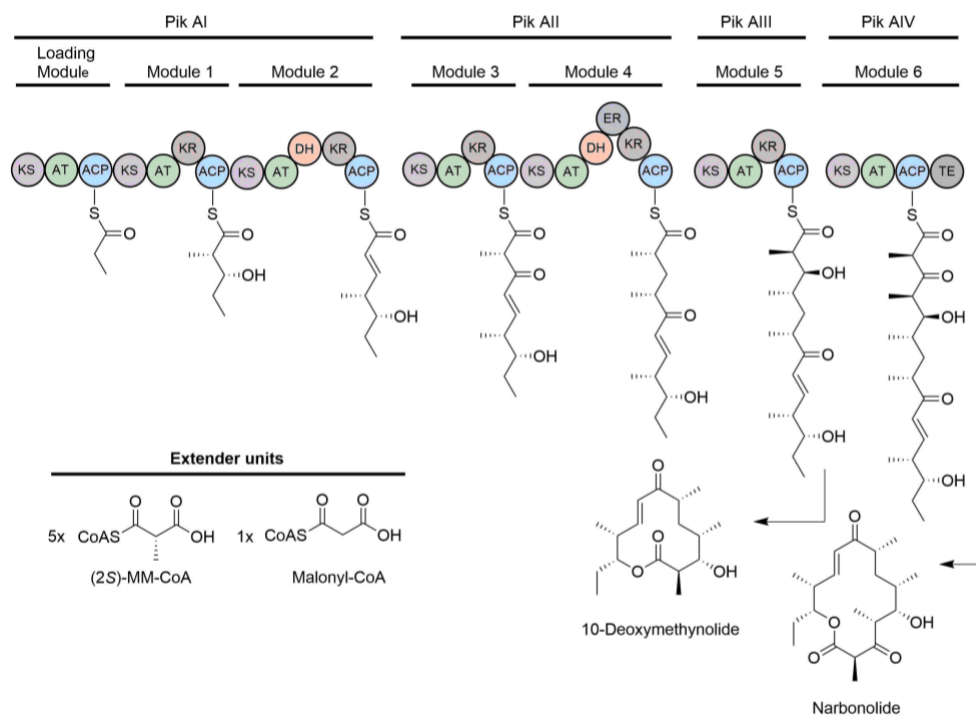
Figure III-1. Modular PKS systems responsible for production of the pikromycin.

Appendix IV-Special Project

Integrating Native Mass Spectrometry with Cryo-Electron Microscopy for High-throughput Structural Biology

Introduction:

Cryo-Electron Microscopy (CryoEM) has emerged as a revolutionary approach to obtain high-resolution protein structures. Despite its advantage, there are many challenges in discovering proper imaging conditions for cryoEM samples, limiting the application of this technology and slowing the protein structure discovery pipelines. One major problem in cryoEM is the optimization of sample conditions needed to keep target proteins stable in order to obtain high resolution images, which normally requires a significant amount of time. Meanwhile, native mass spectrometry (MS) has been developed in order to enable the rapid measurement of protein complex stoichiometry and ligand binding states using small amounts of sample. More importantly, native MS is capable of measuring samples that are mixtures, which make it an ideal companion technology for cryoEM. In addition, recent developments in collision induced unfolding (CIU) technology enables MS to detect subtle stability shifts within protein complexes prepared under different solution conditions. In this preliminary study, we seek to develop a native MS screening paradigm for cryoEM sample preparation, by building on key features detected using the former tool that are able to predict successful high-resolution structure determinations by the latter technology.

124

Prior studies have shown that protein complex stabilization can be achieved in the gas-phase through the addition of Hofmeister anions and cations in solution prior to ESI, where the salts bind as adducts to the surface of the protein and act as a protective "shell" of counterions capable to carrying away excess energy from the protein complex ion upon activation, resulting in significant structural stabilization.[92] Recent studies using IM-MS served to classify the influence of counterions on gas-phase protein structure as structure stabilizers in the absence of bulk solvent.[76] Specifically, anions that bind the protein in high affinity and are released upon dissociation are able to lead to high protein structural stability in the gas-phase; whereas anions that exhibit high-affinity binding but do not dissociate from the protein in the gas phase and low-affinity anions lead to relatively low protein structural stability.

In this study, we established a workflow combining cryoEM and native MS, where MS data was collected before cryoEM experiments, allowing us to screen different buffer condition in order to find those potentially predictive to produce the best cryoEM outputs. Specifically, we use a combination of QE UHMR /orbitrap and IM-MS data to screen sample prepared using different buffer conditions, and search for correlations between either cryoEM or negative stain EM images produced using samples prepared in the same buffers. We collected CIU50 and CID50 values, as well as MS peak widths, which allow us to characterize the stability and anion binding associated with a group of model protein complexes in a range of buffers. Overall, we find that the anions that are classified as the strongest stabilizers using MS approaches are consistent those that produce the highest quality negative stain images.

Experimental Section

Sample Preparation. Alcohol dehydrogenase tetramer (ADH, yeast), beta galactosidase (β-gal), and aldolase were purchased from Sigma. All protein samples were buffer exchanged into 100mM ammonium acetate at pH 7 using Micro Bio-Spin 6 columns (Bio-Rad, Hercules, CA) and prepared to a final concentration of 2 μM (β-gal), 5 μM (ADH, Aldolase). To study the influence of different salts on protein stability by measuring the peak width for ADH and β-gal, and the CID50 for aldolase using the Q Exactive™ UHMR Hybrid Quadrupole-Orbitrap™ Mass Spectrometer, the tested protein samples contained added salt concentration of 100mM. Samples contained added salt concentration of 10mM was used to study the stability difference, by measuring the CIU50 values for ADH and β-gal using IM-MS on a quadrupole-ion mobility-time-of-flight mass spectrometer (Q-IM-ToF MS) instrument (Synapt G2 HDMS, Waters, Milford, MA).

Data Analysis

In terms of peak width analysis for spectrum generated from the UHMR, we smoothed all adduct populations into a single peak (Figure V-1), and recorded its width (FWHM) for two charge states for ADH and β-gal. For CID50 calculation, we recorded the relative abundance of the remaining tetrameric ions ($I_{tet}(\%)$), calculated as a percentage of the total intensity of the intact protein complex and its corresponding fragment ions (i.e., monomer) observed in the mass spectra at different activation energy.

Results & Discussion

In order to achieve higher resolution mass spectra, we explored the upper limit of Hofmeister salt concentration that QE-UHMR could tolerate by using small emitter tips.[93] As shown in Figure IV-1, we were able to resolve the protein-adduct ion peaks from the apo protein ion peaks.
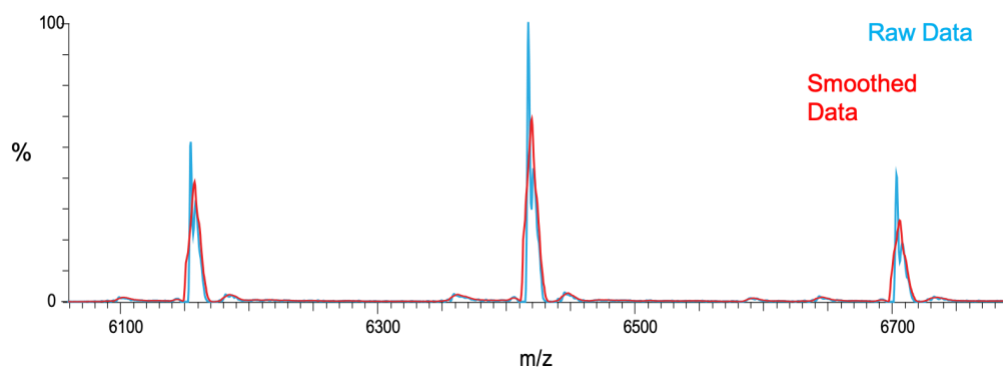
Figure IV-1. Mass Spectra of ADH with 100mM HCO3.

In order to find the most effective stabilizer for ADH tetramers among the different Hofmeister salts, we incubated ADH with different Hofmeister salts. FWHM (m/z) was measured for 25+ and 24+ charge states at 10V HCD energy. Measurement taken for both charge states showed similar results, revealing that SCN, NO_3, Cl are high-affinity anions while F, HCO_3, acetate (control) are weak/non binders. Furthermore, we also quantified the number of the anion adducts lost from complexes when observed between 10 V (low activation) and 150 V (high activation) conditions by tuning the HCD activation energy. The results showed that SCN, NO_3, Cl are relatively easily dissociated from the precursor protein ions upon activation. Therefore, we are able classify SCN, NO_3, Cl as effective stabilizers for ADH. In the meantime, we also quantified the remaining bound SO_4 adducts at 150V (high activation condition). We find significant amount of SO_4 adducts attached to ADH at high activation energies, indicating that SO_4 is a strong binder but not an efficient stabilizer.
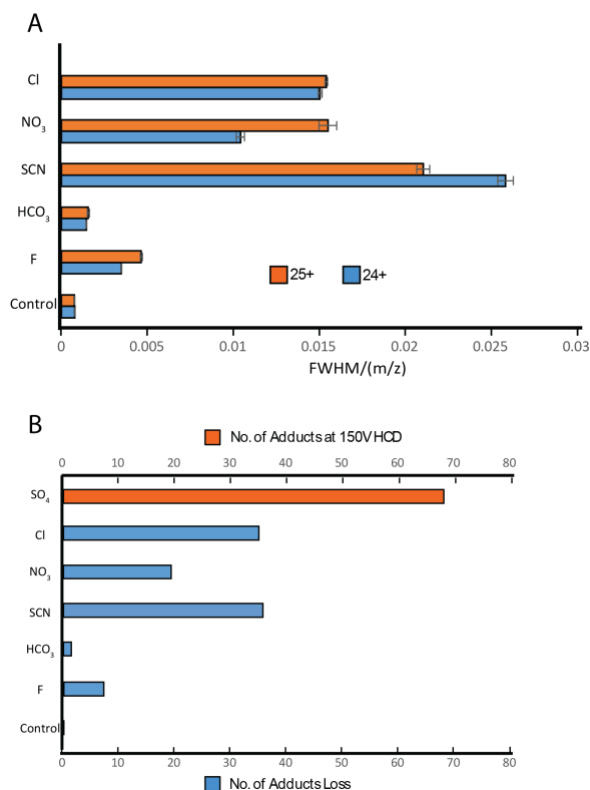
Figure IV-2. (A) Histogram plots charting the FWHM values measured for ADH at 47+ and 48+ at 10V HCD energy and (B) the number of adducts loss between low (10V) and high (150V) activation conditions with added 10mM ammonium salts, compared with the ammonium acetate buffer as control.

Following the analysis of the MS peak widths discussed above, we acquired CIU50 values for ADH bound to different anion populations as stabilizers. As shown in the Figure IV-3A and IV-4B, there are 6 CIU features observed. By comparing the CIU50 values, all five CIU50 values suggest increased stability for ADH bound to NO_3^-. We then used histograms (Figure IV-3C) and Table IV-1 to compared five CIU50 values obtained for ADH in the presence of different stabilizers. CIU50-2, CIU50-3, CIU50-4, and CIU50-5 all showed that ADH bound to SCN, NO_3, Cl are more stable than the those prepared in the presence of F, HCO_3, or acetate (control), leading to a conclusion which is consistent with our MS FWHM dataset.
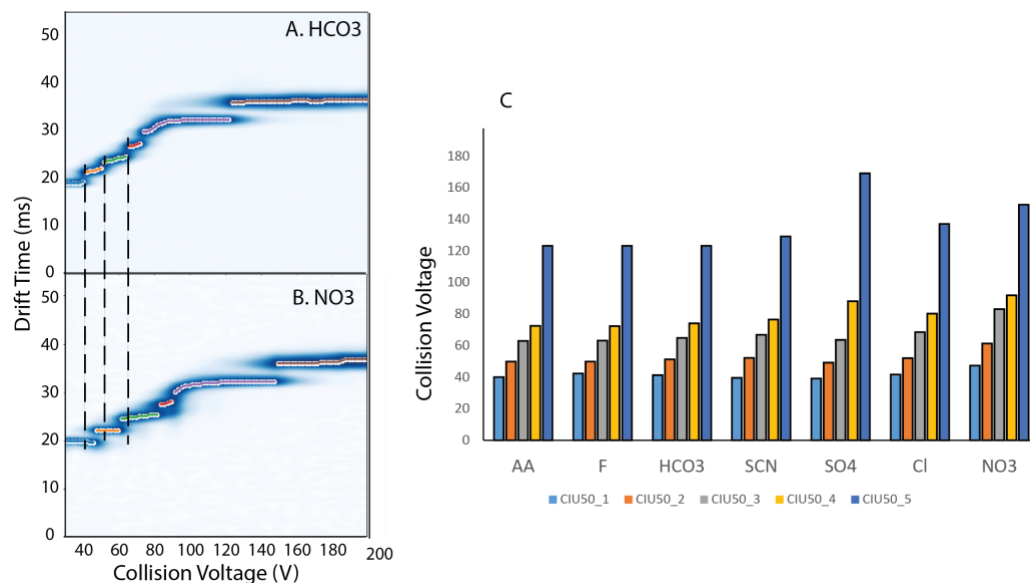
Figure IV-3. (A) CIU fingerprints of ADH with 10mM HCO_3^- salt buffer. (B) CIU fingerprints of ADH with NO_3^-.

(C) Histogram plots charting CIU50 values of ADH protein ion with ammonium salts added in the buffer.

Table IV-1. CIU50 values for ADH with different stabilizers.

	1	2	3	4	5
AA	39.87	49.79	62.82	72.33	123.07
F	42.16	49.8	63	72.23	123.07
HCO3	41.17	51.19	64.74	74.05	123.09
SCN	39.43	52.02	66.67	76.42	128.99
SO4	39.07	49.13	63.62	87.99	169.09
Cl	41.49	52.01	68.49	80.19	136.93
NO3	47.14	61.23	82.98	91.76	149.07

Similar data was as described above for ADH was collected for β-gal. MS peak width data across

two charge states was collected and the quantified adduct populations lost during activation are

shown in Figure IV-4. MS FWHM (m/z) was also measured for the 47+ and 48+ charge states at

100V HCD energy. Measurements taken under both charge states showed similar results, with higher MS FWHM values revealing that SCN, NO_3, Cl, Tartrate are high-affinity anions while lower MS FWHM values indicating that F, HCO_3, acetate (control) are weak/non binders. In addition, the number of the anion adducts lost from complexes when observed between 100 V (low activation) and 150 V (high activation) conditions by tuning the HCD activation energy were quantified. The results indicate that Tartrate, NO_3, Cl, SCN are relatively easily dissociated from the precursor protein ions upon activation. Therefore, we are able to classify tartrate, NO_3, Cl, SCN are effective stabilizers for β-gal.
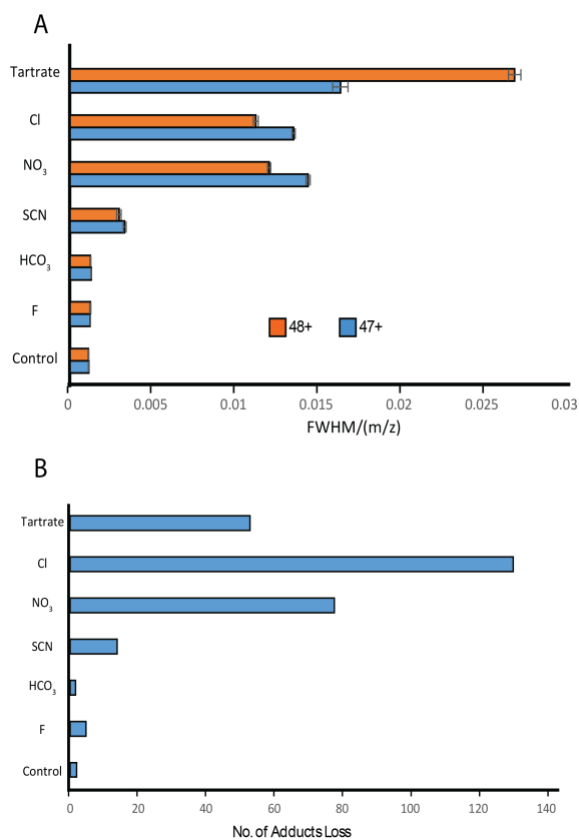

Figure IV-4. (A) Histogram plots charting the FWHM values measured for β-gal at 47+ and 48+ at 100V HCD energy and (B) the number of adducts loss between low (100V) and high (150V) activation conditions with added 100mM ammonium salts, compared with the ammonium acetate buffer as control.

Furthermore, CIU experiments were performed on β-gal 44+ ions prepared with different ammonium salts as stabilizers. As shown in Figure IV-5A, we observe 4 CIU features. We then used a histogram plot (Figure IV-5B) and the data shown in Table IV-2 to compare the data acquired between the five CIU50 values of β-gal bound to different stabilizers. Both CIU50-1 and CIU50-2 showed that β-gal ions attached to SCN, NO_3, Cl are more stable than the ones with F, HCO_3, acetate (control), allowing us to reach a conclusion which is consistent with the MS FWHM dataset.
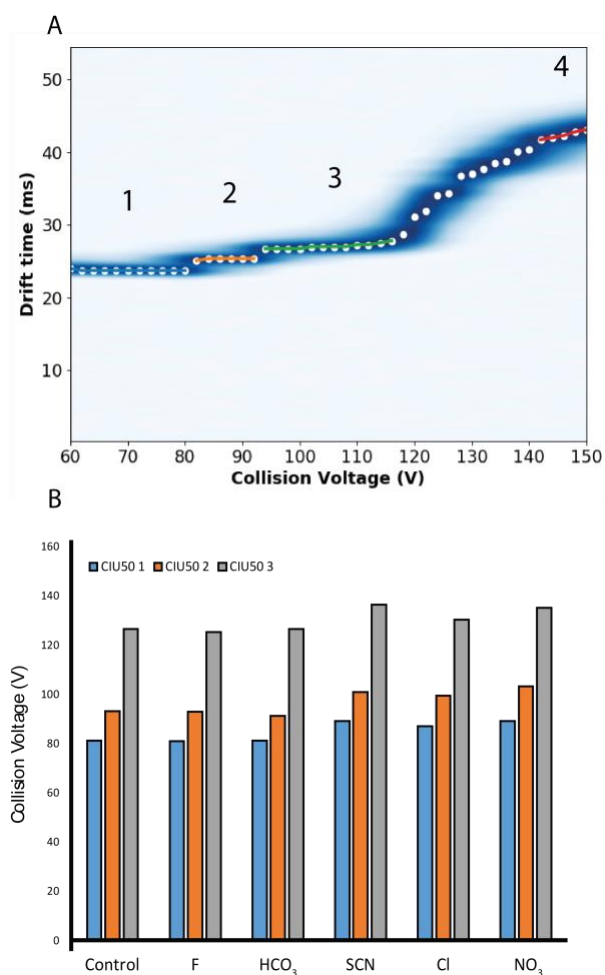
Figure IV-5. (A) CIU fingerprints of β-gal at 44+ with 10mM HCO_3^- salt buffer. (B) Histogram plots charting CIU50 values of β-gal protein ion with ammonium salts added in the buffer.

Table IV-2. CIU50 values for β-gal with different stabilizers.

	CIU50 1	CIU50 2
AA	81.25	93.17
F	81.06	93.01
HCO3	81.07	91.2
SCN	89.03	101
Cl	87.03	99.35
NO3	89.03	103.28

In addition to ADH and β-gal, we studied aldolase tetramers using an HCD approach by measuring CID50 values for samples prepared using different buffer conditions. Activation energy was ramped from 140V till 300V, where the monomer of precursor aldolase was ejected with the trimer remaining as a stripped complex. CID curves for aldolase was generated by plotting the remaining amount of the intact aldolase over the starting quantity, shown in Figure IV-6. Expectedly, CID50 values of β-gal with NO_3, Cl, SCN are much higher than that with F, HCO_3, acetate (control), suggesting that NO_3, Cl, SCN are more effective stabilizers for aldolase with respect to CID.
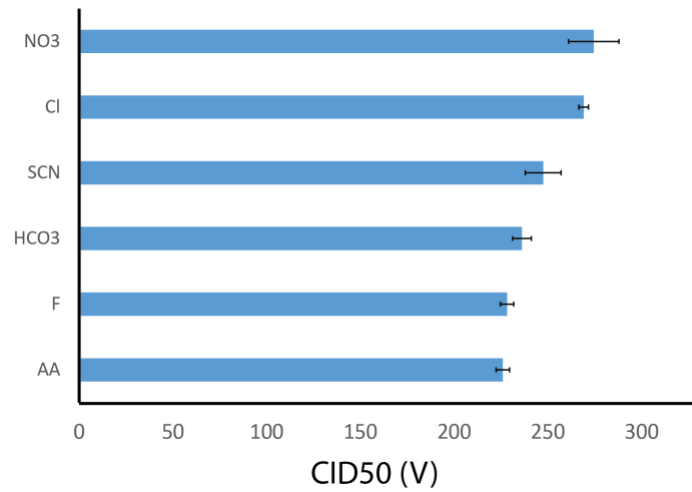
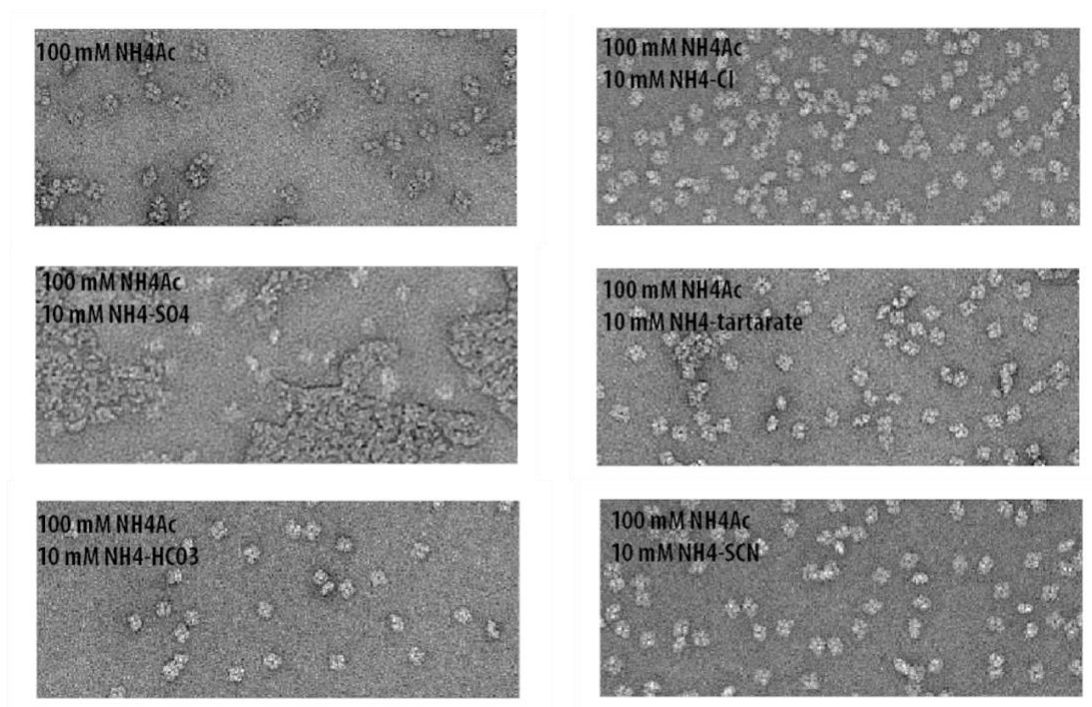**Figure IV-6.** CID50 was determined for Aldolase using HCD approach by UHMR.

Figure IV-7. Negative stain EM data for β-gal with different Hofmeister salts.

Conclusions

The QE UHMR platform allows us to collect high resolving power native MS data, achieving the quantification of individual anion binding events on large (~150kDa) protein complexes. Furthermore, this high-resolution technique combined with the utilization of smaller emitter needles has enable us to record native MS data under high salt (100mM) conditions for the first time.

Our MS FWHM results generated from QE UHMR, along with CIU50 values measured by IM-MS, and the CID50 values quantified from HCD activation all lead to the same conclusion. Anion additives can be sorted into three classes—strong binders and efficient stabilizers, weak/non binders, strong binders but not efficient stabilizers. The strong binders and efficient stabilizers are shown to have high-affinity and easy to dissociate from the precursor protein ions upon activation, therefore have the most potential to stabilize protein complexes. Following the native MS analysis, we compared our findings with negative stain EM images (Figure IV-7). As we expected, efficient stabilizers determined using native MS including SCN, NO_3, Cl are able to help generate better-quality EM images compared to the protein sample with non-binders (F, HCO_3, acetate). Altogether, in this appendix, we present that the combination of IM-MS with high-resolution native MS allows us to screen samples prior to EM analysis, which serves to rapidly reduce the time required for high-quality EM image production.