Next-Generation Computational Methods for the Interpretation of Ion Mobility-Mass Spectrometry Data

by

Chae Kyung Jeon

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2023

Doctoral Committee:

Professor Brandon T. Ruotolo, Chair
Professor Charles L. Brooks
Professor Michael Cianfrocco
Professor Kristina Håkansson

Chae Kyung Jeon

jeonc@umich.edu

ORCID iD: 0009-0008-5292-0198

© Chae Kyung Jeon 2023

Dedication

This thesis is dedicated to all the teachers and mentors I have had during my life.

Acknowledgements

During my journey to getting a PhD, I've come across many wonderful people who all made positive remarks. First, I would like to thank my advisor, Prof. Brandon Ruotolo for his endless support during my PhD. You have mentored and supported me in so many areas which I'm grateful for, and thank you for always being so patient with me. I would like to also thank my committee members Dr. Kristina "Kicki" Håkansson, Dr. Michael Cianfrocco, Dr. Aaron Frank, and Dr. Charles "Charlie" Brooks for all their advice and feedback on my degree progress.

My PhD would not be complete without a very supportive group of scientists in the Ruotolo lab. First, I want to give a huge thank you to Sugyan "Suggie" Dixit for all your knowledge, mentoring, and patience. Without you, I would have not become a modeling and computational expert in the lab. I want to thank all the past members – Varun Gadkari, Daniel Polasky, Sarah Fantin, Chunyi Zhao, Cara D'Amico, Daniel Vallejo, Carolina Rojas Ramirez, Kristine Parson, Yilin Han, and Brock Juliano – for all the thoughtful conversations we had in the lab and creating a safe, fun, and positive working environment. Also, thank you for always hopping in on a Zoom meeting to answer all my questions. Next, I'd like to thank the current members of the lab – Kiani Jeacock, Rosendo "Rosie" Villafuerte-Vega, Iliana Levesque, Devin Makey, Anna Anders, Nicole Rivera Fuentes, Joseph Marte, Kate Hanggi, Addison Bergman, Henry Li, and Ryan Schroeder – for always having my back, supporting me, looking after to make sure I'm not too drunk, and lots of hugs. But most importantly, thank you for always being on board collecting or sharing data for

all my software tantrums. Without all your help and sacrifices, CIUSuite 3 would not have been possible.

In addition to my lab mates, I have to thank my collaborators from Agilent. I had the privilege to work with Bryan Miller and Ruwan Kurulugama on CIUSuite 3 and appreciated all your feedback and ideas on all the cool features we could potentially add to the new CIUSuite. Also huge thanks to Ruwan for being a champ at finding every bug to fix!

A special thanks goes to my undergraduate research advisor, Prof. Elaine Marzluff from Grinnell College. She introduced me to the world of mass spectrometry, and she mentored and challenged me as a scientist. It was a huge pleasure working on a HDX project with her, and thanks to this wonderful opportunity, I knew I wanted to pursue PhD in chemistry, specifically in mass spectrometry.

My friends – old and new – have been an important part of my life especially during the 5 years I've spent in Ann Arbor. Huge thanks to all my friends from Grinnell – Carolyn, Katie, Stella, Grisel, Karin, and Jinna – for sharing all your fun life updates abroad and heartwarming postcards to cheer me up. Philipp Gemmel, my chemistry friend from Grinnell, I'm so glad to have you in Ann Arbor (even though we only overlapped for 6 months). You were so supportive during the stressful time of writing my dissertation and always ready to go out to grab coffee and beer to catch up. I've made many friends from different communities during my time in Ann Arbor. In particular, I want to thank everyone from aUM and the knitting community in Ann Arbor and Seoul for teaching me how to enjoy my time outside of work, giving me a break, and letting me take a step back to relax during stressful times. Lastly, I want to thank all the wonderful community of people that I've connected with during the ASMS conferences.

Finally, I want to thank my family. My mom and dad have made great sacrifices for me throughout their lives. The reason I could pursue higher education was all thanks to your dedication. Thank you for always being supportive of all the life decisions I've made from choosing college to pursuing PhD. Because of your endless support and love, I knew I could succeed in my career even when you are on the other side of the hemisphere. I also want to thank my brother, Paul, for lots of phone calls to keep me companied, visiting me on every possible holiday and especially my birthday, cooking amazing food, and your love. I'm so fortunate to have such an endearing brother, and it is very exciting to hear your joy in working with mass spectrometry. I also want to thank all my family back in Korea – especially my grandma and aunts – for always welcoming me whenever I visit and making me laugh all the time. Lastly, thank you to Camille, for letting me go out on adventures and supporting me emotionally.

Table of Contents

Dedication ii

Acknowledgements iii

List of Tables ix

List of Figures xi

List of Appendices xvii

Abstract xviii

Chapter 1 Introduction 1

1.1 Methods of Protein Structure Determination 1

1.1.1 Experimental Methods for Protein Structure Determination 1

1.1.2 Computation Methods for Protein Structure Determination 4

1.2 Mass Spectrometry in Structural Biology 6

1.2.1 Native Ion Mobility-Mass Spectrometry 7

1.2.2 Ion Mobility-Mass Spectrometry 8

1.2.3 Collision Induced Unfolding 10

1.2.4 Electron Capture Dissociation 11

1.3 The Current State of Integrative Modeling for Native IM-MS 12

1.3.1 Building Three-Dimensional Models for Native IM-MS 12

1.3.2 Charge Assignment for Gas-Phase Protein Ions 14

1.3.3 Molecular Dynamics Simulations for Gas Phase Proteins 15

1.3.4 Methods to Calculate Theoretical Collision Cross Sections 16

1.4 Summary 17

1.5 References 18

Chapter 2 The Gas-Phase Unfolding of Protein Complexes Follows a Largely Domain-Correlated Mechanism 26

2.1 Introduction 26

2.2 Methods 29

2.2.1 Sample Preparation 29

2.2.2 IM-MS Data Collection and Analysis 29

2.2.3 Chemical Probe CID Analysis 30

2.2.4 Sample Preparation & IM-MS Measurements 30

2.2.5 CIU-ECD Experiments 31

2.2.6 CIU-ECD Data Analysis 32

2.2.7 Molecular Dynamics Simulations 34

2.3 Results and Discussion 36

2.4 Conclusions 50

2.5 Acknowledgments 52

2.6 References 52

Chapter 3 Integrative Modeling of Individual Particle Electron Tomography Antibody Models Reveals Gas-Phase Structural Diversity and Memory Effects 56

3.1 Introduction 56

3.2 Methods 58

3.2.1 Sample Preparation and IM-MS Experiment 58

3.2.2 Ion Mobility-Mass Spectrometry 59

3.2.3 Computational Modeling 59

3.3 Results and Discussion 60

Chapter 4 CIUSuite 3: Next-Generation CCS Calibration and Automated Data Analysis Tool for Gas-Phase Protein Unfolding 73

4.1 Introduction 73
4.2 Methods 76
4.2.1 Experimental Section 76
4.2.2 Raw Data Extraction 76
4.2.3 CIUSuite 3 Overview 77
4.2.4 Peak Detection Algorithm 78
4.2.5 Feature Skipping 78
4.2.6 Multi-dimensional Classification 79
4.2.7 Built-in CCS calibration 80
4.3 Results and Discussion 82
4.4 Conclusions 93
4.5 References 93
Chapter 5 Conclusions and Future Directions 98
5.1 Integrative Modeling of Proteins and Protein Complexes 98
5.2 Data Analysis Tools for Native MS and Gas-phase Protein Unfolding 100
5.3 References 101
Appendices 103

List of Tables

Table 1-1. Overview of software for calculating collision cross sections of gas-phase ions..... 17

Appendix Table I-1. Internal fragment matches from CIU-ECD experiments of HSA $D12^{13+}$. 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 111

Appendix Table I-2. Internal fragment matches from CIU-ECD experiments of HSA^{17+}. 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 111

Appendix Table I-3. The number of models within each unfolded feature from steered MD of HSA in domain 3. Three different steered MD simulations were performed within the domain 3 region of HSA to validate the specificity and efficiency of unfolding during the ECD-guided steered MD simulation. The residues that were allowed to unfold during each steered MD are also noted in the table 111

Appendix Table II-1. CCS of all 120 PDB Models in Solution and Gas-phase..... 114

Appendix Table II-2. Clustering Results from IM-filtered Models..... 116

Appendix Table II-3. Changes in Secondary Structures after Gas-phase Compaction 117

Appendix Table III-1. Peak detection results of HSA. The table summarizes the peaks identified in Appendix Figure III-1 along with the low and high m/z bounds which can be used for extracting drift time information during the CIU extraction process..... 123

Appendix Table III-2. Probabilities of class assignments with and without multi-dimensional classification. Based on the classification scheme built using BSA, HSA, RSA, and GSA, probabilities of class assignment were calculated with and without multi-dimensional classification. 123

Appendix Table III-3. Single-field CCS calibration function values using BSA as the calibrant. Linear regressions were performed using the OLS method. The coefficient and t_{fix} (intercept) were the direct output of the OLS method. Beta value is determined by factoring in charge and masses of buffer gas and BSA based on Equation 4-6. 123

Appendix Table III-4. Summary of standard proteins CCS values calibrated using both tunemix and BSA. All protein samples were prepared and ionized to preserve their native structure..... 124

Appendix Table IV-1. Domain-constrained Structural Flips on Aβ42 131

List of Figures

Figure 1-1. Atomic resolution structures of proteins and protein complexes resolved by X-ray crystallography, NMR, and cryo-EM. 3

Figure 1-2. Instrument scheme for native IM-MS. Proteins are introduced into the gas phase by nano ESI. Protein ions are then separated by IM and analyzed by time-of-flight mass analyzer. The drift time and m/z obtained can be plotted in two-dimension to identify protein ions' stoichiometry but also their conformation states (oligomer 2A) 9

Figure 1-3. Collision Induced Unfolding (CIU) of protein ions. (A) Collision voltages are increased to unfold proteins which are monitored using ion mobility drift time. (B) CIU represented as a two-dimensional contour plot with collision voltage on x-axis and drift time on y-axis. Regions where protein conformations remain stable over a range of collision voltages are called features as indicated by a white box as an example 10

Figure 1-4. Fragmentation by electron capture dissociation (ECD). Electrons are captured by the protein ions in the gas-phase which generates radical cations. The radical induces rearrangements causing backbone fragmentation. The most common product ions are c-ions (N-terminal fragments) and z-ions (C-terminal fragments). The less common a-ions are also observed..... 11

Figure 1-5. Nano ESI process places charges on the protein surface. Once the droplets are formed, the size of the droplets decreases as they desolvate causing the charges to be placed on the surface of the protein..... 12

Figure 1-6. Models generated using all-atom and coarse-graining**.** MD simulations can be performed using either (A) all-atom or (B) coarse-graining which represents individual subunits or domains as spheroids. This figure illustrates how models differ for protein complexes between all-atom and coarse-grained where each subunit is represented in different colors and each spheroid represent the domain within the subunit. 13

Figure 2-1. 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. 38

Figure 2-2. 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 40

Figure 2-3. 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 numbered gray boxes in (E) and (F) indicates the CCS ranges. 41

Figure 2-4. Computational Methods for Improved Modeling Accuracy. (A) Charge placement results of HSA D12 monomer at +12 net charge. Red indicates negative charge site, and blue indicates positive charge site. (B) CCS correlation plot between IMPACT and Collidoscope. The best fit line represents TM CCS =0.9591*(IMPACT CCS) + 264.47 with $R^2 = 0.9681$. (C) Example steered MD demonstrates unfolding of protein at specific residues. The input model (right) subjected to heating via SMD with residues in red fixed. The final model successfully kept red residues intact and only the green residues changed in structure. 44

Figure 2-5. CIU-ECD of full-length HSA. (A) CIU Fingerprint of +17 full length HSA. The five colored boxes indicate where in the fingerprint ECD fragmentation was obtained for each of the five features observed. The +17-charge state is shown due to presenting higher fragmentation efficiency and a +5 charge state fragments, unlike +15 HSA fragmentation. For the +15 HSA fingerprint and corresponding spectra see Figure S7. (B) ECD Fragmentation spectrum taken at each voltage selected in A). Three internal peaks are indicated in each spectrum corresponding to domain 1 – domain 2 boundary (D1-D2, red triangle), domain 2 - domain 3 boundary (D2-D3, green square), and domain 3 (blue circle). These fragments were then subsequently used to guide SMD of full-length HSA due to their high matching scores (Table S2), and intensity profiles. (C) Normalized intensity for the three internal fragments indicated in B) at each collision voltage used in B). the Legend indicates the ions types that correspond to each internal fragment as well as the charge state. (D) Full-length HSA models of compact and ECD-guided unfolded structures. Residues of ECD fragments from C are highlighted accordingly in both models. (E) CCS distribution of SMD trajectory using all three internal fragments observed from C. With SMD, HSA undergoes unfolding up to the CCS range (blue box) that agrees with domain 2 and 3 unfolded CCS range 47

Figure 3-1. Solution-phase IgG1 models and CCS distributions from solution-phase models and IM-MS. (A) Overlay of all 120 IPET reconstructions. Blue and Orange represent heavy chains, and green and magenta represent light chains. (B) CCS distribution of 120 IPET models. (C) The smallest, (D) median, and (E) largest models from the IPET models. (F) CCS distribution from native IM-MS experiments with TWIM in blue and DTIM in Green. (G) Overall CCS comparison between IPET models and two IM-MS experiments show significant gas-phase compaction in IgG1 62

Figure 3-2. MD Simulations of Gas-phase Compaction. (A) CCS distribution change from solution to gas-phase using final structures from 120 MD simulations. The green, magenta, and orange bars represent the CCS value at which models from Figure 3-1 C-E are found in each distribution. (B) Overlay of all 120 final structures from MD simulations. (C) The smallest, (D) median, and (E) largest models from final structures from 120 MD simulations. (F) CCS distribution of IgG1 models that are filtered based on IM restraints. Red distribution is filtered based on DTIM CCS centroid and CCS errors. Blue distribution is filtered using the start and the end of the baseline of TWIM and DTIM CCS 64

Figure 3-3. RMSD-based Clustering Analysis. Each dendrogram represents clustering results from (A) 120 IPET models (B) final structures from 120 MD simulations, and (C) IM-filtered gas-phase models 66

Figure 3-4. Detailed Overview of Changes in Structures upon Gas-phase Compaction. (A) The smallest, (B) median, and (C) largest IM-filtered gas-phase models of IgG1 (bottom) and the original solution-phase structures (top). (D) Changes in secondary structures from solution to gas-phase 67

Figure 4-1. CIUSuite 3 overview and workflow. To assess peaks for CIU extraction, raw data in either .RAW or .d file is uploaded. CIUSuite 3 will extract drift time information based on either peak detection results or user input m/z ranges. The users can also directly perform CCS calibration at this stage to generate fingerprint in CCS axis. Following the extraction, modules such as stability shift analysis, classification, and difference analysis can be performed 77

Figure 4-2. Peak detection algorithm output. (A-C) Mass spectra of HSA, IgG1, and GroEL, with peaks detected by the algorithm marked in orange. (D-F) CIU fingerprint of HSA, IgG1, and GroEL at a charge state with the highest relative intensity 84

Figure 4-3. Feature skipping feature comparison. (A) Feature detection results of IgG1, BSA CIU^2, and tRNA without feature skipping (top row) and with feature skipping (bottom row). (B) CIU50 values of the first CIU transition compared without feature skipping and with feature skipping 86

Figure 4-4. Multi-dimensional classification of BSA, HSA, RSA, and GSA. (A) Fingerprints of BSA, HSA, RSA, GSA CIU^1 (top) and CIU^2 (bottom) at charge state +15.a (B) Final classification scheme of multi-dimensional, multi-charge state classification. (C) Probabilities of categorizing all four serum albumins with both CIU^1 and CIU^2 89

Figure 4-5. Single-field CCS calibration methods. (A) Single-field DTIM CCS calibration using BSA as a calibrant at various charge states. Each data point refers to the drift time and CCS of

individual CIU feature. Linear regressions using OLS method are included. All three linear fits resulted r^2 =1. (B) CCS values of various standard proteins compared between the tunemix and BSA as calibrants. The grey dash line refers to $CCS_{tunemix} = CCS_{BSA}$..... 92

Appendix Figure I-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..... 104

Appendix Figure I-2. The CIU fingerprint of D12 monomer of 13with three feature observed 105

Appendix Figure I-3. (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..... 105

Appendix Figure I-4. (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..... 106

Appendix Figure I-5. Experimental CCS values of HSA D12 monomer obtained from Waters G2..... 106

Appendix Figure I-6. Clustering dendrogram of compact HSA D12 monomer at replica 12 (557.384 K) which is the lowest temperature replica at which models that agreed with experimental CCS was observed. Models from each cluster are also displayed and annotated. 107

Appendix Figure I-7. CIU-ECD experiment of +15 HSA. A). CIU fingerprint showing the four features observed in the Synapt G2 dataset. B) Mass spectrum at each voltage where ECD fragmentation was obtained. Data was extracted only at the drift bins containing the precursor and ECD fragmentation. C) Intensity vs Collision Voltage used for Internal fragment +4 c_{174}-$zdot_{436}$, which is located in the boundary of domain 1 and domain 2 108

Appendix Figure I-8. Compaction of full-length HSA (A) CCS distribution of full-length HSA from REMD replica 12 which is the temperature at which the lowest CCS model was observed. Despite the rigor of REMD, CHARMM36 is inefficient at achieving the level of compaction that agree with experimental IM-MS data. Simple equilibration MD simulations at 300K were performed using CHARMM36 (B), and GROMOS9643a2 (C) force fields where gray model represents the original database PDB structure and the colored model represents the final equilibrated HSA model 109

Appedix Figure I-9. Clustering dendrogram of full-length HSA compaction with models from each cluster..... 110

Appendix Figure II-1. Correction plot to convert IMPACT PA CCS into TM N_2 CCS using models from IgG1 monomers and dimers in both solution and gas-phase..... 112

Appendix Figure II-2. Example models from each cluster after clustering analysis of (A) 120 IPET models, (B) Final compacted models after 120 MD simulations, and (C) IM-filtered gas-phase models 113

Appendix Figure II-3. Percent changes in each secondary structure from solution to gas-phase models of IgG1 113

Appendix Figure III-1. Individual steps in peak detection algorithm. (A) Raw mass spectrum of HSA. (B) Mass spectrum of HSA after Savitzky-Golay smoothing. (C) Peaks detected by the algorithm are highlighted in orange within the mass spectrum of HSA 119

Appendix Figure III-2. Peak detection of SIRT5 and CIU fingerprints of various ligand-bound SIRT5. (A) Microfluidics device showing aqueous sample droplets (clear) separated by oil droplets (red). (B) Elution of sample droplets shows differences in intensities between sample droplets and oil droplets. (C) Mass spectrum of apo-SIRT5 with peaks detected by the algorithm annotated in orange. (D) CIU fingerprints of apo and various ligand-bound SIRT5 generated by peak detection outputs 120

Appendix Figure III-3. Application of peak detection algorithm on large-scale CIU data sets of biotherapeutics and biosimilars. (A) Mass spectrum of Remicade (RC) with peaks detected by the algorithm annotated in orange (left); CIU fingerprints generated using the low and high m/z bounds determined by the algorithm. (B) Overview of difference analysis performed between RC and RC, Renflexix (RF), and Inflectra (IF) across multiple charge states and replicates. (C) Total difference analysis plots with respect to RC in native states (top row) and stressed states (bottom row). (D) Heat map of RMSD values from all total difference analyses performed for both native states (left) and stressed states (right) 121

Appendix Figure III-4. Multi-dimensional classification of BSA. (A-C) Fingerprints of three classes used to build the classification scheme. (D-F) Classification results of UFS, cross-validation, and final classification scheme, respectively, when only 1D preprocessing on the CV axis was performed. (G-I) Classification results of UFS, cross-validation, and final classification scheme, respectively, when 2D preprocessing on CV and drift time axes were performed. Overall, the best scoring voltage shifts from 150V to 50V when 2D preprocessing is performed, and the accuracy of training and test data is improved 122

Appendix Figure IV-1. Overview and comparative summary of the effects of the site-specific structural flip of Aβ on cytotoxicity and CCS. ‡Cytotoxicity data for Aβ42-D23S26 are derived from previous literature (J. Org. Chem. 2020, 85, 1385) 132

Appendix Figure IV-2. Chiral differentiation of full-length Aβ42 stereoisomers (WT/dD/dS/dDdS, 4+) by using high-resolution TIMS (A) and cIMS (10 passes) (B) instruments 133

Appendix Figure IV-3. Energetically favorable REMD structures for Aβ42 stereoisomers of 4+ ion species. All structures are filtered by TIMS-CCS(N2) with 3% calibration error; conformations at 300 K (lowest temperature) were clustered and listed are the most confidently identified REMD clusters (population > 1% with well-aligned structures). R0-c3-107: R0 (rep #0), c3 (cluster #3), 107 (total pdb model numbers). N-terminus, blue; C-terminus, red; Ser

residues, magenta; Asp residues, hot pink. Energy for each conformer is also listed right below the cartoon illustration 134

Appendix Figure IV-4. IM-guided REMD simulation results for full-length Aβ42 stereoisomers (WT/dD/dS/dDdS, [Aβ42+4H]4+) and corresponding structural analysis. (A) Lowest energy and most dominant/major REMD structural clusters for Aβ42 stereoisomers. Conformer name, R0-c3-107: R0 (rep #0), c3 (cluster #3), 107 (total structure numbers). N-terminus, blue; C-terminus, red; Ser residues, magenta; Asp residues, hot pink. Energy for each conformer is also listed right below the cartoon illustration. (B-D) Calculated CCS distributions, secondary structures and charge-dipole interactions for REMD-simulated Aβ42 stereoisomers. Charge-dipole interactions are characterized with the spatial distances between certain residues (e.g. S26-E22 (d1), S26-D23 (d2) and S26-K28 (d3)), which is termed Accumulated Distance (Acc. Dis.) Accumulated Distance is calculated from the Euclidian distance between all interacting residues: $d_{acc} = (d_1^2 + d_2^2 + d_3^2)^{0.5}$ 136

Appendix Figure V-1. CCS distributions of (A) IgG1α dimer and (B) IgG1β dimer of 40+ charge state in native, heat-stressed, and pH-stressed conditions..... 140

Appendix Figure V-2. (A) Overlay of 120 IgG1 structures demonstrating the flexibility of mAb Fab regions. (B) Box-and-whisker plot with experimental CCS values from IgG1 α/β and solution and compacted CCS values for 10 monomeric mAb structures. (C) Comparison of the smallest (blue), median (green), and largest (orange) structures for solution-phase (top row) and gas-phase (bottom row) monomeric mAbs. (D) Several hand docked models built using archetypical IgG dimeric structures. (E) CCS measurements of experimental values from IgG1α/β and CCS calculations for compacted dimer models..... 141

List of Appendices

Appendix I: Supporting Information for Chapter 2 104

Appendix II: Supporting Information for Chapter 3.....112

Appendix III: Supporting Information for Chapter 4 119

Appendix IV: Site-Specific Chirality-Conferred Structural Compaction of Aβ42 127

Appendix V: Ion Mobility-Mass Spectrometry Reveals the Structures of Biotherapeutic Antibody Aggregates 140

Abstract

Proteins and their multi-component complexes are dynamic and perform myriad specific functions in cells. Understanding their functions is critical for drug development. Over the last three decades, mass spectrometry (MS) has emerged as a tool able to characterize complex biomolecules rapidly. Developments in native MS have extended such information to the structures of native proteins and complexes. Subsequently, ion mobility-mass spectrometry (IM-MS), in combination with collision induced unfolding (CIU), has further enhanced our understanding of native protein and protein complexes from detailed measurements of protein shapes, sizes, and stabilities, accessible within complex mixtures. While IM-MS generates a plethora of data, a range of computational approaches are typically required to extract information from these technologies. Current challenges include: 1) ambiguity associated with collision cross section (CCS) obtained from IM limiting our ability to accurately assign model structures; 2) increasing demands on IM-MS to provide access to high-throughput assays capable of tracking protein conformation. In this thesis, new computational methods are described that allow more accurate structural assignments based on IM-MS data and automation of CIU data analysis.

In Chapter 2, we developed a steered molecular dynamics (MD) approach that integrates native IM-MS, CIU, and electron capture dissociation (ECD) data to develop a seamless methodology for CIU feature annotation. ECD produces product ions preferentially from disordered regions within protein structures. As CIU progresses, new ECD fragment ions are observed, highlighting those sequences linked to the CIU features observed. Ultimately, this approach is supported by a novel steered MD approach developed and described here, where we

include observed fragmentation data into our simulation methodology to restrain regions within the protein that are not experimentally relevant for CIU allowing only the region identified through ECD as disordered, followed by CCS filtering. Our steered MD has enabled the complete structural assignment of a protein complex CIU fingerprint and has proven that CIU can capture subdomain-level unfolding in protein complexes.

In Chapter 3, we utilized integrative modeling to investigate the gas-phase structural ensemble of immunoglobulin (Ig), in particular, IgG1. The therapeutic benefits of monoclonal antibodies (mAbs) are in many cases linked to their flexibility, however detailed rapid characterization of mAb dynamics, in a manner useful for therapeutic development efforts, remains challenging To study if gas-phase IM measurements can capture mAb ensembles that retain a memory of the range of structures occupied by the protein in solution, we utilized individual particle electron tomography (IPET) mAb models as a starting point for gas-phase MD simulations. Our study reveals that the structural diversity of native IgG1 persists in the gas-phase.

In Chapter 4, we describe the development of CIUSuite 3, a next-generation software package for CIU data analysis, optimized to automate and expand the support various instrument platforms. To accomplish this, we built and validated four new modules and added them to our previous-generation CIUSuite package: 1) On-board, automated CCS calibration supporting multiple IM analyzer types which increases CIU data analysis throughput by a factor of 50-100, 2) A feature skipping algorithm designed to improve our ability to cross-compare complex CIU datasets, 3) Automated peak detection which serves to increase the throughput of CIU data analysis by a factor of 20, and 4) a machine learning classification architecture that supports multi-dimensional CIU (e.g., CIU^2) data types.

Chapter 1 Introduction

Proteins make up the machinery that accomplishes nearly all cellular activities. Thus, much research has focused on determining the structure of proteins as they correlate to their respective functions which can lead to the understanding of disease mechanisms and therapeutic developments.[1] Recently, there have been advances in various structural biology techniques designed to solve structures of protein and protein complexes.[2-5] Despite the plethora of structural biology tools available, structure determination of proteins and protein complexes still lags due to their inherent heterogeneity, dynamics, and their environment.[6,7] In this thesis, I introduce both experimental and computational techniques to determine protein structures with an emphasis on mass spectrometry methods.

1.1 Methods of Protein Structure Determination

Since the establishment of the Protein Data Bank (PDB) in 1971 there has been a continuous growth in the number of PDB entries with structures solved from various techniques such as X-ray, nuclear magnetic resonance (NMR), and electron microscopy. While these techniques give an atomic resolution of protein structures, many proteins are refractory to these high-resolution techniques and require alternative strategies to generate models.

1.1.1 Experimental Methods for Protein Structure Determination

Most PDB structures originate from X-ray crystallography data. This method has the longest history of PDB entries with the first structure resolved by Kendrew et al.[8] This method requires purified proteins to be crystallized which are then subjected to X-rays. The protein in the

crystal diffracts these X-rays generating a pattern that can be analyzed to calculate the electron density of the crystallized protein. This electron density map can then be used to locate the positions of the atoms within the protein. Therefore, protein structures resolved using X-ray crystallography provide detailed, atomic-level structural information.[9,10] However, there are still challenges in this technique mainly stemming from the purification and crystallization of the protein of interest. The protein samples not only have to be homogeneous and high purity, but also large protein quantities are required for crystallization. Furthermore, optimizing for crystallization of the purified proteins can be the limiting step in X-ray crystallography, resulting in long timelines for protein structure determination. This optimization step may include the addition of ligands or small molecules, mutation, or truncation of proteins.[11,12]

NMR spectroscopy is another technique that is used to determine atomic-level protein structures. In NMR, proteins are purified and placed in the magnetic field to expose them to radio frequency excitation pulses to measure nuclear spin resonance. The observed resonances are used to determine the bonding environments of nuclei followed by simulations to build atomic models of proteins.[13-16] One of the advantages of NMR in structural biology is the direct analysis of protein structure in solution which is biologically relevant compared to structure determined from X-ray crystallography. Furthermore, due to its capability to capture structural fluctuations in solution, NMR data yields an ensemble of protein structures.[17-19] Despite these advantages, NMR is limited by its sensitivity to resolve large proteins and protein complexes (~40 kDa) as larger proteins lead to greater spectral complexity with potential broadening and overlapping of NMR signals. However, recent studies have extended the limit of NMR to study both larger sequences and membrane proteins.[20-22]

Lastly, electron microscopy (EM) is another emerging technique in structural biology for high-resolution protein structure determination. In EM, protein particles are placed on a grid or vitrified ice (for cryo-EM) followed by electron beam scattering. These particle images are then filtered and grouped by their orientation to build a three-dimensional electron density map from these two-dimensional images.[23-25] EM has been widely utilized for large protein complexes, macromolecular assemblies, and membrane proteins.[26-28] EM generates structures with resolution comparable to X-ray crystallography without the necessity to purify large amounts of sample; however, the technique still requires conformation homogeneity, and optimization of buffer conditions to produce high-quality EM data remains challenging. Overall, all the techniques outlined above are all capable of generating high-resolution protein structures (Figure 1-1).

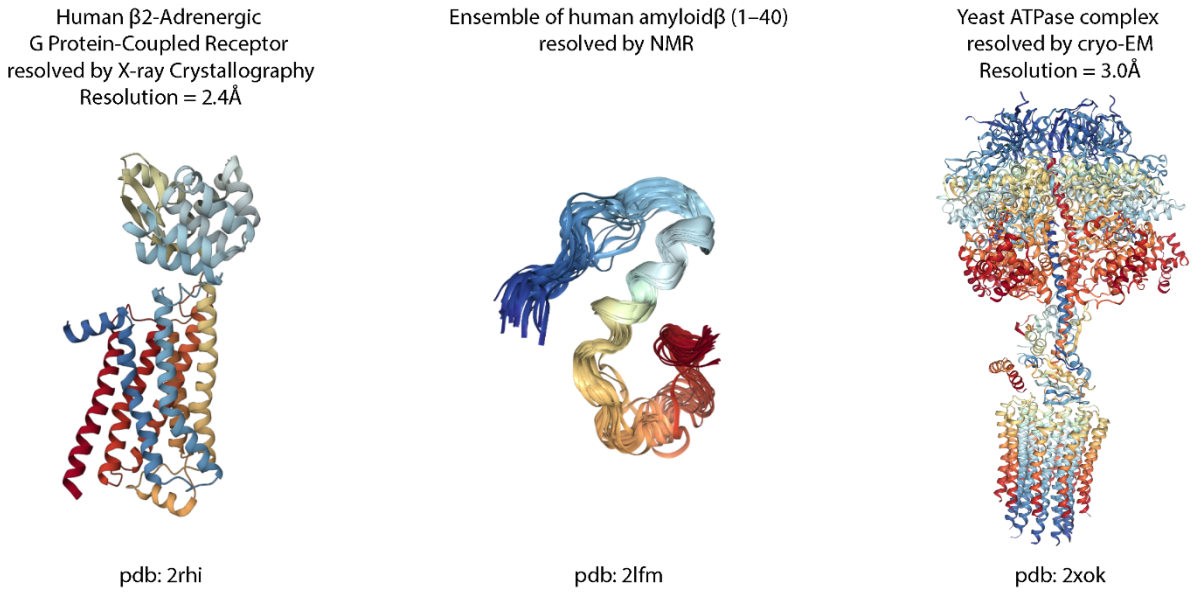


Figure 1-1. Atomic resolution structures of proteins and protein complexes resolved by X-ray crystallography, NMR, and cryo-EM.

While atomic-resolution protein structures are valuable, these high-resolution techniques may not always be ideal for practical reasons such as the amount of time and samples required, and the limitations of each technique outlined above. It is important to note that other structural

biology approaches may not provide high-resolution protein structures but can be integrated to assess the structures of proteins that are refractory to NMR, EM, and X-ray crystallography. Some of the techniques used in integrative structural biology include small angle X-ray scattering (SAXS)[29], circular dichroism (CD)[30], and Fourier transform infrared spectroscopy (FTIR)[31]. While these techniques may not provide atomic resolution, they provide valuable structural restraints with smaller amounts of protein samples without being highly purified. Overall, further development of structural biology tools that allow structural analysis in a high throughput manner with minimal sample preparation would be necessary to gain structural insight to answer and fill the knowledge gap in protein structures.

1.1.2 Computation Methods for Protein Structure Determination

Various computational methods served a critical role in protein structure determination. The initial development of computational methods has focused primarily on the direct application of physics such as statistical mechanics and Newton's law to determine protein structures. In more recent decades, template-based protein structure prediction has been growing fast especially as computational strategies such as machine learning and artificial intelligence are developed and utilized widely. Despite growing interest and technology, predicting and determining accurate protein structures from their sequence remains challenging, especially for intrinsically disordered peptides (IDPs) or structurally heterogeneous proteins.[32,33] In this section, computational methods used for protein structure characterization will be discussed.

Molecular Dynamics (MD) simulations are widely used computational methods in structural biology to characterize protein structures. This is also often referred to as *ab initio* modeling as it performs conformation searches based on fundamental physical properties. MD simulations use a force field (FF), a mathematical function that parameterizes the energy based on

Newton's Law for the system being simulated. The equation below is an example of this physics-based function used in CHARMM (Equation 1).[34] As seen in Equation 1-1, force fields typically consider two components – bonded and nonbonded terms – derived from experimental and theoretical data. Depending on how these energy functions optimized the terms, force fields may vary from one to another. Because energy is calculated at each time step in MD simulations, it has become a useful tool to study protein folding and the dynamic behavior of proteins. Furthermore, several conformational search algorithms have been developed for MD simulations to make the search more robust such as simulated annealing MD and replica exchange MD (REMD).[35] However, it is important to note that computational cost comes with an expense for accurate energy calculations for MD simulations.

$$E_{bonded} = \sum_{bonds} K_b(b - b_o)^2 + \sum_{angles} K_\theta\,(\theta - \theta_0)^2 + \sum_{\substack{improper \\ dihedrals}} K_\varphi(\varphi - \varphi_0)^2$$

$$+ \sum_{dihedrals} \sum_{n=1}^{6} K_{\phi,n}(1 + \cos(n\phi - \delta_n)) \qquad (Equation\ 1-1a)$$

$$E_{nonbonded} = \sum_{\substack{nonbonded \\ pairs\ i,j}} \frac{q_i q_j}{4\pi D \|r_i - r_j\|}$$

$$+ \sum_{\substack{nonbonded \\ pairs\ i,j}} \varepsilon_{ij} \left[\left(\frac{R_{min,ij}}{\|r_i - r_j\|}\right)^{12} - 2\left(\frac{R_{min,ij}}{\|r_i - r_j\|}\right)^{6} \right] \qquad (Equation\ 1-1b)$$

Another common modeling method utilizes sequence information to predict three-dimensional protein structures. Homology modeling is an approach that utilizes knowledge-based searches to predict structures based on the sequence of unknown structures and known three-dimensional structures of related proteins. In recent decades, many homology modeling software tools, such as ROSETTA[36] and I-TASSER[37,38], have combined knowledge-based searches with

physics-based searches to optimize and increase the accuracy of prediction. For instance, ROSETTA uses energy functions based on the physical properties of atoms in the proteins to evaluate and score the energy of predicted protein models[36]. I-TASSER incorporated *ab initio* modeling to generate numerous candidate structures followed by optimization to improve the prediction of protein structures[37,38]. With growing interest in computing algorithms, protein structure prediction has expanded to artificial intelligence (AI). Unlike homology modeling, it does not require information such as evolutionary relatedness or sequence relatedness nor has physics-based modeling and optimization. AI prediction still requires a large amount of data that can be used to train the neural network which will then predict structures based on given protein sequences.[32] AlphaFold2 is one of the most commonly used AI-based protein structure prediction software with high accuracy.[39] The application of AlphaFold2 is also increasing especially for integrative modeling approaches.[40] The growth of computational methods has strived for increasing the accuracy of calculated or predicted protein structures; however, it is important to note that continual development of experimental structural biology tools is necessary to increase the scope of protein structure determination as the prediction of protein structures are trained based on currently resolved proteins. For this reason, AlphaFold2 lacks accuracy for intrinsically disordered peptides and proteins (IDPs) and large protein complexes.

1.2 Mass Spectrometry in Structural Biology

Mass spectrometry (MS) has emerged as a new approach to protein characterization in the past few decades.[41-44] Depending on different MS methods adopted, users can gain insight from the amino acid sequence to full intact proteins and protein complexes. Early development has focused on bottom-up proteomics where proteins are digested with a protease followed by liquid chromatography-mass spectrometry (LC-MS).[45,46] Other proteomics methods have also been

developed such as middle-down and top-down approaches to identify modifications such as post-translational modifications (PTMs) and protein-protein or protein-ligand interactions.[47-51] Combined with proteomics, several labeling techniques for MS are available, such as hydrogen-deuterium exchange (HDX)[52] and cross-linking (XL)[53], which can provide more detailed structural information about proteins and protein complexes such as protein dynamics and binding interface of protein-ligand or protein-protein. Lastly, native MS (nMS) has been widely adopted for its ability to directly probe the structure of protein and protein complexes.[42,54-56] Coupled with various activation and fragmentation techniques, nMS can be a powerful structural biology tool to characterize protein structures with small amounts of samples. This section will discuss MS methods and instrumentation that can provide structural information on proteins from primary to quaternary structure.

1.2.1 Native Ion Mobility-Mass Spectrometry

To study biologically relevant structures and subunit composition of protein and protein complexes, it is important to preserve their native structures as they provide information about their functions and other cellular processes. Native MS allows direct measurement of proteins in the gas phase while preserving the three-dimensional structure of proteins and protein complexes.[57-59] Despite the state at which the measurement is taken, some studies suggest that measurements taken in the gas phase preserve many structural attributes of native proteins and protein complexes.[60,61]

To introduce proteins and protein complexes into the gas phase, samples much be ionized. Electrospray ionization (ESI) is widely used especially for large biomolecules.[62] Due to the high voltage applied at the emitter, charged droplets are formed which undergo droplet fission and evaporation to generate charged analyte ions to be analyzed by MS. ESI-MS also typically uses high source temperature (> 100 ºC) and organic solvents to increase the efficacy of the desolvation

process; however, these conditions are often too harsh to retain native structures of biomolecular analytes. In nMS, proteins are transferred to the gas phase using nano ESI (nESI) using non-denaturing solutions and lower source temperature (20-30 ºC).[63] Many studies have suggested the preservation of native-like structures of proteins in the gas phase using various MS techniques from the comparison of charge state distribution to images of soft-landed ions in the gas phase.[60,61,64,65] However, nMS alone doesn't provide detailed and accurate structural information of proteins and protein complexes in the gas phase which necessitates additional steps to produce structure models.

1.2.2 Ion Mobility-Mass Spectrometry

Ion mobility (IM) is a separation technique where protein ions are introduced into the gas phase and separates the analyte ions based on their shape, size, and charge.[66] During IM separation, packets of protein ions are introduced into a chamber pressurized with inert gas (typically N_2 or He) allowing the ions to be separated under the influence of a weak electric field generated by the ring electrodes (Figure 1-2). The time it takes for protein ions to travel the IM cell will be recorded as drift time which can be converted to collision cross section (CCS) which correlates to the orientationally averaged shape and size of the ions.[66-69] Coupled with nMS, various studies confirmed native-like structures of proteins in the gas phase.[70-73] With native IM-MS, CCS values obtained from IM measurement can be compared with various other experimental data such as X-ray crystallography and EM. Furthermore, these experimental CCSs can be used to filter models from MD simulations and generate experimentally relevant models.[74-76] In this thesis, two types of IM-MS are used: traveling wave ion mobility (TWIM) and drift tube ion mobility (DTIM). In TWIM, radio frequency (RF) and DC pulses are applied sequentially to ring electrodes to guide ions and create a traveling wave.[77,78] This allows ions to be separated by their mobilities as lower mobility ions are overtaken by the waves while higher mobility ions are "surfed" and travel the

IM cell faster. Due to the dynamic nature of the electric field in TWIM, a calibration using standards must be performed to convert TWIM drift time to CCS values. Unlike TWIM, DTIM uses a constant electric field to separate ions, allowing direct quantification of CCS values based on the measured drift time.[79,80]

Many commercial IM-MS instruments couple IM with a time-of-flight (TOF) mass analyzer. TOF mass analyzer is an electric field-free flight tube where ions are pulsed into the flight tube and separates ions according to their masses. This makes heavier ions take a longer time to traverse the flight tube compared to lighter ions.[81] Furthermore, when ion packets are pulsed into the flight tube, broadening can occur which can lower the resolution. To overcome this issue, reflectrons are used to correct the momentum of the ion packets and elongate the path length.[81] TOF mass analyzers are easy to install, have high mass resolving power, and have an unlimited mass range which makes this type of analyzer favorable for various types of analytes, especially proteins and protein complexes with high mass-to-charge values.

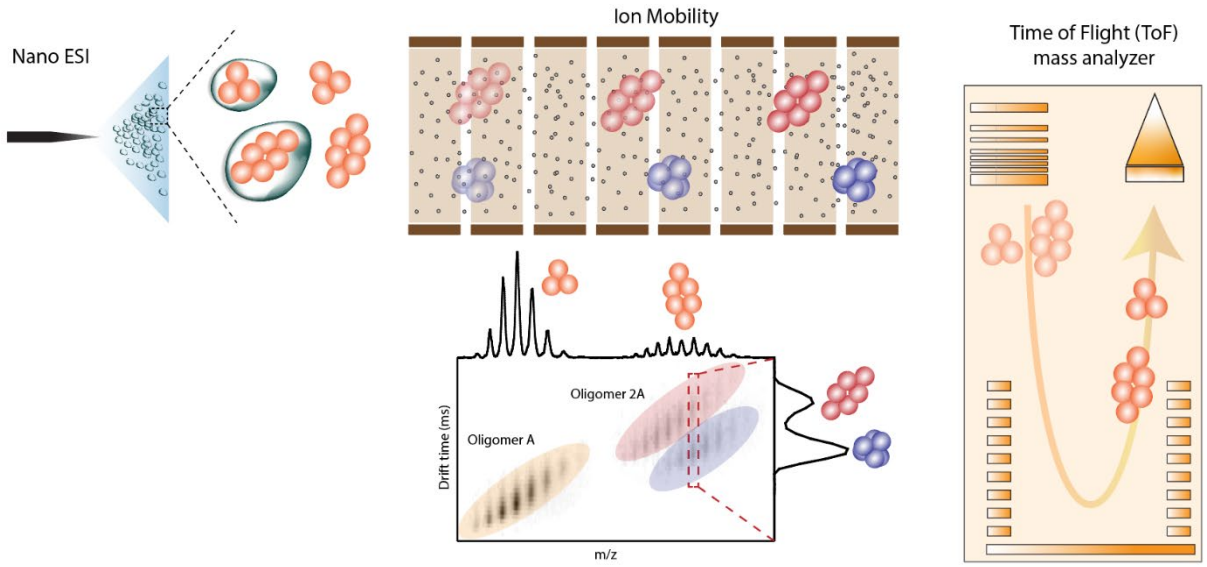


Figure 1-2. Instrument scheme for native IM-MS. Proteins are introduced into the gas phase by nano ESI. Protein ions are then separated by IM and analyzed by time-of-flight mass analyzer. The drift time and m/z obtained can be plotted in two-dimension to identify protein ions' stoichiometry but also their conformation states (oligomer 2A).

1.2.3 Collision Induced Unfolding

Collision induced unfolding (CIU) is another MS technique that is often utilized with IM-MS instrumentation to capture subtle structural differences within proteins in the gas phase. In CIU, protein ions undergo collisional activation with a background gas followed by IM-MS measurement. The collisional activations are performed in a stepwise manner to track the unfolding of proteins measured by increasing IM drift times.[82] The unfolding of proteins represented as IM drift time can be plotted as a function of collision voltage applied to generate a two-dimensional contour plot, or fingerprint (Figure 1-3). CIU fingerprints are extensively utilized to understand protein structures and to differentiate or classify proteins and their variants such as mutants, PTMs, and ligand binding. Early studies of CIU have demonstrated the capability of the technique to determine the number of autonomously folded domains within proteins.[83] Thus, the number of CIU features often correlates to the number of domains present in proteins plus one. In

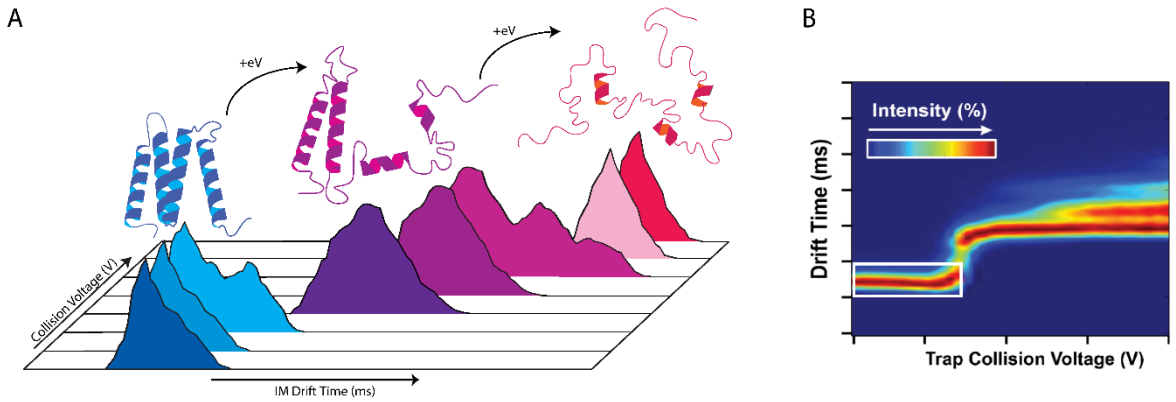


Figure 1-3. Collision Induced Unfolding (CIU) of protein ions. A) Collision voltages are increased to unfold proteins which are monitored using ion mobility drift time. B) CIU represented as a two-dimensional contour plot with collision voltage on x-axis and drift time on y-axis. Regions where protein conformations remain stable over a range of collision voltages are called features as indicated by a white box as an example.

addition, CIU can differentiate the structural differences and stability that are subtle by comparing the 50% convergence point from one CIU feature to the following feature, or CIU-50. This type of CIU analysis has been extensively applied for differentiating various types of biotherapeutics,

and for various types of antibodies that have different numbers of disulfide bonds and glycosylations.[84,85] Lastly, CIU has also been deployed to screen various ligands, and the fingerprints of various ligand-bound proteins can be used to classify the type of ligands being screened.[86]

1.2.4 Electron Capture Dissociation

Electron capture dissociation (ECD) fragments proteins by utilizing low energy electrons (1 eV) to produce backbone fragmentations.[87,88] Two mechanisms of ECD fragmentations are proposed. First, the charges on the protein ions attract these electrons, and the captured electron creates a radical cation followed by a proton transfer rearrangement.[87] Another proposed mechanism suggests direct electron capture of the amide group, forming an aminoketyl radical anion followed by proton transfer.[89,90] The main type of fragment ions produced by ECD are c•-ions (N-terminal fragments) and z•-ions (C-terminal fragments) originating from N-Cα bond (Figure 1-4).[91] Electron-based fragmentations have shown that various structural features are preserved such as higher-order structures, labile PTMs, and reduced disulfide bonds in the gas-phase.[92,93] For this reason, many studies have performed ECD fragmentations on native proteins and protein complexes to gain structural insights. For instance, many studies have utilized ECD to

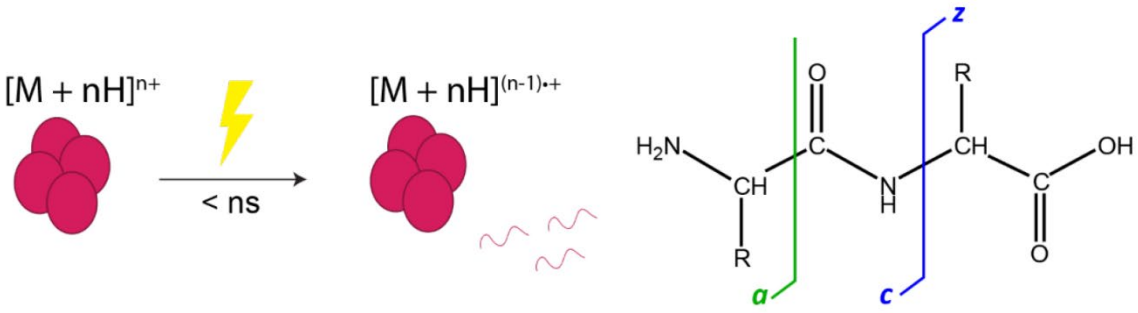


Figure 1-4. Fragmentation by electron capture dissociation (ECD). Electrons are captured by the protein ions in the gas-phase which generates radical cations. The radical induces rearrangements causing backbone fragmentation. The most common product ions are c-ions (N-terminal fragments) and z-ions (C-terminal fragments). The less common a-ions are also observed.

alcohol dehydrogenase (ADH) tetramer ions which generated predominantly N-terminal fragments.[94,95] The fragmentation pattern revealed that these regions correlated with the high B-factors of X-ray crystallography structure, indicating the utility of ECD to probe flexible or disordered regions of protein in the gas-phase.[94]

1.3 The Current State of Integrative Modeling for Native IM-MS

Protein samples studied using native IM-MS are conducted in the absence of solvent. While most structural information is conserved in the gas phase, some minor changes occur on the protein surface that contribute to the compaction of protein ions in the gas phase due to the lack of solvent and charges acquired during the ionization process (Figure 1-5).[96] Experimental data from native IM-MS has provided a wealth of information on the stoichiometry and conformation of proteins and protein complexes based on information such as mass-to-charge, charge state distribution, and IM drift time; however, analysis tools for interpreting and translating these information in the context of three-dimensional structures of proteins remains as a challenge. In this section, computational strategies to analyze native IM-MS data and current challenges will be discussed.

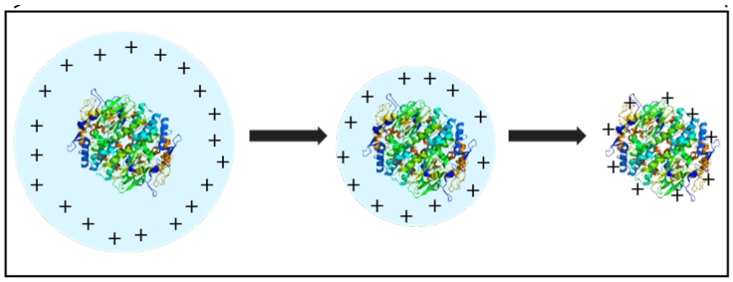


Figure 1-5. Nano ESI process places charges on the protein surface. Once the droplets are formed, the size of the droplets decreases as they desolvate causing the charges to be placed on the surface of the protein.

1.3.1 Building Three-Dimensional Models for Native IM-MS

When building models for protein or protein complexes, it is important to have initial building blocks to generate models that are relevant to IM-MS data. These starting building blocks come in various forms from atomic coordinates from X-ray crystallography to coarse-grained spheroids that represent individual subunits or complexes (Figure 1-6). Furthermore, these initial building blocks can be utilized in further downstream modeling analyses such as MD simulations. The most straightforward way to obtain these building blocks would be using structures that are already experimentally resolved and deposited in the protein structure database. However, not all proteins and protein complexes are resolved currently, and other approaches must be adopted in the case the protein structure is unknown. In cases where the protein structure has not been resolved using other experimental methods, building blocks must be generated using other methods such as homology modeling[36-38] or coarse-graining.[97,98] To obtain atomic resolution structure as a building block, homology modeling or other structure prediction method can be utilized while necessitating sequence information of the protein. However, by having atomic resolution starting structures, it can greatly increase computational costs in downstream analysis and MD simulations. To reduce

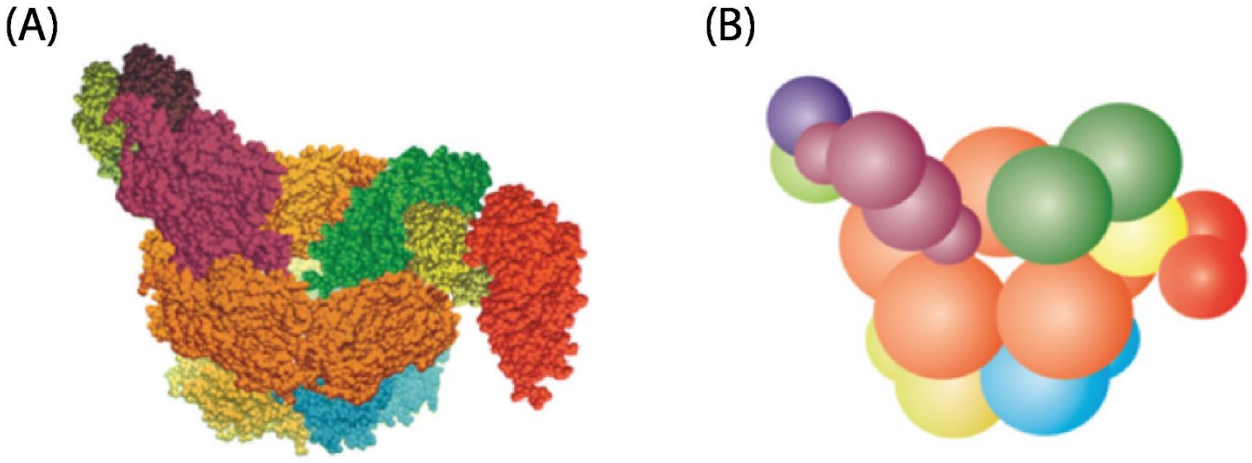


Figure 1-6. Models generated using all-atom and coarse-graining. MD simulations can be performed using either (A) all-atom or (B) coarse-graining which represents individual subunits or domains as spheroids. This figure illustrates how models differ for protein complexes between all-atom and coarse-grained where each subunit is represented in different colors and each spheroid represent the domain within the subunit.

the computational cost of all-atom modeling, coarse-graining can be performed to represent proteins and protein complexes in spheroids. The degree of coarse-graining can vary from subunit level to residue level which greatly affects the computational cost and accuracy of the models. While coarse-graining may be more efficient for downstream analysis, converting coarse-grained models to all-atom models may not straightforward. Furthermore, errors associated with coarse-graining greatly affect the accuracy of the models, especially at the subunit level, coarse-grained models did not accurately represent the structure of the proteins in comparison to the PDB structure deposited in the database.[99] This is an inherent error associated with coarse-graining, although the degree of errors can be reduced by refining the degree of coarse-graining followed by careful assessment of models generated such as restraint-based scoring.

1.3.2 Charge Assignment for Gas-Phase Protein Ions

Because proteins and protein complexes are ionized during the process of nano ESI, they acquire charges on the solvent-exposed area. The distribution of charges on protein is an integral part of modeling for gas-phase protein structures and contributes to the dynamics and accuracy of protein models and following MD simulations. Early studies have suggested charges in the gas phase are mobile and are distributed unequally between product ion populations upon collision induced dissociation (CID), also known as asymmetric charge partitioning.[100] Because of the mobile nature of the protons and the absence of direct observables that correlate to the location of charge sites, charging strategies for protein models in the gas phase need to be considered in order to accurately capture the structures associated with collisionally activated protein ions. Multiple charge isomers are often considered, and the number of possible charge isomers depends on the number of titratable residues present within the system. However, it is important to note that evaluating all charge isomers is not feasible with a plurality of charge isomers and computational

cost.[98] Several approaches to generating and sampling charge isomers are currently available. Monte Carlo (MC) approaches searches for charge isomers by moving protons to protonate or deprotonate basic and acidic residues. Various MC-based charge placement algorithms utilize the physical properties of input proteins such as atomic coordinates, electrostatic repulsion energy, and proton affinity, to evaluate and sample best-scoring charge isomers.[101] Other approaches utilize MD simulations to search for charge locations by searching for the lowest total energy of the protein iteratively.[102] While various charge placement algorithms are present, it is still challenging to locate charge sites that are experimentally relevant as there may be protein ions of the same charge state with different charge locations that are formed during the ESI process.

1.3.3 Molecular Dynamics Simulations for Gas Phase Proteins

With initial building blocks and charge locations determined, MD simulations can be performed to obtain structural information at a higher resolution that cannot be achieved by native IM-MS alone. MD simulations have already been utilized in a wide range of MS data such as solvent effects and ionization mechanisms.[103-105] The use of MD simulations has played an imperative role in interpreting native IM-MS data; however, current MD methods for native IM-MS lag behind those produced for other fields due to the force fields available that are optimized for gas-phase protein simulations and the plurality of charge isomers present in IM-MS experiments. The first challenge in the current MD methodology is that recently-developed force fields are optimized for solution-phase energy calculations.[34] However, this condition may not be suitable for simulating proteins and protein complexes in the gas phase. In the absence of solvent, electrostatic interactions predominate and become more significant for long-range interactions which many force fields are currently not optimized for. Moreover, polarizability optimized for solution-phase calculations may not be suitable for simulating proteins in the gas phase which

could increase the error of the models generated.[98] Another challenge is the multiplicity of charge isomers coexisting in the gas phase after the ionization process and the lack of experimental data to locate the exact charge location. It is also important to consider that charges may translocate in the gas phase, also predicted by the widely used mobile proton model describing peptide and protein CID product ion formation[106], which could contribute to structural changes and dynamics for protein ions in the gas phase. To model proteins in the gas phase that would account for mobile protons, bonds will need to be broken and formed during the simulations. However, in classical MD, breakage of bond is not allowed, which requires quantum mechanics/ molecular mechanics (QM/MM) to fully capture mobile protons and charge movements that continuously occur for protein ions in the gas phase.[107] While it may be favorable to opt for QM/MM to capture these charge movements and simulate protein ions to capture the dynamics that arise from mobile protons, it comes with a high computational cost.

1.3.4 Methods to Calculate Theoretical Collision Cross Sections

Lastly, models generated from MD simulations must be analyzed within the context of native IM-MS by comparing experimentally observed CCSs to theoretical CCSs calculated from the models. The method used for determining theoretical CCSs must be first considered. Two common methods widely used in native IM-MS are projected area (PA) and trajectory method (TM). In PA, the projected area of a model, or its shadow projected upon a plane, will be averaged over various orientations to determine the theoretical CCS.[108] This method is favorable due to its low computational cost. On the other hand, TM calculates CCS by using trajectory calculations which account for various physical properties, such as long-range interactions through Lennard-Jones potentials, which dominate ion-neutral interactions during typical IM-MS experiments.[109,110] Because TM integrates the physics of the trajectories of neutral gas molecules and their interactions

with target ions, it provides accurate theoretical CCS while also providing the most costly method for estimating model CCS. In the past decade, many efforts have been made to accurately and efficiently predict theoretical CCSs. For instance, IMPACT is a software tool developed to quickly calculate PA CCS while also predicting TM CCS using a built-in correction function.[111] Software tools that allow TM CCS calculations include IMoS[112] and Collidoscope[101], which all allow He and N_2 gas molecules for trajectory calculations relevant to the IM-MS experiment (Table 1-1).

Table 1-1. Overview of software for calculating collision cross sections of gas-phase ions.

Software	**Method**	**Scattering Type for Gas Trajectories**	**Charge State-Dependent**	**Buffer Gases**
IMPACT	PA	None	No	He
IMoS	PA & TM	Elastic	No (PA) & Yes ™	He, N_2, Ar, CO_2
Collidoscope	TM	Elastic	Yes	He, N_2

1.4 Summary

This dissertation presents various computational methods for native IM-MS to leverage our understanding of the structures of protein and protein complexes. In Chapter 2, the modeling framework to elucidate the gas phase unfolding mechanism of proteins and protein complexes using native IM-MS, CIU, and electron capture dissociation (ECD) data will be discussed using human serum albumin (HSA) as a model system. The steered MD simulations method is also developed to improve the accuracy of the final models and structures are assigned for each unfolding state. In Chapter 3, the methodology developed in Chapter 2 is used to model gas-phase IgG1 antibodies which revealed its structural diversity and memory effects in the gas-phase.

Similar to the diversity observed in the solution-phase, our gas-phase models also show the preservation of such diversity in the gas-phase. In Chapter 4, algorithms to automate various CIU data analyses are developed and packaged into CIUSuite 3. These algorithms include peak detection for CIU data extraction, multi-dimensional CIU classification, and CCS calibrations.

1.5 References

(1) Osadchy, M.; Kolodny, R. Maps of Protein Structure Space Reveal a Fundamental Relationship between Protein Structure and Function. *Proc. Natl. Acad. Sci.* **2011**, *108* (30), 12301–12306.

(2) Wang, R. Y. R.; Kudryashev, M.; Li, X.; Egelman, E. H.; Basler, M.; Cheng, Y.; Baker, D.; Dimaio, F. De Novo Protein Structure Determination from Near-Atomic-Resolution Cryo-EM Maps. *Nat. Methods*. **2015**, *12* (4), 335–338.

(3) Bai, X. Chen; McMullan, G.; Scheres, S. H. W. How Cryo-EM Is Revolutionizing Structural Biology. *Trends in Biochemical Sciences*. **2015**, 49–57.

(4) Opella, S. J.; Marassi, F. M. Applications of NMR to Membrane Proteins. *Arch. Biochem. Biophys.* **2017**, *628*, 92–101.

(5) Andreas, L. B.; Jaudzems, K.; Stanek, J.; Lalli, D.; Bertarelloa, A.; Marchand, T. Le; De Paepe, D. C.; Kotelovica, S.; Akopjana, I.; Knott, B.; Wegner, S.; Engelke, F.; Lesage, A.; Emsley, L.; Tars, K.; Herrmann, T.; Pintacuda, G. Structure of Fully Protonated Proteins by Proton-Detected Magic- Angle Spinning NMR. *Proc. Natl. Acad. Sci.* **2016**, *113* (33), 9187–9192.

(6) Rózycki, B.; Boura, E. Large, Dynamic, Multi-Protein Complexes: A Challenge for Structural Biology. *J. Phys. Condens. Matter.* **2014**, *26* (46), 463103.

(7) Feig, M.; Yu, I.; Wang, P. H.; Nawrocki, G.; Sugita, Y. Crowding in Cellular Environments at an Atomistic Level from Computer Simulations. *J. Phys. Chem. B.* **2017**, *121* (34), 8009–8025.

(8) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature.* **1958**, *181* (4610), 662–666.

(9) Morales-Rios, E.; Montgomery, M. G.; Leslie, A. G. W.; Walker, J. E. Structure of ATP Synthase from Paracoccus Denitrificans Determined by X-Ray Crystallography at 4.0 Å Resolution. *Proc. Natl. Acad. Sci.* **2015**, *112* (43), 13231–13236.

(10) Gres, A. T.; Kirby, K. A.; KewalRamani, V. N.; Tanner, J. J.; Pornillos, O.; Sarafianos, S. G. X-Ray Crystal Structures of Native HIV-1 Capsid Protein Reveal Conformational Variability. *Science.* **2015**, *349* (6243), 99–103.

(11) Zheng, H.; Handing, K. B.; Zimmerman, M. D.; Shabalin, I. G.; Almo, S. C.; Minor, W. X-Ray Crystallography over the Past Decade for Novel Drug Discovery -Where Are We Heading Next? *Expert Opinion on Drug Discovery*. **2015**, *10* (9), 975–989.

(12) Carpenter, E. P.; Beis, K.; Cameron, A. D.; Iwata, S. Overcoming the Challenges of Membrane Protein Crystallography. *Current Opinion in Structural Biology*. **2008**, *18* (5), 581-586.

(13) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. Protein Structure Determination from NMR Chemical Shifts. *Proc. Natl. Acad. Sci.* **2007**, *104* (23), 9615–9620.
(14) Tang, Y.; Huang, Y. J.; Hopf, T. A.; Sander, C.; Marks, D. S.; Montelione, G. T. Protein Structure Determination by Combining Sparse NMR Data with Evolutionary Couplings. *Nat. Methods.* **2015**, *12* (8), 751–754.
(15) Pan, B. Bin; Yang, F.; Ye, Y.; Wu, Q.; Li, C.; Huber, T.; Su, X. C. 3D Structure Determination of a Protein in Living Cells Using Paramagnetic NMR Spectroscopy. *Chem. Commun.* **2016**, *52* (67), 10237–10240.
(16) Quinn, C. M.; Polenova, T. Structural Biology of Supramolecular Assemblies by Magic-Angle Spinning NMR Spectroscopy. *Quarterly Reviews of Biophysics*. **2017**, *18*.
(17) Bonomi, M.; Heller, G. T.; Camilloni, C.; Vendruscolo, M. Principles of Protein Structural Ensemble Determination. *Current Opinion in Structural Biology*. **2017**, *42*, 106–116.
(18) Kovermann, M.; Rogne, P.; Wolf-Watz, M. Protein Dynamics and Function from Solution State NMR Spectroscopy. *Quarterly reviews of biophysics*. **2016**, *49*.
(19) Lange, O. F.; Rossi, P.; Sgourakis, N. G.; Song, Y.; Lee, H. W.; Aramini, J. M.; Ertekin, A.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Baker, D. Determination of Solution Structures of Proteins up to 40 KDa Using CS-Rosetta with Sparse NMR Data from Deuterated Samples. *Proc. Natl. Acad. Sci.* **2012**, *109* (27), 10873–10878.
(20) Raman, S.; Lange, O. F.; Rossi, P.; Tyka, M.; Wang, X.; Aramini, J.; Liu, G.; Ramelot, T. A.; Eletsky, A.; Szyperski, T.; Kennedy, M. A.; Prestegard, J.; Montelione, G. T.; Baker, D. NMR Structure Determination for Larger Proteins Using Backbone-Only Data. *Science.* **2010**, *327* (5968), 1014–1018.
(21) Religa, T. L.; Sprangers, R.; Kay, L. E. Dynamic Regulation of Archaeal Proteasome Gate Opening as Studied by Trosy NMR. *Science.* **2010**, *328* (5974), 98–102.
(22) Gautier, A.; Mott, H. R.; Bostock, M. J.; Kirkpatrick, J. P.; Nietlispach, D. Structure Determination of the Seven-Helix Transmembrane Receptor Sensory Rhodopsin II by Solution NMR Spectroscopy. *Nat. Struct. Mol. Biol.* **2010**, *17* (6), 768–774.
(23) Fernandez-Leiro, R.; Scheres, S. H. W. Unravelling Biological Macromolecules with Cryo-Electron Microscopy. *Nature*. **2016**, *537* (7620), 339–346. https://doi.org/10.1038/nature19948.
(24) Dimaio, F.; Song, Y.; Li, X.; Brunner, M. J.; Xu, C.; Conticello, V.; Egelman, E.; Marlovits, T. C.; Cheng, Y.; Baker, D. Atomic-Accuracy Models from 4.5-Å Cryo-Electron Microscopy Data with Density-Guided Iterative Local Refinement. *Nat. Methods.* **2015**, *12* (4), 361–365.
(25) Campbell, M. G.; Veesler, D.; Cheng, A.; Potter, C. S.; Carragher, B. 2.8 Å Resolution Reconstruction of the Thermoplasma Acidophilum 20 s Proteasome Using Cryo-Electron Microscopy. *Elife.* **2015**, *2015* (4), 1–22.
(26) Zubcevic, L.; Herzik, M. A.; Chung, B. C.; Liu, Z.; Lander, G. C.; Lee, S. Y. Cryo-Electron Microscopy Structure of the TRPV2 Ion Channel. *Nat. Struct. Mol. Biol.* **2016**, *23* (2), 180–186.
(27) Franken, L. E.; Oostergetel, G. T.; Pijning, T.; Puri, P.; Arkhipova, V.; Boekema, E. J.; Poolman, B.; Guskov, A. A General Mechanism of Ribosome Dimerization Revealed by Single-Particle Cryo-Electron Microscopy. *Nat. Commun.* **2017**, *8* (1), 1–11.
(28) Bai, X. C.; Fernandez, I. S.; McMullan, G.; Scheres, S. H. W. Ribosome Structures to Near-Atomic Resolution from Thirty Thousand Cryo-EM Particles. *Elife.* **2013**, *2* (2), 1-12.

(29) Yang, S. Methods for SAXS-Based Structure Determination of Biomolecular Complexes. *Adv. Mater.* **2014**, *26* (46), 7902–7910.
(30) Miles, A. J.; Wallace, B. A. Circular Dichroism Spectroscopy of Membrane Proteins. *Chemical Society Reviews*. **2016**, *45* (18), 4859–4872.
(31) Titus, J.; Ghimire, H.; Viennois, E.; Merlin, D.; Unil Perera, A. G. Protein Secondary Structure Analysis of Dried Blood Serum Using Infrared Spectroscopy to Identify Markers for Colitis Screening. *J. Biophotonics* **2018**, *11* (3), 1-8.
(32) Kuhlman, B.; Bradley, P. Advances in Protein Structure Prediction and Design. *Nature Reviews Molecular Cell Biology*. **2019**, *20* (11), 681–697.
(33) Moult, J.; Fidelis, K.; Kryshtafovych, A.; Schwede, T.; Tramontano, A. Critical Assessment of Methods of Protein Structure Prediction (CASP)—Round XII. *Proteins Struct. Funct. Bioinforma.* **2018**, *86*, 7–15.
(34) 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. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30* (10), 1545–1614.
(35) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314* (1–2), 141–151.
(36) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; Labonte, J. W.; Pacella, M. S.; Bonneau, R.; Bradley, P.; Dunbrack, R. L.; Das, R.; Baker, D.; Kuhlman, B.; Kortemme, T.; Gray, J. J. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **2017**, *13* (6), 3031–3048.
(37) Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A Unified Platform for Automated Protein Structure and Function Prediction. *Nat. Protoc.* **2010**, *5* (4), 725–738.
(38) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein Structure and Function Prediction. *Nature Methods*. **2014**, *12* (1), 7–8.
(39) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature.* **2021**, *596* (7873), 583–589.
(40) Fischer, E. S.; Yu, C. W. H.; Hevler, J. F.; McLaughlin, S. H.; Maslen, S. L.; Heck, A. J. R.; Freund, S. M. V.; Barford, D. Juxtaposition of Bub1 and Cdc20 on Phosphorylated Mad1 during Catalytic Mitotic Checkpoint Complex Assembly. *Nat. Commun.* **2022**, *13* (1), 1–18.
(41) Van Den Heuvel, R. H. H.; Heck, A. J. R. Native Protein Mass Spectrometry: From Intact Oligomers to Functional Machineries. *Current Opinion in Chemical Biology*. **2004**, *8* (5), 519–526.
(42) Heck, A. J. R. Native Mass Spectrometry: A Bridge between Interactomics and Structural Biology. *Nat. Methods.* **2008**, *5* (11), 927–933.

(43) Benesch, J. L. P.; Ruotolo, B. T. Mass Spectrometry: Come of Age for Structural and Dynamical Biology. *Current Opinion in Structural Biology*. **2011**, *21* (5), 641–649.
(44) Konijnenberg, A.; Butterer, A.; Sobott, F. Native Ion Mobility-Mass Spectrometry and Related Methods in Structural Biology. *Biochimica et Biophysica Acta - Proteins and Proteomics*. **2013**, *1834* (6), 1239–1256.
(45) Tsiatsiani, L.; Heck, A. J. R. Proteomics beyond Trypsin. *FEBS Journal*. **2015**, *282* (14), 2612–2626.
(46) Gundry, R. L.; White, M. Y.; Murray, C. I.; Kane, L. A.; Fu, Q.; Stanley, B. A.; Van Eyk, J. E. Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-up Proteomics Workflow. *Curr. Protoc. Mol. Biol.* **2009**, *90* (88), 10.25.1-10.25.23.
(47) Cannon, J.; Lohnes, K.; Wynne, C.; Wang, Y.; Edwards, N.; Fenselau, C. High-Throughput Middle-down Analysis Using an Orbitrap. *J. Proteome Res.* **2010**, *9* (8), 3886–3890.
(48) Fornelli, L.; Ayoub, D.; Aizikov, K.; Beck, A.; Tsybin, Y. O. Middle-down Analysis of Monoclonal Antibodies with Electron Transfer Dissociation Orbitrap Fourier Transform Mass Spectrometry. *Anal. Chem.* **2014**, *86* (6), 3005–3012.
(49) Toby, T. K.; Fornelli, L.; Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annual Review of Analytical Chemistry*. **2016**, *9*, 499–519.
(50) Cai, W.; Tucholski, T.; Chen, B.; Alpert, A. J.; McIlwain, S.; Kohmoto, T.; Jin, S.; Ge, Y. Top-Down Proteomics of Large Proteins up to 223 KDa Enabled by Serial Size Exclusion Chromatography Strategy. *Anal. Chem.* **2017**, *89* (10), 5467–5475.
(51) Fornelli, L.; Durbin, K. R.; Fellers, R. T.; Early, B. P.; Greer, J. B.; LeDuc, R. D.; Compton, P. D.; Kelleher, N. L. Advancing Top-down Analysis of the Human Proteome Using a Benchtop Quadrupole-Orbitrap Mass Spectrometer. *J. Proteome Res.* **2017**, *16* (2), 609–618.
(52) Houde, D.; Arndt, J.; Domeier, W.; Berkowitz, S.; Engen, J. R. Characterization of IgG1 Conformation and Conformational Dynamics by Hydrogen/Deuterium Exchange Mass Spectrometry. *Anal. Chem.* **2009**, *81* (7), 2644–2651
(53) Sinz, A. Chemical Cross-Linking and Mass Spectrometry to Map Three-Dimensional Protein Structures and Protein-Protein Interactions. *Mass Spectrom. Rev.* **2006**, *25* (4), 663–682.
(54) Erba, E. B.; Petosa, C. The Emerging Role of Native Mass Spectrometry in Characterizing the Structure and Dynamics of Macromolecular Complexes. *Protein Sci.* **2015**, *24* (8), 1176–1192. h
(55) Leney, A. C.; Heck, A. J. R. Native Mass Spectrometry: What Is in the Name? *J. Am. Soc. Mass Spectrom.* **2017**, *28* (1), 5–13.
(56) Li, H.; Nguyen, H. H.; Loo, R. R. O.; Campuzano, I. D. G.; Loo, J. A. An Integrated Native Mass Spectrometry and Topdown Proteomics Method That Connects Sequence to Structure and Function of Macromolecularcomplexes. *Nat. Chem.* **2018**, *10* (2), 139–148.
(57) Loo, J. A. Observation of Large Subunit Protein Complexes by Electrospray Ionization Mass Spectrometry. *J. Mass Spectrom.* **1995**, *30* (1), 180–183.
(58) Robinson, C. V.; Radford, S. E. Weighing the Evidence for Structure: Electrospray Ionization Mass Spectrometry of Proteins. *Structure.* **1995**, *3* (9), 861–865.
(59) Benesch, J. L. P.; Ruotolo, B. T.; Simmons, D. A.; Robinsons, C. V. Protein Complexes in the Gas Phase: Technology for Structural Genomics and Proteomics. *Chemical Reviews*. **2007**, *107* (8), 3544–3567.

(60) Mikhailov, V. A.; Mize, T. H.; Benesch, J. L. P.; Robinson, C. V. Mass-Selective Soft-Landing of Protein Assemblies with Controlled Landing Energies. *Anal. Chem.* **2014**, *86* (16), 8321–8328.
(61) Salome, A. Z.; Lee, K. W.; Grant, T.; Westphall, M. S.; Coon, J. J. Matrix-Landing Mass Spectrometry for Electron Microscopy Imaging of Native Protein Complexes. *Anal. Chem.* **2022**, *94* (50), 17616–17624.
(62) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science*. **1989**, *246* (4926), 64–71.
(63) Wilm, M.; Mann, M. Analytical Properties of the Nanoelectrospray Ion Source. *Anal. Chem.* **1996**, *68* (1), 1–8.
(64) Mann, M.; Meng, C. K.; Fenn, J. B. Interpreting Mass Spectra of Multiply Charged Ions. *Anal. Chem.* **1989**, *61* (15), 1702–1708.
(65) Hamdan, M.; Curcuruto, O. Probing Conformational Changes in Some Proteins by Positive-ion Electrospray Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1994**, *8* (2), 144–148.
(66) Ben-Nissan, G.; Sharon, M. The Application of Ion-Mobility Mass Spectrometry for Structure/Function Investigation of Protein Complexes. *Current Opinion in Chemical Biology*. **2018**, *42*, 25–33.
(67) Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E. The Power of Ion Mobility-Mass Spectrometry for Structural Characterization and the Study of Conformational Dynamics. *Nature Chem*. **2014**, *6* (4), 281–294.
(68) Allen, S. J.; Giles, K.; Gilbert, T.; Bush, M. F. Ion Mobility Mass Spectrometry of Peptide, Protein, and Protein Complex Ions Using a Radio-Frequency Confining Drift Cell. *Analyst.* **2016**, *141* (3), 884–891.
(69) May, J. C.; McLean, J. A. Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Anal. Chem*. **2015**, *87* (3), 1422–1436.
(70) Belov, M. E.; Clowers, B. H.; Prior, D. C.; Danielson, W. F.; Liyu, A. V.; Petritis, B. O.; Smith, R. D. Dynamically Multiplexed Ion Mobility Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2008**, *80* (15), 5873–5883.
(71) Ruotolo, B. T.; Hyung, S. J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V. Ion Mobility-Mass Spectrometry Reveals Long-Lived, Unfolded Intermediates in the Dissociation of Protein Complexes. *Angew. Chemie - Int. Ed.* **2007**, *46* (42), 8001–8004.
(72) Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S. J.; Robinson, C. V. Ion Mobility-Mass Spectrometry Analysis of Large Protein Complexes. *Nat. Protoc.* **2008**, *3* (7), 1139–1152. https://doi.org/10.1038/nprot.2008.78.
(73) Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Anal. Chem.* **2010**, *82* (22), 9557–9565.
(74) Soper, M. T.; DeToma, A. S.; Hyung, S. J.; Lim, M. H.; Ruotolo, B. T. Amyloid-β-Neuropeptide Interactions Assessed by Ion Mobility-Mass Spectrometry. *Phys. Chem. Chem. Phys.* **2013**, *15* (23), 8952–8961.
(75) Ghassabi Kondalaji, S.; Khakinejad, M.; Tafreshian, A.; J. Valentine, S. Comprehensive Peptide Ion Structure Studies Using Ion Mobility Techniques: Part 1. An Advanced Protocol for Molecular Dynamics Simulations and Collision Cross-Section Calculation. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (5), 947–959.

(76) Eschweiler, J. D.; Farrugia, M. A.; Dixit, S. M.; Hausinger, R. P.; Ruotolo, B. T. A Structural Model of the Urease Activation Complex Derived from Ion Mobility-Mass Spectrometry and Integrative Modeling. *Structure.* **2018**, *26* (4), 599-606.
(77) Shvartsburg, A. A.; Smith, R. D. Fundamentals of Traveling Wave Ion Mobility Spectrometry. *Anal. Chem.* **2008**, *80* (24), 9689–9699.
(78) Giles, K.; Williams, J. P.; Campuzano, I. Enhancements in Travelling Wave Ion Mobility Resolution. In *Rapid Communications in Mass Spectrometry*. **2011**, *25* (11), 1559–1566.
(79) Gabelica, V.; Shvartsburg, A. A.; Afonso, C.; Barran, P.; Benesch, J. L. P.; Bleiholder, C.; Bowers, M. T.; Bilbao, A.; Bush, M. F.; Campbell, J. L.; Campuzano, I. D. G.; Causon, T.; Clowers, B. H.; Creaser, C. S.; De Pauw, E.; Far, J.; Fernandez-Lima, F.; Fjeldsted, J. C.; Giles, K.; Groessl, M.; Hogan, C. J.; Hann, S.; Kim, H. I.; Kurulugama, R. T.; May, J. C.; McLean, J. A.; Pagel, K.; Richardson, K.; Ridgeway, M. E.; Rosu, F.; Sobott, F.; Thalassinos, K.; Valentine, S. J.; Wyttenbach, T. Recommendations for Reporting Ion Mobility Mass Spectrometry Measurements. *Mass Spectrometry Reviews*. **2019**, *38* (3), 291–320.
(80) May, J. C.; Goodwin, C. R.; Lareau, N. M.; Leaptrot, K. L.; Morris, C. B.; Kurulugama, R. T.; Mordehai, A.; Klein, C.; Barry, W.; Darland, E.; Overney, G.; Imatani, K.; Stafford, G. C.; Fjeldsted, J. C.; McLean, J. A. Conformational Ordering of Biomolecules in the Gas Phase: Nitrogen Collision Cross Sections Measured on a Prototype High Resolution Drift Tube Ion Mobility-Mass Spectrometer. *Anal. Chem.* **2014**, *86* (4), 2107–2116.
(81) Boesl, U. Time-of-Flight Mass Spectrometry: Introduction to the Basics. *Mass Spectrometry Reviews*. **2017**, *36* (1), 86–109.
(82) Dixit, S. M.; Polasky, D. A.; Ruotolo, B. T. Collision Induced Unfolding of Isolated Proteins in the Gas Phase: Past, Present, and Future. *Current Opinion in Chemical Biology*. **2018**, *42*, 93–100.
(83) Eschweiler, J. D.; Martini, R. M.; Ruotolo, B. T. Chemical Probes and Engineered Constructs Reveal a Detailed Unfolding Mechanism for a Solvent-Free Multidomain Protein. *J. Am. Chem. Soc.* **2017**, *139* (1), 534–540.
(84) Tian, Y.; Han, L.; Buckner, A. C.; Ruotolo, B. T. Collision Induced Unfolding of Intact Antibodies: Rapid Characterization of Disulfide Bonding Patterns, Glycosylation, and Structures. *Anal. Chem.* **2015**, *87* (22), 11509–11515.
(85) Tian, Y.; Ruotolo, B. T. Collision Induced Unfolding Detects Subtle Differences in Intact Antibody Glycoforms and Associated Fragments. *Int. J. Mass Spectrom.* **2018**, *425*, 1–9.
(86) D'Amico, C. I.; Polasky, D. A.; Steyer, D. J.; Ruotolo, B. T.; Kennedy, R. T. Ion Mobility-Mass Spectrometry Coupled to Droplet Microfluidics for Rapid Protein Structure Analysis and Drug Discovery. *Anal. Chem.* **2022**, *94* (38), 13084–13091.
(87) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *Journal of the American Chemical Society*. **1998**, pp 3265–3266.
(88) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. Electron Capture Dissociation for Structural Characterization of Multiply Charged Protein Cations. *Anal. Chem.* **2000**, *72* (3), 563–573.
(89) Sawicka, A.; Skurski, P.; Hudgins, R. R.; Simons, J. Model Calculations Relevant to Disulfide Bond Cleavage via Electron Capture Influenced by Positively Charged Groups. *J. Phys. Chem. B* **2003**, *107* (48), 13505–13511.

(90) Syrstad, E. A.; Tureček, F. Toward a General Mechanism of Electron Capture Dissociation. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (2), 208–224.
(91) Tsiatsiani, L.; Heck, A. J. R. Proteomics beyond Trypsin. *FEBS Journal*. **2015**.
(92) Lermyte, F.; Valkenborg, D.; Loo, J. A.; Sobott, F. Radical Solutions: Principles and Application of Electron-Based Dissociation in Mass Spectrometry-Based Analysis of Protein Structure. *Mass Spectrom. Rev*. **2018**, No. April 2017, 1–22.
(93) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. Electron Capture Dissociation for Structural Characterization of Multiply Charged Protein Cations. *Anal. Chem*. **2000**, *72* (3), 563–573.
(94) Zhang, H.; Cui, W.; Wen, J.; Blankenship, R. E.; Gross, M. L. Native Electrospray and Electron-Capture Dissociation in Fticr Mass Spectrometry Provide Top-down Sequencing of a Protein Component in an Intact Protein Assembly. *J. Am. Soc. Mass Spectrom*. **2010**, *21* (12), 1966–1968.
(95) Zhou, M.; Liu, W.; Shaw, J. B. Charge Movement and Structural Changes in the Gas-Phase Unfolding of Multimeric Protein Complexes Captured by Native Top-Down Mass Spectrometry. *Anal. Chem.* **2020**, *92* (2), 1788–1795.
(96) Wyttenbach, T.; Bowers, M. T. Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-Free Ion Mobility-Mass Spectrometry Environment. *J. Phys. Chem. B.* **2011**, *115* (42), 12266–12275.
(97) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; Kroon, P. C.; Melcr, J.; Nieto, V.; Corradi, V.; Khan, H. M.; Domański, J.; Javanainen, M.; Martinez-Seara, H.; Reuter, N.; Best, R. B.; Vattulainen, I.; Monticelli, L.; Periole, X.; Tieleman, D. P.; de Vries, A. H.; Marrink, S. J. Martini 3: A General Purpose Force Field for Coarse-Grained Molecular Dynamics. *Nat. Methods.* **2021**, *18* (4), 382–388.
(98) Allison, T. M.; Barran, P.; Cianferani, S.; Degiacomi, M. T.; Gabelica, V.; Grandori, R.; Marklund, E. G.; Menneteau, T.; Migas, L. G.; Politis, A.; Sharon, M.; Sobott, F.; Thalassinos, K. Benesch, J. L. P. Computational Strategies and Challenges for Using Native Ion Mobility Mass Spectrometry in Biophysics and Structural Biology. *Anal. Chem*. **2020**, *92 (16),* 10872–10880.
(99) Eschweiler, J. D.; Frank, A. T.; Ruotolo, B. T. Coming to Grips with Ambiguity: Ion Mobility-Mass Spectrometry for Protein Quaternary Structure Assignment. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (10), 1991–2000.
(100) Jurchen, J. C.; Williams, E. R. Origin of Asymmetric Charge Partitioning in the Dissociation of Gas-Phase Protein Homodimers. *J. Am. Chem. Soc.* **2003**, *125* (9), 2817–2826.
(101) Ewing, S. A.; Donor, M. T.; Wilson, J. W.; Prell, J. S. Collidoscope: An Improved Tool for Computing Collisional Cross-Sections with the Trajectory Method. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (4), 587–596.
(102) Popa, V.; Trecroce, D. A.; McAllister, R. G.; Konermann, L. Collision-Induced Dissociation of Electrosprayed Protein Complexes: An All-Atom Molecular Dynamics Model with Mobile Protons. *J. Phys. Chem. B.* **2016**, *120* (23), 5114–5124.
(103) Metwally, H.; McAllister, R. G.; Popa, V.; Konermann, L. Mechanism of Protein Supercharging by Sulfolane and M-Nitrobenzyl Alcohol: Molecular Dynamics Simulations of the Electrospray Process. *Anal. Chem.* **2016**, *88* (10), 5345–5354.
(104) Aliyari, E.; Konermann, L. Formation of Gaseous Proteins via the Ion Evaporation Model (IEM) in Electrospray Mass Spectrometry. *Anal. Chem.* **2020**, *92* (15), 10807–10814.

(105) Aliyari, E.; Konermann, L. Atomistic Insights into the Formation of Nonspecific Protein Complexes during Electrospray Ionization. *Anal. Chem.* **2021**, *93* (37), 12748–12757.
(106) Boyd, R.; Somogyi, Á. The Mobile Proton Hypothesis in Fragmentation of Protonated Peptides: A Perspective. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (8), 1275–1278.
(107) Li, J.; Lyu, W.; Rossetti, G.; Konijnenberg, A.; Natalello, A.; Ippoliti, E.; Orozco, M.; Sobott, F.; Grandori, R.; Carloni, P. Proton Dynamics in Protein Mass Spectrometry. *J. Phys. Chem. Lett.* **2017**, *8* (6), 1105–1112.
(108) Mack Jr, E. Avrage Cross-Sectional Areas of Molecules by Gaseous Diffusion Methods. *J. Am. Chem. Soc* **1925**, *47* (10), 2468–2482.
(109) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem.* **1996**, *100* (40), 16082–16086.
(110) Wyttenbach, T.; Von Helden, G.; Batka, J. J.; Carlat, D.; Bowers, M. T. Effect of the Long-Range Potential on Ion Mobility Measurements. *J. Am. Soc. Mass Spectrom.* **1997**, *8* (3), 275–282.
(111) Marklund, E. G.; Degiacomi, M. T.; Robinson, C. V.; Baldwin, A. J.; Benesch, J. L. P. Collision Cross Sections for Structural Proteomics. *Structure.* **2015**, *23* (4), 791–799.
(112) Larriba, C.; Hogan, C. J. Free Molecular Collision Cross Section Calculation Methods for Nanoparticles and Complex Ions with Energy Accommodation. *J. Comput. Phys.* **2013**, *251*, 344–363.

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

Authors: Zhao, C.; Jeon, C. K.; Rojas Ramirez, C.; and Ruotolo B. T.

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 limitations, leaving a large section of the proteome relatively unknown from a structural biology perspective.

Recently advancements in native ion mobility-mass spectrometry (IM-MS) have been developed to improve the throughput and leverage our understanding of protein structures that weren't amenable to other structural biology techniques. Many studies have successfully characterized protein stoichiometry and topologies that were challenging to study with other structural biology tools. 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 both computational and 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 structural information below the resolution of IM, 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 structures 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 unknown. 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 the 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 allow 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]

Despite various efforts to experimentally characterize protein complexes, there are various challenges associated with computational methods which could be utilized to leverage our understanding of the structural information obtained experimentally. With no direct observables for charge locations, placing charges onto proteins becomes a major challenge. Additionally, many modeling and MD protocols and force fields are optimized for solution-phase, using these methods will affect the accuracy of the models for data obtained in the gas-phase.[15] With various challenges associated with computational methods for gas-phase proteins, the field also lacks general methods for performing MD simulations of gas-phase proteins and protein complexes.

In this study, we used 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 of 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 (ECD) and steered molecular dynamics (MD) simulations to confirm our mechanism and provide additional details. Previous studies have demonstrated that ECD

fragmentation originates from regions with high solvent accessible surface area.[15–17] Therefore, ECD can be used as conformational foot printing method of regions undergoing unfolding. 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 Methods

2.2.1 Sample Preparation

Wild-type (WT) human serum albumin was 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-naphthalene sulfonic 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.

2.2.2 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.[18] 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 at 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.[19] 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 CIUSuite 2.

2.2.3 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 was tracked and normalized to plot the CID curve associated with each probe-bound complex.

2.2.4 Sample Preparation & IM-MS Measurements

Lyophilized 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.20,21 Using static nano electrospray ionization (nESI), ions

were generated from a 10uM solution in 200mM NH4Ac. Gold-coated borosilicated nESI emitters were used with a capillary voltage of 1400 V. Sulfur hexafluoride (SF6) 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 the Agilent tune mix and the Agilent Jet Stream ESI source.

2.2.5 CIU-ECD Experiments

The Agilent 6560 instrument modifications include the addition of a new capillary exit (CE) lens immediately after the ion transfer capillary.[20,21] 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.[21–23] 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 the 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 CIUSuite 2.[19] CIU fingerprints were obtained by plotting ATDs as a function of CE voltage in CIUSuite 2.

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 for ECD.[21,23] ECD fragmentation data was accumulated for 15 min for selected voltage values in a manner correlated with CIU data.

2.2.6 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.[24] 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 averaging method, monoisotopic masses were calculated using a custom Excel spreadsheet.[25] 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 were 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 has been observed in ECD fragmentation spectra previously.[15,26] 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.[27] Another characteristic of ECD is its ability to reduce disulfide bonds.[15,28,29,30] HSA contains 17 disulfide bonds along with two reduced cysteines.[31] 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.[29] 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 involved in a bringing pair were not within the same internal fragment, they were considered to be in their reduced forms or a known modified state. The known cysteine chemical modifications associated with disulfide bond reduction have been previously reported in the literature.[29,32]

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 if all isotopologues were precisely matched to a theoretical isotopic envelope for the intensity and m/z dimensions. 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.

2.2.7 Molecular Dynamics Simulations

Molecular dynamics (MD) simulations were performed using CHARMM and GROMACS on a workstation with 48 CPU core 2.10 Ghz Intel Xeon processor. The CHARMM36 and GROMOS 9643a2 force fields were employed for all MD simulations.[33] To generate input models of HSA for simulations, CHARMM-GUI[34] was used to import the crystal structure of HSA from the PDB (4K2C). Missing N-terminal residues were added. 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 at 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 charge placement, models of HSA D12 were subjected to energy minimization, which included 8 of the 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 utilized 20 replicas ranging from 300K to 800K over a 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 D12 in order to obtain compact models. For models of full-length HSA, the energy-minimized models were equilibrated at 300K over 4 ns simulation time using GROMOS 9643a2 force field on GROMACS. The structures were extracted from the equilibration trajectory which were filtered according to the experimental CCS observed for HSA to obtain the compact models of full-length HSA.

Following MD simulations, IMPACT[36] 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[37] 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 approximated TM CCS for all the structures.

In order to simulate the CIU experiment, we constructed a steered MD (SMD) approach for HSA D12 and full-length HSA, 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 3 ns of simulation time. Structures from the trajectory files were extracted and theoretical CCS was calculated followed by filtering based on experimental CCS values. CCS values recorded for HSA D12 monomer dissociated from the HSA D12 dimer were also used for filtering model structures.

For both the compact and unfolded modeling results, clustering was performed to better identify structural preferences. We employed a Kabsch[38] 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.[39] The lowest energy model of appropriate CCS from each structural ensemble was selected from the filtered models of SMD for docking with compact HSA 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

Previous studies have indicated that the CIU of low charge state multi-domain proteins is correlated with domain level structure.[18] Previous work from our group used HSA as a model

system to further study the domain-correlated unfolding of gas-phase protein ions.[4] Each HSA monomer is a single polypeptide chain consisting of 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.

Previous work has explored the CIU pathways of HSA dimers.[40] CIU of HSA dimers revealed both similarities and differences when compared with HSA monomers with a total of 5 CIU features across all charge states of HSA dimers, one more feature than expected (Appendix Figure I-1). 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 and developed computational methods aimed at probing the level of domain correlation and monomer involvement in the CIU processes observed for HSA dimers.

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 to 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 indicated 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[41,42] Indomethacin[43] was used as the marker for D2 and ANS[44] 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 2-1A) 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-

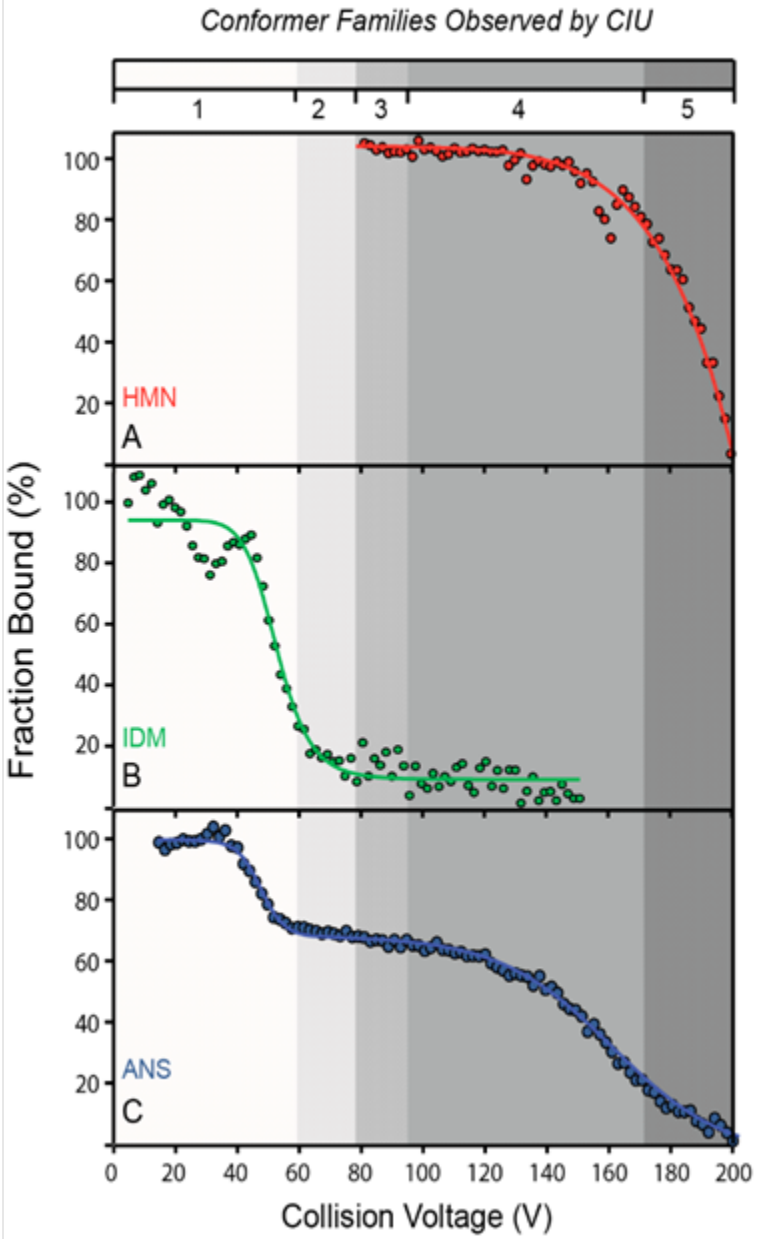


Figure 2-1. Correlation between HSA domain-specific chemical probes CID curves with 22+ HSA dimer CIU fingerprints. CID breakdown curves are fitted with sigmoid functions. Data sets are shown for hemin bound to HSA domain 1 **(A)**, indomethacin bound to HSA domain 2 **(B)**, and ANS bound to HSA domain 3 **(C)**. All plots are overlaid on top of a grey scale indicating the voltage ranges where HSA dimer CIU features are observed.

1B) during the initial phase of the HSA dimer CIU process at relatively 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 2-1C) with a first transition that falls between feature 1 and feature 2 and a second transition that mainly occurs at the end of the feature.[4] Therefore, based on the data shown in Figure 2-1, 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 next.

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 (Appendix Figure I-2), 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 (Figure 2-2A), 1 fewer than detected for the full-length HSA dimer. The number of features detected for domain-deleted 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 Figure 2-2B and Figure 2-2C. 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 2-2B). In contrast, Indomethacin is ejected from the complex primarily across features 2 and 3 (Figure 2-2C). Taken together, we surmised that both features 2 and 3 observed across full-length and domain-deleted HSA dimers are the result of D2 unfolding. Further work has been done using D12-D123 heterodimer and revealed that the unfolding event for this heterodimer is also driven by D12 unfolding.[40]

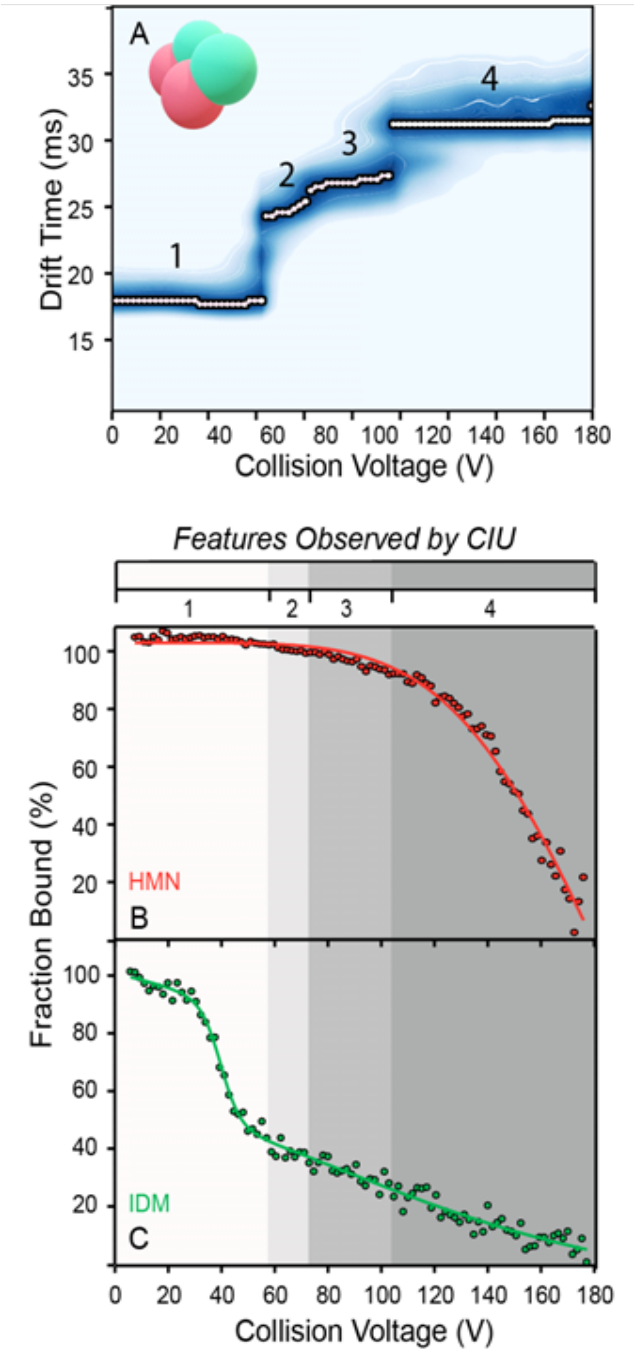


Figure 2-2. Correlation of HSA domain-specific chemical probes dissociation curve with D12 dimer CIU. (A) CIU fingerprints with annotated features for HSA D12 dimer construct. Illustration of CID breakdown curves for D12-ligand complexes. Data sets are shown for domain-specific chemical probe hemin bound 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.

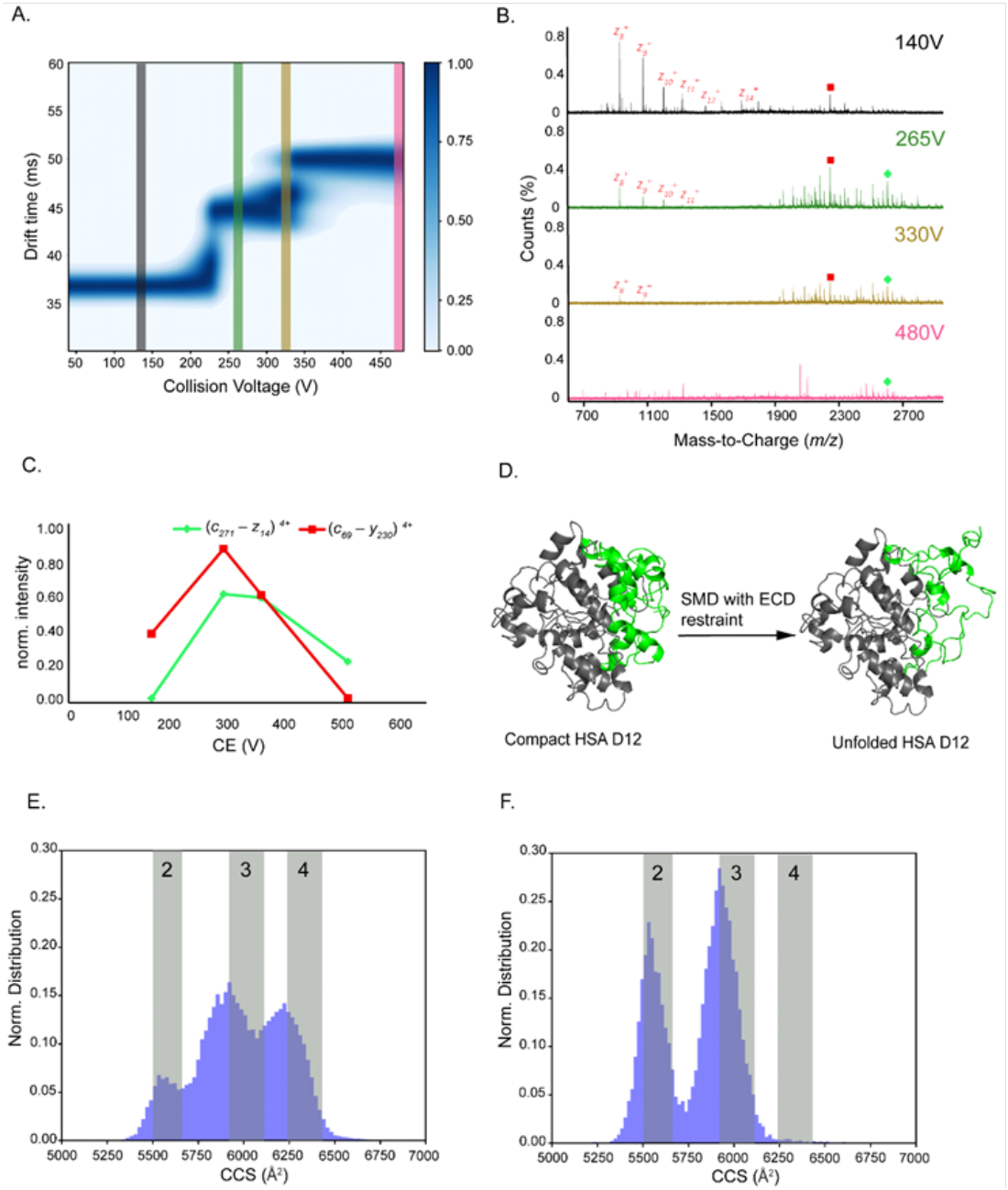


Figure 2-3. CIU-ECD of HSA D12. (A) A CIU fingerprint of the 13+ D12 monomer acquired on drift tube IM-MS is shown. D12 monomer CIU data exhibits four features. For each feature ECD fragmentation data was obtained, as indicated by the color-coded regions of drift time/collision voltage space (dark grey = 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 numbered gray boxes in (E) and (F) indicates the CCS ranges.

Prior reports have indicated that ECD fragment ion production can be linked directly to protein structure, whereby electron capture and bond cleavage are preferential at unstructured sites within a given protein sequence.[45–49] As such, we sought to combine ECD fragmentation with CIU in order to directly annotate the features observed within the latter data type. When transferred to our modified drift tube IM-MS instrument equipped with ECD, we observe four total CIU features for +13 D12 monomers (Figure 2-3A). 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 the quadrupole section of 13+ ions at the following levels of CIU activation: 140, 265, 330, and 480V (Figure 2-3A, 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 (Appendix Figure I-3). We further observed that some z-ions were identified along with their y-ions counterparts (Appendix Figure I-2B), acting as further confirmation of our identifications.[26] These y-ions interestingly were originating from the ECD process, as fragmentation was occurring after quadrupole selection precluding CID originating from the front of the instrument. 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 2-3B). 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 to the terminal fragments observed under these conditions, and the signals in this region were 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 voltage is increased to 480V, 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 (Appendix Figure I-4).

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 (Appendix Table I-1). One corresponded to an internal fragment within the D1 sequence (Figure 2-3B, red square) and the other to a D2 internal fragment (Figure 2-3B, 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 as a restraint in our steered MD simulations.

Performing MD simulations and generating models that best represent gas-phase protein structures remain challenging due to various factors. First, the current charge placement algorithm is built around globular, folded proteins which may not be suitable for simulating proteins that undergo unfolding in the gas-phase as the charges migrate.[37] Thus, to build a model that reflects the instrumental environment of IM-MS, a new charge placement algorithm is necessary to place charges across the entire sequence of the protein. While it is favorable to place charges as accurately as possible to mimic the experiment, there is no direct measurement or observable that could be used as restraints to place charges. Thus, we built a random charge placement algorithm

that places charges randomly and evenly across the entire HSA sequence to search for all possible charge isomers. To model the D12 monomer, the +12 charge state was considered as it is the experimentally observed charge state of HSA D12 dimer CID. Our data suggests that for the +12 charge state D12 monomer, there are almost 4000 possible charge isomers present that are randomly distributed across all 385 residues (Figure 2-4A). Charge isomers were randomly sampled to place charges on the HSA D12 monomer from the PDB database (4K2C) and performed compaction REMD simulations.

Following the simulations, all trajectories from REMD were accessed to extract structures followed by CCS calculations. To improve the throughput of the calculation without sacrificing the accuracy of the

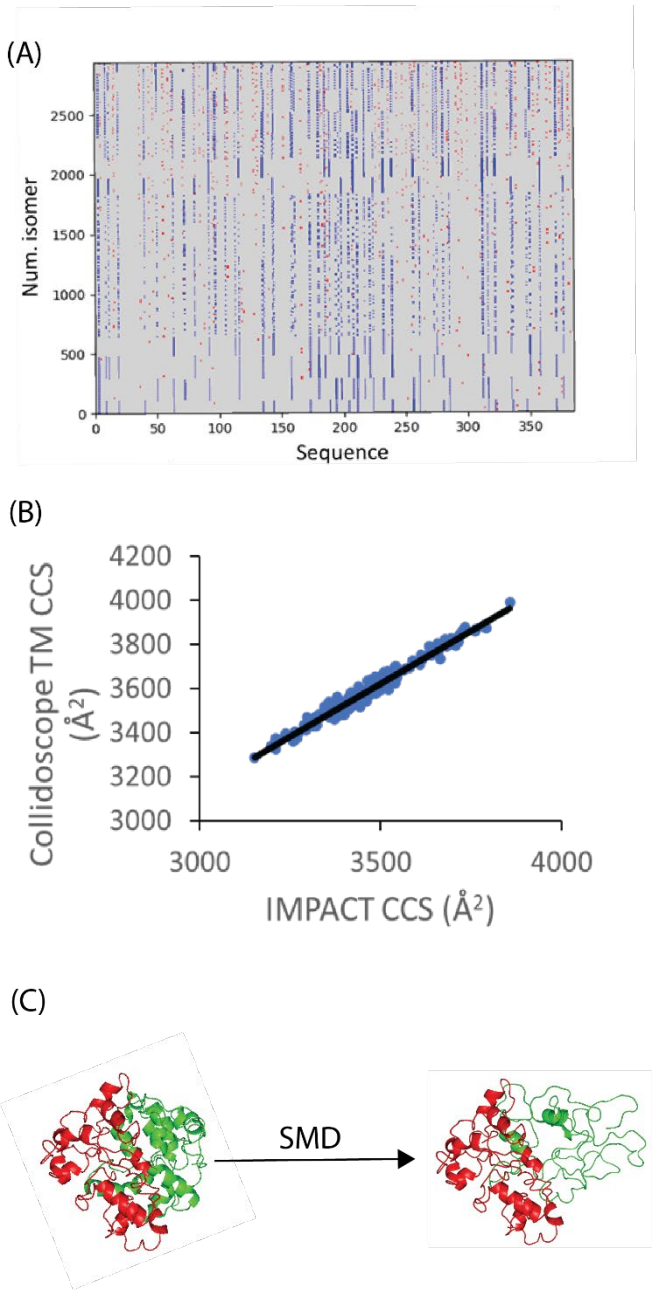


Figure 2-4. Computational Methods for Improved Modeling Accuracy. (A) Charge placement results of HSA D12 monomer at +12 net charge. Red indicates negative charge site, and blue indicates positive charge site. **(B)** CCS correlation plot between IMPACT and Collidoscope. The best fit line represents TM CCS =0.9591*(IMPACT CCS) + 264.47 with R^2 = 0.9681. **(C)** Example steered MD demonstrates unfolding of protein at specific residues. The input model (right) subjected to heating via SMD with residues in red fixed. The final model successfully kept red residues intact and only the green residues changed in structure.

calculations, correlation plot was generated, and all IMPACT CCS were corrected using the correlation equation from IMPACT vs Collidoscope TM CCS (Figure 2-4B). The models were then filtered based on experimentally observed CCS of the D12 monomer (Appendix Figure 1-5). All models that agreed with the experimentally observed CCS of compact D12 monomer were then clustered to separate models by their structural ensemble, and the lowest energy structure from the most populated cluster was chosen for the following SMD simulations (Appendix Figure 1-6).

It is important to note that CCS is considered ambiguous as a single CCS value can represent various protein shapes. With native IM-MS, CCS becomes a major experimental restraint, and due to the resolution achieved by IM, it is inherent that structures generated using CCS values alone will be ambiguous. With CIU-ECD, we obtained additional structural information such as domain-correlated unfolding and specific regions within the protein that are prone to unfolding at the residue level. However, there's no computational method currently available in the field to allow us to perform MD simulations that will unfold the protein of specific residues observed in ECD fragmentation. We developed an SMD simulation that allows users to specify the residues that the user wishes to unfold while maintaining remaining residues intact and constant throughout the simulations. By implementing this simulation approach, users can more accurately model the unfolding of specific regions that are experimentally relevant such as CIU-ECD (Figure 2-4C).

Following the computational generation of compact starting structures for D12 in a range of charge isomeric states, SMD simulations (Figure 2-3D) were performed using two restraint sets: 1) all of the 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. The ECD restraint

used in these simulations is preliminary fragment identification and does not match completely with the updated identification. 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 2-3).

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 2-3E), 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 2-3F). This remarkable level of correlation between experiment and theory, achieved by restraining models using two independent 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.

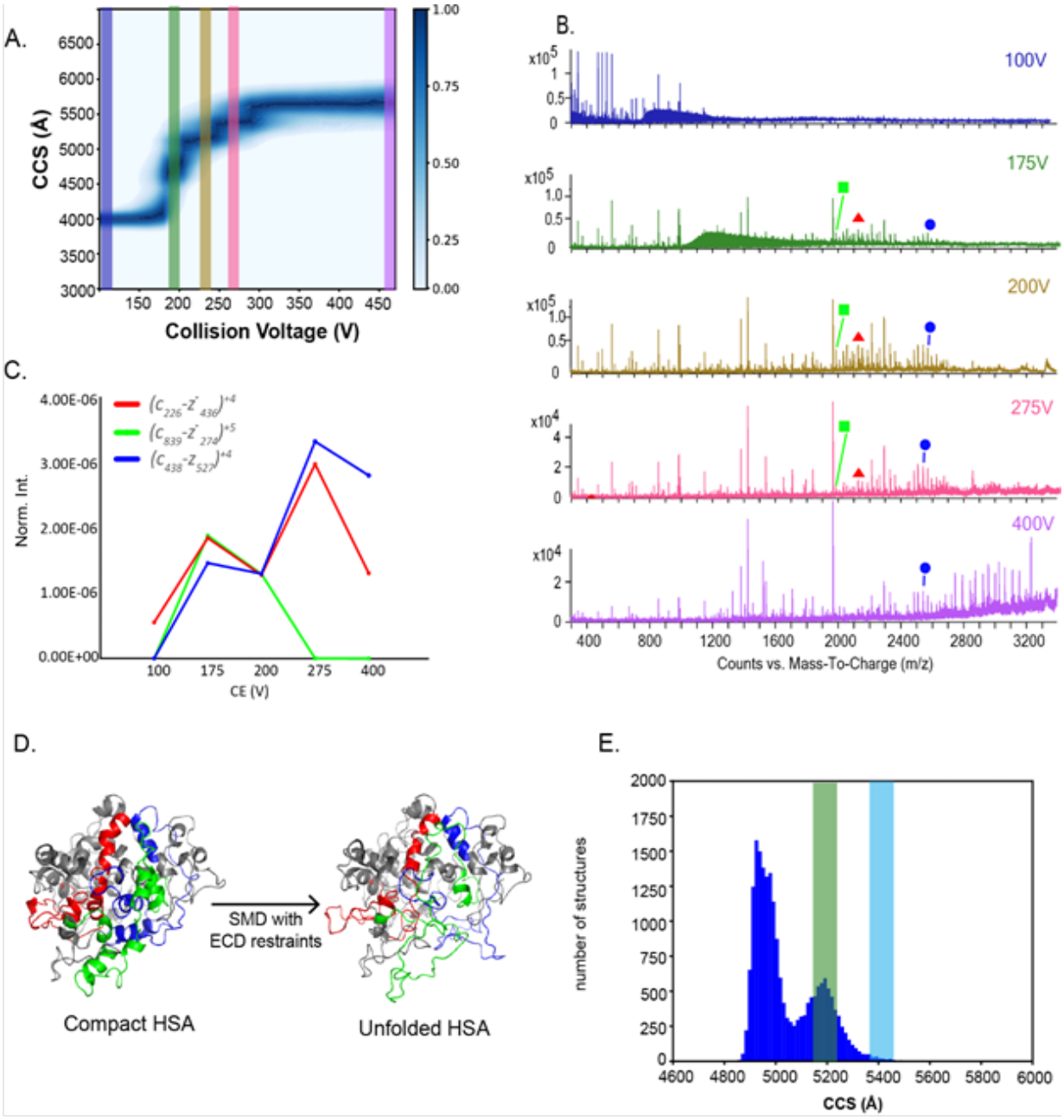


Figure 2-5. CIU-ECD of full-length HSA. (A) CIU Fingerprint of +17 full length HSA. The five colored boxes indicate where in the fingerprint ECD fragmentation was obtained for each of the five features observed. The +17-charge state is shown due to presenting higher fragmentation efficiency and a +5 charge state fragments, unlike +15 HSA fragmentation. For the +15 HSA fingerprint and corresponding spectra see Figure S7. **(B)** ECD Fragmentation spectrum taken at each voltage selected in A). Three internal peaks are indicated in each spectrum corresponding to domain 1 – domain 2 boundary (D1-D2, red triangle), domain 2 - domain 3 boundary (D2-D3, green square), and domain 3 (blue circle). These fragments were then subsequently used to guide SMD of full-length HSA due to their high matching scores (Table S2), and intensity profiles. **(C)** Normalized intensity for the three internal fragments indicated in B) at each collision voltage used in B). the Legend indicates the ions types that correspond to each internal fragment as well as the charge state. **(D)** Full-length HSA models of compact and ECD-guided unfolded structures. Residues of ECD fragments from C are highlighted accordingly in both models. **(E)** CCS distribution of SMD trajectory using all three internal fragments observed from C. With SMD, HSA undergoes unfolding up to the CCS range (blue box) that agrees with domain 2 and 3 unfolded CCS range.

CIU-ECD acquisition for full-length HSA is nearly identical to D12 experiments, except no quad-selection was used. Most of the charge states of HSA laid outside the quadrupole range (> m/z 4000) and transmission of the +17(m/z 3900) was not optimal. Therefore, all charge-state ECD was performed, which has been shown previously to increase ECD fragmentation efficiency.[47] Also, to ensure a sufficient signal-to-noise ratio to obtain good peak shapes for scoring, the signal was accumulated over one hour.

While the ECD fragmentation was acquired without quadrupole selection, fragmentation differences were observed between the two datasets acquired where the higher intensity precursor charge state differ. Four features were observed in the CIU fingerprint of HSA+15, as previously observed in the G2 dataset (Appendix Figure I-7A). When the +15 precursor was at the highest intensity peak, ECD fragmentation was obtained for each feature observed at 200V, 320V, 360V, and 450V (Appendix Figure I-7B). Only a peak was identified with a score higher than 70 (Appendix Figure I-7C). The identified peak was c174-z•436+4 (domain 1 and domain 2). More than half of the internal fragment lay on domain 1, and it followed the expected trend for domain 1: Unfolding towards the end of the fingerprint.

Meanwhile, this same internal fragment was identified along with the other two among fragment ions of HSA+17 data (Appendix Table I-2). The CIU fingerprint of HSA+17 has five features (Figure 2-5A). For each feature, ECD fragmentation was obtained (Figure 2-5B). The internal fragments now able to confidently be identified in HSA+17 are c336-z•274+5 (D2-D3 and c84-z527+4 (D3). The fragment c174-z•436+4 identified in HSA+15 (Appendix Figure I-7C), has a similar yet distinct intensity profile that in HSA+17 (Figure 2-5C). In the latter HSA precursor, the fragment (indicated in red) suddenly increases in intensity then decreases, and once again increases in intensity before finally decreasing. In Figure 2-5C, it is also shown that c336-z•274+5,

increases in intensity in the second feature, but eventually disappears (Figure 2-5C, green). And finally, fragment c84-z527+4 shows two cycles of increasing and decreasing in intensity similar to the D1-D2 fragment (Figure 2-5B, blue). From the three internal fragments identified and their intensity profiles in HSA+17, it can be concluded that domain 2 unfolds accompanied by some unfolding in D1 and D3, but further unfolding occurs in D1 and D3. In order to confirm the validity of our fragmentation assignments and the importance of these regions in the unfolding pathway of HSA, SMD was performed on full-length HSA.

Compact models for full-length HSA were generated using GROMOS9643a2 force field on GROMACS due to insufficient compaction achieved by CHARMM REMD (Appendix Figure I-8A. This specific force field was chosen based on a previous report, and with the full-length HSA, we observed a similar trend where GROMOS9643a2 was able to achieve compaction that is experimentally relevant that CHARMM36 cannot achieve (Appendix Figure I-8 B&C).[49] With GROMACS, 3714 compact models were generated for full-length HSA in an experimentally observed charge state. Similar to HSA D12 modeling, all filtered models were clustered to determine the most populated cluster along with the cluster that preserved most of its secondary structures (Appendix Figure I-9). Using the lowest energy model from these filtered compact HSA models, SMD simulations were performed (Figure 2-5D) sequentially using the preliminary ECD fragments observed in Figure 2-5C as follows: 1) restraint on residues 174-251 and 336-423; 2) restraint on residues 438-527 on top of the two ECD fragments used for restraints in step 1.

Following the SMD, all models were extracted from the SMD trajectory followed by CCS calculation and filtering. Using steered MD, CCS distributions of each MD trajectory were plotted to compare the specificity of ECD-guided steered MD for full-length HSA. With all ECD fragments as restraints, the unfolding of HSA stops with the CCS range observed for domain 2 and

3 unfolded states (Figure 2-5E). This agrees with the experiment as internal fragments used for steered MD correspond to regions in domains 2 and 3. However, once the compact HSA was allowed to unfold without any ECD restraint, models observed during the MD trajectory range past even the most unfolded state of HSA (Figure 2-5F). This again strongly suggests that two experimental data correlate strongly with theoretical data obtained using MD simulations.

To further test the specificity of ECD-guided steered MD simulations, a separate steered MD simulation was performed where instead of unfolding residues 438-527 in the second step of steered MD, we restrained the residues of the same length in a different region within the same domain. The effect of the location of the non-specificity in steered MD was explored as well by performing simulations on both when the restraint is internal or terminal. Between ECD-guided and non-specific internal restraint steered MD simulations, ECD-guided steered MD resulted in over three-fold more models within the target experimental CCS range compared to the steered MD with the non-specific internal fragment (Appendix Table I-3). This suggests that the efficiency of the unfolding event during steered MD is governed by the specificity of residues that are being restrained. Furthermore, when comparing ECD-guided steered MD results to non-specific terminal fragment steered MD results, it is evident that the non-specificity also results in unfolding that is beyond experimentally relevant as the terminal non-specific restraint resulted in models that further agree with the most unfolded state's feature of HSA CIU experiment (Appendix Table I-3). These results further corroborate the level of agreement between experiment and theory which was also observed with HSA D12 dimer data.

2.4 Conclusions

Here, we describe the most comprehensive mechanistic analysis of multiprotein CIU to date. By comparing the CIU fingerprints obtained for HSA monomers 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, given 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 Acknowledgments

This research was supported by R01GM095832 to the Ruotolo research group. C.R.R. was also supported by grant T32 CA140044 from the National Institutes of Health.

2.6 References

(1) Eschweiler, J. D.; Frank, A. T.; Ruotolo, B. T. Coming to Grips with Ambiguity: Ion Mobility-Mass Spectrometry for Protein Quaternary Structure Assignment. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (10), 1991–2000.

(2) Benesch, J. L. P. Collisional Activation of Protein Complexes: Picking Up the Pieces. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (3), 341–348.

(3) Hyung, S. J.; Robinson, C. V.; Ruotolo, B. T. Gas-Phase Unfolding and Disassembly Reveals Stability Differences in Ligand-Bound Multiprotein Complexes. *Chem. Biol.* **2009**, *16* (4), 382–390.

(4) Eschweiler, J. D.; Martini, R. M.; Ruotolo, B. T. Chemical Probes and Engineered Constructs Reveal a Detailed Unfolding Mechanism for a Solvent-Free Multidomain Protein. *J. Am. Chem. Soc.* **2017**, *139* (1), 534–540.

(5) Hopper, J. T. S.; Oldham, N. J. Collision Induced Unfolding of Protein Ions in the Gas Phase Studied by Ion Mobility-Mass Spectrometry: The Effect of Ligand Binding on Conformational Stability. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (10), 1851–1858.

(6) Ruotolo, B. T.; Hyung, S. J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V. Ion Mobility-Mass Spectrometry Reveals Long-Lived, Unfolded Intermediates in the Dissociation of Protein Complexes. *Angew. Chemie - Int. Ed.* **2007**, *46* (42), 8001–8004.

(7) Bornschein, R. E.; Ruotolo, B. T. Ion Mobility-Mass Spectrometry of Charge-Reduced Protein Complexes Reveals General Trends in the Collisional Ejection of Compact Subunits. *Analyst* **2015**, *140* (20), 7020–7029.

(8) Schwartz, B. L.; Bruce, J. E.; Anderson, G. A.; Hofstadler, S. A.; Rockwood, A. L.; Smith, R. D.; Chilkoti, A.; Stayton, P. S. Dissociation of Tetrameric Ions of Noncovalent Streptavidin Complexes Formed by Electrospray Ionization. *J. Am. Soc. Mass Spectrom.* **1995**, *6* (6), 459–465.

(9) Loo, J. A. Observation of Large Subunit Protein Complexes by Electrospray Ionization Mass Spectrometry. *J. Mass Spectrom.* **1995**, *30* (1), 180–183.

(10) Felitsyn, N.; Kitova, E. N.; Klassen, J. S. Thermal Decomposition of a Gaseous Multiprotein Complex Studied by Blackbody Infrared Radiative Dissociation. Investigating the Origin of the Asymmetric Dissociation Behavior. *Anal. Chem.* **2001**, *73* (19), 4647–4661.

(11) Benesch, J. L. P.; Aquilina, J. A.; Ruotolo, B. T.; Sobott, F.; Robinson, C. V. Tandem Mass Spectrometry Reveals the Quaternary Organization of Macromolecular Assemblies. *Chem. Biol.* **2006**, *13* (6), 597–605.

(12) Light-Wahl, K. J.; Schwartz, B. L.; Smith, R. D. Observation of the Noncovalent Quaternary Associations of Proteins by Electrospray Ionization Mass Spectrometry. *J. Am. Chem. Soc.* **1994**, *116* (12), 5271–5278.

(13) Jurchen, J. C.; Garcia, D. E.; Williams, E. R. Further Studies on the Origins of Asymmetric Charge Partitioning in Protein Homodimers. *J. Am. Soc. Mass Spectrom.* **2004**, *15* (10), 1408–1415.
(14) Jurchen, J. C.; Williams, E. R. Origin of Asymmetric Charge Partitioning in the Dissociation of Gas-Phase Protein Homodimers. *J. Am. Chem. Soc.* **2003**, *125* (9), 2817–2826.
(15) Lermyte, F.; Valkenborg, D.; Loo, J. A.; Sobott, F. Radical Solutions: Principles and Application of Electron-Based Dissociation in Mass Spectrometry-Based Analysis of Protein Structure. *Mass Spectrometry Reviews*. John Wiley & Sons, Ltd November 1, 2018, pp 750–771.
(16) Zhang, H.; Cui, W.; Gross, M. L. Native Electrospray Ionization and Electron-Capture Dissociation for Comparison of Protein Structure in Solution and the Gas Phase. *Int. J. Mass Spectrom.* **2013**, *354–355*, 288–291.
(17) Williams, J. P.; Morrison, L. J.; Brown, J. M.; Beckman, J. S.; Voinov, V. G.; Lermyte, F. Top-Down Characterization of Denatured Proteins and Native Protein Complexes Using Electron Capture Dissociation Implemented within a Modified Ion Mobility-Mass Spectrometer. *Anal. Chem.* **2020**, *92* (5), 3674–3681.
(18) Zhong, Y.; Han, L.; Ruotolo, B. T. Collisional and Coulombic Unfolding of Gas-Phase Proteins: High Correlation to Their Domain Structures in Solution. *Angew. Chemie* **2014**, *126* (35), 9363–9366.
(19) Polasky, D. A.; Dixit, S. M.; Fantin, S. M.; Ruotolo, B. T. CIUSuite 2: Next-Generation Software for the Analysis of Gas-Phase Protein Unfolding Data. *Anal. Chem.* **2019**, *91* (4), 3147–3155.
(20) Vallejo, D. D.; Polasky, D. A.; Kurulugama, R. T.; Eschweiler, J. D.; Fjeldsted, J. C.; Ruotolo, B. T. A Modified Drift Tube Ion Mobility-Mass Spectrometer for Charge-Multiplexed Collision-Induced Unfolding. *Anal. Chem.* **2019**, *91* (13), 8137–8146.
(21) Gadkari, V. V.; Ramírez, C. R.; Vallejo, D. D.; Kurulugama, R. T.; Fjeldsted, J. C.; Ruotolo, B. T. Enhanced Collision Induced Unfolding and Electron Capture Dissociation of Native-like Protein Ions. *Anal. Chem.* **2020**, *92* (23), 15489–15496.
(22) Fort, K. L.; Cramer, C. N.; Voinov, V. G.; Vasil'Ev, Y. V.; Lopez, N. I.; Beckman, J. S.; Heck, A. J. R. Exploring ECD on a Benchtop Q Exactive Orbitrap Mass Spectrometer. *J. Proteome Res.* **2018**, *17* (2), 926–933.
(23) Williams, J. P.; Morrison, L. J.; Brown, J. M.; Beckman, J. S.; Voinov, V. G.; Lermyte, F. Top-Down Characterization of Denatured Proteins and Native Protein Complexes Using Electron Capture Dissociation Implemented within a Modified Ion Mobility-Mass Spectrometer. *Anal. Chem.* **2020**, *92* (5), 3674–3681.
(24) Strohalm, M.; Kavan, D.; Novák, P.; Volný, M.; Havlíček, V. MMass 3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data. *Anal. Chem.* **2010**, *82* (11), 4648–4651.
(25) Senko, M. W.; Beu, S. C.; McLaffertycor, F. W. Determination of Monoisotopic Masses and Ion Populations for Large Biomolecules from Resolved Isotopic Distributions. *J. Am. Soc. Mass Spectrom.* **1995**, *6* (4), 229–233.
(26) Zhurov, K. O.; Fornelli, L.; Wodrich, M. D.; Laskay, Ü. A.; Tsybin, Y. O. Principles of Electron Capture and Transfer Dissociation Mass Spectrometry Applied to Peptide and Protein Structure Analysis. *Chem. Soc. Rev.* **2013**, *42* (12), 5014–5030.

(27) Li, X.; Lin, C.; Han, L.; Costello, C. E.; O'Connor, P. B. Charge Remote Fragmentation in Electron Capture and Electron Transfer Dissociation. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (4), 646–656.
(28) Mormann, M.; Eble, J.; Schwöppe, C.; Mesters, R. M.; Berdel, W. E.; Peter-Katalinić, J.; Pohlentz, G. Fragmentation of Intra-Peptide and Inter-Peptide Disulfide Bonds of Proteolytic Peptides by NanoESI Collision-Induced Dissociation. In *Analytical and Bioanalytical Chemistry*; Springer, 2008; Vol. 392, pp 831–838.
(29) Ganisl, B.; Breuker, K. Does Electron Capture Dissociation Cleave Protein Disulfide Bonds. *ChemistryOpen* **2012**, *1* (6), 260–268.
(30) Lee, Y.; Oh, H. Bin. Collisionally-Activated Dissociation of Peptides with a Disulfide Bond: Confirmation of the Mobile-Proton Model Based Explanation. *Mass Spectrom. Lett.* **2010**, *1* (1), 5–8.
(31) Sugio, S.; Kashima, A.; Mochizuki, S.; Noda, M.; Kobayashi, K. Crystal Structure of Human Serum Albumin at 2.5 Å Resolution. *Protein Eng.* **1999**, *12* (6), 439–446.
(32) Lee, M.; Lee, Y.; Kang, M.; Park, H.; Seong, Y.; Sung, B. J.; Moon, B.; Oh, H. Bin. Disulfide Bond Cleavage in TEMPO-Free Radical Initiated Peptide Sequencing Mass Spectrometry. *J. Mass Spectrom.* **2011**, *46* (8), 830–839.
(33) 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. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30* (10), 1545–1614.
(34) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29* (11), 1859–1865.
(35) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314* (1–2), 141–151.
(36) Marklund, E. G.; Degiacomi, M. T.; Robinson, C. V.; Baldwin, A. J.; Benesch, J. L. P. Collision Cross Sections for Structural Proteomics. *Structure* **2015**, *23* (4), 791–799.
(37) Ewing, S. A.; Donor, M. T.; Wilson, J. W.; Prell, J. S. Collidoscope: An Improved Tool for Computing Collisional Cross-Sections with the Trajectory Method. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (4), 587–596.
(38) Kabsch, W. A Solution for the Best Rotation to Relate Two Sets of Vectors. *Acta Crystallogr. Sect. A* **1976**, *32* (5), 922–923.
(39) Chaudhury, S.; Lyskov, S.; Gray, J. J. PyRosetta: A Script-Based Interface for Implementing Molecular Modeling Algorithms Using Rosetta. *Bioinformatics*. Oxford Academic March 1, 2010, pp 689–691.
(40) Zhao, C. *Development of Collision Induced Unfolding for the Characterization of Large Multiprotein Complexes*; 2021.
(41) Buchholz, T. J.; Geders, T. W.; Bartley, F. E.; Reynolds, K. A.; Smith, J. L.; Sherman, D. H. Structural Basis for Binding Specificity between Subclasses of Modular Polyketide Synthase Docking Domains. *ACS Chem. Biol.* **2009**, *4* (1), 41–52.
(42) Zunszain, P. A.; Ghuman, J.; Komatsu, T.; Tsuchida, E.; Curry, S. Crystal Structural Analysis of Human Serum Albumin Complexed with Hemin and Fatty Acid. *BMC Struct. Biol.* **2003**, *3* (1), 1–9.

(43) Metallo, S. J. Intrinsically Disordered Proteins Are Potential Drug Targets. *Current Opinion in Chemical Biology*. Elsevier Current Trends August 1, 2010, pp 481–488.
(44) Bagatolli, L. A. Two Distinguishable Fluorescent Modes of 1-Anilino-8-Naphthalenesulfonate Bound to Human Albumin. *J. Fluoresc.* **1996**, *6* (1), 33–40.
(45) Breuker, K.; Oh, H. Bin; Horn, D. M.; Cerda, B. A.; McLafferty, F. W. Detailed Unfolding and Folding of Gaseous Ubiquitin Ions Characterized by Electron Capture Dissociation. *J. Am. Chem. Soc.* **2002**, *124* (22), 6407–6420.
(46) Horn, D. M.; Breuker, K.; Frank, A. J.; McLafferty, F. W. Kinetic Intermediates in the Folding of Gaseous Protein Ions Characterized by Electron Capture Dissociation Mass Spectrometry. *J. Am. Chem. Soc.* **2001**, *123* (40), 9792–9799.
(47) Lermyte, F.; Sobott, F. Electron Transfer Dissociation Provides Higher-Order Structural Information of Native and Partially Unfolded Protein Complexes. *Proteomics* **2015**, *15* (16), 2813–2822.
(48) Zhang, H.; Cui, W.; Wen, J.; Blankenship, R. E.; Gross, M. L. Native Electrospray and Electron-Capture Dissociation FTICR Mass Spectrometry for Top-down Studies of Protein Assemblies. *Anal. Chem.* **2011**, *83* (14), 5598–5606.
(49) Zhou, M.; Liu, W.; Shaw, J. B. Charge Movement and Structural Changes in the Gas-Phase Unfolding of Multimeric Protein Complexes Captured by Native Top-Down Mass Spectrometry. *Anal. Chem.* **2020**, *92* (2), 1788–1795.
(50) Rolland, A. D.; Prell, J. S. Computational Insights into Compaction of Gas-Phase Protein and Protein Complex Ions in Native Ion Mobility-Mass Spectrometry. *TrAC - Trends in Analytical Chemistry*. Elsevier July 1, 2019, pp 282–291.

Chapter 3 Integrative Modeling of Individual Particle Electron Tomography Antibody Models Reveals Gas-Phase Structural Diversity and Memory Effects

3.1 Introduction

Biotherapeutics are a class of drugs comprised of myriad biological macromolecules, such as proteins and nucleic acids, used to treat various diseases such as cancer and arthritis.[1-3] Monoclonal antibodies (mAbs), or immunoglobulins (Ig), represent the largest and most active type of biotherapeutics in current pharmaceutical research due to their wide range of therapeutic benefits such as high affinity, efficacy, and reduced side effects.[4] However, mAb biotherapeutics are analytically challenging to structurally evaluate as they exist as a heterogeneous mixture of proteoforms due to various post-translational modifications[5] including disulfide bonds, glycosylation, oxidation, and aggregation states.[6-10]

Biotherapeutic mAbs constructed around the typical IgG framework consist of two heavy chains and two light chains with these chains forming two fragment antigen binding arms (Fab) from an N-terminus of heavy chain and light chain.[11] Such mAbs also contain a hinge region where the two heavy chains cross giving the protein its flexibility upon antigen binding. The fragment crystallin domain (Fc) is positioned just after this hinge region and consists of the two C-termini of the heavy chains.[12] The higher-order structures (HOSs) of mAbs dictate much of their antigen-binding characteristics and therapeutic efficacy. Furthermore, it remains challenging to characterize them at the speed necessary to keep pace with typical throughput needs in the pharmaceutical industry. For example, most entries in the Protein Data Bank (PDB), contain mAb fragments, with only three entries currently describing X-ray data for full-length antibodies.[13-15]

A common challenge associated with determining high-resolution structures for full-length mAbs stems from the flexibility of mAbs, with the origin of much of the dynamism in the system originating in the hinge region. Small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS), which generate lower-resolution structural data when compared with X-ray crystallography, have revealed significant fluctuations in IgG HOS primarily associated with Fab arm motion enabled by the hinge.[16,17] Relatively recently, individual-particle electron tomography (IPET) was utilized to obtain single particle images and 3D reconstruction to generate detailed 3D models of individual IgG1 microstates, providing our first detailed look at the ensemble of structures adopted by native mAbs.[18] Overall, 120 individual IgG1 models were generated, providing a framework to study mAb hinge region flexibility.[18]

Many of the technological limitations discussed above are alleviated through the use of native ion mobility-mass spectrometry (nIM-MS), which utilizes an electric field to separate ions according to their size and charge on the millisecond timescale prior to MS analysis, thus enabling the rapid evaluation of mAb conformations.[19] Along with nIM-MS, collision induced unfolding (CIU) experiments have been deployed to further investigate mAb HOS, enabling the resolution of mAb states produced upon glycosylation, alterations to disulfide bonding, drug conjugation, and the engineering of binding epitopes in a manner that exceeds the capabilities of IM separation alone.[20-23] The ability to convert IM drift times to collision cross sections (CCSs) permits direct comparisons between model structures and experimental IM data, and previous studies have revealed evidence of significant compaction of mAb structures in the gas phase, primarily in the flexible hinge region.[24,25] Despite multiple reports utilizing IM-MS to study mAb structure, a detailed evaluation of IM peak widths recorded for mAbs remains under-evaluated, specifically in a manner that quantitatively links native mAb structure ensembles directly to IM peak widths.[24] In

addition, there are various challenges involved in annotating mAb structural ensembles observed by IM: 1) the insufficient peak capacity of IM to fully resolve all mAb microstates; 2) the computational challenges involved in accurately modeling gas-phase mAb structures.[26,27] Due to these challenges, the relationship between native and gas-phase mAb structural ensembles remains opaque.

In this chapter, we deploy molecular dynamics (MD) simulations to model the gas-phase compaction of 120 IPET model IgG1 structures in order to evaluate the fidelity of IM distributions in capturing information regarding the ensemble of mAb microstates that exist in the solution. Some of the methods used in this chapter are adapted from those described in Chapter 2. Our approach uses IM data as restraints for our MD simulations to further enhance our understanding of native-like mAb ions and their compaction in the gas-phase to produce the ensemble of gas-phase conformers observed by IM. Our studies suggest that the structural ensembles adopted in solution by mAbs undergo significant hinge collapse to produce the gas-phase ensemble of mAb structures. Furthermore, our clustering analysis demonstrates that the CCS ordering of structures in solution is not retained in the gas-phase. Lastly, we examined the retention of secondary structural features in our mAb models in the gas-phase and observed evidence of strong memory effects, especially in helical and turn regions.

3.2 Methods

3.2.1 Sample Preparation and IM-MS Experiment

IgG1 monoclonal antibody was purchased from Sigma-Aldrich (St. Louis, MO). The antibody samples were prepared at 2 mg/mL in 200 mM ammonium acetate buffer at pH 7 to preserve their native structure. Prior to the IM-MS experiment, the samples were buffer exchanged into 200 mM ammonium acetate using Micro Bio-spin 30 columns (Bio-Rad, Hercules, CA).

Bovine serum albumin (BSA), alcohol dehydrogenase (ADH), glutamate dehydrogenase (GDH), and GroEL were prepared at 10 uM in 200 mM ammonium acetate and used as protein standards for the CCS calibration on the Waters Synapt G2 using the blend function. The Agilent tunemix was used for the 6560C single-field CCS calibration.

3.2.2 Ion Mobility-Mass Spectrometry

For the traveling wave IM-MS (TWIM-MS) experiment, Waters Synapt G2 HDMS (Milford, MA) was used. The samples were transferred to a gold-coated borosilicate capillary needle (prepared in-house), and ions were generated by direct infusion of a sample using a nanoelectrospray ionization (nESI) source. The capillary voltage was set at voltages of 1.0-1.5 kV with the sampling cone at 40V. The helium cell flow was set to 200 mL/min at 1.40 x 10^3 mbar. The trap ion guide was pressurized to 4.98 x 10^{-2} mbar of argon gas. The traveling wave IM separator was pressurized to 3.4 mbar. IM separation was performed with 30 V wave height and wave velocity of 600 m/s.

For the drift tube IM-MS (DTIM-MS) experiment, Agilent 6560C platform (Santa Clara, CA was used. The samples were introduced by direct infusion using the same method as the TWIM-MS experiment. Ions were generated using the Agilent nESI source using high purity N_2 at 1.5 mL/min at 25 ºC. The capillary voltage was set at 1.5-1.7 kV. The front funnel and trap funnel were operated at 4-4.50 and 3.8 Torr, respectively. The drift tube was operated in 3.95 Torr of nitrogen gas with the electric field at 18.5V/cm.

3.2.3 Computational Modeling

Solution-phase IgG1 structural models were supplied through the Berkeley Molecular Foundry and were generated from IPET experiments.[28] All 120 models generated from IPET were

subject to gas-phase compaction MD simulations. To generate gas-phase models of IgG1, charges were placed to equal net charge of +25 for all 120 models, and the location of charges was determined using the in-house charge placement algorithm (discussed in Chapter 2). Following the charge placement of IgG1, model energies were minimized using the steepest descent method. MD simulations were performed using the GROMOS9643a2 force field on GROMACS.[29] All models were subjected to *in vacuo* MD simulations where models were equilibrated at a constant temperature of 300K over 5 ns simulation time with 0.002 ps timestep. Both solution and gas-phase models' CCS values were calculated using IMPACT to obtain projected approximation (PA) CCS values.[30] Randomly sampled models were further subjected to CCS calculation using Collidoscope to obtain trajectory method (TM) CCS using N_2 gas.[31] A correlation plot of PA vs. TM CCS of these randomly sampled models was generated, and the best-fit regression line was used to correct all IMPACT PA CCS to approximate TM CCS. Following the CCS correction, models from all 120 MD trajectories were filtered based on experimentally observed CCS from IM-MS followed by root-mean-squared deviation (RMSD) clustering analysis.[32]

3.3 Results and Discussion

Figure 3-1A contains an overlay of all 120 IPET models and illustrates the structural diversity adopted by native IgG1 in solution. This structural diversity is driven by the flexible hinge region of IgG1 and is mainly observed through variations in the relative positions of the Fab regions of the structure. We computed the CCSs of all 120 IPET models using IMPACT. In addition, mAb structures were randomly sampled and submitted to Collidoscope to obtain TM N_2 CCS values in an effort to ensure CCS accuracy across less compact mAb structures sampled during MD. A correlation plot was then generated between IMPACT PA CCS and Collidoscope TM N_2 CCS to obtain the best-fit regression equation. This equation was then used to correct all

PA CCS into TM N_2 CCS in all downstream analyses (Appendix Figure II-1). For IPET mAb models of IgG1 analyzed prior to gas-phase MD, we compute an average CCS of 12315 ± 277 $Å^2$ with a CCS variance ranging from 11506 $Å^2$ to 13120 $Å^2$ (Figure 3-2B). To gain insight into the mAb structural features that contribute to such a wide range of CCSs observed, we inspected the smallest, median, and largest IgG1 IPET models, which are shown in Figure 3-1C-E. Despite the prominence of the Fab arm position in the overlayed IPET models shown in Figure 3-1A, our analysis reveals that the extension of the hinge region is a key contributing factor in the range of CCSs exhibited in native mAb structural ensembles (Figure 3-1A, C-E).

We collected IgG1 IM data from two instrument platforms, using TWIM and DTIM analyzers, for comparison with IPET models of solution-phase mAbs. Gas-phase IgG1 CCSs measured using TWIM-MS and DTIM-MS exhibited centroids of 7530 ± 87 $Å^2$ and 8150 ± 40 $Å^2$, respectively. However, the IM peak widths observed were significantly broader than the standard deviations reported for CCS centroids, with CCS widths of 1000 $Å^2$ and 1500 $Å^2$, respectively, at baseline (Figure 3-1F). This CCS peak width is comparable to the CCS width computed for IPET models prior to gas-phase compaction (1613 $Å^2$ base width) (Figure 3-1B). This quantitative alignment of the width of the CCS distributions computed across solution models and gas-phase measurements suggests a strong connection between the ensemble of structures probed in both experiments, and we note that these similarities remain if these width values are normalized to the centroid CCS value observed (e.g., CCS base width expressed as a percent of the centroid CCS), resulting in values of 13%, 13% and 18% for IPET models, TWIM and DTIM datasets respectively. Both IPET models and IM distribution mimic Gaussian distributions (Figure 3-2B, F), and TWIM-MS and DTIM-MS data exhibit evidence of significant gas-phase compaction of

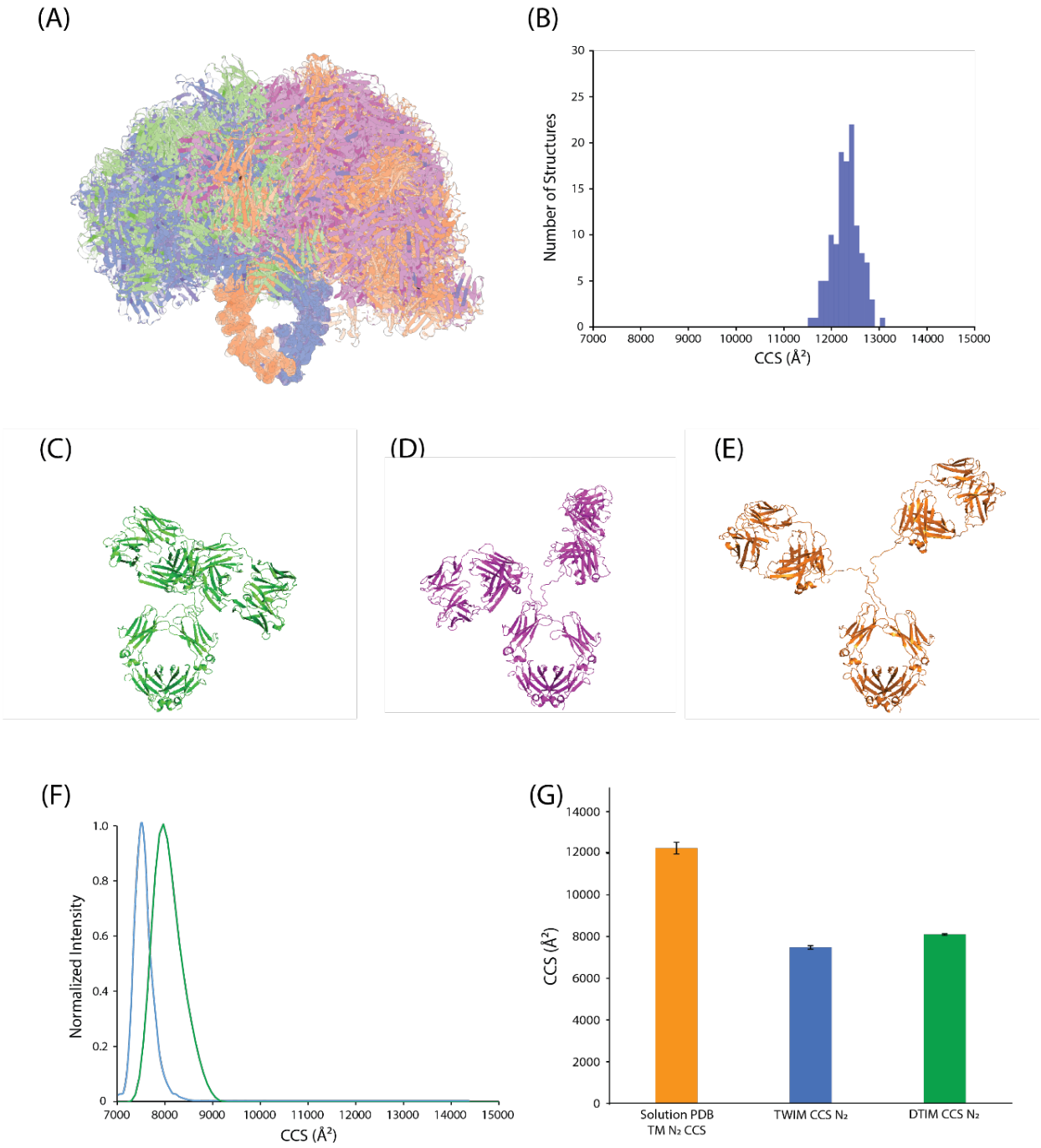


Figure 3-1. Solution-phase IgG1 models and CCS distributions from solution-phase models and IM-MS. **(A)** Overlay of all 120 IPET reconstructions. Blue and Orange represent heavy chains, and green and magenta represent light chains. **(B)** CCS distribution of 120 IPET models. **(C)** The smallest, **(D)** median, and **(E)** largest models from the IPET models. **(F)** CCS distribution from native IM-MS experiments with TWIM in blue and DTIM in Green. **(G)** Overall CCS comparison between IPET models and two IM-MS experiments show significant gas-phase compaction in IgG1.

IgG1, producing centroid CCS values that are up to 38.9% and 33.8% smaller than IPET CCS values respectively (Figure 3-1G).

To survey the structural diversity of gas-phase IgG1, all 120 IPET models were subjected to gas-phase compaction through MD simulations. First, we investigated the distribution and the size of the gas-phase structures with respect to the original IPET structures. Our final gas-phase models ranged from 7706 $Å^2$ to 9796 $Å^2$, thus exhibiting a wider range of CCSs (2090 $Å^2$) when compared to the original IPET distribution (Figure 3-2A). The average CCS of final compacted IgG1 models was 9099.63 ± 370.71 $Å^2$ showing 26.1% compaction, which is less than what is observed experimentally. To investigate the memory effects apparent within our gas-phase ensemble, we tracked the CCS values of IgG1 models between their origin in the starting IPET ensemble, and their final position in the computed gas-phase ensemble. Three example models are shown in Figure 3-1C-E and their position within each ensemble is highlighted. In general, the CCS ordering of IPET models is not retained in our final gas-phase distribution of structures. For instance, the smallest IgG1 structure in our original IPET ensemble does not remain the smallest structure in the final gas-phase distribution. Furthermore, the largest structure observed in the original IPET distribution compacted sufficiently to be among the smallest structures observed in our gas-phase ensemble (Figure 3-2A). All CCS values of 120 IgG1 structures in the solution and gas-phase are summarized in Appendix Table II-1.

An overlay of the final 120 compacted gas-phase IgG1 models illustrates a similar degree of variability in the Fab arms when compared to the original IPET models. The most apparent difference between the two ensembles is related to the hinge region of the IgG1 (Figure 3-2B). Across all IPET structures exhibiting extended hinge regions, we observed significant collapse in all models (Figure 3-2C-E). Furthermore, due to hinge collapse, the Fab domains in some structures fold over towards the Fc region in a manner similar to what is shown in Figure 3-2D. These observations are comparable to previous work on gas-phase mAb modeling.[24,33]

Despite the general conclusions drawn above, a significant number of the structures evaluated in our gas-phase mAb ensemble do not possess CCS values that agree with experimentally determined values. To fully reflect our IM experiments, TWIM-MS and DTIM-MS CCS values (7530 ± 87 Å^2 and 8150 ± 40 Å^2, respectively) were used as restraints to filter the models. Of the 120 IgG1 models, only 5 IPET models reach CCS values that are within the range of experimentally observed CCSs. Furthermore, our MD trajectories did not

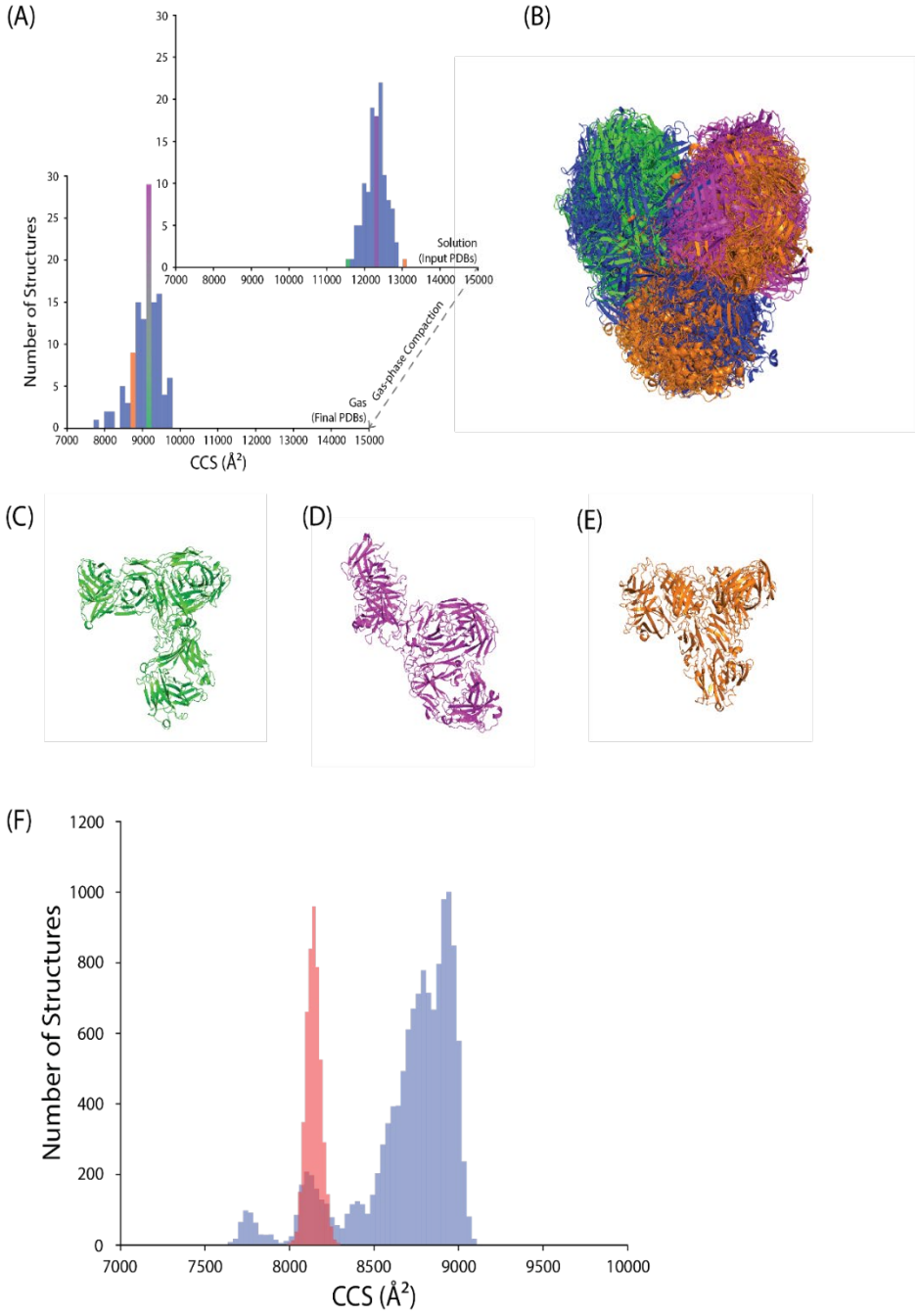


Figure 3-2. MD Simulations of Gas-phase Compaction. (A) CCS distribution change from solution to gas-phase using final structures from 120 MD simulations. The green, magenta, and orange bars represent the CCS value at which models from Figure 3-1 C-E are found in each distribution. **(B)** Overlay of all 120 final structures from MD simulations. **(C)** The smallest, **(D)** median, and **(E)** largest models from final structures from 120 MD simulations. **(F)** CCS distribution of IgG1 models that are filtered based on IM restraints. Red distribution is filtered based on DTIM CCS centroid and CCS errors. Blue distribution is filtered using the start and the end of the baseline of TWIM and DTIM CCS.

reach the level of compaction observed in TWIM-MS data. Using these narrow CCS restraints we arrive at a monomodal Gaussian distribution consistent with our original IPET models and IM data (Figure 3-2F). Clearly, centroid CCS values are insufficient to restrain our efforts to model mAb structural ensembles, so we moved to further filter our models to encompass all CCSs within the experimentally observed IM data. With these altered restraints, 39 out of 120 IPET models reached CCS consistent with experimental IM data, resulting in a total of 131,964 structures with CCS values within the experimental range. Furthermore, multiple Gaussian distributions are observed when we restrain our gas-phase models using this wider range of CCS values, suggesting a variety of structural ensembles that weren't directly observable in our IM experiments alone (Figure 3-2F).

To further evaluate the memory effect between solution and gas-phase models, we performed a clustering analysis to sort each model based on its structural similarities. For 120 starting IPET models, three clusters were identified with cluster 3 being the most populated. It is important to note that the clustering of these models was challenging due to the flexibility of IgG1 structures, necessitating a high cutoff linkage distance of 75 Å (Figure 3-3A). Each model was sorted based on the clustering results, and the largest discrepancy between the identified clusters arose from the angles between the Fab arms (Appendix Figure II-2A). The same clustering analysis was performed with 120 gas-phase compacted models. Unlike solution-phase models, gas-phase compacted models resulted in 4 total clusters with cluster 1 being the most populated (Figure 3-3B). As discussed above, CCS ordering in the original IPET population is not conserved in the gas-phase ensemble, and a similar lack of connection is observed across our IPET and gas-phase clusters. The strongest correlations observed between IPET and gas-phase clusters were found in the positions of the Fab arms, the relative angles for which proved most differentiating in both

structural ensembles. In contrast to the IPET models, however, hinge collapse contributes strongly to structural diversity in our gas-phase ensemble (Appendix Figure II-2B). We then performed clustering analysis using IM-filtered models to better reflect the experimental data. These IM-filtered models have significantly lower cutoff linkage distance within clusters, below 5 Å (significantly different than the 75-80 Å determined for IPET and unrestrained gas-phase models) demonstrating the high structural relatedness within each cluster following IM filtering. Furthermore, all models within each cluster were found to be originating from a single MD trajectory, suggesting that input solution-phase models strongly influence the ability to access sufficiently compact gas-phase structures (Appendix Table II-2). Models from each cluster exhibit highly globular structures, with either one or both Fab arms contacting the Fc region (Appendix Figure II-2C).

Lastly, we compared the domain structures observed between IPET and IM-filtered gas-phase structures to further evaluate potential memory effects present in the gas-

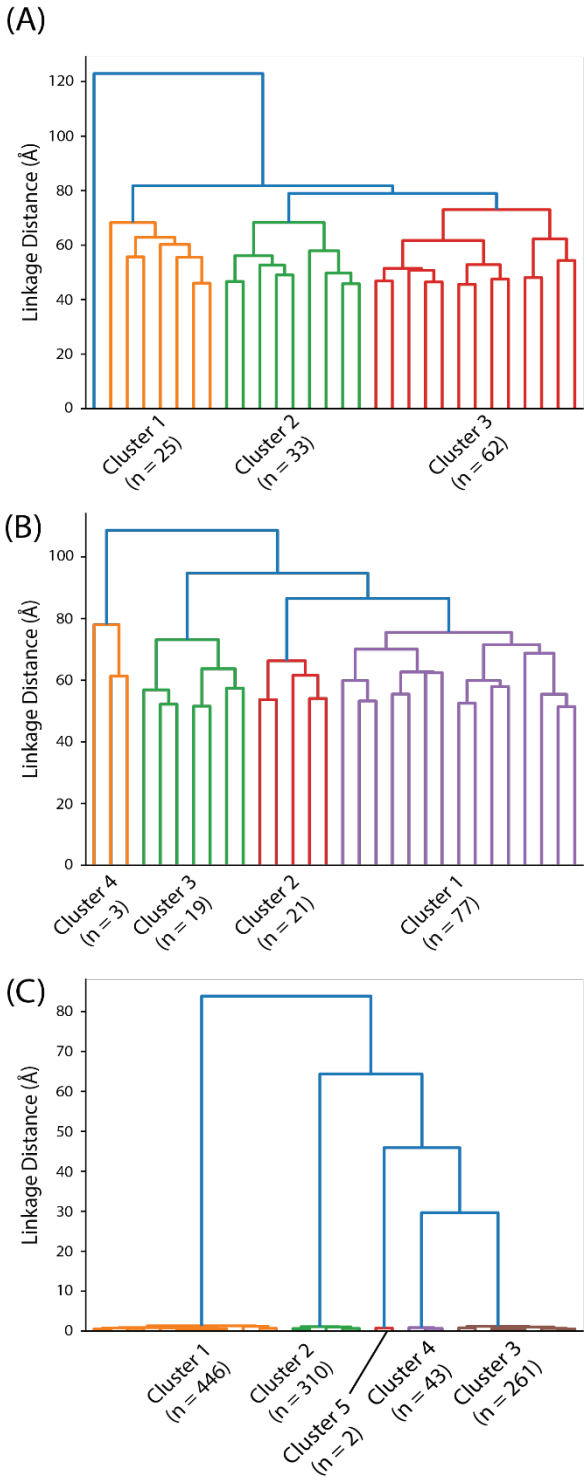


Figure 3-3. RMSD-based Clustering Analysis. Each dendrogram represents clustering results from **(A)** 120 IPET models **(B)** final structures from 120 MD simulations, and **(C)** IM-filtered gas-phase models.

phase structure ensemble. As discussed above, we observe significant compaction and rearrangements of IPET models on the domain level when comparing IPET and gas-phase models. For example, each Fab arm and Fc region are composed of two domains, where each Fab arm has a variable domain and constant domain; and the Fc region has two constant domains.[5] In the IPET ensemble, regardless of their size, Fab arms and Fc regions all have gaps between the domains that constitute these regions (Figure 3-1C-E). However, most of these gaps decrease in size due to gas-phase compaction during our MD simulations (Figure 3-4A).

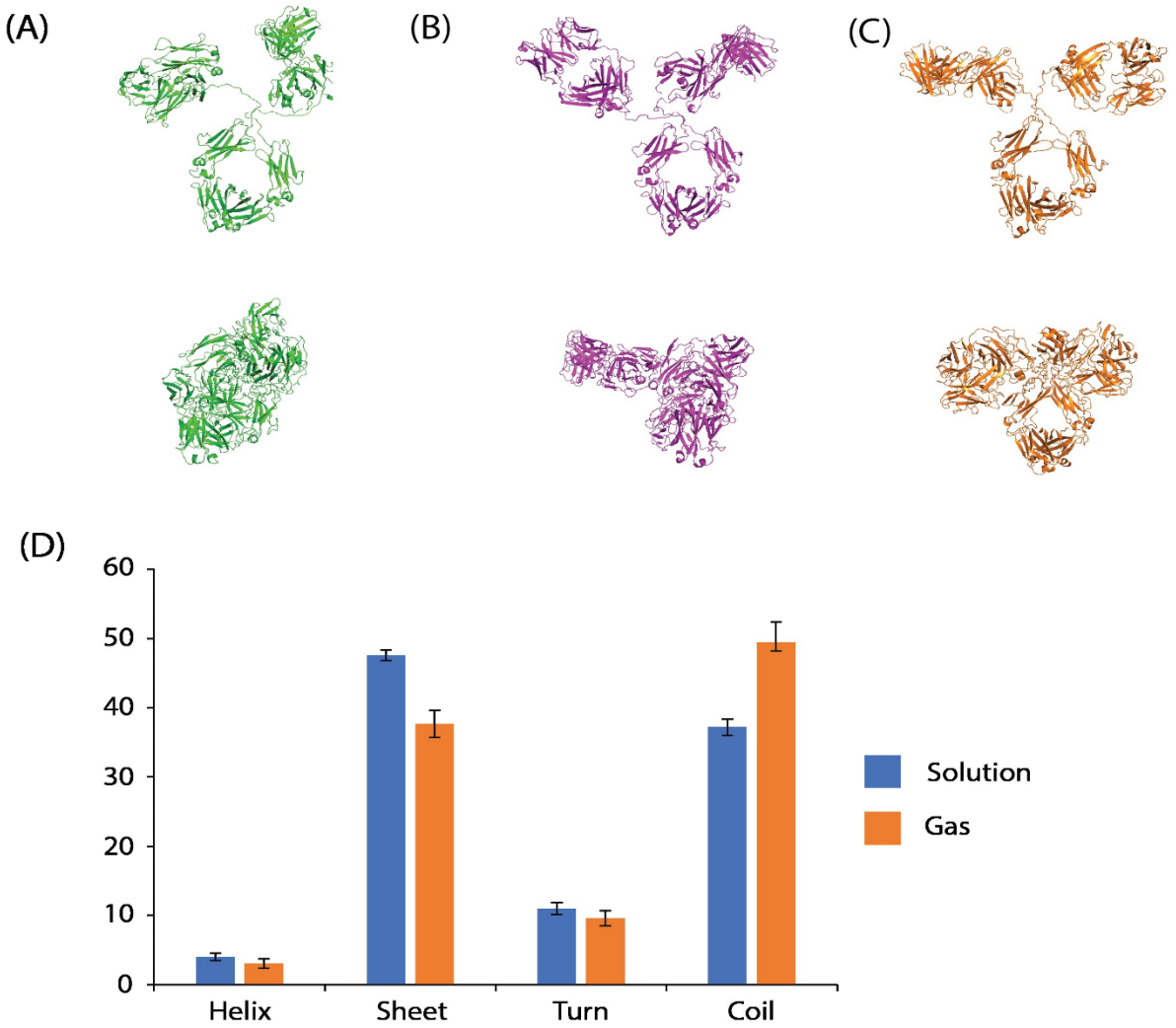


Figure 3-4. Detailed Overview of Changes in Structures upon Gas-phase Compaction. (A) The smallest, **(B)** median, and **(C)** largest IM-filtered gas-phase models of IgG1 (bottom) and the original solution-phase structures (top). **(D)** Changes in secondary structures from solution to gas-phase.

Following our domain-level analysis, we moved to evaluate similarities in the secondary structure elements between the IPET and gas-phase ensembles produced by MD. Overall, we observe strong retention of secondary structural features between both IPET and gas-phase structures. For example, the beta-sheets that compose each domain in the Fc and Fab regions are still intact in our IM-filtered models (Figure 3-4A-C). We performed additional structural analyses to quantify changes in secondary structural features in gas-phase mAbs. In IPET models, beta sheets and coils comprise 84.6% of the total structures present. Similar amounts of both secondary structure elements are observed in our gas-phase ensemble; however, we observe the percentage of beta sheets to be decreased while coils appear increased for all IM-filtered models (Figure 3-4D). The alpha-helices and coils appear to be most strongly retained in our gas-phase structure ensemble when compared to IPET models (Appendix Figure II-3). It is important to note that the values reported in Figure 3-4 represent average values over all structures. In our dataset, there are individual structures that do not comport the general trends noted. For instance, some models saw an increase in alpha helices up to 36.1%, while some saw a decrease of up to 62% (Appendix Table II-3). Overall, based on all the structural analyses of IM-filtered gas-phase models, gas-phase models retain a strong memory of mAb secondary structural elements while experiencing simultaneous rearrangements in their domain structures, driven primarily through hinge compaction.

3.4 Conclusions

In this report, we have compared IPET models of IgG1 flexibility in solution with both IM experimental data and gas-phase structures produced through MD simulation and compaction of the aforementioned IPET models. Our results mirror prior MD and IM results for IgG1 revealing that mAbs compress by ~30% in the gas-phase, although the level of compaction observed

experimentally differs between TWIM and DTIM datasets, an observation that is likely due to the different positions of the IM analyzers relative to the ion source used in both instrument platforms, resulting in disparate ion lifetimes and temperatures. In general, the diversity of structures observed in IPET appears quantitatively similar to what is observed in IM data, as reported by the IM peak widths observed during the experiment.

Our MD results indicate that the CCS ordering of mAb structures in solution is not conserved in the final gas-phase structural ensemble. Furthermore, the diversity of structures observed in the gas-phase appears to be primarily a product of differential hinge compaction and resulting Fab arm arrangements that also do appear to be correlated with the ensemble of mAb structures in solution. Despite this, local mAb structural elements, including secondary structures, remain well conserved with the greatest degree of change observed in beta-sheets, and the highest level of retention observed in helices and coils. Overall, this work is the first to quantitatively assess the memory effects and resulting gas-phase ensemble of structures probed in IM-MS experiments targeting mAbs, and points toward both future utility and challenges in using IM peak widths to quantify mAb structural diversity in solution.

3.5 Acknowledgments

We'd like to thank the Ren group at the Berkeley Molecular Foundry for providing IPET reconstruction of IgG1 models and Daniel D. Vallejo for providing IM-MS data.

3.6 References

(1) Rader, R. A. (Re)Defining Biopharmaceutical. *Nature Biotechnology*. Nature Publishing Group July 2008, pp 743–751.
(2) Sliwkowski, M. X.; Mellman, I. Antibody Therapeutics in Cancer. *Science*. American Association for the Advancement of Science September 13, 2013, pp 1192–1198.
(3) Ecker, D. M.; Jones, S. D.; Levine, H. L. The Therapeutic Monoclonal Antibody Market. *mAbs*. Taylor & Francis January 1, 2015, pp 9–14.

(4) Beck, A.; Wurch, T.; Bailly, C.; Corvaia, N. Strategies and Challenges for the Next Generation of Therapeutic Antibodies. *Nature Reviews Immunology*. Nature Publishing Group May 2010, pp 345–352.
(5) Wang, W.; Singh, S.; Zeng, D. L.; King, K.; Nema, S. Antibody Structure(1) Wang, W.; Singh, S.; Zeng, D. L.; King, K.; Nema, S. Antibody Structure, Instability, and Formulation. Journal of Pharmaceutical Sciences. Elsevier January 1, 2007, Pp 1–26. Https://Doi.Org/10.1002/Jps.20727., Instability, and Formul. *Journal of Pharmaceutical Sciences*. Elsevier January 1, 2007, pp 1–26.
(6) Wypych, J.; Li, M.; Guo, A.; Zhang, Z.; Martinez, T.; Allen, M. J.; Fodor, S.; Kelner, D. N.; Flynn, G. C.; Liu, Y. D.; Bondarenko, P. V.; Ricci, M. S.; Dillon, T. M.; Balland, A. Human IgG2 Antibodies Display Disulfide-Mediated Structural Isoforms. *J. Biol. Chem.* **2008**, *283* (23), 16194–16205.
(7) Dillon, T. M.; Ricci, M. S.; Vezina, C.; Flynn, G. C.; Liu, Y. D.; Rehder, D. S.; Plant, M.; Henkle, B.; Li, Y.; Deechongkit, S.; Varnum, B.; Wypych, J.; Balland, A.; Bondarenko, P. V. Structural and Functional Characterization of Disulfide Isoforms of the Human IgG2 Subclass. *J. Biol. Chem.* **2008**, *283* (23), 16206–16215.
(8) Jefferis, R. Glycosylation of Recombinant Antibody Therapeutics. *Biotechnology Progress*. American Chemical Society (ACS) January 1, 2005, pp 11–16.
(9) Bertolotti-Ciarlet, A.; Wang, W.; Lownes, R.; Pristatsky, P.; Fang, Y.; McKelvey, T.; Li, Y.; Li, Y.; Drummond, J.; Prueksaritanont, T.; Vlasak, J. Impact of Methionine Oxidation on the Binding of Human IgG1 to FcRn and Fcγ Receptors. *Mol. Immunol.* **2009**, *46* (8–9), 1878–1882.
(10) Vázquez-Rey, M.; Lang, D. A. Aggregates in Monoclonal Antibody Manufacturing Processes. *Biotechnology and Bioengineering*. John Wiley & Sons, Ltd July 1, 2011, pp 1494–1508.
(11) Gall, W. E.; Edelman, G. M. The Covalent Structure of a Human YG-Immunoglobulin. X. Intrachain Disulfide Bonds. *Biochemistry* **1970**, *9* (16), 3188–3196.
(12) Jay, J. W.; Bray, B.; Qi, Y.; Igbinigie, E.; Wu, H.; Li, J.; Ren, G. IgG Antibody 3D Structures and Dynamics. *Antibodies*. 2018.
(13) Harris, L. J.; Larson, S. B.; Hasel, K. W.; McPherson, A. Refined Structure of an Intact IgG2a Monoclonal Antibody. *Biochemistry* **1997**, *36* (7), 1581–1597.
(14) Harris, L. J.; Skaletsky, E.; McPherson, A. Crystallographic Structure of an Intact IgG1 Monoclonal Antibody. *J. Mol. Biol.* **1998**, *275* (5), 861–872.
(15) Saphire, E. O.; Parren, P. W. H. I.; Pantophlet, R.; Zwick, M. B.; Morris, G. M.; Rudd, P. M.; Dwek, R. A.; Stanfield, R. L.; Burton, D. R.; Wilson, I. A. Crystal Structure of a Neutralizing Human IgG against HIV-1: A Template for Vaccine Design. *Science (80-.).* **2001**, *293* (5532), 1155–1159.
(16) Rayner, L. E.; Kadkhodayi-Kholghi, N.; Heenan, R. K.; Gor, J.; Dalby, P. A.; Perkins, S. J. The Solution Structure of Rabbit Igg Accounts for Its Interactions with the Fc Receptor and Complement C1q and Its Conformational Stability. *J. Mol. Biol.* **2013**, *425* (3), 506–523.
(17) Sun, Z.; Almogren, A.; Furtado, P. B.; Chowdhury, B.; Kerr, M. A.; Perkins, S. J. Semi-Extended Solution Structure of Human Myeloma Immunoglobulin D Determined by Constrained X-Ray Scattering. *J. Mol. Biol.* **2005**, *353* (1), 155–173.
(18) Tong, H.; Zhang, L.; Kaspar, A.; Rames, M. J.; Huang, L.; Woodnutt, G.; Ren, G. Peptide-Conjugation Induced Conformational Changes in Human IgG1 Observed by Optimized Negative-Staining and Individual-Particle Electron Tomography. *Sci. Rep.* **2013**, *3*.

(19) Terral, G.; Beck, A.; Cianférani, S. Insights from Native Mass Spectrometry and Ion Mobility-Mass Spectrometry for Antibody and Antibody-Based Product Characterization. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2016**, *1032*, 79–90.
(20) Tian, Y.; Han, L.; Buckner, A. C.; Ruotolo, B. T. Collision Induced Unfolding of Intact Antibodies: Rapid Characterization of Disulfide Bonding Patterns, Glycosylation, and Structures. *Anal. Chem.* **2015**, *87* (22), 11509–11515.
(21) Tian, Y.; Ruotolo, B. T. Collision Induced Unfolding Detects Subtle Differences in Intact Antibody Glycoforms and Associated Fragments. *Int. J. Mass Spectrom.* **2018**, *425*, 1–9.
(22) Botzanowski, T.; Erb, S.; Hernandez-Alba, O.; Ehkirch, A.; Colas, O.; Wagner-Rousset, E.; Rabuka, D.; Beck, A.; Drake, P. M.; Cianférani, S. Insights from Native Mass Spectrometry Approaches for Top- and Middle- Level Characterization of Site-Specific Antibody-Drug Conjugates. *MAbs* **2017**, *9* (5), 801–811.
(23) Huang, Y.; Salinas, N. D.; Chen, E.; Tolia, N. H.; Gross, M. L. Native Mass Spectrometry, Ion Mobility, and Collision-Induced Unfolding Categorize Malaria Antigen/Antibody Binding. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (11), 2515–2518.
(24) Pacholarz, K. J.; Porrini, M.; Garlish, R. A.; Burnley, R. J.; Taylor, R. J.; Henry, A. J.; Barran, P. E. Dynamics of Intact Immunoglobulin G Explored by Drift-Tube Ion-Mobility Mass Spectrometry and Molecular Modeling. *Angew. Chemie - Int. Ed.* **2014**, *53* (30), 7765–7769.
(25) Vallejo, D. D.; Jeon, C. K.; Parson, K. F.; Herderschee, H. R.; Eschweiler, J. D.; Filoti, D. I.; Ruotolo, B. T. Ion Mobility-Mass Spectrometry Reveals the Structures and Stabilities of Biotherapeutic Antibody Aggregates. *Anal. Chem.* **2022**.
(26) Ruotolo, B. T. Collision Cross Sections for Native Proteomics: Challenges and Opportunities. *Journal of Proteome Research*. American Chemical Society January 7, 2022, pp 2–8.
(27) Allison, T. M.; Barran, P.; Cianférani, S.; Degiacomi, M. T.; Gabelica, V.; Grandori, R.; Marklund, E. G.; Menneteau, T.; Migas, L. G.; Politis, A.; Sharon, M.; Sobott, F.; Thalassinos, K.; Benesch, J. L. P. Computational Strategies and Challenges for Using Native Ion Mobility Mass Spectrometry in Biophysics and Structural Biology. *Analytical Chemistry*. American Chemical Society August 18, 2020, pp 10872–10880.
(28) Zhang, X.; Zhang, L.; Tong, H.; Peng, B.; Rames, M. J.; Zhang, S.; Ren, G. 3D Structural Fluctuation of IgG1 Antibody Revealed by Individual Particle Electron Tomography. *Sci. Rep.* **2015**, *5* (1), 1–14.
(29) Rolland, A. D.; Prell, J. S. Computational Insights into Compaction of Gas-Phase Protein and Protein Complex Ions in Native Ion Mobility-Mass Spectrometry. *TrAC - Trends in Analytical Chemistry*. Elsevier July 1, 2019, pp 282–291. https://doi.org/10.1016/j.trac.2019.04.023.
(30) Marklund, E. G.; Degiacomi, M. T.; Robinson, C. V.; Baldwin, A. J.; Benesch, J. L. P. Collision Cross Sections for Structural Proteomics. *Structure* **2015**, *23* (4), 791–799.
(31) Ewing, S. A.; Donor, M. T.; Wilson, J. W.; Prell, J. S. Collidoscope: An Improved Tool for Computing Collisional Cross-Sections with the Trajectory Method. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (4), 587–596.
(32) Kabsch, W. A Solution for the Best Rotation to Relate Two Sets of Vectors. *Acta Crystallogr. Sect. A* **1976**, *32* (5), 922–923.

(33) Hansen, K.; Lau, A. M.; Giles, K.; McDonnell, J. M.; Struwe, W. B.; Sutton, B. J.; Politis, A. A Mass-Spectrometry-Based Modelling Workflow for Accurate Prediction of IgG Antibody Conformations in the Gas Phase. *Angew. Chemie* **2018**, *130* (52), 17440–17445.

Chapter 4 CIUSuite 3: Next-Generation CCS Calibration and Automated Data Analysis Tool for Gas-Phase Protein Unfolding

4.1 Introduction

Native mass spectrometry (MS) measurements are of growing importance in structural biology for their ability to rapidly characterize not only proteoforms but also their noncovalent protein-protein and protein-ligand contacts.[1,2] Coupled with ion mobility (IM), native ion mobility-mass spectrometry (IM-MS) can add protein shape and size information in the context of heterogeneous samples for native MS workflows.[3,4] Recently, native MS and IM-MS techniques have seen significant growth in applications associated with biotherapeutic characterization[5], drug discovery and ligand screening[6], and membrane proteins[7] providing both valuable structural insights and integrative opportunities in combination with other structural biology techniques. Despite the value of native IM-MS assays, a key challenge remains the limited resolution of protein IM instruments. However, when protein ions are collisionally heated, unfolded states for protein ions can be produced that can act to differentiate subtle structural differences in precursor ion populations that were not detectable when using native IM-MS alone.[8,9] This method, termed collision induced unfolding (CIU), has now become a widely used technique for native IM-MS studies of protein conformation and stability.

Recent reports have illustrated the application of CIU to a wide array of applications in the pharmaceutical sciences, leveraging the high-throughput nature of the experiment. For example, CIU has been applied to evaluate protein-ligand binding, a variety of protein structural isoforms,

the structural consequences of post-translational modifications, the higher-order structure of biotherapeutics, and the structures of membrane proteins.[10-17] Specifically, CIU has been used to probe and assess the domain-level unfolding patterns produced by protein ions, yielding a variety of structural insights that remain obscured in standard native IM-MS data.[16,18-20] Many studies have leveraged CIU data to reveal important information regarding the stability, cooperativity, binding location, and allostery of ligand binding within proteins and multiprotein complexes.[9,21-25] Similar studies targeting biotherapeutics have characterized various antibody isoforms, glycosylation states, antibody-drug conjugates, binding epitopes, and biosimilars using CIU.[26-30] Lastly, CIU uncovered important structural consequences associated with membrane protein variants and lipid-bound states across a wide array of membrane mimetics.[7,31-35] Overall, CIU continues to provide critical data across a range of samples touching a wide array of applications, primarily associated with biosciences.

Given the wide adoption of CIU methods, several software packages have been developed to process and analyze CIU data. Indeed, many of the studies mentioned above utilize software packages to process CIU data, extract quantitative values, and accommodate the high-throughput nature of modern CIU experiments.[36-40] In particular, CIUSuite 2 has served to automate various analytical tasks associated with CIU data processing.[37] Features of this software package include pre-processing of CIU data (i.e., smoothing, cropping, interpolation, CIU fingerprint generation), stability shift (or CIU50) analysis (which determines the inflection point of sigmoidal fits between two adjacent CIU features), root-mean-squared deviation (RMSD) analysis, and machine learning-based classification of CIU datasets. All these CIUSuite2 features can be performed in an automated fashion within a single platform to advance the rapidity of CIU data analysis overall.[37]

Recent advances in IM instrumentation have allowed users to explore and develop next-generation CIU techniques. For instance, the development of cyclic IM-MS (cIM-MS) allows for the acquisition of a primary CIU data, followed by a timed ion selection event to capture unfolded intermediates of interest produced during the primary unfolding experiment. The selected ions can be subjected to a second CIU experiment (CIU^2), where ions are further activated in the pre-array region of the cIM-MS platform.[41] Furthermore, a wide range of higher resolution IM-MS instrument platforms have been developed and made commercially available, many of which enable the rapid acquisition of CIU data, necessitating a software solution capable of automating various aspects of CIU analyses in a manner that is cross-platform accessible.[41,42] Current CIU analysis software packages do not support higher-order CIU (*e.g.,* CIU^2) and the seamless automation of CIU data analysis,[37] which serves to limit the applications of CIU assays in the areas of biomarker analysis, pharmaceutical chemistry, and protein engineering.

Here, we describe the development of a new software package for CIU data analysis, CIUSuite 3, that allows users to automate not only data analysis but also the evaluation and extraction of raw data for downstream CIU analysis. First, we have developed algorithms that enable the automated extraction of raw data for both TWIM-MS and DTIM-MS platforms has been developed by deploying built-in MS peak detection. The detected peaks are then used to extract drift time data in a batch mode to automate CIU and increase the throughput of the analysis workflow. Furthermore, we constructed a built-in CCS calibration algorithm for both TWIM-MS and DTIM-MS data to enable CIU data analysis automation and the faster conversion of experimental drift time to CCS. In addition, we have found that cross-comparisons between CIU datasets are often aided by the ability to record stability values (CIU50s) between non-adjacent features. To support such CIU50 workflows, we have produced a feature-skipping function for

CIUSuite 3 that allows users to record CIU50 values between non-adjacent CIU features. Lastly, we have constructed algorithms capable of producing machine learning-based classifiers that support multi-dimensional CIU data, as well as mixtures of CIU^1 and CIU^2 data, as produced by cIM-MS. These capabilities are added to the prior features of CIUSuite 2, with CIUSuite 3 aimed at minimizing the manual analysis of raw IM-MS data, leveraging next-generation IM-MS and CIU datasets, and allowing for greater automation and robustness in CIU data analysis.

4.2 Methods

4.2.1 Experimental Section

Lyophilized proteins were purchased from Sigma Aldrich (St. Louis, MO). GroEL was prepared following a previously published protocol.[43] All other proteins were prepared in 200 mM ammonium acetate. All proteins were buffer-exchanged using Bio-Rad BioSpin 6 and 30 columns (Hercules, CA) into 100-200 mM ammonium acetate solution prior to IM-MS measurements. All IM-MS experiments were carried out at a protein concentration of 5-10 uM. IM-MS experiments were performed on Waters Synapt G2, cIM-MS (Milford, MA), and Agilent 6560C (Santa Clara, CA). Ions were generated using nanoelectrospray ionization (nESI) to preserve the proteins' native structures. CCS calibrations were performed for both Waters and Agilent instrument platforms for every IM-MS experiment. For Waters, poly-alanine (poly-A) and bovine serum albumin (BSA) were used as calibrants. The Agilent instrument was calibrated with the Agilent tunemix and BSA.

4.2.2 Raw Data Extraction

All raw data were converted to a text-based format (".csv"). For Waters, the conversion from .raw to .csv was performed using TWIMExtract.[44] For Agilent MS data, data were viewed with MassHunter IM-MS (Agilent, Santa Clara, CA), and drift time data were extracted at each

CIU voltage in C# that utilizes the MIDAC API as previously described.[42] Both IM-MS data extraction toolkits are built into CIUSuite 3.

4.2.3 CIUSuite 3 Overview

CIUSuite 3 was developed in Python 3.9 that utilizes the SciPy ecosystem which includes NumPy, Matplotlib, and Scitkit-learn.[45-49] The graphical user interface was built using the Pygubu GUI builder. All extracted IM-MS data are accessed as _raw.csv (text) format files with each file consisting of a single CIU data set. All files may undergo smoothing and normalization. Smoothing is performed using a 2D Savitzky-Golay filter with a user-specified window size. Other preprocessing includes a 1D Savitzky-Golay filter or no smoothing options. Other data processing features include cropping, interpolation of data along one or both axes, averaging multiple data sets, and signal cutoffs to remove noises. A separate .ciu file is created once preprocessing is complete (the Python pickle module is used to save and store CIU data with processing results). CIU fingerprint generation, RMSD comparison, stability shift analysis including feature detection, Gaussian fitting, and classification modules described in the previous CIUSuite version are also included (Figure 4-1).[37] New

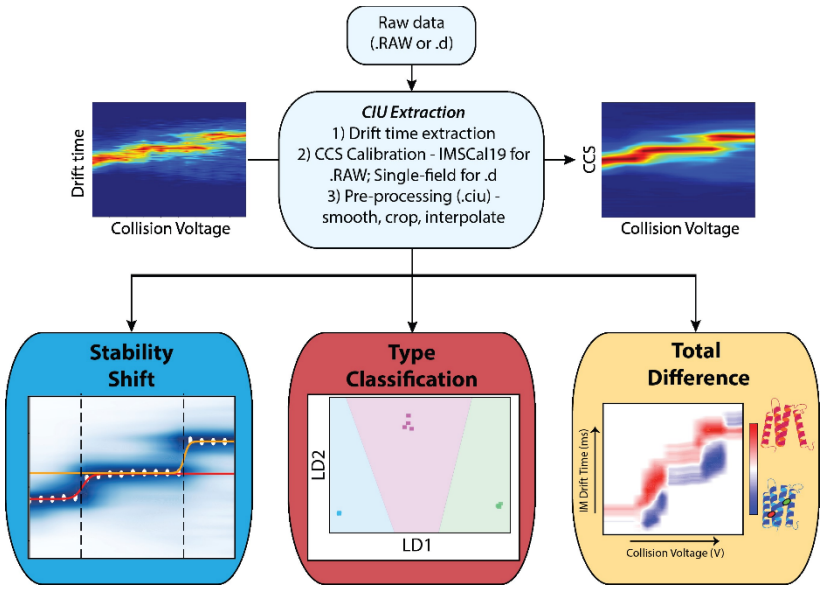


Figure 4-1. CIUSuite 3 overview and workflow. To assess peaks for CIU extraction, raw data in either .RAW or .d file is uploaded. CIUSuite 3 will extract drift time information based on either peak detection results or user input m/z ranges. The users can also directly perform CCS calibration at this stage to generate fingerprint in CCS axis. Following the extraction, modules such as stability shift analysis, classification, and difference analysis can be performed

features that are unique to CIUSuite 3 to further automate CIU data analysis in a high-throughput manner will be described below.

4.2.4 Peak Detection Algorithm

To extract IM drift time data of each protein charge state, the mass spectrum must be evaluated to determine the m/z range to extract drift time data. To automate and increase the throughput of CIU analysis, we built a peak detection algorithm using the SciPy toolkit for inclusion in CIUSuite 3. First, all intensities across the collision voltages are summed to maximize the accuracy of the peak detection. Following the intensity summation, a 1D Savitzky-Golay filter is applied with a user-specified window size to smooth the mass spectrum. Using SciPy local maxima searches, the algorithm searches for local maxima by comparing each data point to neighboring values. Once the initial search of local maxima is complete, filters are applied to finalize the detected peak list. These filters include peak prominence and peak width. To determine the peak prominence and width, the baseline is first calculated for all local maxima searched. Once the baseline for each local maximum is determined, peak prominence and width are determined by calculating the height and length of the peak at its baseline respectively. The filtering of peaks using peak prominence and width is performed based on the user-defined values where peaks with prominence and width smaller than the values defined will be removed from the final peak list. Once the final list of peaks detected is established, the algorithm will then determine the lower and upper bound m/z values for each peak which are used for the extraction of drift time data and downstream CIU analyses.

4.2.5 Feature Skipping

To perform global stability analysis or to ignore a CIU feature not relevant to the study, we constructed a feature skipping capability for inclusion in CIUSuite 3 for users to quantify global stability values from CIU data and to aid cross-comparisons between dissimilar CIU datasets. To perform feature skipping, drift time peaks present across all collision voltages in the CIU data set must be evaluated first to determine the features within the CIU data as described previously.[37] Following the feature detection, the algorithm determines the smallest and the largest drift time or CCS features in the feature list and removes the intermediate features from the feature list. Lastly, stability shift, or CIU50, analysis is performed by fitting the sigmoid between the first and the last features and calculating the voltage at which the sigmoid inflects to determine the 50% conversion point between the two features.

4.2.6 Multi-dimensional Classification

We previously described a classification module that allows users to differentiate and sort CIU fingerprint data using statistical approaches and supervised machine learning.[37] The classification module found in CIUSuite2 builds a classifier based on user input training data for each group of CIU data sets. First, univariate feature selection (UFS) is performed to determine the voltage at which the difference in drift time is most significant.[50] Following the UFS, "leave one out" cross-validation is performed to train the machine learning algorithm and to test the accuracy of classification along with a linear discriminant analysis (LDA) to build the final classifier.[51,52] This methodology has been tested and implemented successfully in CIUSuite2.

To allow multi-dimensional classification of CIU^1 and CIU^2 data, additional data preprocessing before the UFS and cross-validation is necessary to build a classifier. Both collision voltage and drift time axes of CIU^1 and CIU^2 require preprocessing for the classification scheme to compare and quantify the difference between CIU data sets followed by building the final classifier. The

collision voltage of the trap cell in cIM-MS ranges from 0 – 200 V while the pre-array store cell ranges from 0 – 140 V. Furthermore, voltage steps used may differ between the CIU^1 and CIU^2 data sets. To properly build a classifier, the classification preprocessing module will compare all the voltages that are within the CIU data selected and modify to only consider voltages that are common across the entire CIU data sets. During the acquisition of cIM-MS data, despite the reinjection of ions from the pre-array store cell when performing the CIU^2 experiment, the drift time measured does not reset to 0, causing a shift in output drift time. As such, to build CIU datasets that can be readily compared, identical traveling wave and height must be applied for the acquisition of both CIU^1 and CIU^2 data, and both the start and end times of IM measurement should be normalized to the same reference value. Thus, the classifier algorithm we constructed for CIUSuite 3 corrects all CIU^2 data according to reference CIU^1 data and normalizes the IM drift time axis across all CIU data sets to maximize the accuracy of the resulting classifier produced.

4.2.7 Built-in CCS calibration

To automate and increase the throughput of CCS values from experimental drift times, CCS calibration modules for both Waters TWIM-MS and Agilent DTIM-MS instrument platforms have been constructed and included in CIUSuite 3. For Waters TWIM-MS, a new calibration approach, IMSCal19, that includes a radial correction to improve the overall calibration error was adopted.[53] Using a template text .csv file which includes .RAW data file names and instrument settings for each sample, CIUSuite 3 subprocesses IMSCal19 and automatically inputs values into IMSCal19 to calibrate all the files in the template .csv file.

For Agilent DTIM-MS, BSA CIU data is used as a calibrant. Since the drift tube is linear, the single-field CCS calibration can be directly derived from Mason-Schamp equation[54,55] (Equation 4-1)

$$K_O = \frac{(18\pi)^{1/2}}{16}\frac{ze}{(k_bT)^{1/2}}\left[\frac{1}{m_I}+\frac{1}{m_B}\right]^{1/2}\frac{1}{N}\frac{1}{\Omega} \quad \text{(Equation 4-1)}$$

Where K_O is the reduced mobility, z is the charge of the analyte ion, k_b is the Boltzmann constant, T is the temperature, m_I is the mass of the ion, m_B is the mass of the buffer gas molecule, and N is the number density of the buffer gas. The reduced mobility can also be expressed with respect to pressure and temperature as follows:

$$K_O = \frac{L}{t_d E}\frac{P}{760}\frac{273.2}{T} \quad \text{(Equation 4-2)}$$

Where E is the electric field applied, P is the pressure within the drift tube cell, L is the length of the drift tube cell and the t_d is the drift time (in ms) of the ion through the drift tube cell. By combining Equations 4-1 and 4-2 and rearranging, CCS can be calculated using Equation 4-3 while preserving the dimensionality and universality of the Mason-Schamp equation.

$$\Omega = \frac{(18\pi)^{1/2}}{16}\frac{ze}{(k_bT)^{1/2}}\left[\frac{1}{m_I}+\frac{1}{m_B}\right]^{1/2}\frac{t_d E}{L}\frac{760}{P}\frac{T}{273.2}\frac{1}{N} \quad \text{(Equation 4-3)}$$

While t_d is the drift time of the ion through the drift tube cell, it is important to note that the time measured and output by the instrument is t_{total}, which is the time of the ion to travel from the trapping gate, located immediately before the drift tube cell, to the ion pulser which is located before the ToF analyzer.[56] Thus, t_{total} can be defined as follows:

$$t_{total} = t_d + t_0 \quad \text{(Equation 4-4)}$$

The measured value of the IM experiment is t_{total}, while t_d and t_0 can't be measured accurately with the instrument. Based on the previous study and development of DTIM-MS, variables T, E, L, P, and N are instrument settings that are well controlled and constant Thus, Equation 4-3 can be rewritten with respect to t_{total} as follows:

$$\Omega = \frac{(18\pi)^{1/2}}{16}\frac{e}{(k_bT)^{1/2}}\frac{1}{N}\frac{t_{total}}{273.2}\frac{E}{L}\frac{760}{P}m_B^{1/2}\times z\times\left[\frac{m_I}{m_B+m_I}\right]^{1/2} \quad \text{(Equation 4-5)}$$

Which can be simplified as:

$$\Omega = \beta \times t_{total} \times z \times [\frac{m_I}{m_B + m_I}]^{1/2} + t_{fix} \quad \text{(Equation 4-6)}$$

Where β is a coefficient that considers all constants including k_b, T, e, N, E, L, T, and P and t_{fix} is a correction factor that arises from using t_{total} instead of t_d.[56]

To convert DTIM drift time data into CCSs from the CIU data sets, β and t_{fix} are necessary for accurate CCS conversion. The built-in CCS calibration script is written with a database of BSA CIU information such as the CCS of each CIU feature for each charge state. Based on the user-input charge state of BSA and drift time measured (t_{total}) at each CIU feature, the calibration algorithm will evaluate the input t_{total} value with respect to the BSA CCS at each feature by performing ordinary least squares (OLS) method of linear regression analysis.[57] First, the mean values of t_{total} and BSA CCS will be calculated followed by the variance of t_{total} and covariance of the two variables. The variance of t_{total} and covariance of t_{total} and BSA CCS are used to determine the coefficient and intercept of the linear regression fit. The intercept determined using the OLS method is equivalent to t_{fix}. The coefficient from this analysis accounts not only for β but also for other terms such as z, m_I, and m_B, which are all constant. Thus, further calculations to determine β are built into the script to correctly output this value. Using β and t_{fix} values determined from BSA CIU data can be used to convert DTIM drift time into CCS in the Agilent Extractor.

4.3 Results and Discussion

To generate CIU fingerprints and perform further downstream analyses such as stability shift and classification, it is necessary to extract drift times for each charge state captured in our raw data files. Previous work from our laboratory has developed tools to extract such information in an automated fashion which includes TWIMExtract and Agilent Extractor.[37] Despite this, the

extraction of such information still requires users to open and evaluate MS data found in the raw data files using vendor-specific software. During this process, users will manually determine lower and upper bounds in both m/z for drift time space needed to extract specific signals of interest. Considering that there are multiple such signals (e.g., charge states) that require extraction from a single raw data file, the throughput of CIU analysis is often limited by such manual analysis steps.

CIUSuite 3 includes a peak detection algorithm designed to omit such manual data manipulation steps and allows users to load raw CIU data files and automate peak evaluation and extraction procedures. Raw CIU data (.RAW or .d) is loaded directly into CIUSuite 3 to begin the peak detection process. We used human serum albumin (HSA) data initially to test each step of our peak detection algorithm. Appendix Figure III-1 illustrates the data processing steps built into the peak detection algorithm such as smoothing and final peak detection after filtering. As expected, the smoothing step plays an important role in removing baseline noise (Appendix Figure III-1B). The final output of the peak detection algorithm is a .csv summary written in a format that is suitable for batch extraction of CIU data (Appendix Table III-1).

Following our initial tests, our CIUSuite 3 peak detection algorithm was tested using various protein datasets. For instance, several standard proteins across a wide mass range were tested to validate the efficacy and accuracy of the peak detection algorithm. HSA (66 kDa), IgG1 (150 kDa), and GroEL (58 kDa monomer, 812 kDa 14-mer) were tested as they are proteins that are well characterized with native IM-MS and CIU from previous studies.[20,26,58] In each case, the final peaks detected are evaluated to determine the accuracy of our CIUSuite 3 peak detection algorithm. Across the three standard proteins noted above, all the peaks detected in our HSA and IgG1 datasets were manually confirmed and assigned to known protein charge states that aligned with the expected molecular weights of the proteins (Figure 4-2A and B). The peak detection

results from these two proteins also demonstrate the sensitivity of the peak detection algorithm, as HSA trimer signals are detected despite their low (<1%) relative intensities (Figure 4-2A). For GroEL, all peaks detected for the 14-mer were readily assigned to charge states that aligned with the expected molecular weight for the native GroEL complex; however, likely due to buffer components adhered to the protein during nESI, some adduct peaks were also detected for monomeric GroEL (Figure 4-2C). Our peak detection algorithm is further tested by monitoring the mass range output from the CIU fingerprint generation algorithm. The most intense peaks detected by the CIUSuite 3 algorithm were used to generate fingerprints, and overall, across all the model systems tested here, the fingerprints obtained using automatic peak detection through CIUSuite 3 were nearly identical to those obtained in previous studies (Figure 4-2D-F).[20,26,58]

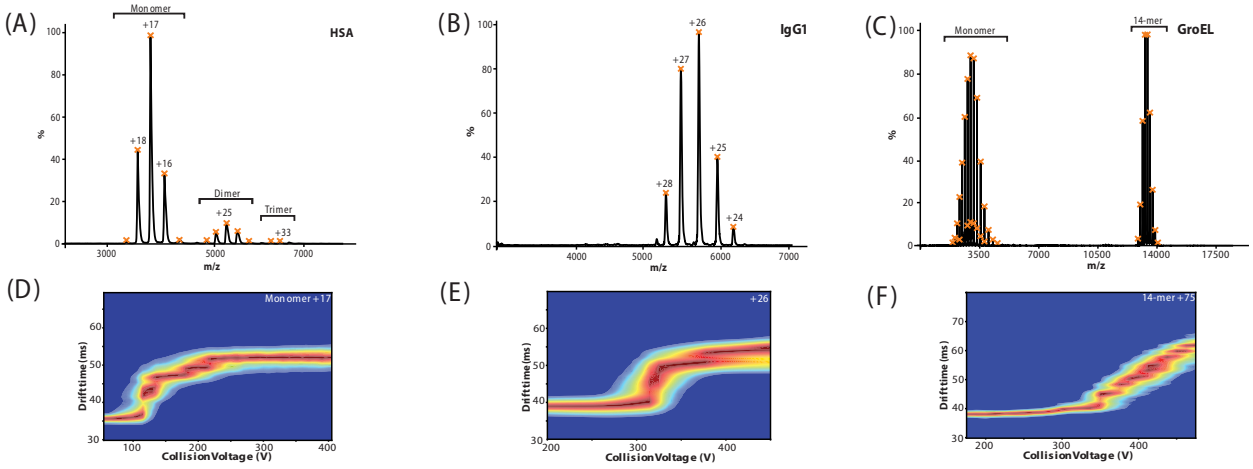


Figure 4-2. Peak detection algorithm output. (A-C) Mass spectra of HSA, IgG1, and GroEL, with peaks detected by the algorithm marked in orange. **(D-F)** CIU fingerprint of HSA, IgG1, and GroEL at a charge state with the highest relative intensity.

We further tested our peak detection algorithm using a significant library of CIU data obtained for Sirtuin 5-ligand complexes (SIRT5), a deacetylase that has elicited significant pharmaceutical interests due to its links to cellular aging, which we previously collected through high-throughput screening (Appendix Figure III-2A & B).[6] While this novel approach for

compound screening has improved the throughput, data analysis of such large number of CIU data sets are limited by manual extraction of CIU data for downstream analyses. By implementing automated peak detection through CIUSuite 3, we were able to automate CIU data extraction and fingerprint generation for the entire dataset, including control data and 288 experimental CIU datasets (96 samples analyzed in triplicate). CIU features detected for SIRT5 discussed in our prior report were also detected with the CIUSuite 3 algorithm (Appendix Figure III-2C). A deeper analysis of the CIU fingerprints generated based on automated CIUSuite 3 peak detection was plotted for both apo and ligand-bound SIRT5 and compared quantitatively to those from our previous study, revealing excellent agreement (Figure III-2D).[7] Overall, the manual CIU data analysis performed in our previous work, which took approximately a week to complete, was reduced to a single day of data processing using CIUSuite 3.

A CIU dataset containing Remicade (RC), an approved biotherapeutic used to treat Crohn's disease, ulcerative colitis, and rheumatoid arthritis, along with its biosimilars – Renflexis (RF) and Inflectra (IF), were used to further test the throughput advantages of CIU data extraction and analysis using CIUSuite 3 automated peak detection. This experiment not only compared CIU fingerprints generated across three different biotherapeutic samples, but also included data collected under standard and multiple different stressed conditions, charge states, and replicates, resulting in a large CIU data set which totals analysis of 144 fingerprints.[59] Using automated peak detection as implemented in CIUSuite 3, peaks were detected and annotated in a manner consistent with our manual data annotation reported previously (Appendix Figure III-3A). Our comparison included conducting a total difference CIU fingerprint analysis comparing RC to the RF and IF biosimilars (Appendix Figure III-3B), resulting in a nearly identical heat map of root-mean-square-deviation (RMSD) values (Appendix Figure III-3C) as what has been reported previously for this

dataset. Our prior manual analysis of this dataset required 40 days, with CIUSuite 3 completing the same tasks in 20-fold less processing time (2 days total).

CIUSuite 2 included a robust feature detection algorithm that enabled the quantitative analysis of CIU50 values. With growing interest and application of CIU, stability shift analyses can become less intuitive depending on the system being studied. For instance, some charge states of αSynuclein have been observed to exhibit parallel CIU features across multiple collision voltages, complicating robust stability comparisons.[60] In addition, CIU of nucleic acids, especially RNAs, typically undergo slight compaction transitions prior to unfolding during CIU, necessitating non-standard approaches for the evaluation of CIU50 values.[61] CIUSuite 2 feature detection only registers CIU features only when they become the most intense signal in at least one voltage slice, resulting in many lower intensity features being unanalyzed. In CIUSuite 3, we developed a feature skipping function to allow global stability analyses that track the transition between CIU features that appear at the shortest and longest drift times (or the smallest and largest CCSs).

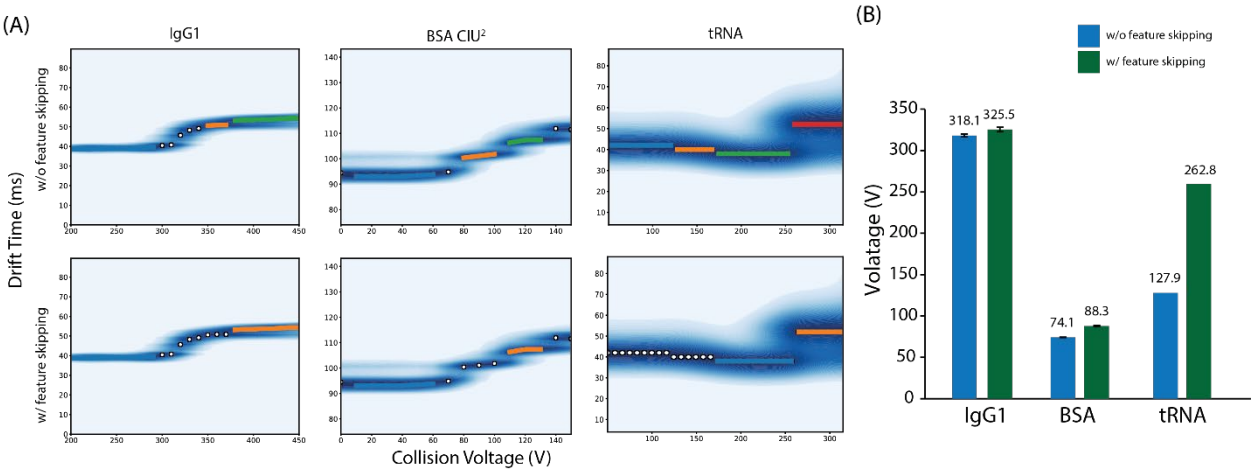


Figure 4-3. Feature skipping feature comparison. (A) Feature detection results of IgG1, BSA CIU^2, and tRNA without feature skipping (top row) and with feature skipping (bottom row). **(B)** CIU50 values of the first CIU transition compared without feature skipping and with feature skipping.

Our feature skipping function in CIUSuite 3 was tested across both standard proteins and mitochondrial tRNA samples. CIU feature detection compared between IgG1, BSA CIU^2, and

tRNA data analyzed with and without feature skipping active (Figure 4-3A). Across all three samples, features were properly omitted from the feature list and plotted accordingly when feature skipping was engaged. For example, intermediate features are skipped for both IgG1 and BSA CIU^2 data (Figure 4-3A), while lower energy features are skipped for the tRNA CIU analyzed as these features populate intermediate IM drift times. In addition, we retain the definition of the CIU feature in our CIUSuite 3 feature skipping algorithm, which results in the highest energy, low-intensity feature which appears at ~115ms to be omitted from our feature skipped CIU analysis. Overall, across all samples tested, feature skipping properly identifies the CIU features with the shortest and longest drift times and omits all other CIU features from the final feature list. As a result, sigmoidal fitting performed as part of CIU50 workflows (Figure 4-3B), provides a more global insight into biomolecular stabilities that is often amenable for cross-comparisons between CIU datasets.

The growth in cIM-MS applications targeting biomolecular structure and biophysics is in part due to the ability to select and store ions in the pre-array, followed by further collisional activation to perform CIU^2. Such multi-dimensional CIU data provides a range of new useful options for building CIU assays with increased structural sensitivity and selectivity; however, the different activation voltage and drift time ranges produced in CIU^2 data are not compatible with the CIUSuite 2 classifier function as originally implemented. In CIUSuite 3, a 2D CIU preprocessing feature is included to resolve the issues outlined above for users to successfully build mixed multi-dimensional CIU classification schemes.

In order to evaluate our ability to build classifiers from both CIU^1 and CIU^2 data, we evaluated such data for BSA to test multi-dimensional classification featuring 2D preprocessing (Appendix Figure III-4A-C). As intended, with the preprocessing step included before the

classification of CIU^1 and CIU^2, only the voltages that were common across all CIU data sets selected were considered for UFS analysis. 1D preprocessing on the collision voltage axis resulted in the higher voltages (151 – 200V) collected in CIU^1 data sets being disregarded, and a successful UFS analysis was performed, with the highest scoring voltage identified as 150V (Appendix Figure III-4D). Further downstream classification modules demonstrated the ability of successive 1D preprocessing steps to determine additional high-scoring voltage values in the dataset (Appendix Figure III-4E & F). It is also important to note that the BSA data sets used for these tests also contained different voltage step sizes between the CIU^1 and CIU^2 data included (5V and 10V respectively) to allow preprocessing on a broader scale.

We then moved on to evaluate our complete 2D preprocessing algorithm, which is performed on both the collision voltage and drift time axes. First, the highest scoring voltage from the UFS differed slightly between 2D and 1D preprocessed data. Furthermore, at high-scoring voltages, the errors associated with UFS decreased when additional correction was applied to the drift time axis (Appendix Figure III-4D&G). Second, we compared the cross-validation results obtained from 1D and 2D preprocessed data to determine whether the accuracy of the classification scheme improves with the additional correction offered in the latter data format. We noted in the testing described above that for 1D preprocessed data, the cross-validation revealed poorer classifier accuracy in the training data as the number of features increased (Appendix Figure III-4E), typically the hallmark of a weaker classifier. Conversely, the cross-validation results obtained for 2D preprocessed data indicated significantly improved cross-validation results for both the training and test data, crucially including those classifiers that utilized a greater number of features (Appendix Figure III-4H). Furthermore, the final classification scheme obtained from LDA differs significantly between the 1D and 2D preprocessed data, with the latter producing a classifier of

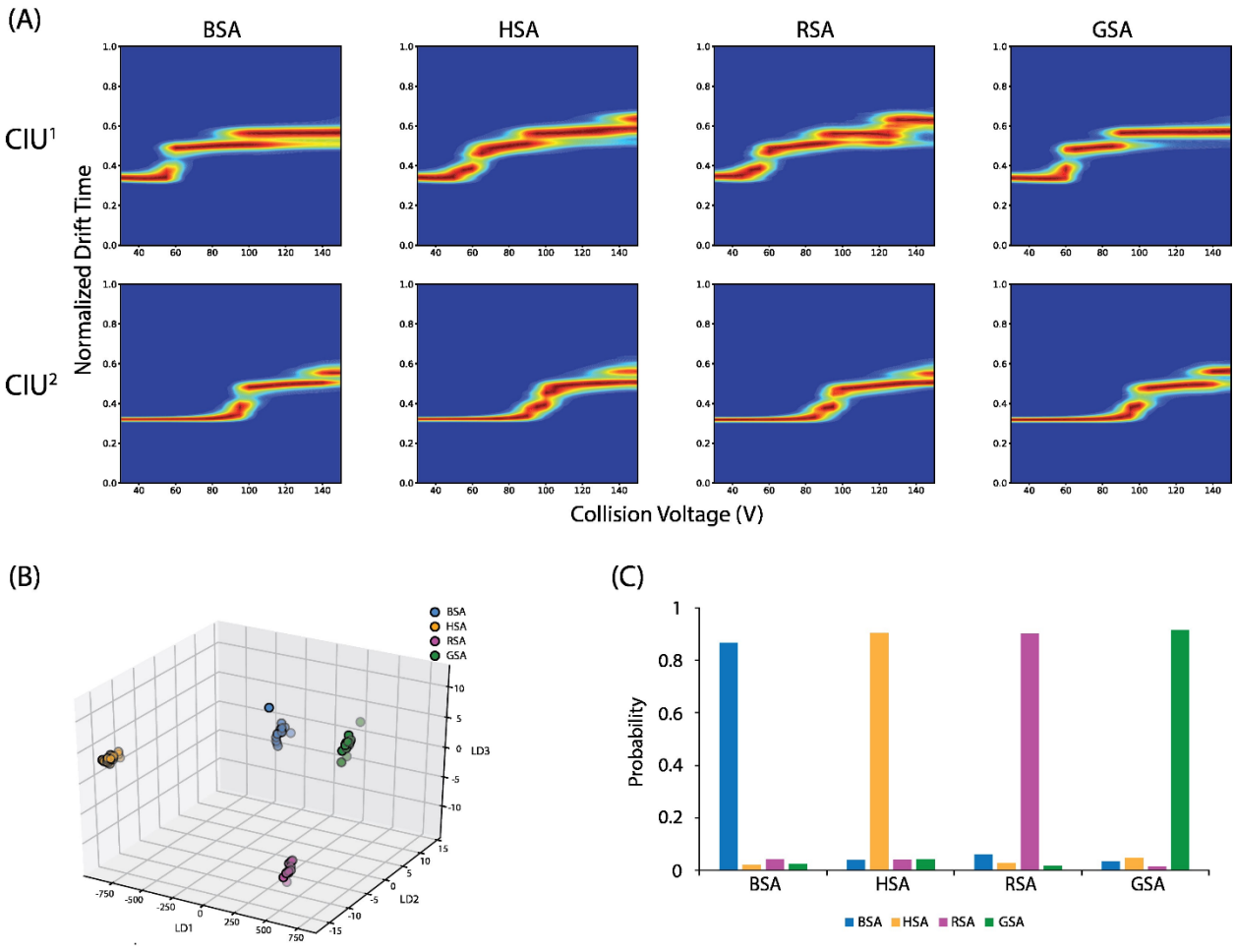


Figure 4-4. Multi-dimensional classification of BSA, HSA, RSA, and GSA. (A) Fingerprints of BSA, HSA, RSA, GSA CIU^1 (top) and CIU^2 (bottom) at charge state +15.a **(B)** Final classification scheme of multi-dimensional, multi-charge state classification. **(C)** Probabilities of categorizing all four serum albumins with both CIU^1 and CIU^2.

higher precision, most apparent in the CIU^1 acquired for +15 BSA (Appendix Figure III-4F&I). Overall, the 2D preprocessing algorithm not only allows for the successful construction of machine learning classifiers from mixed CIU^N data streams, and also improves the accuracy of the final classifier.

In order to demonstrate the utility of mixed CIU^1/CIU^2 classifiers, we utilized CIUSuite 3 to construct a multi-dimensional classification capable of differentiating serum albumins from various species, specifically bovine, human, and rabbit. The implementation of the new classification algorithm is shown in Figure 4-4, utilizing CIU^1/CIU^2 data acquired from BSA, human serum albumin (HSA), rabbit serum albumin (RSA), and goat serum albumin (GSA).

Classification is performed using multi-charge state and multi-dimensional CIU to test the efficacy and improvement in classification results. The fingerprints of 4 classes of serum albumin tested are shown in Figure 4-4A to demonstrate the difference between CIU^1 and CIU^2 data which mostly show higher activation energy necessary to reach the first unfolding state for all 4 classes in CIU^2. The final classification scheme resulted in the successful identification of all 4 classes (Figure 4-4B). We further evaluated the probability of each data set's classification into each class. The probability values further confirmed the accuracy of our final classification scheme with probability values ranging from 86.7 to 91.6% (Figure 4-4C, Appendix Table III-2). The probability of classification into each class was evaluated and compared using a classification scheme built with and without multi-dimensional CIU classification and confirmed the accuracy of the classification scheme. The probability of classification into the correct class without multi-dimensional classification (i.e. only CIU^1 data were used to build the classifier) remained roughly around 84% for all four classes. However, with the inclusion of CIU^2 and performing multi-dimensional classification resulted in a higher probability of correct assignment of each class and increased the probability up to 7% (Appendix Table III-2). Overall, multi-dimensional classifications of serum albumins corroborate and demonstrate the accuracy and applicability of the feature in CIU analysis.

The final new feature of CIUSuite 3 focuses on the inclusion of built-in CCS calibration schemes to support multiple instrument platforms. While the detailed workflow for CCS calibration may vary between instrument platforms, both TWIM-MS and DTIM-MS benefit from calibration for the generation of accurate CCS values.

For onboard calibration in CIUSuite 3, we aimed primarily to increase the throughput of CCS values in the context of CIU data analysis workflows. Our CIUSuite 3 calibration workflow

begins by utilizing direct input from raw data (.RAW or .d) rather than an extracted .csv calibrant data file. For TWIM-MS data, a template .csv file is loaded to initialize the CCS calibration by subprocessing our previously developed software package utilizing blended-form TWIM calibration functions (IMSCal19).[53] CIUSuite 3 then automatically inputs information from the template file to extract, calibrate, and convert IM drift time into CCS in a single step automatically until all data files in the template are calibrated (Figure 4-1). This approach allows the extraction of CIU data in a batch mode where multiple data sets are analyzed in an automated fashion, significantly improving the throughput of CIU data analysis in CCS mode. For instance, our CIUSuite 3 tests revealed that the manual CCS calibration of 81 TWIM CIU requires up to 3 days to complete. In contrast, the built-in calibration feature in CIUSuite 3, performs the same task within 30 minutes.

DTIM-MS calibration typically involves the infusion of small molecules in an organic solvent (e.g., Agilent tunemix) to determine β and t_{fix} values. To streamline the DTIMS calibration process for CIU data processing, we developed a DTIMS workflow using CIU data acquired for BSA, a widely used standard protein in native MS, for our CIUSuite 3 DTIMS CCS calibration procedure. BSA CIU calibrant data can be acquired under native MS conditions, further streamlining our CCS calibration workflow. Figure 4-4A shows single-field CCS calibration functions generated using t_{total} and known CCS values of each CIU feature observed for BSA. The linear regressions obtained all exhibit high correlation coefficients indicating their suitability for

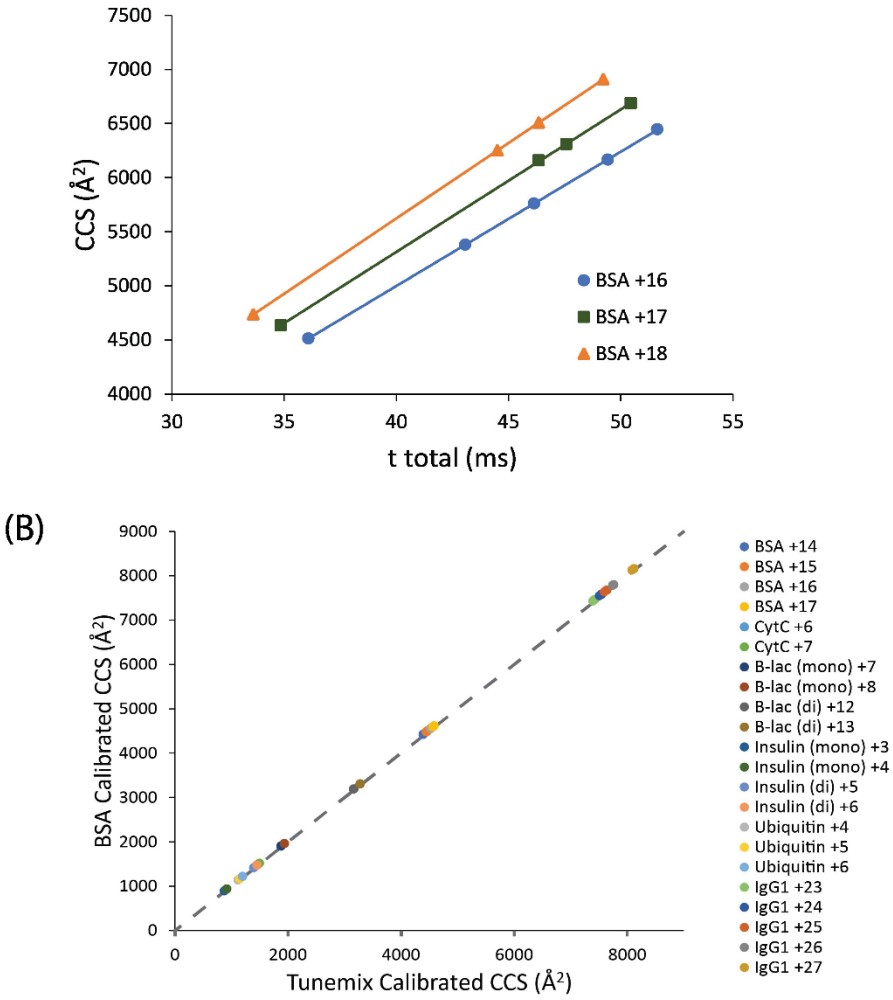


Figure 4-5. Single-field CCS calibration methods. (A) Single-field DTIM CCS calibration using BSA as a calibrant at various charge states. Each data point refers to the drift time and CCS of individual CIU feature. Linear regressions using OLS method are included. All three linear fits resulted r^2 =1. **(B)** CCS values of various standard proteins compared between the tunemix and BSA as calibrants. The grey dash line refers to $CCS_{tunemix} = CCS_{BSA}$.

generating high-quality DTIMS CCS calibration and determination of β and t_{fix} values (Appendix Table III–2).

We evaluated our DTIMS CCS calibration procedure using BSA CIU calibrant data by comparing with one generated using a standard tunemix. Our unknowns included various proteins that have been well-characterized by native MS previously, having sizes ranging from 5.8 kDa to 150 kDa. In general, CCS values calibrated using BSA CIU data were comparable to those calibrated using tunemix (Figure 4-5B). Overall, our BSA CIU procedure and our CIUSuite 3 DTIMS calibration algorithm serve to increase CIU analysis throughput when workflows require CCS information while retaining the CCS accuracy typically obtained in standard methods (Figure 4-5B; Appendix Table III-3).

4.4 Conclusions

Here, we present CIUSuite 3, a new CIU data analysis package that is optimized to improve CIU data analysis throughput and the support of multi-dimensional modes of CIU data acquisition. To accommodate higher throughput modes of CIU data collection, we have constructed and validated built-in peak detection and CCS calibration modules into CIUSuite 3. To enable the improved comparisons of complex CIU datasets, we have developed a feature skipping algorithm that allows users to perform global stability analyses. Finally, CIUSuite 3 contains support for multi-dimensional classification where a mix of CIU^1 and CIU^2 data sets can be used to differentiate unknown data populations. Preprocessing CIU data in both the activation voltage and drift time dimension improves the overall accuracy of the final classifier generated. Overall, we envision that CIUSuite 3 will enable the generation of wide-ranging CIU assays that further push the boundaries of what can be accomplished with gas-phase unfolding datasets.

4.5 References

(1) Leney, A. C.; Heck, A. J. R. Native Mass Spectrometry: What Is in the Name? *J. Am. Soc. Mass Spectrom.* **2017**, *28* (1), 5–13.

(2) Mehmood, S.; Allison, T. M.; Robinson, C. V. Mass Spectrometry of Protein Complexes: From Origins to Applications. *Annu. Rev. Phys. Chem.* **2015**, *66*, 453–474.

(3) Reading, E.; Liko, I.; Allison, T. M.; Benesch, J. L. P.; Laganowsky, A.; Robinson, C. V. The Role of the Detergent Micelle in Preserving the Structure of Membrane Proteins in the Gas Phase. *Angew. Chemie - Int. Ed.* **2015**, *54* (15), 4577–4581.

(4) Zhong, Y.; Hyung, S. J.; Ruotolo, B. T. Ion Mobility-Mass Spectrometry for Structural Proteomics. *Expert Review of Proteomics*. Taylor & Francis February 2012, pp 47–58.

(5) Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E. The Power of Ion Mobility-Mass Spectrometry for Structural Characterization and the Study of Conformational Dynamics. *Nature Chemistry*. Nature Publishing Group March 21, 2014, pp 281–294.

(6) Terral, G.; Beck, A.; Cianférani, S. Insights from Native Mass Spectrometry and Ion Mobility-Mass Spectrometry for Antibody and Antibody-Based Product Characterization. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2016**, *1032*, 79–90.

(7) D'Amico, C. I.; Polasky, D. A.; Steyer, D. J.; Ruotolo, B. T.; Kennedy, R. T. Ion Mobility-Mass Spectrometry Coupled to Droplet Microfluidics for Rapid Protein Structure Analysis and Drug Discovery. *Anal. Chem.* **2022**, *94* (38), 13084–13091.

(8) Fantin, S. M.; Parson, K. F.; Yadav, P.; Juliano, B.; Li, G. C.; Sanders, C. R.; Ohi, M. D.; Ruotolo, B. T. Ion Mobility–Mass Spectrometry Reveals the Role of Peripheral Myelin Protein Dimers in Peripheral Neuropathy. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (17), e2015331118.

(9) Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. Protein Structure in Vacuo: Gas-Phase Conformations of BPTI and Cytochrome C. *J. Am. Chem. Soc.* **1997**, *119* (9), 2240–2248.

(10) Hyung, S. J.; Robinson, C. V.; Ruotolo, B. T. Gas-Phase Unfolding and Disassembly Reveals Stability Differences in Ligand-Bound Multiprotein Complexes. *Chem. Biol.* **2009**, *16* (4), 382–390.

(11) Zhang, H.; Liu, H.; Lu, Y.; Wolf, N. R.; Gross, M. L.; Blankenship, R. E. Native Mass Spectrometry and Ion Mobility Characterize the Orange Carotenoid Protein Functional Domains. *Biochim. Biophys. Acta - Bioenerg.* **2016**, *1857* (6), 734–739.

(12) Zhao, Y.; Singh, A.; Xu, Y.; Zong, C.; Zhang, F.; Boons, G. J.; Liu, J.; Linhardt, R. J.; Woods, R. J.; Amster, I. J. Gas-Phase Analysis of the Complex of Fibroblast GrowthFactor 1 with Heparan Sulfate: A Traveling Wave Ion Mobility Spectrometry (TWIMS) and Molecular Modeling Study. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (1), 96–109.

(13) Byrne, D. P.; Vonderach, M.; Ferries, S.; Brownridge, P. J.; Eyers, C. E.; Eyers, P. A. CAMP-Dependent Protein Kinase (PKA) Complexes Probed by Complementary Differential Scanning Fluorimetry and Ion Mobility-Mass Spectrometry. *Biochem. J.* **2016**, *473* (19), 3159–3175.

(14) Chan, D. S. H.; Kavanagh, M. E.; McLean, K. J.; Munro, A. W.; Matak-Vinković, D.; Coyne, A. G.; Abell, C. Effect of DMSO on Protein Structure and Interactions Assessed by Collision-Induced Dissociation and Unfolding. *Anal. Chem.* **2017**, *89* (18), 9976–9983.

(15) Beynon, R. J.; Armstrong, S. D.; Claydon, A. J.; Davidson, A. J.; Eyers, C. E.; Langridge, J. I.; Gómez-Baena, G.; Harman, V. M.; Hurst, J. L.; Lee, V.; McLean, L.; Pattison, R.; Roberts, S. A.; Simpson, D. M.; Unsworth, J.; Vonderach, M.; Williams, J. P.; Woolerton, Y. E. Mass Spectrometry for Structural Analysis and Quantification of the Major Urinary Proteins of the House Mouse. *Int. J. Mass Spectrom.* **2015**, *391*, 146–156.

(16) Han, L.; Hyung, S. J.; Mayers, J. J. S.; Ruotolo, B. T. Bound Anions Differentially Stabilize Multiprotein Complexes in the Absence of Bulk Solvent. *J. Am. Chem. Soc.* **2011**, *133* (29), 11358–11367.

(17) Zhong, Y.; Han, L.; Ruotolo, B. T. Collisional and Coulombic Unfolding of Gas-Phase Proteins: High Correlation to Their Domain Structures in Solution. *Angew. Chemie - Int. Ed.* **2014**, *53* (35), 9209–9212.

(18) Samulak, B. M.; Niu, S.; Andrews, P. C.; Ruotolo, B. T. Ion Mobility-Mass Spectrometry Analysis of Cross-Linked Intact Multiprotein Complexes: Enhanced Gas-Phase Stabilities and Altered Dissociation Pathways. *Anal. Chem.* **2016**, *88* (10), 5290–5298.

(19) Wagner, N. D.; Russell, D. H. Defining Noncovalent Ubiquitin Homodimer Interfacial Interactions through Comparisons with Covalently Linked Diubiquitin. *J. Am. Chem. Soc.* **2016**, *138* (51), 16588–16591.

(20) Wagner, N. D.; Clemmer, D. E.; Russell, D. H. ESI-IM-MS and Collision-Induced Unfolding That Provide Insight into the Linkage-Dependent Interfacial Interactions of Covalently Linked Diubiquitin. *Anal. Chem.* **2017**, *89* (18), 10094–10103.

(21) Eschweiler, J. D.; Martini, R. M.; Ruotolo, B. T. Chemical Probes and Engineered Constructs Reveal a Detailed Unfolding Mechanism for a Solvent-Free Multidomain Protein. *J. Am. Chem. Soc.* **2017**, *139* (1), 534–540.
(22) Hopper, J. T. S.; Oldham, N. J. Collision Induced Unfolding of Protein Ions in the Gas Phase Studied by Ion Mobility-Mass Spectrometry: The Effect of Ligand Binding on Conformational Stability. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (10), 1851–1858.
(23) Niu, S.; Ruotolo, B. T. Collisional Unfolding of Multiprotein Complexes Reveals Cooperative Stabilization upon Ligand Binding. *Protein Sci.* **2015**, *24* (8), 1272–1281.
(24) Rabuck, J. N.; Hyung, S. J.; Ko, K. S.; Fox, C. C.; Soellner, M. B.; Ruotolo, B. T. Activation State-Selective Kinase Inhibitor Assay Based on Ion Mobility-Mass Spectrometry. *Anal. Chem.* **2013**, *85* (15), 6995–7002.
(25) Beveridge, R.; Migas, L. G.; Payne, K. A. P.; Scrutton, N. S.; Leys, D.; Barran, P. E. Mass Spectrometry Locates Local and Allosteric Conformational Changes That Occur on Cofactor Binding. *Nat. Commun.* **2016**, *7* (1), 1–9.
(26) Rabuck-Gibbons, J. N.; Keating, J. E.; Ruotolo, B. T. Collision Induced Unfolding and Dissociation Differentiates ATP-Competitive from Allosteric Protein Tyrosine Kinase Inhibitors. *Int. J. Mass Spectrom.* **2018**, *427*, 151–156.
(27) Tian, Y.; Ruotolo, B. T. The Growing Role of Structural Mass Spectrometry in the Discovery and Development of Therapeutic Antibodies. *Analyst*. Royal Society of Chemistry May 29, 2018, pp 2459–2468.
(28) Tian, Y.; Han, L.; Buckner, A. C.; Ruotolo, B. T. Collision Induced Unfolding of Intact Antibodies: Rapid Characterization of Disulfide Bonding Patterns, Glycosylation, and Structures. *Anal. Chem.* **2015**, *87* (22), 11509–11515.
(29) Tian, Y.; Ruotolo, B. T. Collision Induced Unfolding Detects Subtle Differences in Intact Antibody Glycoforms and Associated Fragments. *Int. J. Mass Spectrom.* **2018**, *425*, 1–9.
(30) Botzanowski, T.; Erb, S.; Hernandez-Alba, O.; Ehkirch, A.; Colas, O.; Wagner-Rousset, E.; Rabuka, D.; Beck, A.; Drake, P. M.; Cianférani, S. Insights from Native Mass Spectrometry Approaches for Top- and Middle- Level Characterization of Site-Specific Antibody-Drug Conjugates. *MAbs* **2017**, *9* (5), 801–811.
(31) Huang, Y.; Salinas, N. D.; Chen, E.; Tolia, N. H.; Gross, M. L. Native Mass Spectrometry, Ion Mobility, and Collision-Induced Unfolding Categorize Malaria Antigen/Antibody Binding. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (11), 2515–2518.
(32) Laganowsky, A.; Reading, E.; Allison, T. M.; Ulmschneider, M. B.; Degiacomi, M. T.; Baldwin, A. J.; Robinson, C. V. Membrane Proteins Bind Lipids Selectively to Modulate Their Structure and Function. *Nature* **2014**, *510* (7503), 172–175.
(33) Allison, T. M.; Reading, E.; Liko, I.; Baldwin, A. J.; Laganowsky, A.; Robinson, C. V. Quantifying the Stabilizing Effects of Protein-Ligand Interactions in the Gas Phase. *Nat. Commun.* **2015**, *6* (1), 1–10.
(34) Liu, Y.; Cong, X.; Liu, W.; Laganowsky, A. Characterization of Membrane Protein–Lipid Interactions by Mass Spectrometry Ion Mobility Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (4), 579–586.
(35) Fantin, S. M.; Parson, K. F.; Niu, S.; Liu, J.; Polasky, D. A.; Dixit, S. M.; Ferguson-Miller, S. M.; Ruotolo, B. T. Collision Induced Unfolding Classifies Ligands Bound to the Integral Membrane Translocator Protein. *Anal. Chem.* **2019**, *91* (24), 15469–15476.

(36) Eschweiler, J. D.; Rabuck-Gibbons, J. N.; Tian, Y.; Ruotolo, B. T. CIUSuite: A Quantitative Analysis Package for Collision Induced Unfolding Measurements of Gas-Phase Protein Ions. *Anal. Chem.* **2015**, *87* (22), 11516–11522.
(37) Polasky, D. A.; Dixit, S. M.; Fantin, S. M.; Ruotolo, B. T. CIUSuite 2: Next-Generation Software for the Analysis of Gas-Phase Protein Unfolding Data. *Anal. Chem.* **2019**, *91* (4), 3147–3155.
(38) Sivalingam, G. N.; Yan, J.; Sahota, H.; Thalassinos, K. Amphitrite: A Program for Processing Travelling Wave Ion Mobility Mass Spectrometry Data. *Int. J. Mass Spectrom.* **2013**, *345–347*, 54–62.
(39) Sivalingam, G. N.; Cryar, A.; Williams, M. A.; Gooptu, B.; Thalassinos, K. Deconvolution of Ion Mobility Mass Spectrometry Arrival Time Distributions Using a Genetic Algorithm Approach: Application to A1-Antitrypsin Peptide Binding. *Int. J. Mass Spectrom.* **2018**, *426*, 29–37.
(40) Migas, L. G.; France, A. P.; Bellina, B.; Barran, P. E. ORIGAMI: A Software Suite for Activated Ion Mobility Mass Spectrometry (AIM-MS) Applied to Multimeric Protein Assemblies. *Int. J. Mass Spectrom.* **2018**, *427*, 20–28.
(41) Giles, K.; Ujma, J.; Wildgoose, J.; Pringle, S.; Richardson, K.; Langridge, D.; Green, M. A Cyclic Ion Mobility-Mass Spectrometry System. *Anal. Chem.* **2019**, *91* (13), 8564–8573.
(42) Vallejo, D. D.; Polasky, D. A.; Kurulugama, R. T.; Eschweiler, J. D.; Fjeldsted, J. C.; Ruotolo, B. T. A Modified Drift Tube Ion Mobility-Mass Spectrometer for Charge-Multiplexed Collision-Induced Unfolding. *Anal. Chem.* **2019**, *91* (13), 8137–8146.
(43) Hogan, C. J.; Ruotolo, B. T.; Robinson, C. V.; Fernandez De La Mora, J. Tandem Differential Mobility Analysis-Mass Spectrometry Reveals Partial Gas-Phase Collapse of the GroEL Complex. *J. Phys. Chem. B* **2011**, *115* (13), 3614–3621.
(44) Haynes, S. E.; Polasky, D. A.; Dixit, S. M.; Majmudar, J. D.; Neeson, K.; Ruotolo, B. T.; Martin, B. R. Variable-Velocity Traveling-Wave Ion Mobility Separation Enhancing Peak Capacity for Data-Independent Acquisition Proteomics. *Anal. Chem.* **2017**, *89* (11), 5669–5672.
(45) Millman, K. J.; Aivazis, M. Python for Scientists and Engineers. *Computing in Science and Engineering*. March 2011, pp 9–12.
(46) Oliphant, T. E. Python for Scientific Computing. *Comput. Sci. Eng.* **2007**, *9* (3), 10–20.
(47) Dubois, P. F.; Hinsen, K.; Hugunin, J. Numerical Python. *Comput. Phys.* **1996**, *10* (3), 262.
(48) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9* (3), 90–95.
(49) Pedregosa, F.; Varoquax, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Dubourg, V.; Vanderplas, J. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12* (Oct), 2825-2830
(50) Dowdy, S.; Wearden, S.; Chilko, D. *Statistics for Research*, 3rd Ed.; Wiley-Interscience: Hoboken, NJ, **2004**.
(51) Arlot, S.; Celisse, A. A Survey of Cross-Validation Procedures for Model Selection. *Stat. Surv.* **2010**, *4*, 40–79.
(52) Xanthopoulos, P.; Pardalos, P. M.; Trafalis, T. B. *Robust Data Mining*; Springer: New York, NY, **2013**.
(53) Richardson, K.; Langridge, D.; Dixit, S. M.; Ruotolo, B. T. An Improved Calibration Approach for Traveling Wave Ion Mobility Spectrometry: Robust, High-Precision Collision Cross Sections. *Anal. Chem.* **2021**, *93* (7), 3542–3550.

(54) Mason, E. A.; McDaniel, E. W. *Transport Properties of Ions in Gases*; Wiley, **1988**.
(55) Siems, W. F.; Viehland, L. A.; Hill, H. H. Improved Momentum-Transfer Theory for Ion Mobility. 1. Derivation of the Fundamental Equation. *Anal. Chem.* **2012**, *84* (22), 9782–9791.
(56) Mordehai, A.; Kurulugama, R. T.; Klein, C.; Fjeldsted, J. C. Fast Method for Measuring Collision Cross Section of Ions Utilizing Ion Mobility Spectrometry. 20150219598, **2014**.
(57) Chumney, E.; Simpson, K. N. *Methods and Designs for Outcomes Research*; **2006**.
(58) Gadkari, V. V.; Ramírez, C. R.; Vallejo, D. D.; Kurulugama, R. T.; Fjeldsted, J. C.; Ruotolo, B. T. Enhanced Collision Induced Unfolding and Electron Capture Dissociation of Native-like Protein Ions. *Anal. Chem.* **2020**, *92* (23), 15489–15496.
(59) Vallejo, D. D.; Kang, J.; Coghlan, J.; Ramírez, C. R.; Polasky, D. A.; Kurulugama, R. T.; Fjeldsted, J. C.; Schwendeman, A. A.; Ruotolo, B. T. Collision-Induced Unfolding Reveals Stability Differences in Infliximab Therapeutics under Native and Heat Stress Conditions. *Anal. Chem.* **2021**, *93* (48), 16166–16174.
(60) Jeacock, K.; Chappard, A.; Gallagher, K. J.; Mackay, C. L.; Kilgour, D. P. A.; Horrocks, M. H.; Kunath, T.; Clarke, D. J. Determining the Location of the α-Synuclein Dimer Interface Using Native Top-Down Fragmentation and Isotope Depletion-Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2022**.
(61) Abi-Ghanem, J.; Rabin, C.; Porrini, M.; Rosu, F.; Gabelica, V. Compaction of RNA Hairpins and Their Kissing Complexes in Native Electrospray Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (10), 2035–2043.
(62) Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Anal. Chem.* **2010**, *82* (22), 9557–9565.

Chapter 5 Conclusions and Future Directions

In this thesis, various computational methods were developed for structural assignments of protein and protein complexes and a software tool that automates IM-MS data analysis including newly released instrument platforms such as cyclic MS. While these methods have improved the throughput and accuracy of data analysis to model structures of proteins, there are still limitations as there are various other IM-MS platforms such as structure for lossless ion manipulations (SLIM) which can not only provide more detailed structural information but also necessitates the analysis tool that can automate the data analysis.

5.1 Integrative Modeling of Proteins and Protein Complexes

In Chapters 2 and 3, integrative modeling was developed and utilized for the structural characterization of proteins in the gas-phase. In both steered MD method development for gas-phase protein unfolding and gas-phase structural assignment of IgG1, one of the challenges associated with generating accurate models arises from charge placements. Myriads of charge isomers are generated while the locations of these charges remain unknown making charge placement necessary for gas-phase MD simulations challenging. To mitigate this challenge, efforts have been made to place charges on the protein surface using a Monte Carlo simulation.[1] Despite various charge placement algorithms available in the field, it is still questionable whether the charges placed are experimentally relevant. Furthermore, these charge placement methods are built for native proteins, and for structural assignments of protein complexes for gas-phase protein unfolding, charges need to be migrated towards the unfolded region within the protein. In Chapter

2, we've demonstrated the random charge placement algorithm to resolve such issues regarding unknown charge locations and migrating charges during unfolding by generating and sampling all possible charge isomers. However, several experimental data do suggest charge localization using fragmentation results from UV photodissociation (UVPD).[2,3] The study has also demonstrated how charge location can impact the gas-phase model structures, highlighting the importance of charge localization.[2,4] Thus, to improve the accuracy of models from integrative modeling, additional UVPD data to filter charge isomers from the charge placement algorithm could be beneficial. Furthermore, tracking changes in fragmentation patterns by activating the protein could be utilized to localize charges as the protein unfolds which would improve the accuracy of the structural assignment of protein unfolding observed during CIU.

Despite all the efforts to accurately model the structures, it is inevitable that there remains ambiguity in structural modeing in the IM-MS community. This is largely due to the nature of MS that it does not generate direct observables that can be used to build high resolution models. Study has been done where the ambiguity associated with CCS is evaluated and suggested the more refined simulations set up (i.e all-atom over coarse-grained MD) reduce the ambiguity of the structures determined using IM-MS data.[5] However, this doesn't completely eliminate the ambiguity. Thus, detailed analysis approach is necessary to determine the structure that are relevant to the study. Some studies utilize the simulation energy to select the final model. More widely used method utilized RMSD-based structural clustering to sort the models based on their structural similarities. This method, although still uncertainty to the structures, provide more credibility as the structures are sorted based on their similarities.

While this thesis has focused on utilizing activation and fragmentation data from native IM-MS and CIU-ECD as restraints for the modeling workflow, the uncerntainty still exists. A new

experimental technique has emerged to leverage the gap of uncertainty in gas-phase protein structures. The method that the field can take advantage of is the soft-landing of gaseous protein ions.[6-8] Studies have demonstrated that the protein ions landed in the matrix preserved native structures in the gas-phase. Furthermore, the 3D reconstructions from individual particles provide additional visualization which can be compared with models generated from MD simulations. In this thesis, we've utilized RMSD-based clustering to assign structural ensembles to our models. While this method is successful in sorting models based on structural similarity, there haven't been any direct observations that can be utilized to compare our clustered models. Incorporating the reconstruction from soft-landing native MS into our integrative modeling approach could enhance our understanding of gas-phase protein structures with confidence. Specifically, the addition of such data could greatly improve our understanding of the structural diversity that we observed with IgG1 and could confirm whether the hinge collapse observed in our models is consistent with the reconstruction from soft-landing data.

5.2 Data Analysis Tools for Native MS and Gas-phase Protein Unfolding

In Chapter 4, we've developed and demonstrated the automation of CIU data analysis to improve the data analysis throughput including various raw data processing features, such as automated peak detection and CCS calibration, greatly improved the data analysis time. Furthermore, multi-dimensional classification using cyclic IM-MS data has shown better classification prediction than single-dimension classification. While we aimed to build a software tool to accommodate various instrument platforms, more updates could be beneficial to keep up with growing IM technology. For instance, the CCS calibration approaches built into CIUSuite 3 are only feasible with DTIM and linear drift cell TWIM. However, high-resolution IM techniques are developed such as cyclic IM-MS[9] and structures for lossless ion manipulations (SLIM)[10,11]

necessitating the new CCS calibration approaches that are accurate and compatible with these IM technologies.

With growing interest and advancement in CIU technologies, various systems are being explored including non-standard protein samples expressed in the lab, RNA, and DNA samples. With such a wide range of sample types being studied using CIU, further optimization, and inclusion of user-interactive feature skipping could increase the information content being extracted and analyzed. The current feature skipping only allows analysis of ground and the most unfolded features to be considered for downstream CIU50 analysis. However, any fingerprints with 4 or more features could utilize user-selective feature skipping where one particular feature within the fingerprint can be skipped. This could avoid skipping the entire intermediate features and provide more information content regarding gas-phase protein stability. Furthermore, variable thresholding for feature detection could be useful for fingerprints that exhibit parallel features across consecutive collision voltages. With the current feature detection algorithm, only the highest intensity drift time feature is considered and selected as a feature. However, the inclusion of multiple detection by varying drift time thresholds could fit features that appear in parallel which could provide valuable information regarding how different unfolding states vary in stability.

5.3 References

(1) Ewing, S. A.; Donor, M. T.; Wilson, J. W.; Prell, J. S. Collidoscope: An Improved Tool for Computing Collisional Cross-Sections with the Trajectory Method. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (4), 587–596.

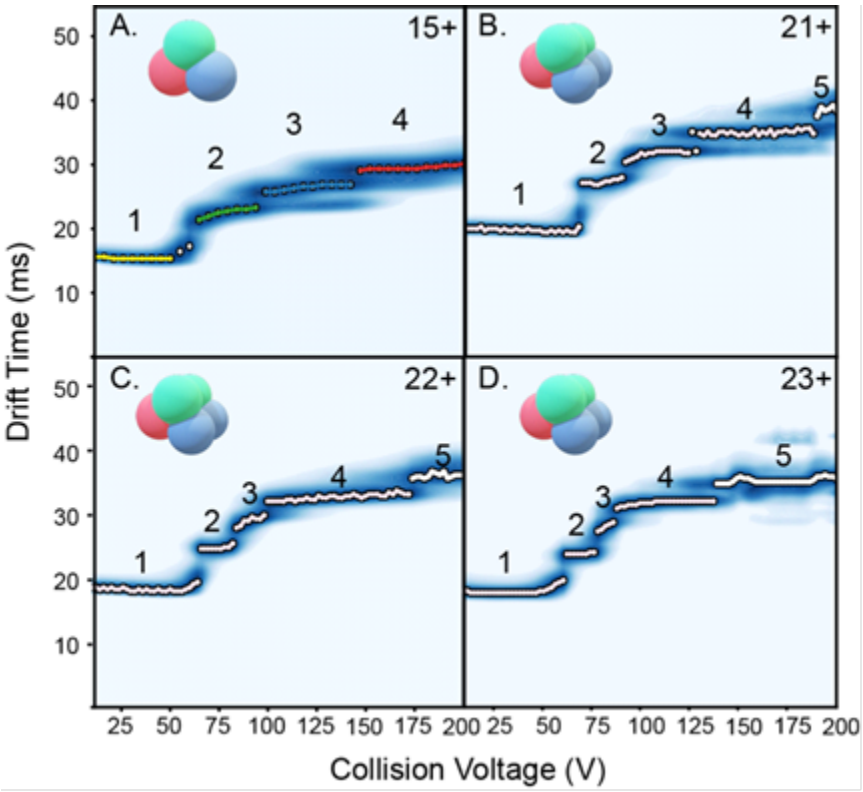
(2) Morrison, L. J.; Brodbelt, J. S. Charge Site Assignment in Native Proteins by Ultraviolet Photodissociation (UVPD) Mass Spectrometry. *Analyst* **2016**, *141* (1), 166–176.

(3) Zhou, M.; Liu, W.; Shaw, J. B. Charge Movement and Structural Changes in the Gas-Phase Unfolding of Multimeric Protein Complexes Captured by Native Top-Down Mass Spectrometry. *Anal. Chem.* **2020**, *92* (2), 1788–1795.

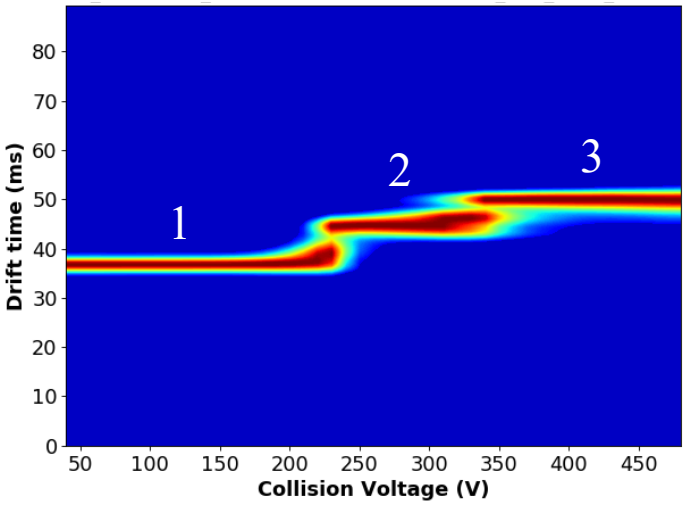
(4) Rolland, A. D.; Biberic, L. S.; Prell, J. S. Investigation of Charge-State-Dependent Compaction of Protein Ions with Native Ion Mobility-Mass Spectrometry and Theory. *J. Am. Soc. Mass Spectrom.* **2022**, *33* (2), 369–381.
(5) Eschweiler, J. D.; Frank, A. T.; Ruotolo, B. T. Coming to Grips with Ambiguity: Ion Mobility-Mass Spectrometry for Protein Quaternary Structure Assignment. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (10), 1991–2000.
(6) Westphall, M. S.; Lee, K. W.; Salome, A. Z.; Lodge, J. M.; Grant, T.; Coon, J. J. Three-Dimensional Structure Determination of Protein Complexes Using Matrix-Landing Mass Spectrometry. *Nat. Commun.* **2022**, *13* (1), 1–6.
(7) Salome, A. Z.; Lee, K. W.; Grant, T.; Westphall, M. S.; Coon, J. J. Matrix-Landing Mass Spectrometry for Electron Microscopy Imaging of Native Protein Complexes. *Anal. Chem.* **2022**, *94* (50), 17616–17624.
(8) Lee, K. W.; Salome, A. Z.; Westphall, M. S.; Grant, T.; Coon, J. J. Onto Grid Purification and 3D Reconstruction of Protein Complexes Using Matrix-Landing Native Mass Spectrometry. *J. Proteome Res.* **2023**, *22* (3), 851–856.
(9) Giles, K.; Ujma, J.; Wildgoose, J.; Pringle, S.; Richardson, K.; Langridge, D.; Green, M. A Cyclic Ion Mobility-Mass Spectrometry System. *Anal. Chem.* **2019**, *91* (13), 8564–8573.
(10) Dodds, J. N.; Baker, E. S. Ion Mobility Spectrometry: Fundamental Concepts, Instrumentation, Applications, and the Road Ahead. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (11), 2185–2195. https://doi.org/10.1007/s13361-019-02288-2.
(11) Deng, L.; Ibrahim, Y. M.; Hamid, A. M.; Garimella, S. V. B.; Webb, I. K.; Zheng, X.; Prost, S. A.; Sandoval, J. A.; Norheim, R. V.; Anderson, G. A.; Tolmachev, A. V.; Baker, E. S.; Smith, R. D. Ultra-High Resolution Ion Mobility Separations Utilizing Traveling Waves in a 13 m Serpentine Path Length Structures for Lossless Ion Manipulations Module. *Anal. Chem.* **2016**, *88* (18), 8957–8964.

Appendices

Appendix I: Supporting Information for Chapter 2



Appendix Figure I-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.



Appendix Figure I-2. The CIU fingerprint of D12 monomer of 13with three feature observed.

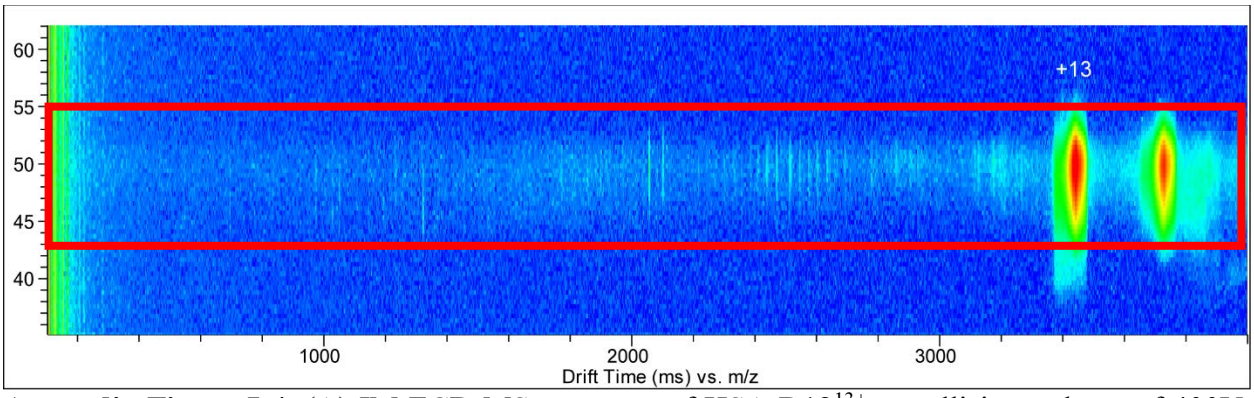
(A)

DAHKSEVAHRFKDLGEENFKALVLIAFAQYLQQCPFEDHVKLVNEVTEFAKTCVADES
AENCDKSLHTLFGDKLCTVATLRETYGEMADCCAKQEPERNECFLQHKDDNPNLPRLV
RPEVDVMCTAFHDNEETFLKKYLYEIARRHPYFYAPELLFFAKRYKAAFTECCQAADK
AACLLPKLDELRDEGKASSAKQGLKCASLQKFGERAFKAWAVARLSQRFPKAEFAEVS
KLVTDLTKVHTECCHGDLLECADDRADLAKYICENQDSISSKLKECCEKPLLEKSHCIAE
VENDEMPADLPSLAADFVGSKDVCKNYAEAKDVFLGMFLYEYARRHPDYSVVLLLRL
AKTYETTLEKCCAAADPHECYAKVFDEFKPLVEEPQ

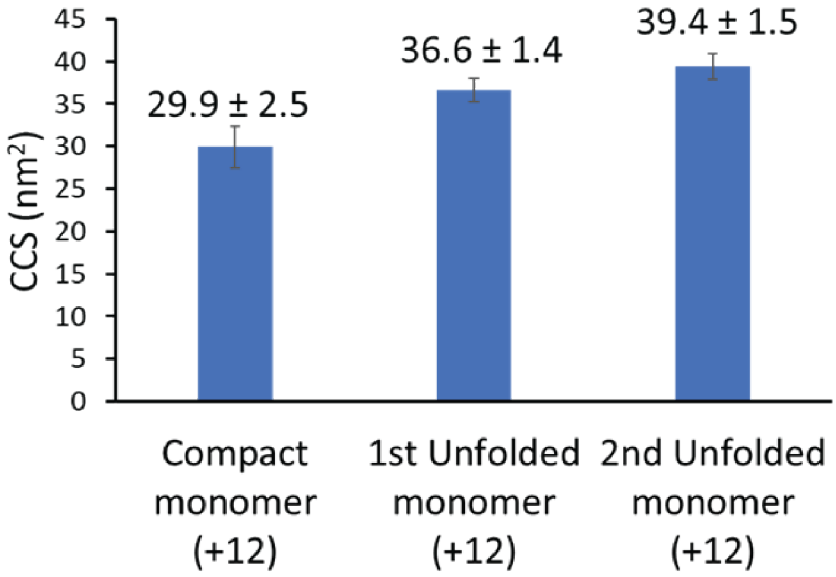
(B)

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

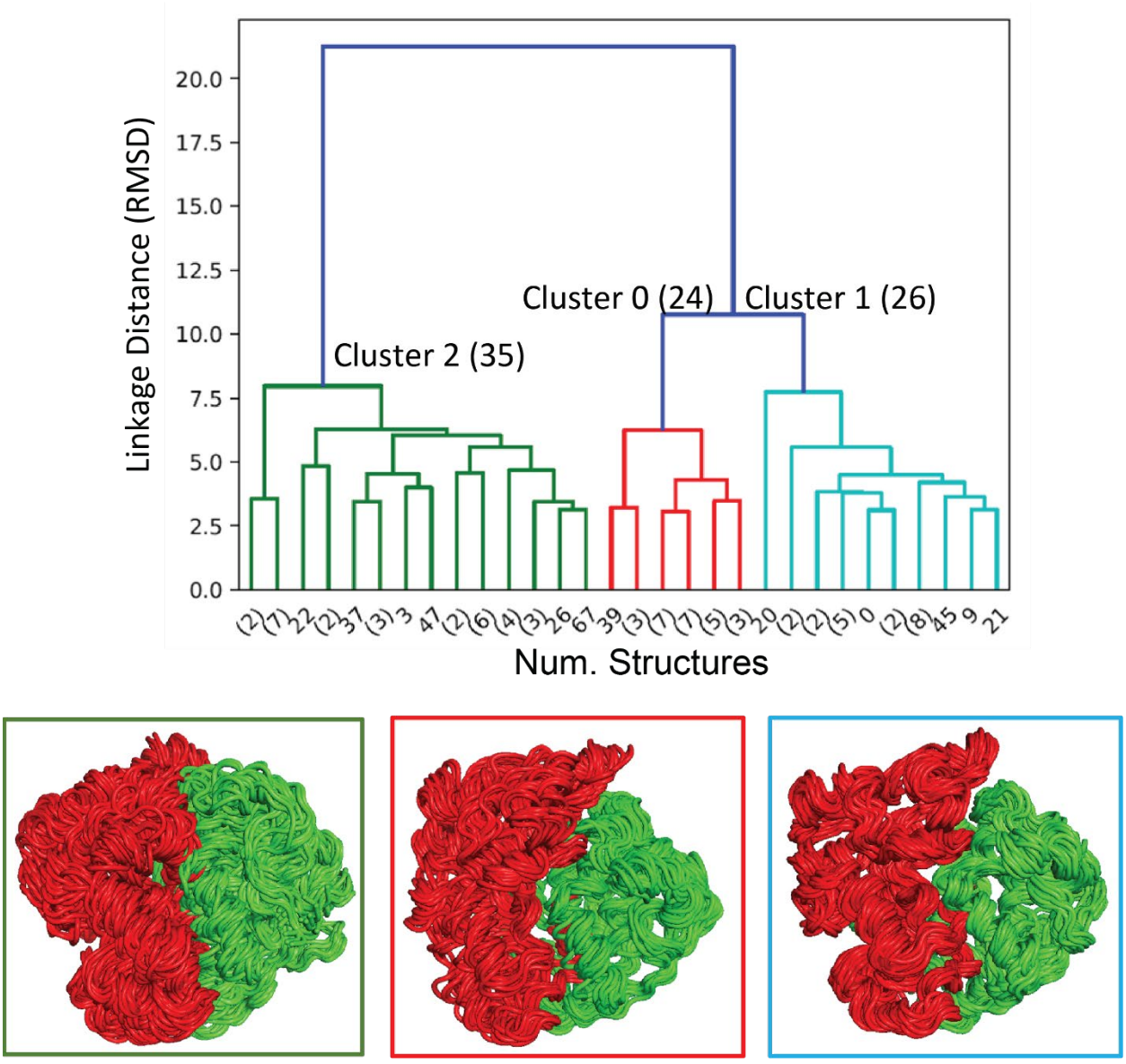
Appendix Figure I-3. (A) HSA D12 sequence. The construct encompasses residues 1-385 of HSA. **(B)** Terminal fragments matches with sequence, m/z value, charge, error tolerance and ion type.



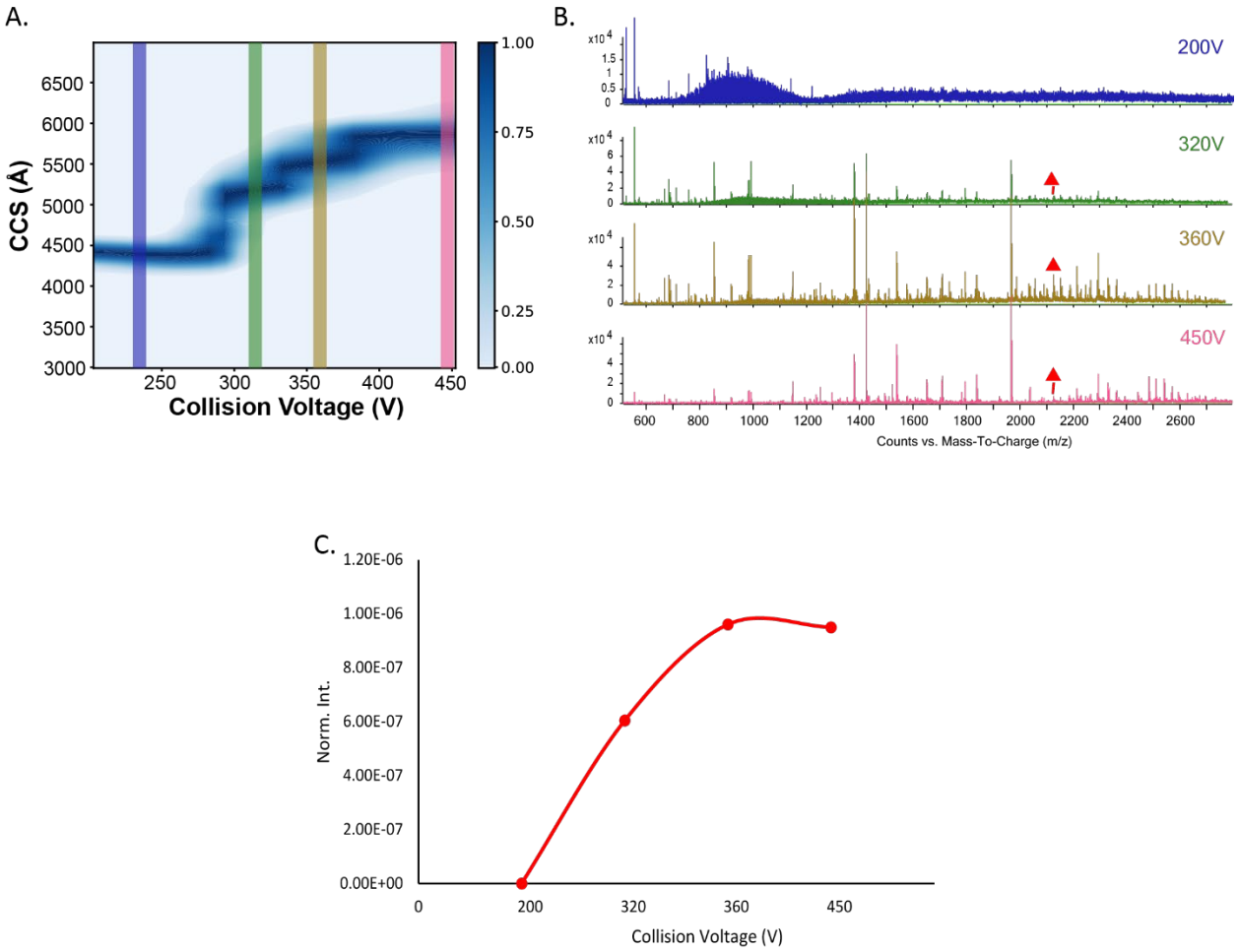
Appendix Figure I-4. (A) IM-ECD-MS spectrum of HSA D12^{13+} at collision voltage of 400V. The new signals that appeared at this voltage do not come from in-source CID. If that was the case the CID fragment would have been separated in the IM dimension. However, in the spectrum above the fragments occupy the same drift bins as the precursor ion.



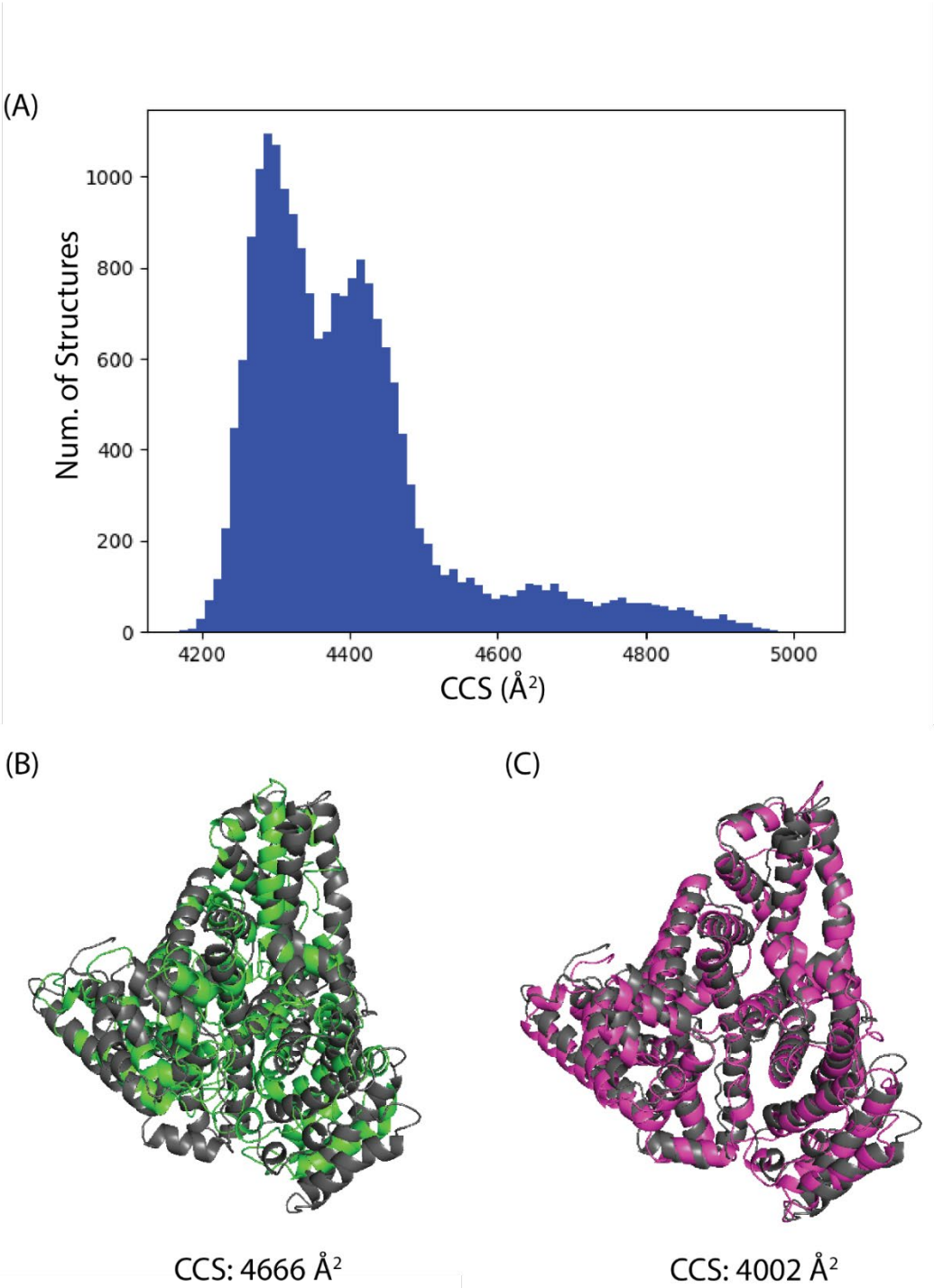
Appendix Figure I-5. Experimental CCS values of HSA D12 monomer obtained from Waters G2.



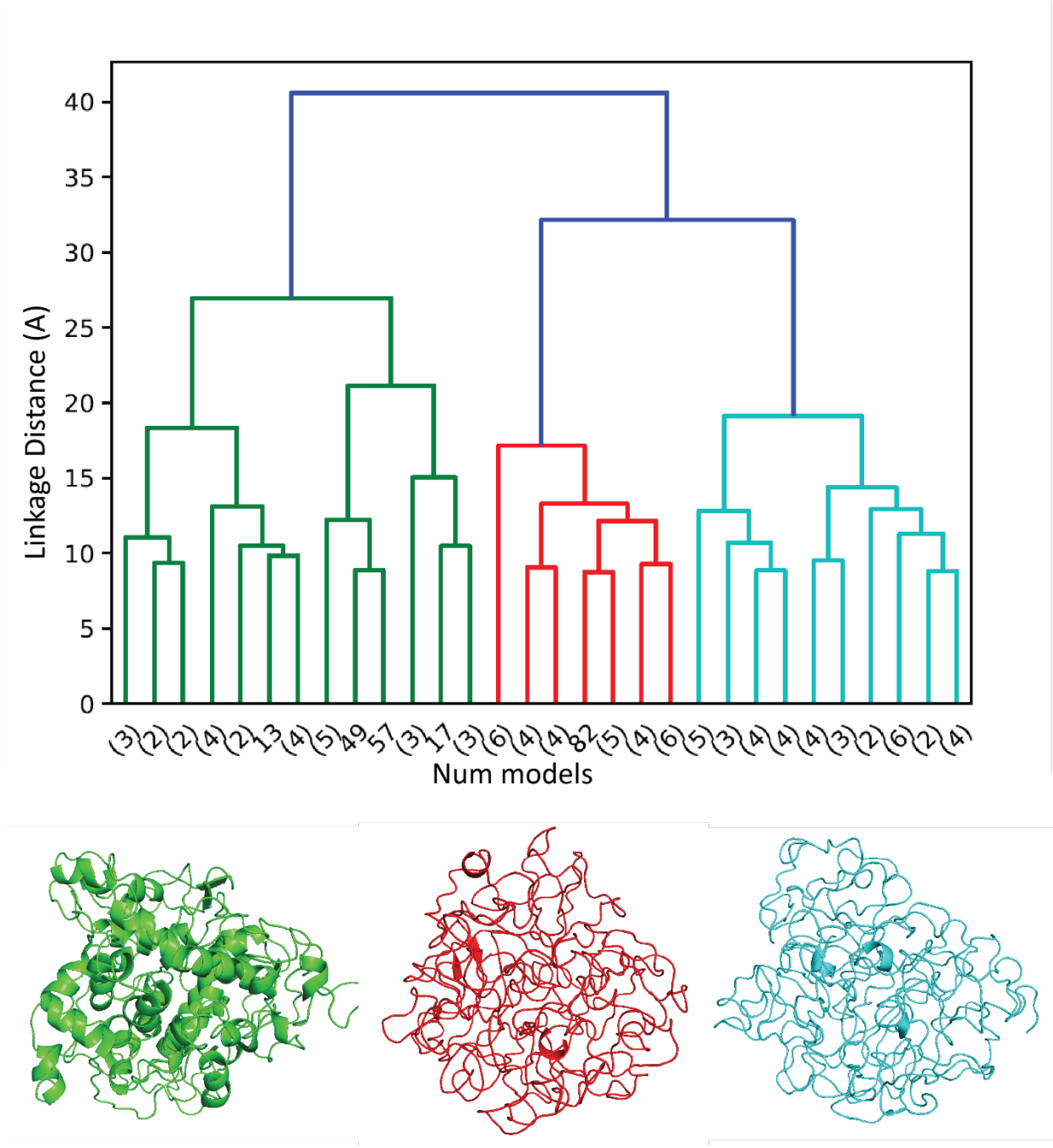
Appendix Figure I-6. Clustering dendrogram of compact HSA D12 monomer at replica 12 (557.384 K) which is the lowest temperature replica at which models that agreed with experimental CCS was observed. Models from each cluster are also displayed and annotated.



Appendix Figure I-7. CIU-ECD experiment of +15 HSA. A). CIU fingerprint showing the four features observed in the Synapt G2 dataset. B) Mass spectrum at each voltage where ECD fragmentation was obtained. Data was extracted only at the drift bins containing the precursor and ECD fragmentation. C) Intensity vs Collision Voltage used for Internal fragment +4 c_{174}-$zdot_{436}$, which is located in the boundary of domain 1 and domain 2



Appendix Figure I-8. Compaction of full-length HSA (A) CCS distribution of full-length HSA from REMD replica 12 which is the temperature at which the lowest CCS model was observed. Despite the rigor of REMD, CHARMM36 is inefficient at achieving the level of compaction that agree with experimental IM-MS data. Simple equilibration MD simulations at 300K were performed using CHARMM36 **(B)**, and GROMOS9643a2 **(C)** force fields where gray model represents the original database PDB structure and the colored model represents the final equilibrated HSA model.



Appedix Figure I-9. Clustering dendrogram of full-length HSA compaction with models from each cluster.

Appendix Table I-1. Internal fragment matches from CIU-ECD experiments of HSA D12^{13+}. 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	-0.358	100	67	72
4	2241.295	c_{69}-y_{230}	chhsshl	3	-0.486	100	75	78

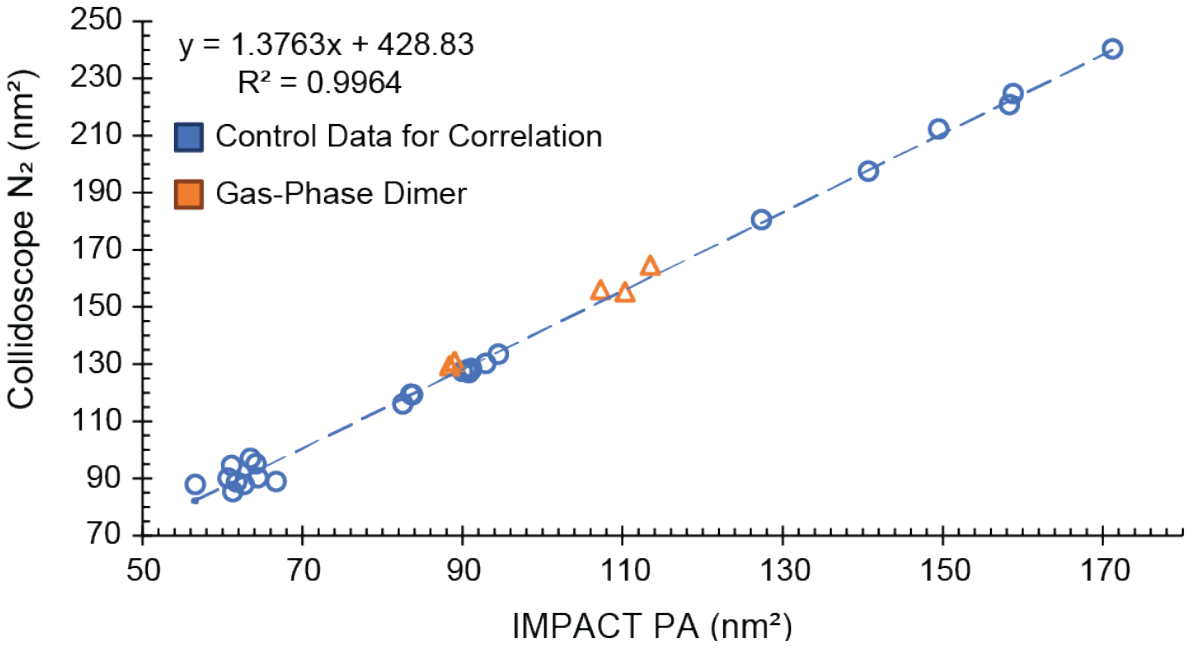
Appendix Table I-2. Internal fragment matches from CIU-ECD experiments of HSA^{17+}. The following information for each internal fragment identified is shown: *m/z* values, charge state, error tolerance, the number of disulfides bonds (ss_count), and the modification for the reduced cysteines present (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
5	2100.776	c_{76}-$z^{\bullet}_{168}$	shl, hl	0	0.76	100	44	72
4	2291.608	c_{267}-$z^{\bullet}_{352}$	sshl, hl	0	0.836	100	67	83

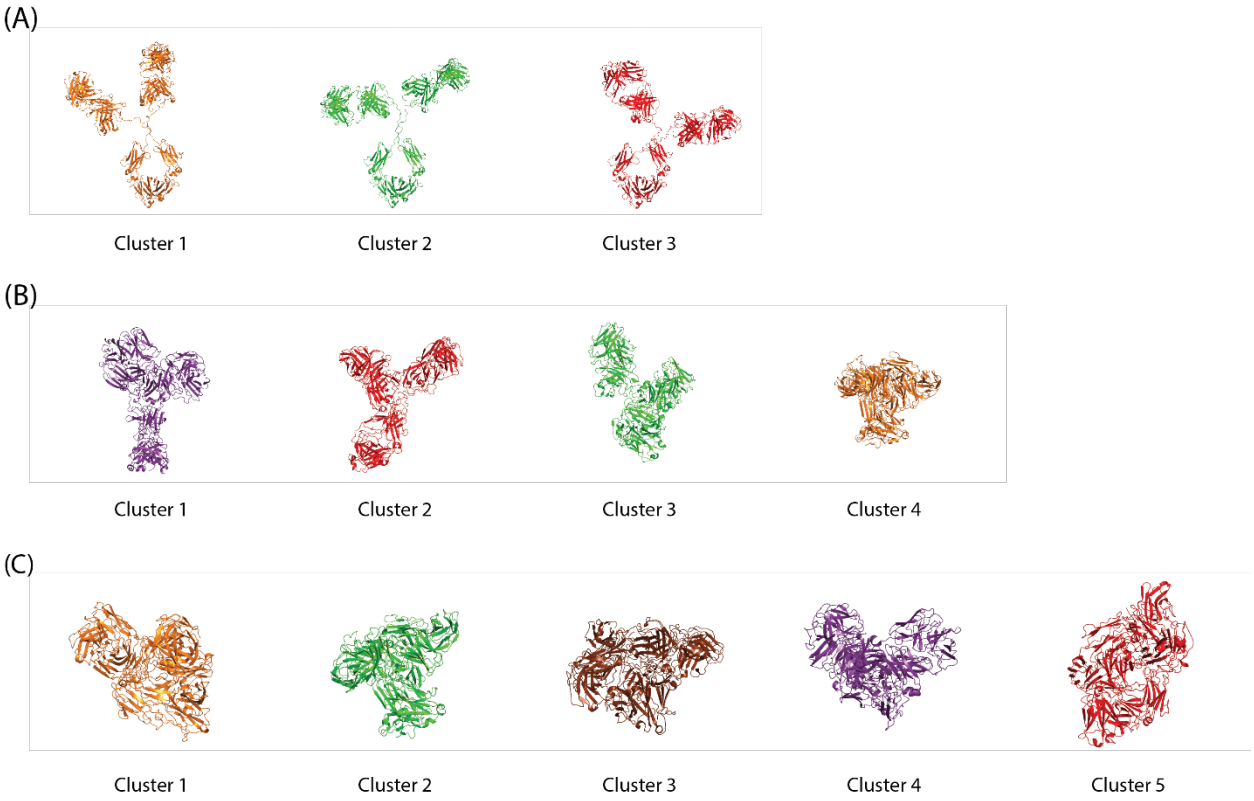
Appendix Table I-3. The number of models within each unfolded feature from steered MD of HSA in domain 3. Three different steered MD simulations were performed within the domain 3 region of HSA to validate the specificity and efficiency of unfolding during the ECD-guided steered MD simulation. The residues that were allowed to unfold during each steered MD are also noted in the table.

Feature	2 (5174Å^2)	3 (5445Å^2)	4 (5922Å^2)
Domain 3 ECD-guided (438-527)	6104	259	0
Domain 3 internal non-specific (458-547)	6762	79	0
Domain 3 terminal non-specific (496-585)	4792	1950	5

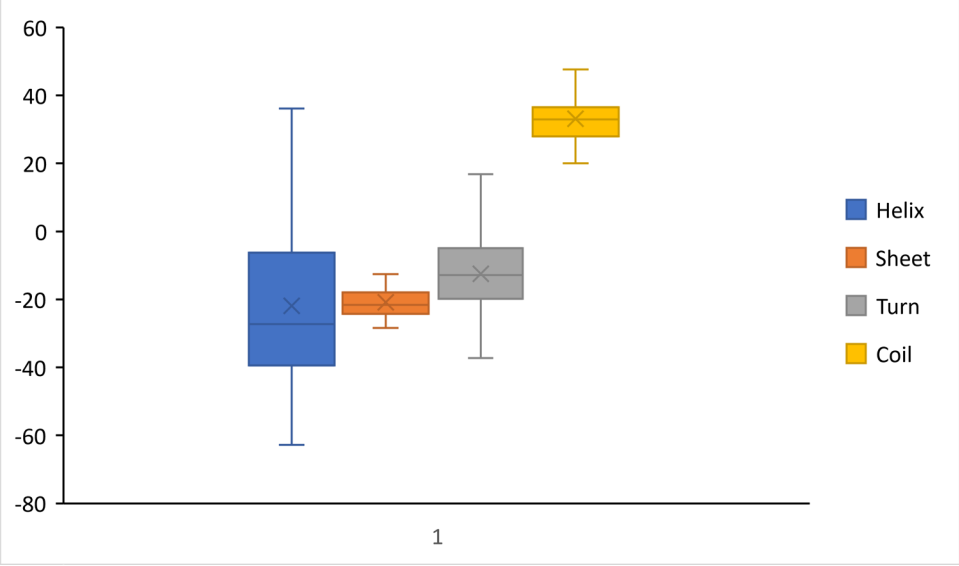
Appendix II: Supporting Information for Chapter 3



Appendix Figure II-1. Correction plot to convert IMPACT PA CCS into TM N_2 CCS using models from IgG1 monomers and dimers in both solution and gas-phase.



Appendix Figure II-2. Example models from each cluster after clustering analysis of (A) 120 IPET models, **(B)** Final compacted models after 120 MD simulations, and **(C)** IM-filtered gas-phase models.



Appendix Figure II-3. Percent changes in each secondary structure from solution to gas-phase models of IgG1.

Appendix Table II-1. CCS of all 120 PDB Models in Solution and Gas-phase.

IgG1 ID	Solution TM CCS	Gas TM CCS	% Compaction
p1	12670.14	9225.986	27.18327
p2	12459.83	9065.627	27.24115
p3	12382.72	8863.88	28.41735
p4	12709.64	9054.069	28.76221
p5	12620.5	9133.66	27.62839
p6	12015.95	8910.11	25.84763
p7	12501.84	9360.838	25.12429
p8	12252.96	9376.252	23.47769
p9	12474.3	9757.438	21.77969
p10	12094.41	9152.925	24.32102
p11	12280	9094.819	25.93795
p12	12337.7	9579.119	22.35897
p13	12318.87	9134.762	25.84741
p14	12509.75	9409.886	24.77962
p15	12315.14	8944.599	27.36909
p16	12273.02	8523.188	30.55346
p17	12390.9	8960.129	27.68784
p18	12006.68	9473.345	21.09935
p19	12689.82	8686.489	31.54756
p20	12382.98	8894.371	28.17264
p21	12220.61	9063.814	25.83171
p22	12498.82	9149.211	26.79938
p23	12200.16	8651.353	29.08821
p24	12173.34	9011.949	25.96982
p25	11991.85	8706.347	27.39782
p26	11855.71	9450.975	20.28336
p27	12008.06	8916.721	25.74388
p28	12460.99	9556.634	23.30757
p29	12523.04	8515.465	32.00163
p30	12033.75	9140.535	24.04251
p31	12137.28	9184.359	24.32935
p32	12547.99	9231.863	26.42757
p33	12220.79	9753.006	20.19333
p34	12207.74	9398.473	23.01218
p35	12233.92	8617.617	29.55964
p36	11506.46	9102.188	20.89497
p37	12257.93	9173.393	25.16362
p38	11798.55	8846.144	25.02347
p39	12063	9236.02	23.43512
p40	12057.97	8772.038	27.25112
p41	12366.08	9323.632	24.60316

p42	12213.69	9508.73	22.14698
p43	11854.93	9268.574	21.81672
p44	12144.97	9448.886	22.19916
p45	12346.25	8509.651	31.07504
p46	12360.63	9219.955	25.40872
p47	12219.44	9446.848	22.69004
p48	12533.07	8186.46	34.68112
p49	12605.73	9291.237	26.29357
p50	12577.82	9226.009	26.64858
p51	12357.56	9147.28	25.97826
p52	12432.36	9347.899	24.80994
p53	12396.55	9087.998	26.6893
p54	12429.85	9796.211	21.18804
p55	12414.97	8951.389	27.89844
p56	12568.74	9150.815	27.19383
p57	12322.15	9235.812	25.04709
p58	13119.88	8736.298	33.41177
p59	12605.15	8410.689	33.2758
p60	12564.62	9397.791	25.20432
p61	12405.6	9401.702	24.21403
p62	12474.51	9016.03	27.72437
p63	12579.89	8905.636	29.20735
p64	11783.9	8959.633	23.96715
p65	11986.11	9335.101	22.11733
p66	12426.37	8737.292	29.68748
p67	12403.6	9206.349	25.7768
p68	11818.92	8699.714	26.39167
p69	12761.99	9040.73	29.15893
p70	12610.9	9776.958	22.47215
p71	12103.77	9520.974	21.3388
p72	12323.01	9345.335	24.16353
p73	12256.23	8793.968	28.24901
p74	12194.03	9141.988	25.02901
p75	12229.79	9285.439	24.07526
p76	12043.63	7706.447	36.01224
p77	12425.21	7995.99	35.64702
p78	12238.25	9036.892	26.15864
p79	12137.1	8642.823	28.79007
p80	12470.86	9136.786	26.73493
p81	12038.44	9407.966	21.85062
p82	12414.82	9714.453	21.75114
p83	12231.44	9341.553	23.6267
p84	11836.68	8760.671	25.98708

p85	12324.69	9228.042	25.12555
p86	12170.63	9209.972	24.32624
p87	12018.74	9497.02	20.98156
p88	12244.47	9228.679	24.62983
p89	12446.17	9486.404	23.78054
p90	11687.11	9183.417	21.42266
p91	12121.6	8950.089	26.16413
p92	12719.17	9283.219	27.01395
p93	11796.07	8890.617	24.63071
p94	12334.24	9151.387	25.80499
p95	12808.07	8839.685	30.9835
p96	12449.03	9267.821	25.55388
p97	12643.71	9389.899	25.73465
p98	12271.93	9182.048	25.17848
p99	12259.63	9107.84	25.70869
p100	12618.04	8992.771	28.73084
p101	12848.86	9104.669	29.14025
p102	12785.26	8477.874	33.69023
p103	12783.64	8915.672	30.25715
p104	12435.82	9297.129	25.23913
p105	12343.96	8837.538	28.40596
p106	11926.89	9478.101	20.53168
p107	12255.37	8771.45	28.42767
p108	12134.5	9498.656	21.72193
p109	12407	9051.784	27.04291
p110	12386.03	9489.633	23.38441
p111	12285.99	9294.165	24.35149
p112	12525.06	9138.939	27.03476
p113	11811.66	9521.251	19.39108
p114	12326.86	8124.909	34.08777
p115	11979.34	9692.656	19.08856
p116	12846.55	9180.253	28.53918
p117	12232.25	8113.189	33.6738
p118	11908.94	8985.753	24.54617
p119	12761.14	9312.513	27.02443
p120	12636.85	9098.037	28.00391

Appendix Table II-2. Clustering Results from IM-filtered Models.

Cluster Number	Number of Structures	IgG1 ID
Cluster 1	446	p117
Cluster 2	310	p114
Cluster 3	261	p77
Cluster 4	43	p48

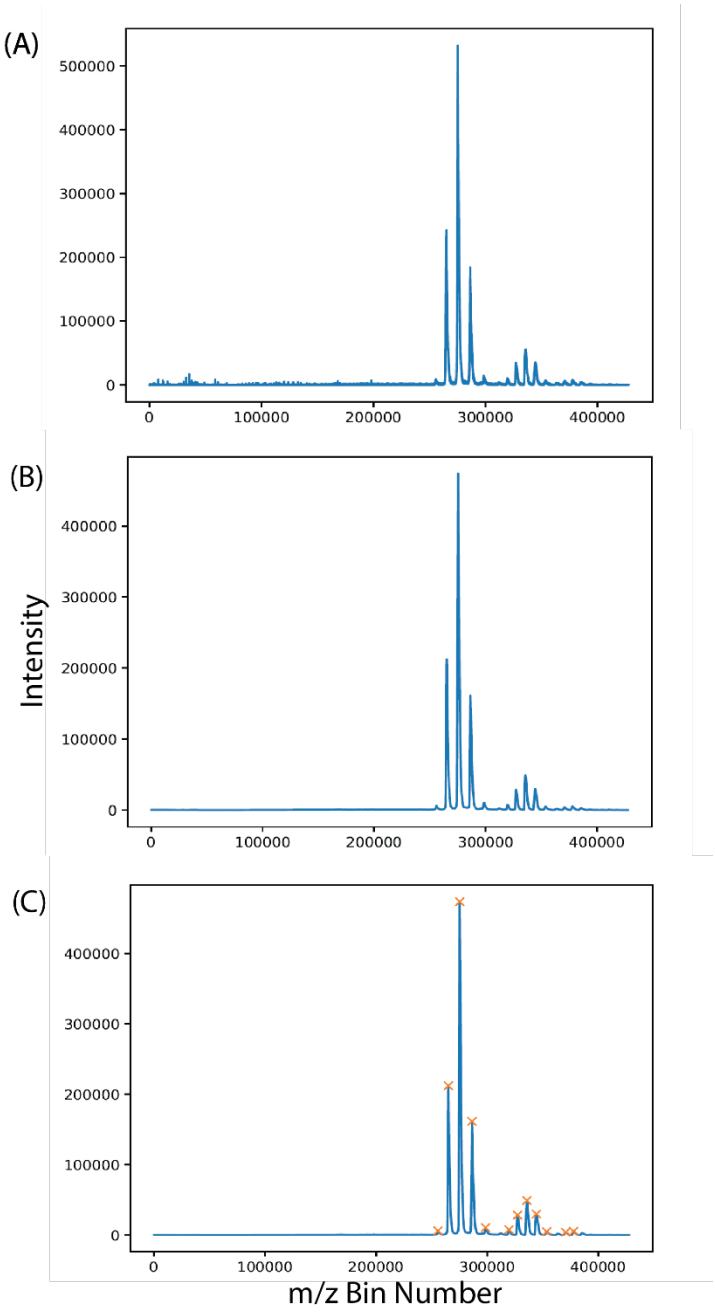
Cluster 5	2	p76

Appendix Table II-3. Changes in Secondary Structures after Gas-phase Compaction

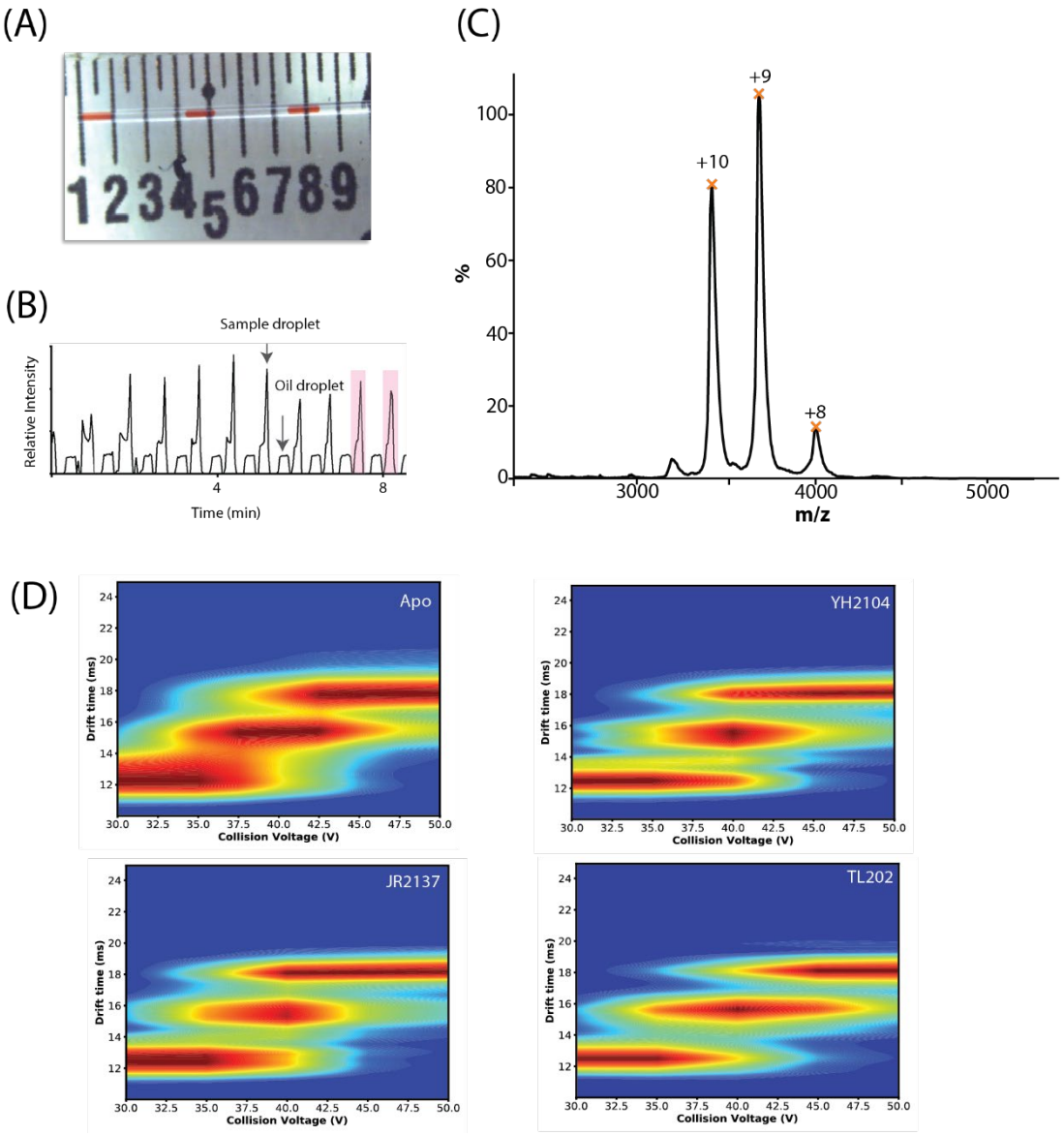
IgG1 ID	Helix	Sheet	Turn	Coil
p3	-12.1	-18.4	-20.5	29.6
p6	-27.5	-18.7	-3.7	26.4
p15	-62.7	-23.3	-12.1	42.8
p16	-30.2	-28.4	-15.0	47.6
p17	-37.2	-18.1	-23.1	33.0
p19	-55.6	-21.7	-13.0	36.2
p20	-10.0	-15.6	-26.9	28.5
p23	-32.5	-22.8	-6.5	35.1
p25	-3.1	-18.3	-24.6	32.2
p27	-12.5	-27.4	0.0	32.9
p29	-5.7	-23.9	-18.2	39.9
p35	-39.4	-14.8	-14.9	27.2
p38	26.5	-21.9	14.4	20.1
p40	12.9	-22.1	-33.6	37.7
p45	-25.0	-27.2	3.9	35.1
p48	-39.1	-24.7	-17.0	43.1
p55	-15.0	-15.7	-6.5	24.0
p58	-45.5	-15.2	-4.5	26.6
p59	-47.8	-22.0	-11.2	36.5
p63	-6.3	-15.9	-29.0	27.5
p64	-34.8	-18.0	-12.9	31.1
p66	-9.1	-19.1	-13.1	28.4
p68	-12.2	-14.3	-14.2	25.6
p73	-10.0	-22.7	-11.6	33.3
p76	-27.3	-25.7	-4.9	35.5
p77	0.0	-19.2	-12.8	29.6
p79	-36.7	-13.2	-25.0	30.6
p84	-2.7	-21.6	-4.1	27.4
p91	-42.5	-20.7	-22.4	37.3
p93	-30.8	-21.3	2.0	29.5
p95	36.1	-20.1	-37.3	33.6
p100	-32.5	-24.3	-12.8	36.3
p102	-41.9	-12.6	-19.8	27.9
p103	-44.2	-25.4	16.8	35.0
p105	20.0	-16.7	-10.2	26.5
p107	9.1	-26.3	-17.9	41.1
p114	-18.4	-22.4	-10.3	30.6
p117	-57.5	-27.3	-15.5	46.6

p118	-51.9	-24.7	0.0	42.5

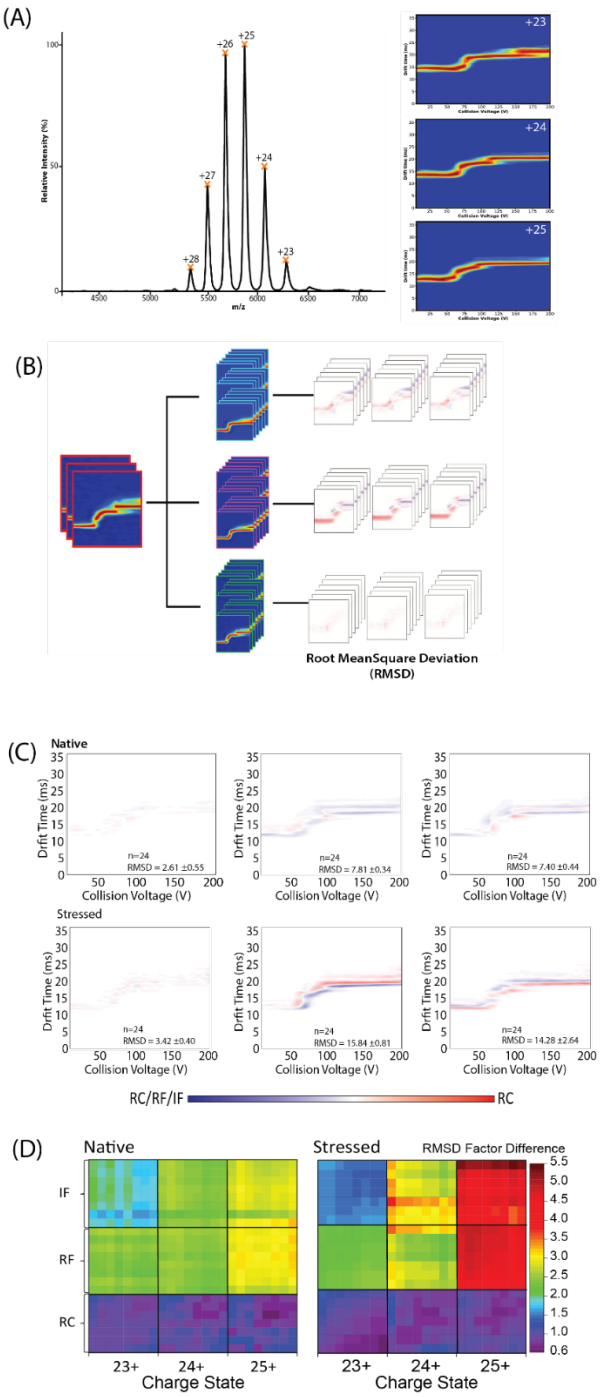
Appendix III: Supporting Information for Chapter 4



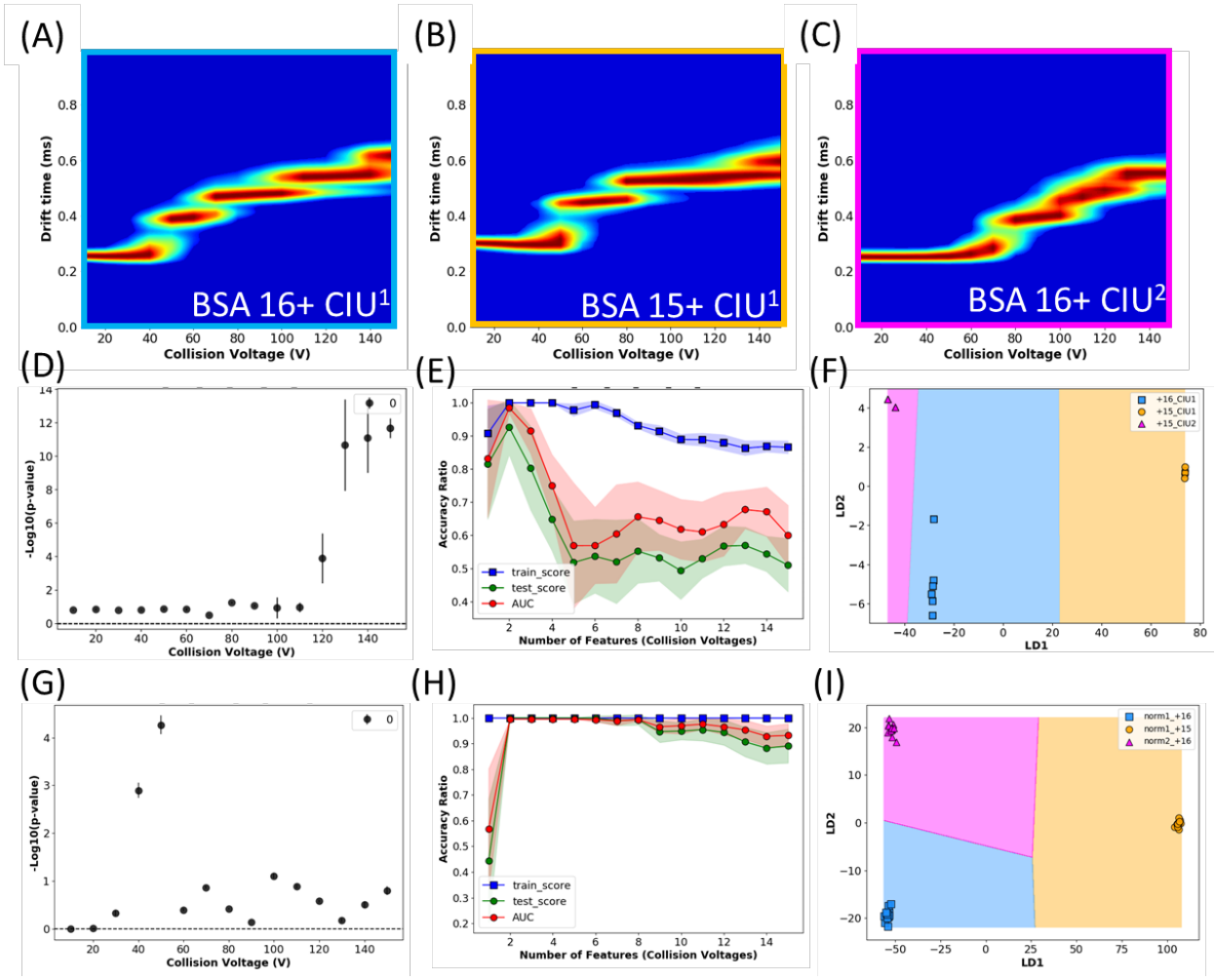
Appendix Figure III-1. Individual steps in peak detection algorithm. (A) Raw mass spectrum of HSA. **(B)** Mass spectrum of HSA after Savitzky-Golay smoothing. **(C)** Peaks detected by the algorithm are highlighted in orange within the mass spectrum of HSA.



Appendix Figure III-2. Peak detection of SIRT5 and CIU fingerprints of various ligand-bound SIRT5. (A) Microfluidics device showing aqueous sample droplets (clear) separated by oil droplets (red). **(B)** Elution of sample droplets shows differences in intensities between sample droplets and oil droplets. **(C)** Mass spectrum of apo-SIRT5 with peaks detected by the algorithm annotated in orange. **(D)** CIU fingerprints of apo and various ligand-bound SIRT5 generated by peak detection outputs.



Appendix Figure III-3. Application of peak detection algorithm on large-scale CIU data sets of biotherapeutics and biosimilars. (A) Mass spectrum of Remicade (RC) with peaks detected by the algorithm annotated in orange (left); CIU fingerprints generated using the low and high m/z bounds determined by the algorithm. **(B)** Overview of difference analysis performed between RC and RC, Renflexix (RF), and Inflectra (IF) across multiple charge states and replicates. **(C)** Total difference analysis plots with respect to RC in native states (top row) and stressed states (bottom row). **(D)** Heat map of RMSD values from all total difference analyses performed for both native states (left) and stressed states (right).



Appendix Figure III-4. Multi-dimensional classification of BSA. (A-C) Fingerprints of three classes used to build the classification scheme. **(D-F)** Classification results of UFS, cross-validation, and final classification scheme, respectively, when only 1D preprocessing on the CV axis was performed. **(G-I)** Classification results of UFS, cross-validation, and final classification scheme, respectively, when 2D preprocessing on CV and drift time axes were performed. Overall, the best scoring voltage shifts from 150V to 50V when 2D preprocessing is performed, and the accuracy of training and test data is improved.

Appendix Table III-1. Peak detection results of HSA. The table summarizes the peaks identified in Appendix Figure III-1 along with the low and high m/z bounds which can be used for extracting drift time information during the CIU extraction process.

Peak m/z	Peak m/z bin	Low m/z bound	High m/z bound	Low m/z bin bound	High m/z bin bound
3506.712	255692	3492.005	3521.449	254976.5	256407.5
3701.764	265050	3687.623	3715.931	264379.6	265720.4
3919.661	275217	3903.917	3935.437	274491.9	275942.1
4164.916	286328	4147.389	4182.48	285544.8	287111.2
4442.409	298512	4424.1	4460.757	297720.2	299303.8
4943.621	319591	4921.65	4965.642	318689.6	320492.4
5132.867	327270	5108.815	5156.975	326302	328238
5344.333	335685	5316.625	5372.113	334592.1	336777.9
5567.811	344399	5537.1	5598.607	343212.5	345585.5
5806.339	353509	5777.483	5835.267	352416.9	354601.1
6277.781	370962	6247.034	6308.603	369842.8	372081.2
6469.124	377861	6434.68	6503.66	376626	379096

Appendix Table III-2. Probabilities of class assignments with and without multi-dimensional classification. Based on the classification scheme built using BSA, HSA, RSA, and GSA, probabilities of class assignment were calculated with and without multi-dimensional classification.

		Assigned Classes			
	Input Data Class	BSA	HSA	RSA	GSA
CIU[1] only	BSA	0.837	0.0319	0.058	0.073
	HSA	0.033	0.840	0.052	0.075
	RSA	0.066	0.071	0.835	0.029
	GSA	0.059	0.074	0.027	0.839
CIU[1] and CIU[2] (multi-dimensional)	BSA	0.868	0.039	0.060	0.033
	HSA	0.021	0.904	0.028	0.048
	RSA	0.042	0.041	0.902	0.048
	GSA	0.026	0.042	0.017	0.916

Appendix Table III-3. Single-field CCS calibration function values using BSA as the calibrant. Linear regressions were performed using the OLS method. The coefficient and t_{fix} (intercept) were the direct output of the OLS method. Beta value is determined by factoring in charge and masses of buffer gas and BSA based on Equation 4-6.

BSA Charge State	Coefficient	Beta	T_{fix}
+16	124.246	7.767	28.230
+17	131.863	7.758	38.838
+18	139.464	7.750	44.393

Appendix Table III-4. Summary of standard proteins CCS values calibrated using both tunemix and BSA. All protein samples were prepared and ionized to preserve their native structure.

Protein	Mass (Da)	Charge State	$CCS_{tunemix}$ (Å^2)	CCS_{BSA} (Å^2)
BSA	66870	14	4408.3	4445.33
BSA	66870	14	4387.7	4424.674
BSA	66870	14	4384.5	4421.413
BSA	66870	14	4386.6	4423.587
BSA	66870	14	4396.4	4433.371
BSA	66870	15	4448.6	4485.943
BSA	66870	15	4453.5	4499.92
BSA	66870	15	4456.7	4494.096
BSA	66870	15	4441.6	4478.954
BSA	66870	15	4446.2	4483.613
BSA	66870	16	4531.6	4569.421
BSA	66870	16	4520.9	4559.481
BSA	66870	16	4520.4	4558.238
BSA	66870	16	4520.4	4558.238
BSA	66870	16	4521.7	4559.481
BSA	66870	17	4581.4	4619.585
BSA	66870	17	4559.3	4598.463
BSA	66870	17	4556.3	4594.503
BSA	66870	17	4573.4	4611.664
BSA	66870	17	4580	4618.265
Cytochrome C	12360	6	1450.2	1473.12
Cytochrome C	12360	6	1449.1	1473.12
Cytochrome C	12360	6	1445.7	1476.378
Cytochrome C	12360	6	1443.1	1473.12
Cytochrome C	12360	7	1492.3	1519.514
Cytochrome C	12360	7	1488.3	1511.911
Cytochrome C	12360	7	1487.1	1512.997
Cytochrome C	12360	7	1485.1	1513.54
B-lac (monomer)	18370	7	1881	1912.516
B-lac (monomer)	18370	7	1873.9	1905.448
B-lac (monomer)	18370	7	1874.4	1905.992
B-lac (monomer)	18370	7	1874.4	1905.992
B-lac (monomer)	18370	8	1932.9	1964.861
B-lac (monomer)	18370	8	1932.3	1964.24
B-lac (monomer)	18370	8	1931.7	1963.619
B-lac (monomer)	18370	8	1937.7	1963.619
B-lac (dimer)	36740	12	3163.7	3197.822
B-lac (dimer)	36740	12	3158.1	3192.232
B-lac (dimer)	36740	12	3156.2	3190.369

B-lac (dimer)	36740	12	3160.9	3195.027
B-lac (dimer)	36740	13	3274	3308.537
B-lac (dimer)	36740	13	3268.9	3303.491
B-lac (dimer)	36740	13	3276.3	3313.584
B-lac (dimer)	36740	13	3274	3308.537
Insulin (monomer)	5810	3	869.7	895.9675
Insulin (monomer)	5810	3	873.5	899.6867
Insulin (monomer)	5810	4	915.5	941.9152
Insulin (monomer)	5810	4	913.3	939.7457
Insulin (dimer)	11620	5	1389.2	1417.626
Insulin (dimer)	11620	5	1391.1	1419.566
Insulin (dimer)	11620	6	1447.1	1475.809
Insulin (dimer)	11620	6	1448.5	1477.205
Ubiquitin	8560	4	1113.07	1140.51
Ubiquitin	8560	4	1115.56	1142.992
Ubiquitin	8560	4	1115.56	1142.992
Ubiquitin	8560	4	1115.56	1142.992
Ubiquitin	8560	4	1115.56	1142.992
Ubiquitin	8560	5	1140.36	1168.116
Ubiquitin	8560	5	1140.36	1168.116
Ubiquitin	8560	5	1140.36	1168.116
Ubiquitin	8560	5	1140.36	1168.116
Ubiquitin	8560	5	1140.36	1168.116
Ubiquitin	8560	6	1194.08	1222.086
Ubiquitin	8560	6	1194.08	1222.086
Ubiquitin	8560	6	1193.61	1221.621
Ubiquitin	8560	6	1193.14	1221.155
Ubiquitin	8560	6	1195.47	1223.482
IgG1	147800	23	7389.6	7433.984
IgG1	147800	23	7416.3	7460.778
IgG1	147800	23	7423.5	7467.923
IgG1	147800	23	7418.1	7462.564
IgG1	147800	23	7419.9	7464.35
IgG1	147800	24	7504.2	7549.08
IgG1	147800	24	7537.7	7582.63
IgG1	147800	24	7545.1	7590.086
IgG1	147800	24	7539.5	7584.494
IgG1	147800	24	7539.5	7584.494
IgG1	147800	25	7591.9	7637.227

IgG1	147800	25	7630.6	7676.058
IgG1	147800	25	7636.5	7681.883
IgG1	147800	25	7636.5	7681.883
IgG1	147800	25	7632.6	7678
IgG1	147800	26	7756.4	7802.26
IgG1	147800	26	7758.4	7804.279
IgG1	147800	26	7746.3	7792.164
IgG1	147800	26	7742.2	7788.126
IgG1	147800	26	7738.2	7784.087
IgG1	147800	27	8113.3	8159.974
IgG1	147800	27	8115.4	8162.071
IgG1	147800	27	8081.9	8128.521
IgG1	147800	27	8117.5	8164.168
IgG1	147800	27	8107	8153.684

Appendix IV: Site-specific Chirality-conferred Structural Compaction of Aβ42

Partially adapted from **Li, G.; Jeon, C. K.; Ma, M.; Jia, Y.; Zheng, Z.; Delafield, D. G.; Lu, G.; Romanova, E. V.; Sweedler, J. V.; Ruotolo, B. T.; Li, L. Site-Specific Chirality-Conferred Structural Compaction Differentially Mediates the Cytotoxicity of Aβ42. *Chem. Sci.* 2023, *14* (22), 5936–5944.**

Introduction

Naturally occurring D-amino acid substitution, also known as amino acid D-epimerization, has been observed in many disease-associated peptides and proteins, including amyloid beta (Aβ), one of the putative biomarkers and drug targets for Alzheimer's disease (AD).[1-6] Aβ is of significant interest due to the prevalence of post-translational D-epimerization in AD brain samples.[7] In brain samples extracted from AD patients, the two most common residues isomerized are aspartic acid and serine.[5,8,9] Although a systematic chirality survey of Asp and Ser residues within Aβ42 is still lacking, emerging data support the prospects of a chiral regulation mechanism capable of rescuing Aβ in cytotoxicity,[3,10-12] which may be further probed in the context of enzymatic regulation and phage display.[13,14] Remaining challenges confronting a systematic survey concerning the chiral effects of Asp and Ser residues of Aβ42 consist not only of those associated with the analytical discrimination and preparation of Aβ42 stereoisomers but also of the inadequate molecular understanding of the combinatorial structural consequences of D-epimerization at all Asp and Ser residues within different domains of Aβ42.

While many prior reports have described technical advancements associated with the chiral discrimination and separation of D-amino acid-containing peptides,[15-23] there remains a dearth of tools capable of targeting Aβ42 stereochemistry. Traditional biophysical techniques frequently suffer from an inability to reconcile the mixture states created from the relatively low structural stability and high aggregation propensity of Aβ42. These challenges are amplified when one maps the subtle structural impacts induced by D-epimerization at a few sites onto the mixture of oligomeric states created within most Aβ samples. Ion mobility-mass spectrometry (IM-MS) has increasingly become an important alternative for the chiral separation of Aβ stereoisomers.[24-31] IM-MS offers high analytical speed, low sample consumption, and the ability to resolve small structural differences in peptide analytes, driven by recent technological advancements in IM-MS.[28-33] While molecular dynamics (MD) simulations have been combined with IM-MS to reveal the structural consequences of D-epimerization within small neuropeptides,[34] Li et al. have recently reported a multi-dimensional IM-MS (md-IM-MS)-based structural analysis strategy to facilitate the study of the chiral effects on monomer structure, oligomeric propensity and receptor binding for Aβ fragments.[28,35] However, results from truncated Aβ (e.g. Aβ N-terminus and core region fragments) cannot be reliably extrapolated to predict those for the full-length bioactive Aβ forms (e.g. Aβ42).

To fully probe the chiral effects in full-length, bioactive Aβ42 including the differential roles of the N terminal and core regions, we herein provide an improved md-IM-MS[28,35] approach for the systematic study of the structural consequences of domain-specific Asp/Ser D-epimerization within Aβ42 (Table 1 & Scheme S1). Analytical challenges facing the MS characterization of full-length Aβ42 have been successfully tackled through an improved sample preparation workflow.

Moreover, IM-MS-guided replica exchange MD (REMD) simulations were employed to construct and refine the 3D structures of different Aβ42 stereoisomers.

Methods

Sample Preparation: Aβ peptides with D- and L-forms as designed were purchased from GenScript USA, Inc. (Piscataway, NJ). For each Aβ peptide, 5 mg of the peptide was dissolved into 500 uL 1% ammonium hydroxide followed by lyophilization at -50 ºC and ~50 mTorr overnight. The lyophilized peptides were dissolved in 10mM ammonium acetate, and the concentration of each sample was determined by UV measurements on a Thermo NanoDrop 2000c instrument (ε coefficient = 1490 for Aβ42). To avoid rapid degradation and/or aggregation, Aβ peptides were diluted 10-fold with ammonium acetate buffer followed by flash-freezing aliquots.

IM-MS Experiment: For the traveling wave IM-MS experiment (TWIM-MS), Waters cyclic IM-MS (Milford, MA) instrument was used. The samples were transferred to a gold-coated borosilicate capillary needle (prepared in-house), and ions were generated by direct infusion of a sample using a nanoelectrospray ionization (nESI) source. Capillary voltages ranged between 1.0-2.0 kV and the sampling cone was used at 30 V. The MS cone temperature was 75 ºC. The traveling-wave ion mobility separator was operated at a pressure of ~3.5 mbar, and DC voltage waves (30 V wave height traveling at 400 m/s) to generate ion mobility separation.

Trapped ion mobility spectrometry (TIMS) measurements were performed on timsTOF Pro (Bruker Corp.) operated in MS1 mode, with TIMS on, over the m/z range 100-1700, ramp time 100 ms. Low-concentration of electrospray ionization (ESI) tune mix (Agilent Technologies, Santa Clara, CA) was directly infused using a standard ESI source for both quadratic TOF calibration and linear calibration of the TIMS analyzer over a 1/k0 window of (0.6, 1.6) V·s·cm-

2. After charge states were assigned to dominant masses in the mobilogram using dataAnalysis 5.2 (Bruker Corp.), the experimental peptide CCSs were deduced based on tuning mix ions 622.0, 922.0, and 1222.0 as external calibrants.

REMD Simulations: Molecular dynamics (MD) simulations were performed with CHARMM on a workstation with 48 CPU core 2.10 GHz Intel Xeon Processor. The CHARMM36 force field[36] was employed for all MD simulations. The input models were generated on CHARMM with experimentally observed peptide sequences. Charges were placed on titratable residues (Lys, Arg, Glu, Asp) by protonating or deprotonating so that the net charge of the peptides corresponds to the experimentally observed charge state of peptides. These charge-placed peptides were subjected to energy minimization via steepest descent (SD) over 1000 steps. The energy-minimized peptide model was used as the input for subsequent equilibration MD simulation. The equilibration was performed at 300 K over a 200 ps timescale with a simulation time step of 2 fs. The final equilibrated structure was then used as a starting structure for REMD simulations.[37] The REMD runs utilized 12 replicas ranging from 300 K to 600 K over a 150 ns simulation timescale. The exchange frequency was set to 1 ps. Structures from REMD were extracted from all REMD trajectory files. These structures were filtered based on the experimental CCSs.

The theoretical CCS values of all models were calculated using IMPACT[38] due to the large number of models generated from MD simulations. To increase the accuracy of CCS calculations, 2000 structures were randomly selected from the extracted structures, and Collidoscope[39] was used for trajectory method (TM) CCS calculations. The IMPACT projected area (PA) CCS and Collidoscope TM CCS were plotted, and a regression was used to build a calibration function to correct the rapid IMPACT PA CCS values into approximated TM CCS values for all the structures.

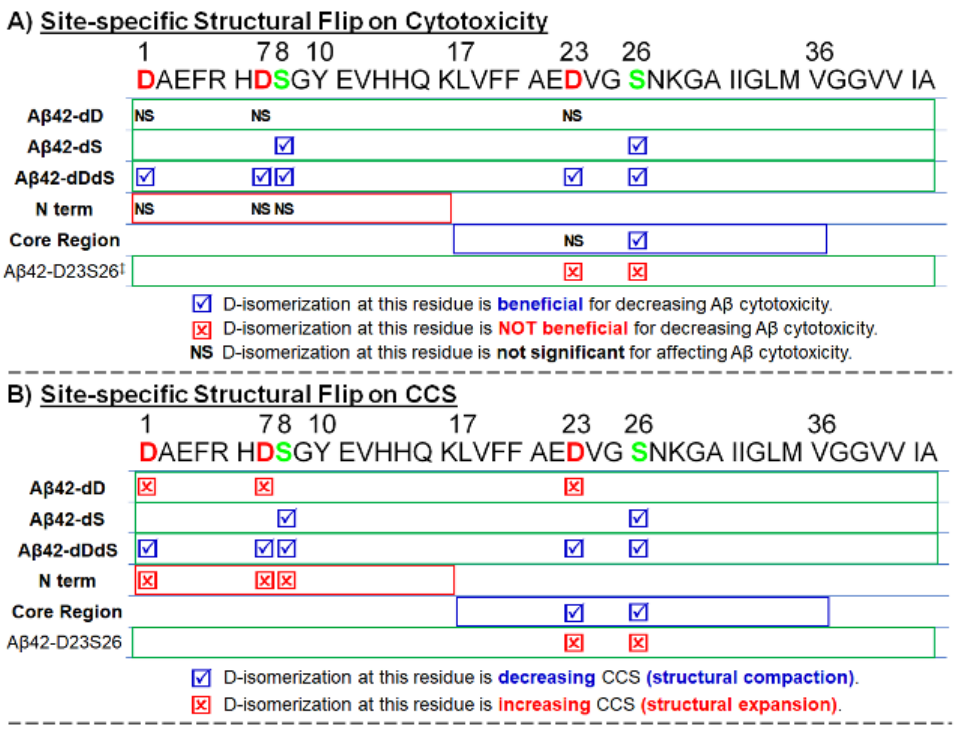
All extracted structures were filtered based on the experimental CCSs. The filtered models were then clustered for their structural similarities based on Kabsch's RMSD analysis.[40]

Results and Discussion

First, we customized a variety of Aβ stereoisomers with D-Asp and/or D-Ser residues to systematically construct site-specific, domain-constrained isomer models. As shown in Appendix Table IV-1, dD Aβ refers to Aβ isoforms with all Asp D-isomerized, dS Aβ refers to Aβ isoforms with all Ser D-isomerized, while dDdS Aβ refers to Aβ with all Asp and Ser co-D-isomerized. Appendix Figure IV-1 shows the sequence and chemical structure for Aβ42, where N-terminal and core regions are highlighted in different colors. To be more specific, Aβ10/Aβ16 and Aβ(17-36) are designated as N-terminal and core peptides, respectively, all of which are synthesized in four isoforms, WT, dD, dS, and dDdS (Appendix Table IV-1). In total, we studied 12 Aβ stereoisomers in three groups to inspect domain-constrained chiral effects.

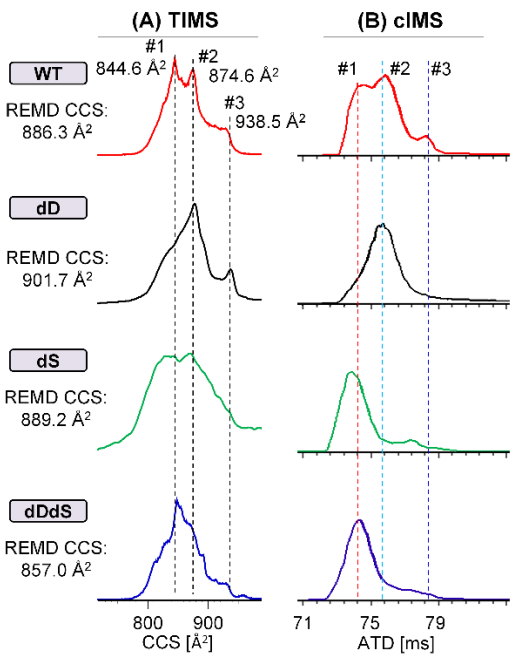
Appendix Table IV-1. Domain-constrained Structural Flips on Aβ42.

D-residues	**N-terminus (Aβ10/Aβ16)**	**Core region (Aβ17-36)**	**Full-length Aβ (Aβ42)**
WT	–	–	–
dD	$Asp_{1/7}$	Asp_{23}	$Asp_{1/7/23}$
dS	Ser_{8}	Ser_{26}	$Ser_{8/26}$
dDdS	$Asp_{1/7}Ser_{8}$	$Asp_{23}Ser_{26}$	$Asp_{1/7/23}Ser_{8/26}$



Appendix Figure IV-1. Overview and comparative summary of the effects of the site-specific structural flip of Aβ on cytotoxicity and CCS. ‡Cytotoxicity data for Aβ42-D23S26 are derived from previous literature (J. Org. Chem. 2020, 85, 1385).

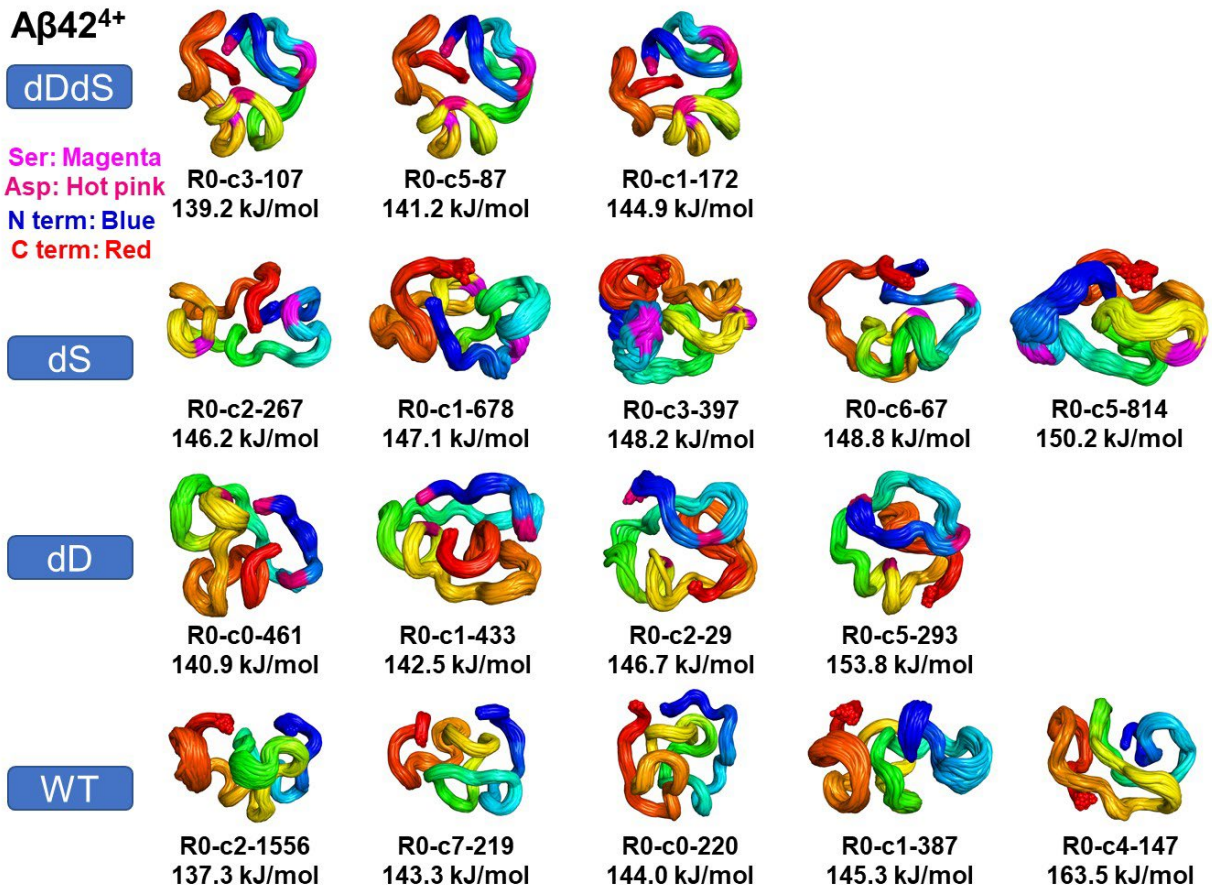
To characterize the structural impacts of chiral inversion of Aβ, it is important to deploy an analytical technique that can discriminate subtle structural differences of Aβ stereoisomers. The challenges often associated with analyzing full-length Aβ arise from 1) the propensity of Aβ42 to form aggregates in solution limiting the ability to introduce the sample into the gas-phase via nESI; 2) inadequate collision cross section (CCS) resolution of IM-MS for Aβ stereoisomers when using a single IM. The first challenge was resolved through the optimization of sample preparation where resuspended peptides were diluted 10-fold to avoid aggregation.



Appendix Figure IV-2. Chiral differentiation of full-length Aβ42 stereoisomers (WT/dD/dS/dDdS, 4+) by using high-resolution TIMS **(A)** and cIMS (10 passes) **(B)** instruments.

To overcome the second challenge, the trapped ion mobility spectrometry (TIMS), which provides higher resolving power (R>200) was utilized to reveal subtle CCS changes between Aβ42 isomers. Our IM data shows that the D-epimerization of Asp and Ser residues induces a moderate reduction of CCS (Appendix Figure IV-2A). The observation of multiple mobility peaks with similar distribution windows for four stereoisomers suggests high conformational dynamics of Aβ42, with varied dominate conformers across stereoisomers that might contribute to their differences in activity. For WT Aβ42, at least three conformers were observed; cross-sectional measurements were calculated as 844.6 $Å^2$ (conf #1), 874.6 $Å^2$ (conf #2), and 938.5 $Å^2$ (conf #3). While dD Aβ42 holds mainly conf #2 and #3, both dS Aβ42 and dDdS Aβ42 adopt primarily conf #1 and/or conf #2. D-epimerization at Asp and Ser residues generally makes Aβ42 more structurally compact. To further resolve and confirm the CCS differences amongst these Aβ42 stereoisomers, we incorporated analyses from a cyclic IM instrument (cIMS), representative of a new generation of high-resolution IMS. As shown in Appendix Figure IV-2B, gradual separation was achieved at higher pass numbers, where the highest resolution measurements (10-passes) confirmed the negative trend in CCS that was observed in our TIMS measurements. These data allow us to definitively conclude that simultaneous chiral inversion of all Asp and Ser residues compacts the overall structure of full-length Aβ42 by ~3.4% (CCS differences between conf #1 and conf #2).

With the successful characterization of Aβ42 stereoisomers, we performed IM-MS-guided molecular dynamics (MD) simulations to gain structural insight into these peptides. CCS values based on TIMS measurements were used to guide REMD. REMD results in distinct cluster models that are energetically favorable for Aβ42 stereoisomers (Appendix Figure IV-3). Appendix Figure IV-4A illustrates the lowest energy clusters for WT, dD, dS, and dDdS Aβ42, ranging from 137.3 kJ/mol to 146.2 kJ/mol. These chosen clusters of lowest energy are considered representative conformers, as they possess significant percentages of total structure numbers, ranging from 107 (29% of all cluster models) to 1556 (61% of all cluster models). Aβ42 cluster models generated

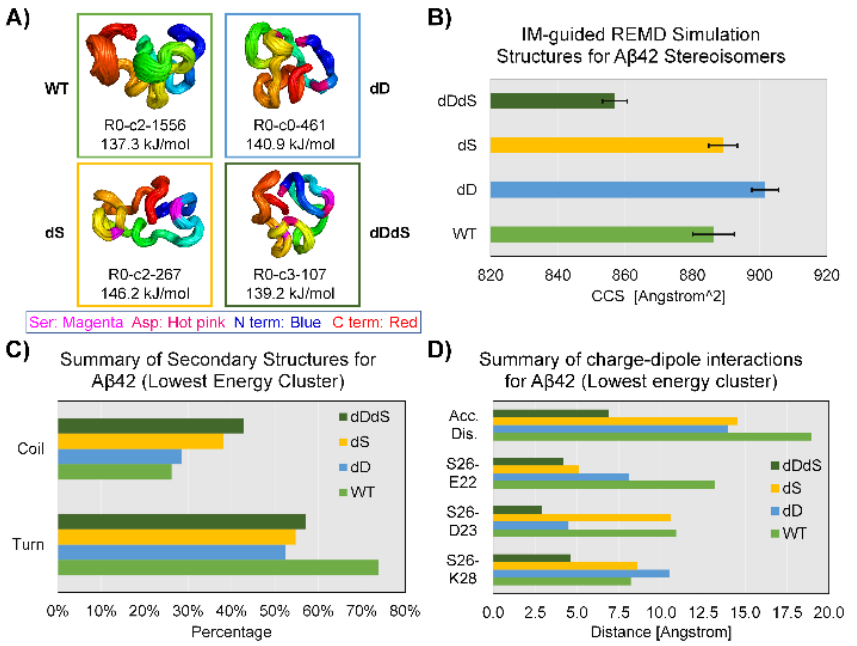


Appendix Figure IV-3. Energetically favorable REMD structures for Aβ42 stereoisomers of 4+ ion species. All structures are filtered by TIMS-CCS(N2) with 3% calibration error; conformations at 300 K (lowest temperature) were clustered and listed are the most confidently identified REMD clusters (population > 1% with well-aligned structures). R0-c3-107: R0 (rep #0), c3 (cluster #3), 107 (total pdb model numbers). N-terminus, blue; C-terminus, red; Ser residues, magenta; Asp residues, hot pink. Energy for each conformer is also listed right below the cartoon illustration.

from 300 K REMD simulations displayed mean CCS values of 886.3 ± 3.2 $Å^2$ (4+), 901.7 ± 2.0 $Å^2$ (4+), 889.2 ± 2.1 $Å^2$ (4+) and 857.0 ± 1.9 $Å^2$ (4+), for WT, dD, dS, and dDdS, respectively (Appendix Figure IV-4B). The generated WT, dD, and dS Aβ42 models match well with our experimental results (Appendix Figure IV-4A), especially conf #2, with CCS deviations spanning 1.3%-3.0%. The *in silico* dDdS Aβ42 model is in accordance with our experimentally observed conf #1, deviating by only 1.5%. From these results, we can conclude not only a strong agreement between our experimental results and theoretical models but also that our hypothesis regarding structural compaction post-D-epimerization can be rigorously tested.

In order to distill more structural detail, we further evaluated the REMD-generated Aβ42 stereoisomer models. Secondary structure analysis, as shown in Appendix Figure IV-4C, suggests an increased coil fraction but decreased turn content in Aβ42 after chiral inversion. The D-isomerized Aβ42 clusters exhibit a coil fraction range of 29% - 43% while the turn fraction ranges between 52% - 57%. These observed changes in secondary structural elements appear to be linked to the structural compaction observed in Aβ42 isomers. Taking Ser26 as an example, we found that charge-dipole interactions[12] are effectively altered at both long (WT: S26-E22 distance up to 13 Å, d_1, Appendix Figure IV-4D) and short ranges (dDdS: S26-E22 distance of 4.2 Å). This type of alteration in charge-dipole attraction is also prevalent across two other pairs of residues, S26-D23 (d_2) and S26-K28 (d_3). To characterize these overall charge-dipole interactions, we cumulatively evaluated the distances of these three pairs (d_{acc}, Appendix Figure IV-4D). In total, the d_{acc} values for Ser26 were diminished from ~19 Å (WT Aβ42) to 6.9 Å (dDdS Aβ42), indicative of the elevated charge-dipole attractions upon D-epimerization. Therefore, REMD

simulations suggest that the major consequences of D-epimerization are the enhanced S26-related dipole-charge attractions and increased coil fraction at the C-terminus, which together result in the compaction of full-length Aβ42.



Appendix Figure IV-4. IM-guided REMD simulation results for full-length Aβ42 stereoisomers (WT/dD/dS/dDdS, [Aβ42+4H]4+) and corresponding structural analysis. (A) Lowest energy and most dominant/major REMD structural clusters for Aβ42 stereoisomers. Conformer name, R0-c3-107: R0 (rep #0), c3 (cluster #3), 107 (total structure numbers). N-terminus, blue; C-terminus, red; Ser residues, magenta; Asp residues, hot pink. Energy for each conformer is also listed right below the cartoon illustration. **(B-D)** Calculated CCS distributions, secondary structures and charge-dipole interactions for REMD-simulated Aβ42 stereoisomers. Charge-dipole interactions are characterized with the spatial distances between certain residues (e.g. S26-E22 (d1), S26-D23 (d2) and S26-K28 (d3)), which is termed Accumulated Distance (Acc. Dis.) Accumulated Distance is calculated from the Euclidian distance between all interacting residues: $d_{acc} = (d_1^2 + d_2^2 + d_3^2)^{0.5}$

Conclusion

Collectively, we present integrated, comprehensive IM-MS datasets on the analyses of Aβ42 stereochemistry, with a focus on the two most frequently observed D-epimerization residues in nature, Asp and Ser. Our efforts in both IM-MS and REMD give rise to refined structural models of all stereoisomers in the gas phase, providing unprecedented information regarding local structural features that confirm our hypothesis surrounding Aβ42 structural compaction-based toxicity modulation. As shown, domain-constrained amino acid isomerization alters and generally compacts Aβ42 while promoting differential, cooperative, and domain-specific structural changes. Our results suggest that enhanced interactions between Ser26 and neighboring residues, as well as the increased amount of coil structure at the C-terminus, are accommodated in Aβ42 chiral variants. This study provides a comprehensive insight into the

molecular mechanisms behind the naturally occurring D-epimerization of Ser/Asp in Aβ42 and serves to inform on the molecular foundation that enables amino acid-level chiral control that can reduce Aβ42 cytotoxicity[3,10-12] through enzymatic regulation and phage display.[13,14]

References

(1) Ritz-Timme, S.; Collins, M. J. Racemization of Aspartic Acid in Human Proteins. Ageing Research Reviews. Ageing Res Rev February 2002, pp 43–59.

(2) Sugiki, T.; Utsunomiya-Tate, N. Site-Specific Aspartic Acid Isomerization Regulates Self-Assembly and Neurotoxicity of Amyloid-β. Biochem. Biophys. Res. Commun. 2013, 441 (2),493–498.

(3) Dutta, S.; Foley, A. R.; Warner, C. J. A.; Zhang, X.; Rolandi, M.; Abrams, B.; Raskatov, J. A. Suppression of Oligomer Formation and Formation of Non-Toxic Fibrils upon Addition of Mirror-Image Aβ42 to the Natural l-Enantiomer. Angew. Chemie - Int. Ed. 2017, 56 (38), 11506–11510.

(4) Warner, C. J. A.; Dutta, S.; Foley, A. R.; Raskatov, J. A. Introduction of D-Glutamate at a Critical Residue of Aβ42 Stabilizes a Prefibrillary Aggregate with Enhanced Toxicity. Chem. - A Eur. J. 2016, 22 (34), 11967–11970.

(5) Kubo, T.; Kumagae, Y.; Miller, C. A.; Kaneko, I. β-Amyloid Racemized at the Ser26 Residue in the Brains of Patients with Alzheimer Disease: Implications in the Pathogenesis of Alzheimer Disease. J. Neuropathol. Exp. Neurol. 2003, 62 (3), 248–259.

(6) Banreti, A.; Bhattacharya, S.; Wien, F.; Matsuo, K.; Réfrégiers, M.; Meinert, C.; Meierhenrich, U.; Hudry, B.; Thompson, D.; Noselli, S. Biological Effects of the Loss of Homochirality in a Multicellular Organism. Nat. Commun. 2022, 13 (1).

(7) Mukherjee, S.; Perez, K. A.; Lago, L. C.; Klatt, S.; McLean, C. A.; Birchall, I. E.; Barnham, K. J.; Masters, C. L.; Roberts, B. R. Quantification of N-Terminal Amyloid-β Isoforms Reveals Isomers Are the Most Abundant Form of the Amyloid-β Peptide in Sporadic Alzheimer's Disease. Brain Commun. 2021, 3 (2).

(8) Fujii, N.; Kaji, Y.; Fujii, N. D-Amino Acids in Aged Proteins: Analysis and Biological Relevance. J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci. 2011, 879 (29), 3141–3147.

(9) Kuo, Y. M.; Emmerling, M. R.; Woods, A. S.; Cotter, R. J.; Roher, A. E. Isolation, Chemical Characterization, and Quantitation of Aβ 3-Pyroglutamyl Peptide from Neuritic Plaques and Vascular Amyloid Deposits. Biochem. Biophys. Res. Commun. 1997, 237 (1), 188–191.

(10) Dutta, S.; Finn, T. S.; Kuhn, A. J.; Abrams, B.; Raskatov, J. A. Chirality Dependence of Amyloid β Cellular Uptake and a New Mechanistic Perspective. Chembiochem 2019, 20 (8), 1023–1026.

(11) Dutta, S.; Finn, T. S.; Kuhn, A. J.; Abrams, B.; Raskatov, J. A. Chirality Dependence of Amyloid β Cellular Uptake and a New Mechanistic Perspective. ChemBioChem 2019, 20 (8), 1023–1026.

(12) Foley, A. R.; Lee, H. W.; Raskatov, J. A. A Focused Chiral Mutant Library of the Amyloid β 42 Central Electrostatic Cluster as a Tool to Stabilize Aggregation Intermediates. J. Org. Chem. 2020, 85 (3), 1385–1391.

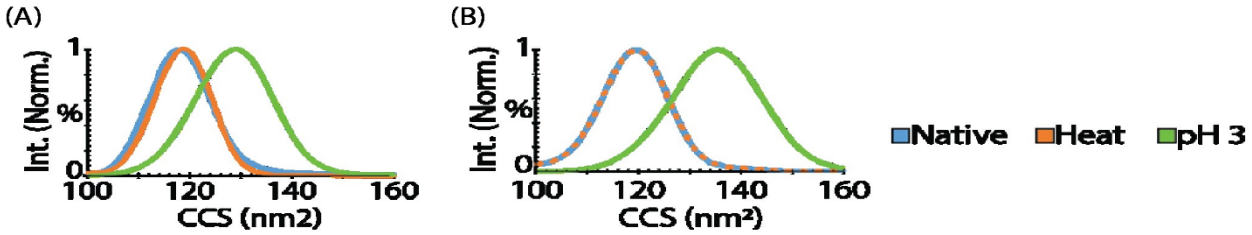
(13) Takagi, S.; Balu, D. T.; Coyle, J. T. Factors Regulating Serine Racemase and D-Amino Acid Oxidase Expression in the Mouse Striatum. Brain Res. 2021, 1751.
(14) Zhou, X.; Zuo, C.; Li, W.; Shi, W.; Zhou, X.; Wang, H.; Chen, S.; Du, J.; Chen, G.; Zhai, W.; Zhao, W.; Wu, Y.; Qi, Y.; Liu, L.; Gao, Y. A Novel D-Peptide Identified by Mirror-Image Phage Display Blocks TIGIT/PVR for Cancer Immunotherapy. Angew. Chemie - Int. Ed. 2020, 59 (35), 15114–15118.
(15) Mast, D. H.; Checco, J. W.; Sweedler, J. V. Differential Post-Translational Amino Acid Isomerization Found among Neuropeptides in Aplysia Californica. ACS Chem. Biol. 2020, 15 (1), 272–281.
(16) Mast, D. H.; Checco, J. W.; Sweedler, J. V. Advancing D-Amino Acid-Containing Peptide Discovery in the Metazoan. Biochimica et Biophysica Acta - Proteins and Proteomics. Biochim Biophys Acta Proteins Proteom January 1, 2021.
(17) Patel, A. V.; Kawai, T.; Wang, L.; Rubakhin, S. S.; Sweedler, J. V. Chiral Measurement of Aspartate and Glutamate in Single Neurons by Large-Volume Sample Stacking Capillary Electrophoresis. Anal. Chem. 2017, 89 (22), 12375–12382.
(18) Livnat, I.; Tai, H. C.; Jansson, E. T.; Bai, L.; Romanova, E. V.; Chen, T. T.; Yu, K.; Chen, S. A.; Zhang, Y.; Wang, Z. Y.; Liu, D. D.; Weiss, K. R.; Jing, J.; Sweedler, J. V. A D-Amino Acid-Containing Neuropeptide Discovery Funnel. Anal. Chem. 2016, 88 (23), 11868–11876.
(19) Bai, L.; Romanova, E. V.; Sweedler, J. V. Distinguishing Endogenous D-Amino Acid-Containing Neuropeptides in Individual Neurons Using Tandem Mass Spectrometry. Anal. Chem. 2011, 83 (7), 2794–2800.
(20) Bai, L.; Sheeley, S.; Sweedler, J. V. Analysis of Endogenous D-Amino Acid-Containing Peptides in Metazoa. Bioanal. Rev. 2009, 1 (1), 7–24.
(21) Ewing, M. A.; Wang, J.; Sheeley, S. A.; Sweedler, J. V. Detecting D-Amino Acid-Containing Neuropeptides Using Selective Enzymatic Digestion. Anal. Chem. 2008, 80 (8), 2874–2880.
(22) Sheeley, S. A.; Miao, H.; Ewing, M. A.; Rubakhin, S. S.; Sweedler, J. V. Measuring D-Amino Acid-Containing Neuropeptides with Capillary Electrophoresis. Analyst 2005, 130 (8), 1198–1203.
(23) Mast, D. H.; Liao, H. W.; Romanova, E. V.; Sweedler, J. V. Analysis of Peptide Stereochemistry in Single Cells by Capillary Electrophoresis-Trapped Ion Mobility Spectrometry Mass Spectrometry. Anal. Chem. 2021, 93 (15), 6205–6213.
(24) Lambeth, T. R.; Riggs, D. L.; Talbert, L. E.; Tang, J.; Coburn, E.; Kang, A. S.; Noll, J.; Augello, C.; Ford, B. D.; Julian, R. R. Spontaneous Isomerization of Long-Lived Proteins Provides a Molecular Mechanism for the Lysosomal Failure Observed in Alzheimer's Disease. ACS Cent. Sci. 2019, 5 (8), 1387–1395.
(25) Zheng, X.; Deng, L.; Baker, E. S.; Ibrahim, Y. M.; Petyuk, V. A.; Smith, R. D. Distinguishing D- and l-Aspartic and Isoaspartic Acids in Amyloid β Peptides with Ultrahigh Resolution Ion Mobility Spectrometry. Chem. Commun. 2017, 53 (56), 7913–7916.
(26) Nagy, G.; Kedia, K.; Attah, I. K.; Garimella, S. V. B.; Ibrahim, Y. M.; Petyuk, V. A.; Smith, R. D. Separation of β-Amyloid Tryptic Peptide Species with Isomerized and Racemized L-Aspartic Residues with Ion Mobility in Structures for Lossless Ion Manipulations. Anal. Chem. 2019, 91 (7), 4374–4380.

(27) Shvartsburg, A. A.; Berthias, F.; Baird, M. A. Differential Ion Mobility Separations of D/L Peptide Epimers. Anal. Chem. 2021, 93 (8), 4015–4022.
(28) Li, G.; DeLaney, K.; Li, L. Molecular Basis for Chirality-Regulated Aβ Self-Assembly and Receptor Recognition Revealed by Ion Mobility-Mass Spectrometry. Nat. Commun. 2019, 10 (1).
(29) Li, G.; Delafield, D. G.; Li, L. Improved Structural Elucidation of Peptide Isomers and Their Receptors Using Advanced Ion Mobility-Mass Spectrometry. TrAC - Trends in Analytical Chemistry. 2020, p 115546.
(30) Jeanne Dit Fouque, K.; Garabedian, A.; Porter, J.; Baird, M.; Pang, X.; Williams, T. D.; Li, L.; Shvartsburg, A.; Fernandez-Lima, F. Fast and Effective Ion Mobility-Mass Spectrometry Separation of d -Amino-Acid-Containing Peptides. Anal. Chem. 2017, 89 (21), 11787–11794.
(31) Jia, C.; Lietz, C. B.; Yu, Q.; Li, L. Site-Specific Characterization of d-Amino Acid Containing Peptide Epimers by Ion Mobility Spectrometry. Anal. Chem. 2014, 86 (6), 2972–2981.
(32) Tolmachev, A. V.; Webb, I. K.; Ibrahim, Y. M.; Garimella, S. V. B.; Zhang, X.; Anderson, G. A.; Smith, R. D. Characterization of Ion Dynamics in Structures for Lossless Ion Manipulations. Anal. Chem. 2014, 86 (18), 9162–9168.
(33) Stow, S. M.; Causon, T. J.; Zheng, X.; Kurulugama, R. T.; Mairinger, T.; May, J. C.; Rennie, E. E.; Baker, E. S.; Smith, R. D.; McLean, J. A.; Hann, S.; Fjeldsted, J. C. An Interlaboratory Evaluation of Drift Tube Ion Mobility-Mass Spectrometry Collision Cross Section Measurements. Anal. Chem. 2017, 89 (17), 9048–9055.
(34) Do, T. D.; Checco, J. W.; Tro, M.; Shea, J. E.; Bowers, M. T.; Sweedler, J. V. Conformational Investigation of the Structure-Activity Relationship of GdFFD and Its Analogues on an Achatin-like Neuropeptide Receptor of: Aplysia Californica Involved in the Feeding Circuit. Phys. Chem. Chem. Phys. 2018, 20 (34), 22047–22057.
(35) Xu, X.; Han, L.; Zheng, Z.; Zhao, R.; Li, L.; Shao, X.; Li, G. Composite Multidimensional Ion Mobility-Mass Spectrometry for Improved Differentiation of Stereochemical Modifications. Anal. Chem. 2023, 95 (4), 2221–2228.
(36) 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. CHARMM: The Biomolecular Simulation Program. J. Comput. Chem. 2009, 30 (10), 1545–1614.
(37) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. Chem. Phys. Lett. 1999, 314 (1–2), 141–151.
(38) Marklund, E. G.; Degiacomi, M. T.; Robinson, C. V.; Baldwin, A. J.; Benesch, J. L. P. Collision Cross Sections for Structural Proteomics. Structure 2015, 23 (4), 791–799.
(39) Ewing, S. A.; Donor, M. T.; Wilson, J. W.; Prell, J. S. Collidoscope: An Improved Tool for Computing Collisional Cross-Sections with the Trajectory Method. J. Am. Soc. Mass Spectrom. 2017, 28 (4), 587–596.
(40) Kabsch, W. A Solution for the Best Rotation to Relate Two Sets of Vectors. Acta Crystallogr. Sect. A 1976, 32 (5), 922–923.

Appendix V: Ion Mobility-Mass Spectrometry Reveals the Structures of Biotherapeutic Antibody Aggregates

Partially adapted from **Vallejo, D. D.; Jeon, C. K.; Parson, K. F.; Herderschee, H. R.; Eschweiler, J. D.; Filoti, D. I.; Ruotolo, B. T. Ion Mobility-Mass Spectrometry Reveals the Structures and Stabilities of Biotherapeutic Antibody Aggregates.** ***Anal. Chem.*** **2022.**

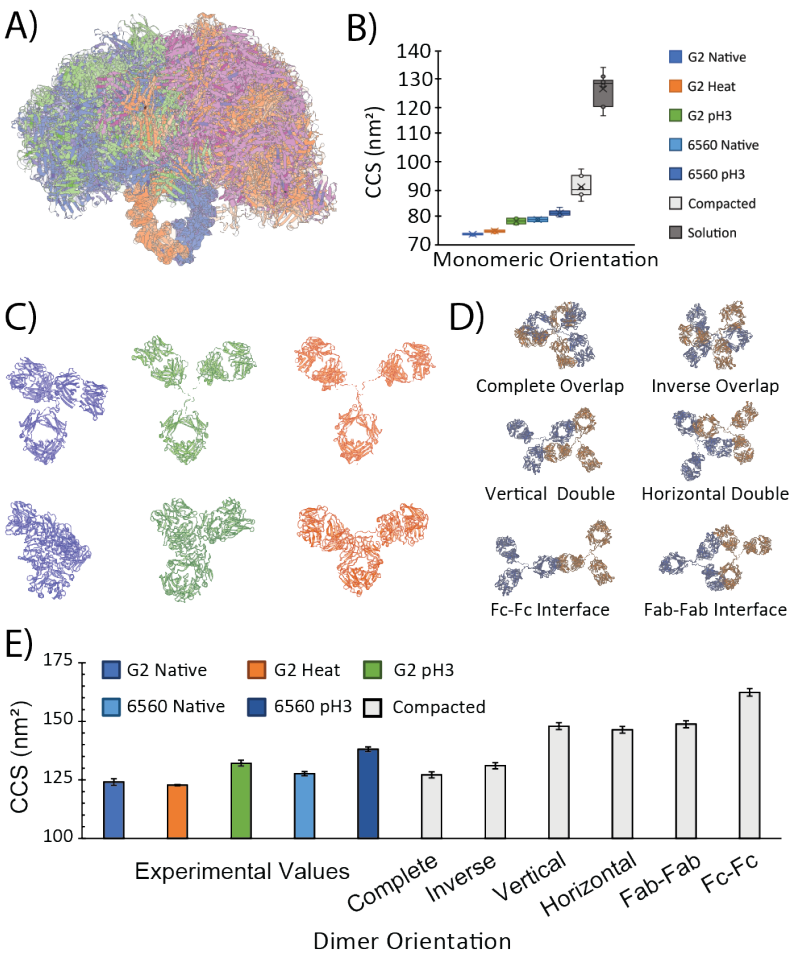
Stability is one of the critical quality attributes (CQAs) which are essential for evaluating the efficacy and safety of biotherapeutics such as monoclonal antibodies (mAbs). Stability-based CQAs include aggregation and fragmentation of mAbs.[1] Aggregation, in particular, is a concern for manufacturers as it can induce immunogenicity in patients if not controlled properly.[2,3] There are several stress factors throughout the production that may induce antibody aggregation such as heating, pH, process, and light. A previous study has shown that these stresses induce dimerization with a wide range of conformations.[4] In this study, we investigated the dimerization of two model IgG1 (α and β) antibodies, specifically the orientations at which the dimers are formed using ion mobility-mass spectrometry (IM-MS)-guided molecular dynamics (MD) simulations.



Appendix Figure V-1. CCS distributions of (A) IgG1α dimer and (B) IgG1β dimer of 40+ charge state in native, heat-stressed, and pH-stressed conditions.

IM experiments were performed on both traveling wave IM-MS (TWIM-MS) and drift tube IM-MS (DTIM-MS) using Waters Synapt G2 (Milford, MA) and Agilent 6560C (Santa Clara, CA). Three different IgG1 conditions are studied: native, heat-stressed, and pH-stressed. All sample preparation and IM-MS experimental methods follow a previous study.[5] For MD simulations, methods from Chapter 3 were adopted to build gas-phase models of IgG1 monomer, and only ten out of 120 models were chosen to proceed with MD simulations. The final structure of MD simulations was then docked using PyMOL using the "measurement wizard" function so that the distance between each subunit is within 1 to 6 Å.

We compared the collision cross sections (CCSs) of IgG1α/β across three different conditions. The $^{TW}CCS_{N2}$ recorded for IgG1α was 117.9 ± 0.06, 119.2 ± 0.06, and 129.5 ± 0.04 nm^2 for native, heat, and pH 3 conditions, respectively (Appendix Figure V-1A). Measurements of $^{TW}CCS_{N2}$ for IgG1β dimers range from 119.3 ± 0.51 nm^2, for those created under native conditions, to 119.4 ± 0.71



Appendix Figure V-2. (A) Overlay of 120 IgG1 structures demonstrating the flexibility of mAb Fab regions. (B) Box-and-whisker plot with experimental CCS values from IgG1 α/β and solution and compacted CCS values for 10 monomeric mAb structures. (C) Comparison of the smallest (blue), median (green), and largest (orange) structures for solution-phase (top row) and gas-phase (bottom row) monomeric mAbs. (D) Several hand docked models built using archetypical IgG dimeric structures. (E) CCS measurements of experimental values from IgG1α/β and CCS calculations for compacted dimer models.

and 136.0 ± 0.84 nm^2 generated through heat and pH 3 degradation conditions (Appendix Figure V-1B). Overall, heat conditions have a small impact on CCS in both IgG1 models. In contrast, pH 3 conditions increased CCSs by 7.5% and 9% for IgG1α and IgG1β, respectively, and broadened the CCS distributions. To evaluate the inter-instrument comparability of the measurements, IgG1β IM-MS experiments were performed on DTIM-MS. Our measurements with DTIM-MS resulted in a 7-8% CCS offset which is similar to the differences observed for the mAb samples between G2 and 6560C previously.[6]

In order to investigate the structure of mAb dimers, we began by mining over 120 monomeric IgG1 structural models generated through electron tomography (Appendix Figure V-2A).[7] When overlayed, the structures demonstrate the overall flexibility of the mAb hinge region, which has been implicated previously as the primary driver for antibody compaction during transfer to the gas phase.[8] When the CCS values for ten of these model structures were calculated, we found that trajectory method (TM) median CCS measurements for solution phase models ranged between 35 and 45%, larger than our experimental gas-phase measurements (Appendix Figure V-2B). Strikingly, when we compacted these ten models in the gas phase, we obtain CCS measurements that range between 74 and 84 nm^2, corresponding to differences of 1–7% between our models and experimental data (Appendix Figure V-2B). We note that while our models indicate that a significant amount of gas-phase compaction has occurred in our mAb ions when compared to other native globular proteins, the extent of compaction computed remains consistent with recent MD simulations aimed at tracking the details of this phenomenon.[9,10] In Appendix Figure V-2C, we project the smallest (blue), median (green), and largest (orange) IgG models from the solution phase (top row) and gas phase (bottom row). It can be clearly seen in all three structures that there is considerable compaction of the hinge region. The F(ab)2 region in the

smallest structures, in particular, seems to be "folding over" the Fc region, which would support previous computational models for mAbs centered around ~70 nm^2 in their computational CCS measurements.[8] We constructed several dimer orientations based on archetypical structures described in the literature using the mAb model associated with the median to evaluate the structures of the mAb dimers in our experiments (Appendix Figure V-2 5D). Due to the large size of dimeric antibodies, IMPACT was used for theoretical projected area (PA) CCS calculations, followed by the approximation of TM CCS to reduce the computational cost without losing the fidelity of the theoretical CCS values. The CCSs for these solution-phase dimer models range between 190 and 257 nm^2 for TJM approximations (Appendix Figure V-2E). By factoring in the offset observed when we compared our monomer mAb measurements to the models shown in Appendix Figure V-2B, which were generated by docking and equilibrating compact monomers, and assuming a similar level of offset associated with comparisons to our experimental gas-phase dimer models (~35–45%), it is clear that under native and heat degradation conditions, mAb dimers agree most closely with completely overlapped structures, whereas dimers created using pH 3 degradation conditions may adopt inverted overlapped quaternary structures, although completely overlapped configurations comprising significantly elongated mAb monomers cannot be ruled out. Our data likely rule out the presence of larger quaternary structures that rely upon interfacial contact at either the Fab or Fc regions of the mAbs involved, or the double-type structure archetypes, in our mAb samples. Further computational analysis is required to generate a relaxed, gas-phase version of the model structures shown here to evaluate their fidelity.

References

(1) Le Basle, Y.; Chennell, P.; Tokhadze, N.; Astier, A.; Sautou, V. Physicochemical Stability of Monoclonal Antibodies: A Review. *J. Pharm. Sci.* **2020**, *109* (1), 169–190.

(2) Roberts, C. J. Therapeutic Protein Aggregation: Mechanisms, Design, and Control. *Trends in Biotechnology*. Elsevier Current Trends July 1, 2014, pp 372–380.

(3) Moussa, E. M.; Panchal, J. P.; Moorthy, B. S.; Blum, J. S.; Joubert, M. K.; Narhi, L. O.; Topp, E. M. Immunogenicity of Therapeutic Protein Aggregates. *Journal of Pharmaceutical Sciences*. Elsevier February 1, 2016, pp 417–430.
(4) Paul, R.; Graff-Meyer, A.; Stahlberg, H.; Lauer, M. E.; Rufer, A. C.; Beck, H.; Briguet, A.; Schnaible, V.; Buckel, T.; Boeckle, S. Structure and Function of Purified Monoclonal Antibody Dimers Induced by Different Stress Conditions. *Pharm. Res.* **2012**, *29* (8), 2047–2059.
(5) Vallejo, D. D.; Jeon, C. K.; Parson, K. F.; Herderschee, H. R.; Eschweiler, J. D.; Filoti, D. I.; Ruotolo, B. T. Ion Mobility-Mass Spectrometry Reveals the Structures and Stabilities of Biotherapeutic Antibody Aggregates. *Anal. Chem.* **2022**.
(6) Gadkari, V. V.; Ramírez, C. R.; Vallejo, D. D.; Kurulugama, R. T.; Fjeldsted, J. C.; Ruotolo, B. T. Enhanced Collision Induced Unfolding and Electron Capture Dissociation of Native-like Protein Ions. *Anal. Chem.* **2020**, *92* (23), 15489–15496.
(7) Zhang, X.; Zhang, L.; Tong, H.; Peng, B.; Rames, M. J.; Zhang, S.; Ren, G. 3D Structural Fluctuation of IgG1 Antibody Revealed by Individual Particle Electron Tomography. *Sci. Rep.* **2015**, *5* (1), 1–14.
(8) Hansen, K.; Lau, A. M.; Giles, K.; McDonnell, J. M.; Struwe, W. B.; Sutton, B. J.; Politis, A. A Mass-Spectrometry-Based Modelling Workflow for Accurate Prediction of IgG Antibody Conformations in the Gas Phase. *Angew. Chemie - Int. Ed.* **2018**, *57* (52), 17194–17199.
(9) Rolland, A. D.; Prell, J. S. Computational Insights into Compaction of Gas-Phase Protein and Protein Complex Ions in Native Ion Mobility-Mass Spectrometry. *TrAC - Trends in Analytical Chemistry*. Elsevier July 1, 2019, pp 282–291.
(10) Rolland, A. D.; Biberic, L. S.; Prell, J. S. Investigation of Charge-State-Dependent Compaction of Protein Ions with Native Ion Mobility-Mass Spectrometry and Theory. *J. Am. Soc. Mass Spectrom.* **2022**, *33* (2), 369–381.