Protein Primary and Quaternary Structure Elucidation by Mass Spectrometry

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Yang Song

Graduate Program in Chemistry

The Ohio State University

2015

Dissertation Committee:

Vicki H. Wysocki, advisor

Mark Foster

Abraham Badu-Tawiah
Abstract

Mass spectrometry (MS) has become an indispensable tool for characterizing proteins, from primary structures to quaternary structures. This dissertation shows efforts to expand the role of MS in protein characterization.

Chapter 2 is focused on protein primary structure determination. Blood collected from 33 bird species that are common in the U.S. as hosts for ticks but that have unreported hemoglobin sequences was provided to the Wysocki group. The sequence information for the bird hemoglobin is important for tick-borne disease prevention and intervention. A top-down-assisted bottom-up MS approach with a customized searching database, based on variability in known bird hemoglobin sequences, was devised to facilitate fast and complete sequencing of hemoglobin from birds with unknown sequences and hemoglobin from all 33 species was sequenced.

Besides sequencing primary structures of protein complexes in denaturing conditions, MS can be utilized to study intact non-covalent protein complexes in the gas phase, retaining memory of their solution phase structures. Tandem MS is applied to dissociate non-covalent protein complexes, probing their subunit connectivity by generating structurally informative subcomplexes. The most widely used gas-phase dissociation method is collision induced dissociation (CID), which generally does not provide enough substructural information. An alternative gas-phase disassembly method,
surface induced dissociation (SID), has been shown to reveal substructure information. Especially in recent years, with the incorporation of an SID device into a Waters SYNAPT G2 or G2-S mass spectrometer with ion mobility capability, the conformation of SID products and remaining precursor were studied to improve our understanding of the SID process. In Chapter 3, 4 and 5, fundamental studies on the gas-phase dissociation behavior of homodimers (enolase, α-lactalbumin, and β-lactoglobulin), inter-chain disulfide reduced monoclonal antibody, and homo-hexamer protein complexes (bovine GDH, bacterial GDH, GCH and insulin) are performed. SID has been shown to be beneficial for protein complex interface analysis, because SID minimizes intra-protein conformational disruptions (unfolding) of subunits in the dissociation process. In addition, an excellent correlation of the SID behaviors and in silico analysis of interfacial areas of known systems demonstrates that SID can be very useful in structural characterization of unknown systems, especially in predicting the interfaces and relative interfacial strength.

Multiple MS based approaches were then utilized to predict quaternary structure of toyocamycin nitrile hydratase (TNH) (Chapter 6). Unfortunately, there are no structures solved by either NMR or x-ray crystallography on this important enzymatic complex. By coupling SID with ion mobility mass spectrometry (SID/IM), a complete connectivity map with relative interfacial strengths is obtained. Collisional cross-sections (CCS) measured from IM experiments are used as constraints for postulating a coarse-grained complex model. Protein complex homology modeling with all the constraints from connectivity, CCS and covalent labeling is utilized to propose likely structures with
atomic coordinates for TNH hexamer. A similar comprehensive mass spectrometric structural analysis was performed in Chapter 8 on a multicopper oxidase protein complex that is has no crystal structure or NMR results due to its heterogeneity and large size (over 200 kDa).

Overall, this dissertation has expanded the role of MS in protein characterization.
Dedication

This document is dedicated to my parents, Meishuang Song and Qiuying Wang

And all my educators.
Acknowledgments

I am extremely grateful for those who have helped me to reach this point during the past five years. I am afraid that it is impossible to express all my gratitude here for their generous guidance, encouragement, and support.

First and foremost, I must thank my advisor, Vicki Wysocki, who gave me the opportunity to do research in her lab and continuously provides assistance to me. I thank her for the freedom she allows me to think independently and improve myself in trials and errors. The hardworking attitude, problem solving skills, and presentation skills that I have learned from her will definitely benefit me for the rest of my life.

I would like to thank all of the Wysocki group members, past and present, for their help with my research. Particularly, Wenzhou Li taught me about computer coding for bottom-up peptide identification and Mowei Zhou taught me about native mass spectrometry of protein complexes. Also, Sophie Harvey, Andrew VanSchoiack, Lindsay Morrison, Matthew Bernier, Yan Jing, Akiko Tanimoto, Royston Quintyn, Yun Zhang and Chengsi Huang are thanked for their advices and supports on my different projects.

I am grateful to my collaborators, who are responsible and passionate about our research projects. In particular, I would like to thank Micah Nelp and Vahe Bandarian for the TNH project, Marcos Sotomayor for introducing me to molecular dynamics simulation, Mowei Zhou and Christine Romano for the Mnx project. Additionally, I
would like to thank Liwen Zhang, Kari Basso, Nanette Kleinholz and Jeremy Keirsey for the enjoyable working experience in the facility.

Finally, I must thank my parents for their love. It is impossible to finish my schooling without their endless encouragement and support. I want to specially thank Akiko Tanimoto, who is my labmate, roommate and best friend. I am so thankful for her company through the ups and downs of my graduate life.
Vita

2006 to 2010 ................................................. B.S. Chemistry, Peking University, China

2010 to 2012 ................................................. Graduate Assistant, Department of Chemistry and Biochemistry, The University of Arizona

2012 to 2015 ................................................. Graduate Assistant, Department of Chemistry and Biochemistry, The Ohio State University

Publications


Fields of Study

Major Field: Chemistry
Table of Contents

Abstract.................................................................................................................................................. ii
Dedication............................................................................................................................................... v
Acknowledgments.................................................................................................................................... vi
Vita.......................................................................................................................................................... viii
Table of Contents................................................................................................................................... x
List of Tables ........................................................................................................................................... xvii
List of Figures ........................................................................................................................................... xix

Chapter 1. Introduction........................................................................................................................... 1
  1.1   Protein Structure............................................................................................................................ 1
  1.2   Methods Used in Studying Protein Structure............................................................................... 2
  1.3   Basics of Mass Spectrometry.......................................................................................................... 4
    1.3.1 Overview..................................................................................................................................... 4
    1.3.2 Ionization.................................................................................................................................... 5
    1.3.3 Mass Analysis.............................................................................................................................. 8
    1.3.4 Gas Phase Dissociation............................................................................................................. 19
1.4 Intact and Top-down Mass Spectrometry based protein structural study 31

1.4.1 Native Mass Spectrometry ................................................................. 31
1.4.2 Ion Mobility .................................................................................. 33
1.4.3 Top-down Mass Spectrometry ......................................................... 41

1.5 Bottom-up Mass Spectrometry based protein structural study .......... 42

1.5.1 Covalent Labeling ........................................................................ 43
1.5.2 Non-covalent Labeling .................................................................. 45
1.5.3 Protein Digestion ........................................................................... 47
1.5.4 Structure modeling ......................................................................... 48

Chapter 2. Primary structure determination: Bottom-up with a customized
database assisted Top-down approach for avian hemoglobin sequencing .......... 51

2.1 Introduction ........................................................................................ 51

2.2 Experimental ...................................................................................... 55

2.2.1 Full scan mass spectrometry analysis of bird blood ....................... 55
2.2.2 Bottom-up experiments ................................................................. 56
2.2.3 Bottom-up data analysis ............................................................... 58
2.2.4 Top-down experiments and data analysis ...................................... 60

2.3 Results and Discussions ................................................................... 61

2.3.1 Intact mass of hemoglobin chains .............................................. 61
2.3.2 Customizing a database for searching ........................................... 63

2.3.3 Bottom-up approach ........................................................................... 66

2.3.4 Top-down completion of partial bottom-up sequences ....................... 69

2.4 Conclusions .......................................................................................... 74

Chapter 3. Effect of Conformational Flexibility on Gas-Phase Unfolding of
Noncovalent Protein Homodimers Probed by CID and SID .......................... 76

3.1 Introduction ............................................................................................. 76

3.2 Experimental .......................................................................................... 77

3.3 Results and Discussions ........................................................................ 78

3.3.1 Drastically different dissociation behavior of enolase dimer by CID and SID .................................................................................. 78

3.3.2 Dissociation of disulfide intact and disulfide reduced α-lactalbumin dimer ....................................................................................... 84

3.3.3 Dissociation of disulfide intact and disulfide reduced β-lactoglobulin dimer ....................................................................................... 93

3.4 Conclusions ............................................................................................ 95

Chapter 4. Top-down mass spectrometric study of monoclonal IgG by CID and SID 97

4.1 Introduction ............................................................................................. 97
4.2 Experimental ............................................................... 100

4.3 Results and Discussion.................................................. 101

4.3.1 Release of glycans..................................................... 101

4.3.2 Top-down +29 Refimab by CID and SID mainly releases light chain and b ions..................................................... 104

4.3.3 Gently reduced Refimab has light chain and complementary partner non-covalent interacting together ........................................... 106

4.3.4 Gently reduced Refimab releases light chain and heavy chain together in SID........................................................................................................ 108

4.4 Conclusions ........................................................................ 112

Chapter 5. SID behavior matches with in silico disassembly for several homo-hexamer protein complexes with known structures........................................... 114

5.1 Introduction ........................................................................ 114

5.2 Experimental ...................................................................... 116

5.3 Results and Discussions .................................................. 117

5.4 Conclusions ........................................................................ 129

Chapter 6. Quaternary structure determination: Structure Refinement of a Hetero-hexamer protein by multiple mass spectrometry approaches ..................... 131

6.1 Introduction ........................................................................ 131
### 6.2 Experimental

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2.1</td>
<td>TNH expression and purification</td>
<td>135</td>
</tr>
<tr>
<td>6.2.2</td>
<td>Complex analysis by native mass spectrometry</td>
<td>136</td>
</tr>
<tr>
<td>6.2.3</td>
<td>Collisional cross-section measurement and spherical subunit modeling</td>
<td>138</td>
</tr>
<tr>
<td>6.2.4</td>
<td>Bottom-up approach for surface mapping modification localization</td>
<td>142</td>
</tr>
<tr>
<td>6.2.5</td>
<td>Molecular modeling</td>
<td>144</td>
</tr>
</tbody>
</table>

### 6.3 Results and Discussions

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.3.1</td>
<td>Solution-disruption experiments reveal TNH subcomplexes $\alpha\beta\gamma$ and $\alpha\beta$</td>
<td>145</td>
</tr>
<tr>
<td>6.3.2</td>
<td>SID/IM and SID/IM/SID reveal increased information on subunit arrangement</td>
<td>148</td>
</tr>
<tr>
<td>6.3.3</td>
<td>Predicting the TNH structure via collisional cross-section constraints on coarse-grained models</td>
<td>152</td>
</tr>
<tr>
<td>6.3.4</td>
<td>Phenylglyoxal surface mapping of TNH helps determining models of $\alpha\beta\gamma$ trimer</td>
<td>156</td>
</tr>
<tr>
<td>6.3.5</td>
<td>ZDOCK helps determining models of TNH hexamer $\alpha_2\beta_2\gamma_2$</td>
<td>167</td>
</tr>
</tbody>
</table>

### 6.4 Conclusions

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4</td>
<td>Conclusions</td>
<td>170</td>
</tr>
</tbody>
</table>
Chapter 7. The Effect of Charge and His-tag on the Gas Phase Dissociation of Toyocamycin Nitrile Hydratase Hexamer Probed by CID and SID.......................... 172

7.1 Introduction .................................................................................................................. 172
7.2 Experimental .................................................................................................................. 175
  7.2.1 Reagents and surface mapping reactions ......................................................... 175
  7.2.2 Instrumentation for native mass spectrometry analysis ......................... 176
  7.2.3 Bottom-up approach for surface mapping modification localization .................................................. 177
7.3 Results and Discussions ............................................................................................... 179
  7.3.1 Surface mapping suggests that His-tag is on the protein surface and does not significantly alter the quaternary structure of TNH .......... 179
  7.3.2 His-tagging or charge states did not influence SID behavior of TNH significantly .............................................................................................................. 182
  7.3.3 His-tagging or charge states influence CID behavior of TNH significantly .............................................................................................................. 186
7.4 Conclusions ..................................................................................................................... 197

Chapter 8. SID/IM/MS as a Tool for Analysis of Composition and Subunit Organization of a Bacterial Multicopper Oxidase ............................................. 199

8.1 Introduction ................................................................................................................. 199
8.2 Experimental ............................................................................................................... 200
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.3</td>
<td>Results and Discussion</td>
<td>201</td>
</tr>
<tr>
<td>8.4</td>
<td>Conclusions</td>
<td>211</td>
</tr>
<tr>
<td>Chapter 9</td>
<td>Conclusions and Future Directions</td>
<td>213</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>220</td>
</tr>
</tbody>
</table>
List of Tables

Table 1.1. Reagents used in covalent labeling of proteins ........................................ 44
Table 1.2. Enzymes used for protein digestion ................................................................. 48
Table 2.1. Peptide components in the customized database (Variable sites are highlighted) ................................................................. 66
Table 3.1. Dimer complexes studied ............................................................................. 84
Table 5.1. Interface area, number of salt bridges and hydrogen bonds of different interfaces and those associated with different pathways ........................................ 118
Table 6.1. SID/IM experiment, trap DC bias and trap SID lenses voltage at different SID acceleration voltages. (Unit: V) ............................................................................. 138
Table 6.2. Experimentally determined CCS of subunits and subcomplexes, radius of subunits .................................................................................................................... 153
Table 6.3. GDH arginine exposed information and labeling results.............................. 158
Table 6.4. Table of PGO labeling results and extent of solvent exposer by trimer model based on 3HHT. .............................................................. 166
Table 7.1. Sequence information of TNH α, β, γ subunit ............................................ 175
Table 8.1 Sequence information of subunits MnxE, MnxF, MnxG ............................... 203
Table 8.2. Theoretical masses of hexamers composed of different combination of MnxE and F.
List of Figures

Figure 1.1. Block diagram of basic components in a mass spectrometer. ...................... 4
Figure 1.2. Summary of ion release mechanisms .......................................................... 6
Figure 1.3. Schematic of a quadrupole mass filter.......................................................... 8
Figure 1.4. Voltages applied on the electrodes on the xz plane and yz plane ................. 9
Figure 1.5. The a-q stability diagram.............................................................................. 10
Figure 1.6. Schematic of a linear ion trap................................................................. 12
Figure 1.7. Schematics of LTQ Velos mass spectrometer........................................... 13
Figure 1.8. Ion trapping efficiency changes as helium pressure changes in ion trap. ...... 13
Figure 1.9. CID efficiency changes as helium pressure changes in ion trap. .............. 14
Figure 1.10. Schematics of an instrument coupling ESI and TOF in an orthogonal way. 15
Figure 1.11. Design of a Kingdon trap. ........................................................................ 16
Figure 1.12. Design of a Knight quadrupole.............................................................. 17
Figure 1.13. View of Orbitrap with half cutaway. Red trace shows the ions orbiting around the central electrode. ................................................................. 18
Figure 1.14. Ions are ejected from C-trap to Orbitrap. .............................................. 19
Figure 1.15. Nomenclature of b and y type ion generated in CID for a heptapeptide. [31]
.................................................................................................................................. 20

Figure 1.16. a) Schematics of Waters QTOF II instrument. Hex$_2$ ion guide was removed and collision cell was moved towards TOF to accommodate SID device in between the quadrupole and the collision cell. b) Side view and 3D view of the SID device, composed of three entrance lenses (1A, B, C), entrance top (1D) and bottom (1E) deflector, surface (3A), middle bottom deflector (2A), exit top (2C) and bottom (2B) deflector, and two exit lenses (2D, E).......................................................................................................................... 23

Figure 1.17. Schematic representations of CID versus SID. ........................................... 25

Figure 1.18. Schematic of the Waters SYNAPT G2 with a surface induced dissociation cell incorporated after the CID cell and before the ion mobility chamber (SID/IM mode). .................................................................................................................................................. 26

Figure 1.19. Schematic of the Waters SYNAPT G2 with a surface induced dissociation cell incorporated after the ion mobility chamber and before the transfer CID cell (IM/SID mode). .................................................................................................................................................. 26

Figure 1.20. Nomenclature of c and z type ion generated in ECD/ETD for a heptapeptide. ............................................................................................................................................................................. 27

Figure 1.21. Cornell mechanism for c and z$^-$ ion observed in ECD. ......................... 28

Figure 1.22. Utah-Washington mechanism for c and z$^-$ ion observed in ECD. .......... 29

Figure 1.23. Diagram of Velos Pro dual linear ion trap, Orbitrap analyzer and ETD system in Orbitrap Elite. .................................................................................................................................................. 30

xx
Figure 1.24. Stepwise evolution after ESI of the structure of a globular protein .......... 32
Figure 1.25. Working Principle of drift tube, travelling wave and field-asymmetric ion mobility spectrometry. ........................................................................................................ 34
Figure 1.26. Plot of CCS versus m/z for different classes of biomolecules. ............ 37
Figure 1.27. SID (ΔV=100 V) spectrum of +15 TRAP 11mer with overlapping peaks in m/z space (up) and SID/IM plot with m/z overlapping peaks separated in the drift time space (down). ........................................................................................................ 38
Figure 1.28. The CCS of different models and experimental values of Aβ42 oligomers (left); A plausible mechanism for aggregation (right) .............................................. 41
Figure 1.29. Workflow of hydrogen/deuterium exchange monitored by mass spectrometry......................................................................................................................... 46
Figure 2.1. Flow chart of bottom-up data analysis ...................................................... 59
Figure 2.2. Intact mass of hemoglobin chains ............................................................ 62
Figure 2.3. Tandem MS of the top peptide hit for A2 region identified by SQID with Customized Database for the species Zonotrichia atricapilla. ......................... 69
Figure 2.4. MS/MS product ion map of Hb αA chain (a) and β chain (b) from Zonotrichia atricapilla. (c) Legend for fragmentation symbols ................................................. 70
Figure 2.5. Peptides aligned from a SQID search with a customized database in the bottom-up method for Zonotrichia atricapilla αA chain (a) and β chain (b). ............ 72
Figure 2.6. Phylogenetic tree of birds based on sequences of their Hb αA chains and generated by using phylogeny.fr .................................................................................. 74
Figure 2.7. Top-down-assisted bottom-up approach ................................................................. 75

Figure 3.1. Mass spectra obtained by spraying Enolase solution: a) super charged by
adding 2.5% m-NBA in 100 mM AmAc; b) 100 mM AmAc; c) Charge reduced by
adding 30 mM TEAA to 100mM AmAc .................................................................................. 79

Figure 3.2. Dissociation of enolase dimer at different charge states by CID and SID. .... 80

Figure 3.3. CCS of monomer products from +21 enolase dimer precursor by either CID
or SID .................................................................................................................................. 81

Figure 3.4. CCS profiles of remaining +21 enolase dimer upon CID (a) and SID (b) ..... 83

Figure 3.5 CID spectra of +11 α-lactalbumin dimer .............................................................. 85

Figure 3.6 SID spectra of +11 α-lactalbumin dimer .............................................................. 86

Figure 3.7 Energy resolved CID curves (a,b) and SID curves (c,d) for +11 charged native
α-lactalalbumin dimer (a,c) and disulfide-reduced α-lactalalbumin dimer (b,d) ............... 88

Figure 3.8. Tandem mass spectra of charge-reduced α-lactalalbumin +9 native dimer (a,c)
and disulfide-reduced dimer (b,d) with different dissociation methods ......................... 89

Figure 3.9. Collision induced unfolding of disulfide-intract (a) and disulfide-reduced (b)
+9 α-lactalalbumin dimer ....................................................................................................... 90

Figure 3.10. α-lactalalbumin monomer calcium binding information ............................... 91

Figure 3.11. CCS profiles of the monomers of different charge states released from
disulfide-intact or disulfide-reduced α-lactalalbumin dimer by either CID 900 eV or SID
1260 eV ................................................................................................................................ 92
Figure 3.12 Tandem mass spectra of β-lactoglobulin +11 native dimer (a,c) and disulfide-reduced dimer (b,d) with different dissociation methods ................................................................. 94

Figure 3.13. CCS profiles of the monomers of different charge states released from disulfide-intact or disulfide-reduced β-lactoglobulin dimer by either CID 880 eV or SID 330 eV ...................................................................................................................................................................................... 95

Figure 4.1. Structures of commonly used therapeutic monoclonal antibodies: IgG1, IgG2 and IgG4................................................................................................................................................................................................................................................................. 98

Figure 4.2. Sequence information for the heavy chain of Refimab ............................................. 99

Figure 4.3. Sequence information for the light chain of Refimab ............................................. 100

Figure 4.4. Release of N-glycan from protein by PNGase F ................................................... 101

Figure 4.5. Mass spectrum full scan of Refimab reacted with PNGase F. ......................... 102

Figure 4.6. Structure of released G0F glycosylamine.......................................................... 103

Figure 4.7.a) Mass spectrum full scan of Refimab reacted with PNGase F and small molecules removed by buffer exchange. b) Mass spectrum full scan of Refimab buffer exchanged to 100 mM AmAc. ................................................................................................................................................................................................................................. 104

Figure 4.8. a) SID 140 V and b) CID 180 V tandem mass spectra performed on +29 Refimab.................................................................................................................................................................................................. 105

Figure 4.9. Collision induced dissociation at an acceleration voltage of 150 V, performed on +29 a) disulfide-reduced Refimab by DTT and b) intact Refimab......................... 107
Figure 4.10. Zoom in y96 and y97 ion from light chain by fragmenting the +29 a) disulfide reduced Refimab with CID 190 V, compared with the same m/z region of the b) intact Refimab.

Figure 4.11. Ion mobility - mass spectrum of all the products released from isolated +29 reduced Refimab a) and intact Refimab b) by SID (ΔV= 140 V).

Figure 4.12. Extracted mass spectra from SID/IM 2D plot of +29 reduced Refimab.

Figure 4.13. Mobiligram of +29 disulfide-reduced Refimab CID at an acceleration voltage of 150 V.

Figure 5.1. Ball-and-stick representation for a D3 hexamer and it’s in silico disassembly pathways.

Figure 5.2. CID and SID tandem mass spectra of normal charged and reduced charged bovine GDH.

Figure 5.3. CID (ΔV=125 V) mobiligram of +36 Proteus GDH.

Figure 5.4. SID (ΔV=100 V) mobiligram of +36 Proteus GDH.

Figure 5.5. CID (ΔV=125 V) mobiligram of +27 GTP cyclohydrolase.

Figure 5.6. SID (ΔV=140 V) mobiligram of +27 GTP cyclohydrolase.

Figure 5.7. Extracted tandem mass spectra of products with different oligometric states from SID (ΔV=140 V) mobiligram of +27 GTP cyclohydrolase (The intensity is normalized to +6 monomer, the highest intensity product).

Figure 5.8. Insulin hexamer +11 CID (ΔV=70 V).
Figure 5.9. Insulin hexamer +11 SID ($\Delta V=60$ V) ................................................................. 127

Figure 5.10. Extracted tandem mass spectra for trimer, tetramer and pentamer from Insulin hexamer +11 CID ($\Delta V=70$ V) ........................................................................................................... 128

Figure 5.11. Extracted tandem mass spectra for trimer, tetramer and pentamer from Insulin hexamer +11 SID ($\Delta V=60$ V) ........................................................................................................... 128

Figure 6.1. A) The flow chart of the surface mapping experiment. B) Labeling reagent PGO and its reaction with arginine rendering a mass increase of 134.0368 or 116.0262. ........................................................................................................................................ 142

Figure 6.2. Methanol-induced solution-disruption of TNH hexamer ..................... 146

Figure 6.3. Increasing ionic strength from F) 0.1M ammonium acetate to A) 1M ammonium acetate doesn’t induce the dissociation of TNH hexamer. ....................... 147

Figure 6.4. A) SID/IM/MS plot (700 eV) and B) SID/IM/MS plot (1680 eV) for +14 TNH hexamer. The intensity of spots shows the normalized abundance of the ion species in square root scale. C) A cartoon showing the connectivity of subunits in TNH by SID/IM/MS results ........................................................................................................... 150

Figure 6.5 The energy resolved SID results for dissociation of +8 (A) and +9 (B) TNH $\alpha\beta\gamma$ trimer generated in the trap-SID and separated in the ion mobility cell. ............ 151

Figure 6.6. Coarse-grained model of TNH trimer and hexamer ......................... 155

Figure 6.7. Sensitivity and specificity of phenylglyoxal surface labeling on glutamate dehydrogenase hexamer ........................................................................................................ 159
Figure 6.8. Higher concentration of the labeling reagent disruptions the structure of TNH hexamer to subcomplexes ........................................... 161

Figure 6.9. RMSD of backbone atoms over simulation time for trimer model candidates ........................................................................................................... 163

Figure 6.10. Collisional cross-section for trimer model candidates during simulation compared with the experimental value .................................................. 164

Figure 6.11. Modeled TNH $\alpha\beta\gamma$ trimer structure based on homolog PDB 3hht and equilibrated in periodic water box for 10ns. Subunits $\alpha$, $\beta$ and $\gamma$ are in red, blue and green respectively. The exposed arginines are in stick representation. Labeled arginines are in cyan color. Arginine $\beta_2$ and $\gamma_26$ in gray are not labeled and thus postulated to be involved in trimer-trimer interaction to form hexamer .......................................... 167

Figure 6.12. RMSD of backbone atoms over simulation time for TNH hexamer model candidates ........................................................................................................... 168

Figure 6.13. Collisional cross-section for TNH hexamer model candidates during simulation compared with the experimental value ........................................... 169

Figure 6.14. Possible TNH hexamer structures (Hexamer 1 in green, 2 in magenta and 3 in blue) with docking of $\alpha\beta\gamma$ trimer structure based on homolog PDB 3hht and equilibrated in periodic water box for 10ns. .................................................. 170

Figure 6.15. A summary of TNH structure elucidation by multiple mass spectrometric techniques ........................................................................................................... 171

Figure 7.1. Dose-dependent average number of labels of different subunits. .......... 180
Figure 7.2. Surface mapping results. Sequence information of wild type TNH is shown on
the left and His-tagged TNH on the right. ................................................................. 181

Figure 7.3. Energy-resolved SID results of (a)+19, (b)+17, (c)+14 TNH wild type
hexamer; (d)+19, (e) +17, (f)+14 His-tagged TNH hexamer. ................................. 183

Figure 7.4. Unfolding of wild type (a-f) and His-tagged (g-l) THN hexamer with different
charges by CID. .................................................................................................................. 185

Figure 7.5. CID spectra at low collision energy (purple spectra) and high collision energy
(black spectra) of (a)+19, (b)+18, (c)+17, (d)+16, (e)+15, (f)+14 TNH wild type
hexamer; (g)+19, (h) +18, (i) +17, (j) +16, (k)+15, and (l)+14 His-tagged TNH hexamer.
........................................................................................................................................ 187

Figure 7.6. Energy-resolved CID results of (a)+19, (b)+18, (c)+17, (d)+16, (e)+15, (f)+14
TNH wild type hexamer; (g)+19, (h) +18, (i) +17, (j) +16, (k)+15, (l)+14 His-tagged
TNH hexamer.................................................................................................................... 188

Figure 7.7. Tandem mass spectrum of +19 wild type(a,d), His-tagged (b,e) and chimera
(c,f) hexamer by various collision energy. Zoom in monomers (m/z=1000-2500, yellow
shadow) are shown in d, e and f. His-tagged γ subunit is preferentially ejected versus wild
type γ subunit ........................................................................................................................ 189

Figure 7.8. Tandem mass spectrum of +17 wild type(a,d), His-tagged (b,e) and chimera
(c,f) hexamer by various collision energy. Zoom in monomers (m/z=1200-2700, yellow
shadow) are shown in d, e and f. His-tagged γ subunit is preferentially ejected versus wild
type γ subunit ........................................................................................................................ 190
Figure 7.9. Tandem mass spectrum of +14 wild type (a,d), His-tagged (b,e) and chimera (c,f) hexamer by various collision energy. Zoom in monomers (m/z=1100-4000, yellow shadow) are shown in d, e and f.......................... 191

Figure 7.10. Average charges on monomer products by a) CID of different charge state hexamer precursor; b) SID of different charge state hexamer precursor............... 193

Figure 7.11. Mobiligram of dissociating +19 chimera hexamer by 1520 eV SID....... 195

Figure 7.12. Energy-resolved CID results of (a)+19, (b)+17, (c)+14 TNH chimera hexamer, which is composed of two α, two β, one γ and one His-tagged γ subunits..... 196

Figure 8.1. Native full mass spectrum of Mnx, the major complex is EₓF₆₋ₓG. .......... 202

Figure 8.2. Ion mobility - mass spectrum of all the products released from isolated +29 of Mnx complex by collision induced dissociation (ΔV=150 V).......................... 204

Figure 8.3. Ion mobility - mass spectrum of all the products released from isolated +29 of Mnx complex by surface induced dissociation with low energy (ΔV= 40 V).......... 205

Figure 8.4. Ion mobility - mass spectrum of all the products released from isolated +29 of Mnx complex by surface induced dissociation with high energy (ΔV= 100 V)........... 206

Figure 8.5. a) Energy-resolved SID results for fragmentation of MnxE₃F₃G, focusing on E₃F₃ hexamer, and normalized to total EF oligomer intensity. b) Zoom in low intensity oligomers.............................................................. 207

Figure 8.6. Ball-and-stick representation for E₃F₃ hexamer and several of its in silico disassembly pathways (E is blue and F is red) ......................................................... 208

Figure 8.7. +4 monomer metal binding information. ............................................. 209
Figure 8.8. The heterogeneity of the Mnx complex.................................................. 211
Chapter 1. Introduction

1.1 Protein Structure

The knowledge of protein structures is essential for understanding how proteins function, and the roles they play in living organisms. Generally, protein structure is described in three definitive levels with most proteins then forming a fourth level[1]. Proteins are composed of amino acids. The primary structure is the amino acid sequence, while the secondary structure refers to the local conformation of a protein covering partial sequence, such as α-helix and β-sheets. The next level is the tertiary structure which is the three-dimensional organization of all atoms in a polypeptide chain. Many proteins also have two or more chains or subunits associated non-covalently to function. Quaternary structure refers to the arrangement of these subunits. It is widely believed that primary structure directs the secondary structure, and secondary structure drives the tertiary structure. Quaternary structure can be considered to be developed through some common guiding principles such as the use of symmetry, the need to keep overall nucleic acid sequence lengths minimized, and decreasing translation error frequency.

The experimentally solved three-dimensional coordinates of proteins are stored in the online Protein Data Bank. As of 2014, at least 60% of those entries are dimers or larger assemblies.[2] This shows that most of proteins do not exist as simple monomeric entities, thus the investigation of the structure of protein quaternary complexes is crucial for understanding their functions.
1.2 Methods Used in Studying Protein Structure

X-ray crystallography has the resolution to determine the three-dimensional structure information at the atomic level for analytes in the solid phase. The process [1] requires generating a crystal and placing the crystal between an X-ray beam (wavelength 0.7-1.5 Å) and a detector. The X-rays are diffracted by the crystal and a regular array of spots on the detector is formed. The diffraction pattern is further converted to an electron-density map by a Fourier transform. Then the final structure is built by the guidance of the electron-density map and multiple refinements of the structure. The derived protein structure represents the spatial and temporal average of the structures within the crystal, which limits the method in providing information about the dynamic nature of proteins. Additionally, X-ray crystallography requires a purified and crystallized sample, which may be difficult to achieve for some important proteins.

Nuclear magnetic resonance (NMR) [1] can measure three-dimensional structures of molecules in aqueous solution. NMR makes use of nuclear spin energy splitting in atoms, such as \(^1\)H, \(^{13}\)C, \(^{15}\)N, \(^{19}\)F and \(^{31}\)P, placed in a static magnetic field. The frequency of the electromagnetic radiation for energy resonance is determined by the chemical environment of the atoms, and further converted to a chemical shift in NMR spectra. Analysis of large molecules like proteins requires two-dimensional NMR. The method measures distance-dependent coupling of nuclear spin (the nuclear Overhauser effect, NOE) or nuclear spin coupling of atoms that are covalently connected (total correlation spectroscopy, TOCSY).[3] The analysis of large protein complexes by NMR is challenging. One reason is the line broadening as the size of the molecules increase.
Another reason is signal overlapping due to the increased number of residues.[4] In addition, NMR has difficulties in analyzing mixtures. With the development of NMR-based techniques, such as chemical shift perturbations (CSPs) [5] and paramagnetic relaxation enhancements (PREs) [6], information about subunit interfacial interactions can be obtained and together with three-dimensional high resolution data for each subunit, an overall structure for large complexes can be built.[4]

Compared to the classical structural tools, X-ray crystallography and NMR, there are also lower resolution structural techniques, such as cryo-electron microscopy (Cryo-EM), small-angle X-ray scattering (SAXS) and mass spectrometry. Samples studied by cryo-EM are frozen hydrated specimens. The two-dimensional projections of the three-dimensional structure in all possible orientation are acquired and analyzed to construct the structure.[7] The resolution of Cryo-EM can reach near-atomic resolution (≈3.8–4.5 Å) for several macromolecular complexes, such as GroEL, cytoplasmic polyhedrosis virus and epsilon 15 bacteriophage.[8] SAXS provides overall size and dimensions for large biomolecules. Monochromatic X-rays with wavelength 1-7 Å impinge on the sample in aqueous solution. The scattering pattern is observed as small angle deviations from the primary beam.[9] The isotropic orientation of biomolecules in the solution makes this technique low resolution compared to X-ray crystallography. But shape information for large protein assemblies can be obtained easily and used as restraints for structural modeling. SAXS is a powerful technique, especially in combination with the domain/subunit high resolution structural data measured from NMR or other techniques.
Numerous experiments have shown that under careful manipulation, mass spectrometric analysis can reflect the protein complex structure in the native condition. The information includes composition and stoichiometry of different components. Analytes with low purity can be directly analyzed in mass spectrometry due to the fact that ions of interest can be isolated in the gas phase based on their unique mass-to-charge ratio, \( m/z \). The amount of sample required to perform mass spectrometric analysis is in the range of \( \mu g \), which is very small compared to other techniques. Furthermore, the coupling of ion mobility with mass spectrometry provides size and shape information about the analytes. Labeling techniques coupled with mass spectrometry provide more structural information that can be used to postulate the three-dimensional structures.

1.3 Basics of Mass Spectrometry

1.3.1 Overview

Mass spectrometry deals with ions in the gas phase, because charged particles are easy to guide and detect.[10] A simple schematic of a simple single stage mass spectrometer is shown in Figure 1.1.

![Figure 1.1. Block diagram of basic components in a mass spectrometer.](image)
The sample inlet transfers the sample to the ion source. The ion source converts analytes into gas-phase ions. The mass analyzer then separates the ions based on different $m/z$ and the detector detects the ions and amplifies the ion current. A data system converts the electrical signals into mass spectra. The mass analyzer and detector are kept under vacuum to maintain the lifetime of ions, because the charges on ions get quenched or the ions are deflected off path by collision with molecules in air. Most mass spectrometers use combinations of roughing pumps and turbomolecular pumps to maintain the required vacuum for a specific type of mass analyzer. Activation and dissociation of the ions can be performed by coupling stages of mass analysis in time (ion trap) or in space (Q-TOF, QQQ).

1.3.2 Ionization

Electrospray ionization (ESI) and nano-electrospray ionization (nano-ESI) were almost exclusively used in my research. Both of the methods involving applying high voltage (1 - 4 kilovolts range) at the tip of a capillary filled with the analyte solution. Electrospray ionization is performed at a higher liquid flow rate (usually a few $\mu$L/min) with organic content while nano-ESI is performed at much lower flow rate (tens of nL/min). Usually in nano-ESI, there is no pump driving the liquid flow, the electrostatic force applied at the capillary will drive the liquid flow. Moreover, nebulizing gas is usually supplied in ESI which is not always supplied in nano-ESI. The advantages of nano-ESI are low sample consumption, increased sensitivity and higher salt tolerance than ESI.[11] In ESI, the solution at the capillary tip forms a Taylor cone that emits charged droplets in the micrometer range. In nano-ESI, these initial charged droplets are
believed to be one order of magnitude smaller.[11] In both ionization techniques, the initial charged droplets undergo repeated evaporation and fission events until the radii reach a diameter of a few nanometers in size. Analyte ions are generated from these highly charged nanodroplets.[12]

There are three models which describe the process of analyte ions being released from the highly charged nanodroplets: ion evaporation model (IEM), charged residue model (CRM) and chain ejection model (CEM).[12] A simple drawing (Figure 1.2) shows the major differences among the three models.

![Figure 1.2. Summary of ion release mechanisms](image)

Reproduced with permission from reference [12]. Copyright (2013) American Chemical Society
It is believed that low molecular weight analyte ions follow the IEM. Molecular dynamic simulations (MD) were performed on water/methanol droplets with NH$_4^+$ ions.[13] The results show that a NH$_4^+$ is ejected with an extended solvent bridge formation and rupture. The initial product is a small cluster with the ammonium ion and several solvent molecules. The solvent molecules are then eliminated when the initial cluster travels through the early stages of the mass spectrometer and collides with gas.[14]

Large globular analyte ions releasing from the highly charged nanodroplets are believed to follow the CRM. In this process, solvent in the nanodroplets evaporate to leave charges on the analyte ions. The experimental support for the CRM model is that globular proteins have charge states similar to the Rayleigh charge of water droplets of the same size as a given protein.[15]

Unfolded protein chains are released from the highly charged nanodroplets by the CEM. Molecular simulation performed on extended chains shows that the macromolecule is quickly expelled from the droplet surface sequentially.[16, 17] The CEM takes place when a protein is unfolded because the hydrophobic interior is exposed to the solvent. This makes the protein unstable in the droplet, so that the unfolded chain immediately migrates to the droplet surface and one terminus is ejected to the gas phase, followed by the remaining chain which gets expelled. This mechanism is used to explain the charge state distribution of unfolded proteins in ESI spectra. The wide range of charge states result from different radii of the highly charged nanodroplets.
1.3.3 Mass Analysis

Once the analyte is ionized and passed to the vacuum, its \( m/z \) information needs to be measured by a mass analyzer. There are different types of mass analyzers that vary in measuring principle, speed, accuracy, resolution and cost. The mass analyzers being used for the research reported in this dissertation are introduced here.

1.3.3.1 Quadrupole Mass Filter

A quadrupole mass filter consists of four parallel electrodes in a radial arrangement (Figure 1.3). The cross section of the electrodes is parabolic in an ideal world but is often round due to lower manufacturing cost.[18] Two electrodes in the \( xz \) plane (with the ions transmitted down the \( z \) axis) share the same voltage, which consist of a direct current (DC) voltage \( (U) \) and a radio-frequency (RF) voltage \( (V \cos \omega t) \). The other two orthogonal electrodes (yz plane) have an equal and opposite applied voltage \( (-U-V \cos \omega t) \).

![Figure 1.3. Schematic of a quadrupole mass filter](image)

Modified drawing from JEOL and Waters websites

Understanding the principle of how a quadrupole mass filter works can be approached from a qualitative perspective. The average voltage applied on the \( xz \) plane is
positive due to the positive DC voltage applied. If an ion is relatively heavy, it will tend to be influenced by the average of a varying voltage. The average positive voltage provided by electrodes in the xz plane will make heavy ions pass through. However, if an ion is relatively light, it may be influenced by RF voltage more significantly, especially in the negative cycle the ions may have been attracted to collide with the negative electrodes already. Thus the electrodes in the xz plane can serve as a high mass pass filter. The RF voltage applied on the electrodes on yz plane is 180° out of phase of that applied on the xz electrodes and DC voltage applied on yz plane is the negative of that applied on the xz electrodes (Figure 1.4). These electrodes serve as a low mass pass filter. The combination of high mass filter and low mass filter results in passing of ions with a specific m/z.

Figure 1.4. Voltages applied on the electrodes on the xz plane and yz plane
Dark dash lines show zero voltage. Blue traces show how the voltage varies along with the time in the two cases. Red traces show the average or DC voltage, it is positive for the electrodes in the xz plane and negative for the electrodes in the yz plane.

Understanding the principle of how quadrupole mass filter work can also be approached from a quantitative perspective.[10] The motion of an ion in the xy plane of a quadrupole is described by the Mathieu equation:
\[
\frac{d^2u}{d^2(\omega^2)} + (a_u - 2q_u \cos \omega t)u = 0 \quad \text{Equation 1-1}
\]

Where \( u \) is \( x \) or \( y \) from the center of the field, \( \omega \) is the angular frequency of the RF voltage applied, \( t \) is time, and \( a_u, q_u \) are shown as the following equations:

\[
a_u = a_x = -a_y = \frac{8zeU}{m \omega^2 r_0^2} \quad \text{Equation 1-2}
\]

\[
q_u = q_x = -q_y = \frac{4zeV}{m \omega^2 r_0^2} \quad \text{Equation 1-3}
\]

Where \( e \) is the unit of charge \( 1.602 \times 10^{-19} \) C, \( U \) is the DC voltage, \( V \) is the amplitude of the RF voltage applied and \( r_0 \) is the inscribed radius of the quadrupole (Figure 1.3).

---

Figure 1.5. The a-q stability diagram.
Reproduced from ref.[10] by permission of John Wiley and Sons

Once plotted as \( a_u \) versus \( q_u \), a stable region can be obtained and it is shown in Figure 1.5. The \( m/z \) can be scanned at the summit of the stable region by varying \( U \) and \( V \), but keeping the ratio constant. (Scan line in Figure 1.5). At the apex, the \( a_u \) value is
0.237 and \( q_u \) is 0.706. Given that a limited amplitude of voltages can be applied, the highest \( m/z \) that can be measured is limited based on the equation.

\[
\frac{m}{z} = \frac{eV}{q\pi^2f^2r_0^2} \quad \text{Equation 1-4}
\]

In order to obtain good transmission for large \( m/z \) ions, the frequency of the RF voltage needs to be reduced. However, the mass resolution decreases with decreased frequency, according to Equation 1-5.[19, 20]

\[
\text{Resolution} = \frac{m_f^2L^2}{\pi^2f^2V_{\text{acc}}} \quad \text{Equation 1-5}
\]

where \( V_{\text{acc}} \) is the acceleration voltage, \( f \) is the frequency of the RF voltage and \( L \) is the length of the rods. A balance point is needed for adjusting the RF frequency.

Quadrupoles in many occasions are operated in RF only mode. The value of \( a_u \) is zero and the scan line is the \( q_u \) axis in the stable diagram (Figure 1.5). The ions of a wide \( m/z \) range are then stable within the quadrupole, as long as \( q_u < 0.908 \). RF mode quadrupoles are operated as ion guides or collision cells in almost all the commercial hybrid mass spectrometers.

1.3.3.2 Linear Ion Trap

A linear ion trap consists of three segments of four hyperbolic rods (Figure 1.6).[10] Small slits are cut along the center \( x \) electrodes to eject ions for detection. Ions are trapped in the axial direction (z direction) by applying DC voltages on the separate sections. The radial direction (xy direction) trapping of ions is achieved by a main RF voltage applied on the \( x \) and \( y \) electrodes. For separation, supplementary RF voltages are applied to the \( x \) electrodes for ion isolation, activation and ejection. To perform \( m/z \) analysis, the main RF voltage is increased gradually which causes ions with increasing
m/z to become unstable successively. The unstable ions are then ejected in the radial direction and get detected at two electron multiplier detectors.

The ion trap type instrument used in this dissertation is Thermo Scientific LTQ Velos Pro. It is a dual linear ion trap instrument with the first trap acting as a high pressure cell (HPC) and the second one acting as a low pressure cell (LPC). The schematic of an LTQ Velos is shown in Figure 1.7. The high pressure cell has pressure around 5 mTorr compared to 2.5 mTorr that is normally used in ion trap instruments. The increase in ion trap pressure leads to improved trapping efficiency and higher dissociation efficiency as shown in Figure 1.8 and Figure 1.9, respectively. Therefore, the ions can be more effectively trapped and fragmented in the HPC. Unfortunately, increased pressure causes decreased resolution in ion traps. [22] To counteract this problem, a second trap was added. Thus, the LPC with pressure around 1 mTorr is used for scanning the ions to obtain the m/z information.
Figure 1.7. Schematics of LTQ Velos mass spectrometer. Reproduced from Thermo PSB-127-Dual-Pressure-Linear-Ion-Trap-Technology

Figure 1.8. Ion trapping efficiency changes as helium pressure changes in ion trap. Reproduced from Thermo PSB-127-Dual-Pressure-Linear-Ion-Trap-Technology
1.3.3.3 Time of Flight

The principle of m/z measuring in a time of flight (TOF) mass analyzer is shown in the following equations.

The ions with different mass (m) and charge (q) are accelerated with the same voltage (V), thus the kinetic energy (KE) obtained is:

\[ KE = Vq \quad \text{Equation 1-6} \]

The velocity (v) of an ion can be obtained from rearranging the expression

\[ KE = \frac{mv^2}{2} \]

\[ v = \left( \frac{2KE}{m} \right)^{1/2} = \left( \frac{2Vq}{m} \right)^{1/2} \quad \text{Equation 1-7} \]

If an ion with the velocity described above enters a field-free flight tube (length L), the time (t) needed to arrive at the detector is:

\[ t = \frac{L}{v} = L\left( \frac{m}{2Vq} \right)^{1/2} \quad \text{Equation 1-8} \]

Thus, the m/z of an ion can be derived from its arrival time.
In TOF measurements, all ions are required to enter the flight tube at the same time. A pulse-mode ion source like MALDI readily meets this requirement. To couple a continuous ion source, such as electrospray ionization, to a batch-mode mass analyzer such as a TOF, a pulse of accelerating voltage in a region prior to the flight tube is required as shown in Figure 1.10. The ion beam from ESI is orthogonal to the acceleration electric field to reduce dispersion in space and kinetic energy spread for the initially generated ions from the source. In order to further decrease the kinetic energy spread and increase TOF resolution, a reflectron [23] is utilized as shown in Figure 1.10. A reflection is an electrostatic mirror composed of a stack of ring electrodes with increasingly repelling potentials. The direction of the ion path is reversed at this mirror. The ions with more kinetic energy will penetrate deeper in this mirror and the longer distance for traveling compensates for the higher initial velocity. Thus, the reflectron corrects the kinetic energy spread of ions analyzed in TOF.

Figure 1.10. Schematics of an instrument coupling ESI and TOF in an orthogonal way. Reproduced from ref.[10] by permission of John Wiley and Sons
1.3.3.4 Orbitrap

The very first model of related to Orbitrap trapping was proposed by K. H. Kingdon [24] as shown in Figure 1.11. A DC voltage is applied between the outer and inner electrodes, producing a radial logarithmic potential (Φ) shown in the following equation,

\[ \phi = A \ln(r) + B \quad \text{Equation 1-9} \]

where \( r \) is the radial coordinate, A and B are constants.

![Figure 1.11. Design of a Kingdon trap. Modified from Reference [24] with permission. Copyright © 1923, American Physical Society](image)

A modified version of the Kingdon trap is a Knight quadrupole (Figure 1.12). R. D. Knight[25] modified the shape of the outer electrode of the Kingdon trap to produce an axial quadrupole potential term in addition to the radial logarithmic potential. The potential (Φ) is expressed in the following equation:

\[ \phi = A(z^2 - \frac{r^2}{2} + B \ln r) \quad \text{Equation 1-10} \]

The Knight quadrupole has RF voltage applied between the split outer electrodes. This allows the observation of ion motion in both the axial and radial directions.
The modern orbitrap was devised by Alexander Makarov.[26, 27, 28] It is a modified ‘Knight-style’ Kingdon trap with spindle shaped inner and outer electrodes. (Figure 1.13) Only DC voltage is applied on the Orbitrap. The m/z analysis is obtained by converting axial oscillation frequency of ions.

The ‘Quadro-logarithmic’ electrostatic potential between the electrodes is expressed as follows,

$$U(r, z) = \frac{k}{2} \left( z^2 - \frac{r^2}{2} \right) + \frac{k}{2} (R_m)^2 \ln \left[ \frac{r}{R_m} \right] + C$$  \hspace{1cm} \text{Equation 1-11}$$

where \((r, z)\) shows the position between the electrodes, \(k\) is the field curvature, \(R_m\) is the characteristic radius and \(C\) is a constant.

The electric field along the \(z\) direction can be derived from the above equation to be the following equation:

$$E_z = \frac{dU(r,z)}{dz} = k z$$ \hspace{1cm} \text{Equation 1-12}$$
From Newton’s second law of motion, the following equation can be obtained:

\[ q \cdot E_z = m\ddot{z} = qkz \quad \text{Equation 1-13} \]

\[ m\ddot{z} = qkz \] is a harmonic oscillation which has a solution:

\[ z = \sin\left( \sqrt{\frac{q}{m}} k \cdot t \right) \quad \text{Equation 1-14} \]

Therefore, the angular frequency (\( \omega \)) of axial oscillation is

\[ \omega = \sqrt{\frac{q}{m}} k \quad \text{Equation 1-15} \]

Figure 1.13. View of Orbitrap with half cutaway. Red trace shows the ions orbiting around the central electrode.
Reproduced from Reference [28] with permission of John Wiley and Sons

In our hybrid instrument with an Orbitrap analyzer, the Orbitrap Elite, the ions are injected to the Orbitrap via a curved linear trap (C-trap, Figure 1.14). The entrance on the Orbitrap is off the center plane (where \( z=0 \) in Figure 1.13). The ions sharing the same m/z will be squeezed along the axial direction and move in coherent axial oscillation. The detection of an ion’s axial oscillation frequency is through an image current detected on split outer electrodes and amplified by a differential amplifier. The signal transient in the frequency domain is than Fourier-transformed to m/z information of the ions.
1.3.4 Gas Phase Dissociation

To obtain structure-specific information by MS, it is crucial to perform tandem mass spectrometry analysis (MS/MS). The basic operation of MS/MS is selection of a desired precursor ion with a specific m/z from all ions produced in the ion source (MS1), followed by the selected ions undergoing activation and dissociation in an intermediate region. The products produced are then mass analyzed (MS2). For some mass analyzers, such as ion trap, the described process can be continued many times to produce MS\textsuperscript{n} analysis (fragmenting successive fragments of fragments...). Several activation and dissociation methods that are used in this dissertation are briefly introduced here.

1.3.4.1 Collision Induced Dissociation

Collision Induced Dissociation (CID) is the most popular dissociation method that is readily accessible in most commercial instruments. CID is performed by accelerating analyte ions to collide with neutral gas atoms or molecules (helium, nitrogen, argon, etc).
[29, 30] During the collision events, kinetic energy from the collision is partially converted to internal energy of the analyte, leading to the activation and dissociation of the analyte ion. The maximum amount of kinetic energy for the conversion ($E_{com}$) during collision depends on the mass of the precursor ($m_p$), mass of the gas molecule ($m_g$), dissociation energy in the laboratory fame ($E_{lab}$), and is given by the following equation:

$$E_{com} = \frac{m_g}{m_p + m_g} E_{lab}$$

Equation 1-16

Because the energy is deposited gradually through multiple collisions, CID is ascribed as a multistep activation process and the dissociation occurs via the lowest energy pathway. The lowest energy pathway for the dissociation of protonated peptides and small proteins by CID is the cleavage of the amide bonds to produce b and y type ions. [31] The nomenclature of b and y type ions are illustrated in Figure 1.15 with CID of heptapeptide “STELLAR” shown as an example.

![Figure 1.15. Nomenclature of b and y type ion generated in CID for a heptapeptide. [31]](image)

The mass of b and y ions follows the following equations:

b ion mass = $\Sigma$residuemass (N-terminus) + H-e
y ion mass = $\Sigma$residuemass (C-terminus) + O + 3H-e

Equation 1-17
The intensity of b and/or y fragments generated by cleaving various amide bonds shows that the cleavage at some amide bonds are preferred over cleavage at other amide bonds. The observation of preferred cleavages in peptides can be explained by the “mobile proton” model [32].

Typically, large protein complexes in CID experience tens of thousands of collisions during the time spent in the collision cell, which is about 200-400μs.[33] The use of high gas pressure and heavy collision gas atoms or molecules increases the conversion of kinetic energy to internal energy.[33] CID has been applied to dissociate non-covalent protein complexes to confirm composition and stoichiometry of the complexes. [34] Many groups have reported that in collision induced dissociation of large protein complexes, an unfolded monomer is ejected and carries away about half of the charge of the complex’s precursor ion. Sobott and Robinson [35] observed the ejection of highly charged monomer from GroEL 14mer by CID. The authors attributed the asymmetric dissociation to a highly excited state of the complex, which aids the migration of charges onto the monomer, followed by the ejection of the monomer. The collision induced dissociation behavior of homodimers, the simplest protein complexes, was investigated by Jurchen and Williams. [36, 37] They found that asymmetric charge partitioning is influenced by energy input, charges on precursors and conformational flexibility. In addition, Hall et al. [38] found that subunit flexibility, charge density, inter-subunit salt bridges and interface area influence symmetry of charge partitioning by studying a set of protein complexes.
1.3.4.2 Surface Induced Dissociation

Ion activation and dissociation can be also achieved by colliding ions with a surface.[39, 40] The surface used can be either a graphite surface, other surfaces (diamond, dirty metal) or self-assembled monolayers (SAM) adhered to a metal surface.[41] Surface-induced dissociation (SID) is more efficient in energy deposition as shown in Equation 1-16 due to the large mass of the surface compare to that of gas molecules in CID. SID has shown its application in fundamental studies of peptide fragmentation.

In the recent decade, SID has been applied in the Wysocki lab on QTOF type mass spectrometers for studying quaternary structures of proteins. The incorporating of SID device to a Waters Micromass QTOF II was described in the Reference [42] and a schematic of the modified instrument is shown in Figure 1.16.
On the modified QTOF II instrument with SID capability, Jones et al. [43] reported symmetric charge partitioning by dissociating noncovalent cytochrome C dimer by SID (+11dimer $\rightarrow$ +6 and +5 monomer), whereas, highly asymmetric dissociation
behavior is observed by CID (+11dimer $\rightarrow$ +8 and +3 monomer). Beardsley et al. [44] observed monomers carrying charges proportional to their mass fraction in several protein homotetramer and homopentamer by SID (1/4 or 1/5 of the charges from the precursors). Blackwell et al. [45] utilized SID on the QTOF II platform to dissect a heterohexamer protein complex composed of two copies of each α, β and γ subunit. The dominant αβγ trimer products in SID highly suggest that the hexamer is composed of two trimers, while by CID, only single chain ejection of α and β has been observed.

The different activation processes of CID versus SID illustrated in Figure 1.17 serve as simple explanations for the drastically different spectra seen between CID and SID. Large protein complexes in CID experience tens of thousands of collisions during the time spent in the collision cell, thus the energy is added stepwise. The favorable pathway involves subunit gradually unfolding, charges migrating to the unfolded subunit to reduce Coulombic repulsion, disrupting intersubunit interaction. In contrast, SID is proposed to be deposit a higher amount of energy in a single step. Thus, SID can favor the fast, direct dissociation pathway, yielding more symmetric charge-to-mass dissociation products.
In CID, energy is deposited step by step and single subunit unfolding and ejection (A pathway) is more favorable. In SID, the high energy is deposited in one step results in more kinetically favored direct dissociation to folded subunits (B pathway).

Reproduced from Ref. [46] Copyright © 2014 American Chemical Society

The Wysocki lab has also incorporated SID devices into Waters SYNAPT G2 and G2-S quadrupole/ion mobility/TOF instruments. The SID device can be placed either in front of the ion mobility cell (Figure 1.18) or after the ion mobility cell (Figure 1.19). Recently, we have also performed experiments with two SID devices placed in the G2-S mass spectrometer, one before and one after the ion mobility cell (SID/IM/SID mode). In SID/IM mode, numerous fragments generated by SID can be easily interpreted by the mobility separation [47] and their corresponding structurally characteristic, collisional cross-sections, can be obtained [48, 49]. In IM/SID mode, SID can be performed on precursor ions with similar m/z but different conformations that separated in the drift time space. [50] In the SID/IM/SID mode, we can perform pseudo-MS$^3$ type experiment. This
proves especially useful to further dissociate some subcomplexes uniquely generated by SID experiments, such as TNH αβγ trimer from TNH α₂β₂γ₂ hexamer.

Figure 1.18. Schematic of the Waters SYNAPT G2 with a surface induced dissociation cell incorporated after the CID cell and before the ion mobility chamber (SID/IM mode).

Reproduced from Ref.[49] with permission of John Wiley and Sons. Copyright © 2012 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Figure 1.19. Schematic of the Waters SYNAPT G2 with a surface induced dissociation cell incorporated after the ion mobility chamber and before the transfer CID cell (IM/SID mode).

Reproduced from Ref. [50] with permission of American Chemical Society. Copyright © 2012 American Chemical Society

SID has been shown to disrupt the weaker interfaces in protein complexes selectively to release subcomplexes, which reflect the native structures of protein complexes. Ma et al.[51] has shown that SID of glutamate dehydrogenase, a dimer of trimers protein complex, results in trimer products. SID of 20S proteasome with α₇β₇α₇ symmetry showed the generation of α₇β₇ products, which is consistent with the stacked ring topology of the complex. [52] Similarly, GroEL, which is composed of two stacked heptamer rings, was dissociated to heptamers upon SID.[47]
A part of my dissertation covers some fundamental studies of protein complexes with standards used as model systems to characterize the SID performance. Besides those studies, there are two exciting SID applications revealing structures of TNH and Mnx without assistance of any high-resolution structural data. These studies demonstrate the power of SID in structural biology.

### 1.3.4.3 Electron Capture and Transfer Dissociation

Electron capture dissociation (ECD) involves interaction of multiply protonated analytes with low energy (0.2 eV) electrons.[53] The dissociation method was first introduced in 1998 by Zubarev and coworkers.[54] Electron transfer dissociation (ETD), which was introduced later in 2004 by Hunt, Coon, Syka and coworkers, involves reaction of radical anions instead of electrons with multiply charged analyte cations.[55] Different from CID and SID, the fragmentation of a peptide by ECD or ETD is dominated by the N-Cα bond cleavage and results in generation of c and z´ ions (Figure 1.20) rather than b and y ions. ECD/ETD has the advantage of sequencing large and post-translationally modified (PTM-containing) peptides.[56]

![Figure 1.20. Nomenclature of c and z type ion generated in ECD/ETD for a heptapeptide.](image-url)
The mass of c and \( z^- \) ions follows the following equations:

\[
c \text{ ion mass} = \Sigma \text{residue mass} (\text{N-terminus}) + N + 4\text{H} - e \\
\]

\[
z^- \text{ ion mass} = \Sigma \text{residue mass} (\text{C-terminus}) + O - N + \text{H} - e \\
\text{Equation 1-18}
\]

One side reaction is hydrogen transfer.\[57\] It leads to the formation of [c-1]- and \[z+1\] ion products. Hydrogen transfer is believed to happen when the c and \( z^- \) ions remain in a complex after the cleavage of the N-C\( \alpha \) bond. Precursors with low charge states tend to have more hydrogen transfer side reaction due to less electrostatic repulsion in the complex of c and \( z^- \) ions. The mass shifting caused by this side reaction needs to be considered while interpreting the ETD spectra.

There are several proposed mechanisms explaining how c and \( z^- \) ions are generated by ECD. The Cornell mechanism\[54\] and the Utah-Washington mechanism\[58, 59, 60, 61\] are briefly introduced here. In the Cornell mechanism, because a protonated carbonyl has higher electron affinity than a protonated nitrogen in an amine or amide, the electrons attach to the protonated carbonyl and induce the cleavage of N-C\( \alpha \) bond as shown in Figure 1.21.

![Figure 1.21. Cornell mechanism for c and \( z^- \) ion observed in ECD.
Reproduced from Ref [54] with permission. Copyright © 1998, American Chemical Society](image-url)
In the Utah-Washington mechanism, the electron directly attaches to the amide $\pi^*$ orbital forming a radical anion. The neutralization of the anion is achieved by cleaving of the N-Cα bond. The mechanism is illustrated in Figure 1.22.

![Utah-Washington mechanism](image)

Figure 1.22. Utah-Washington mechanism for c and z$^\prime$ ion observed in ECD. Reproduced from Ref. [58] with permission. Copyright © 2003, American Chemical Society

ECD is almost exclusively realized in Fourier transform ion cyclotron resonance mass spectrometer due to the requirement of electron trapping. ETD is applied to observe similar fragmentation in other types of mass spectrometers. The ETD experiments in this dissertation are performed in an Orbitrap Elite. The diagram of the Velos Pro dual linear ion trap, Orbitrap analyzer and ETD system is shown in Figure 1.23.
ETD reactions occur in the high pressure linear ion trap. Charge-sign independent trapping (CSIT) enables trapping of both positive ions (multiply charged analyte ions) and negative ions (ETD reagent radical anions) in the same space at the same time. High-purity nitrogen (99.999%) is supplied to the ETD module. The filament is the source of 70eV electrons. Nitrogen molecules collide with the 70eV electrons in the ion volume, resulting in the generation of secondary electrons from the nitrogen molecules. These secondary electrons, which have near thermal kinetic energies, are captured by the ETD reagent fluoranthene (supplied as reagent 1) to form reagent radical anions that are transferred into the high pressure linear ion trap via the transfer multipole, HCD cell, C-Trap, quadrupole mass filter and low pressure linear ion trap. The quadrupole mass filter is used to remove the adduct ions of the fluoranthene radicals with molecular nitrogen ($m/z$ 216) and to transfer the fluoranthene radical anions ($m/z$ 202).
1.4 Intact and Top-down Mass Spectrometry based protein structural study

1.4.1 Native Mass Spectrometry

Native mass spectrometry refers to the macromolecule ions retaining their “native-like” structures and interactions in the solution phase after being ionized and analyzed by mass spectrometry. In order to be analyzed by mass spectrometry, the protein/protein complex is first buffer exchanged from their original buffer solution, such as Tris buffer, to a mass spectrometry compatible buffer solution, such as an ammonium acetate solution, with a similar ionic strength. Mass spectrometry compatible buffer means the buffer components need to be volatile, such as ammonia and acetic acid in the ammonium acetate solution, thus the solute can be efficiently removed in the early stages of the vacuum as an important desolvation step, without degrading mass spectrometer performance by salt buildup on surface.[34] Native MS almost exclusively utilizes nano-ESI, which has been shown to preserve non-covalent interactions, such as H-bonds and salt bridges, of the protein complex or protein-ligand complex.[62, 63, 64] Different mechanisms of ESI/Nano-ESI are introduced earlier in 1.3.2.

It is still in debate whether native-like structures are preserved in the gas phase. Soluble proteins have been evolving in aqueous environment, and the interaction of water molecules with proteins play an important role in the folding of protein. [65, 66] While transferred into a vacuum system, it is undeniable that structures can be altered in some cases [67, 68, 69], especially the experiments that trap the ions in the gas phase for tens of seconds. Bruker and McLafferty described the structural evolution of a protein after ESI into the gas phase.[66] The structural modification may involve side-chain collapse,
unfolding (loss of hydrophobic bonding and electrostatic interactions), and refolding into non-native, energy minima structures (Figure 1.24).

![Diagram](https://via.placeholder.com/150)

**Figure 1.24.** Stepwise evolution after ESI of the structure of a globular protein
Reproduced from Ref. [66]. Copyright © by the National Academy of Sciences

However, a number of experiments have shown that gas-phase structures have sufficient memory of solution-phase structures, including preservation of non-covalent interactions and overall dimensions, especially in the typical millisecond time scale of analysis used in QTOF instruments. This is the result, in part, of the fact that in nanoESI, some solvent and buffer salts are left on the complexes as they enter the MS. The rapid removal of solvent during the formation of the gas phase protein/protein complex is likely to have a primary effect on surface residues and hydrophobic interactions, while buried portion of the protein/protein complex and electrostatic interactions are more likely to be less effected by the fast desolvation. [70] Thus, native MS allows people to
study inter-subunit interaction within large protein complexes and native protein-ligand binding.

The charge state distribution in the spectrum produced by nano-ESI of a protein can be informative about its solution phase conformation.[71, 72] Compact structures have a narrow charge distribution of lower charge states, while extended structures have a wide charge state distribution with relatively higher charge states. Thus, charge state distributions can be used to reveal conformational changes [73] and estimate the solvent-exposed surface area in solution [74, 75].

1.4.2 Ion Mobility

Ion mobility spectrometry (IMS) is a gas phase separation technique based on the size, shape and charge of the gaseous analyte ions. The coupling of IMS and mass spectrometry (IM-MS) has become popular recently due to the commercialization of the instrumentation setup. Travelling-wave IM-MS by Waters was released as a commercial instrument in 2006. A commercial drift tube IM-MS by Agilent was on the market since 2014. Both the travelling-wave IMS (TWIMS) and drift tube IMS (DTIMS) are temporally dispersive IMS techniques. There are also spatially dispersive IMS, such as field-asymmetric IMS (FAIMS) and differential mobility analyzer (DMA).[76] The working principle of DTIMS, TWIMS and FAIMS are introduced briefly here.

1.4.2.1 Instrumentation
In drift tube ion mobility spectrometry (DTIMS), as shown in Figure 1.25a, ions introduced into the drift tube are propelled by a static electric field force and lag by interaction with the drift gas in the drift tube.[78] The typical drift gas is helium. The time for an ion to drift through the tube is determined by its collisional cross-section (CCS), a two-dimension projection of the three-dimensional shape in free rotation. An equation can directly relate the CCS ($\Omega$) and the drift time ($t_D$) [79]:

$$\Omega = \frac{(18\pi)^{1/2}}{16} \frac{ze}{(k_BT)^{1/2}} \left[ \frac{1}{m_I} + \frac{1}{m_N} \right]^{1/2} E \frac{760}{L} \frac{T}{273\cdot2} \frac{1}{N}$$  

Equation 1-19

where $k_b$ is the Boltzmann constant, $T$ is Temperature in Kelvin, $z$ is the ion charge, $e$ is the elementary charge in C, $m_I$ is the mass of the ion in kg, $m_N$ is the mass of the bath gas in kg, $E$ is the electric field strength (V/m), $L$ is the length of the drift region (m), $P$ is the pressure in torr and $N$ is the neutral gas number density (m$^{-3}$).
In travelling wave ion mobility spectrometry (TWIMS, Figure 1.25b), a dynamic electric force field is applied on stacked ring ion guides (SRIG).[80] The same radio-frequency (RF) voltage is applied on alternating ring electrodes and the adjacent ring electrodes have opposite sign RF voltage. The RF voltage is applied to confine the ions radially. Additionally, a transient direct current (DC) voltage is applied on the electrodes in succession as a wave to drive the ions. Ions with low mobility roll over the wave and arrive at the detector later than the ions with high mobility. Due to the dynamic electric filed applied in TWIMS, the direct relationship of the CCS and drift time no longer exists. Fortunately, the CCS of an analyte ion can be calibrated by measuring the drift time of calibrants with known CCS and similar chemical properties under the same conditions. Commercial traveling wave ion mobility on Waters SYNPAT G2/G2-S instruments is used to perform the research reported in this dissertation. A total of 200 time-of-flight scans corresponding to the full time for all injected ions to exit the IM cell make up a “mobiligram”, and for a typical acquisition of m/z up to 14,000, a mobiligram represents a drift time range of 0-32.27 ms.

Field-asymmetric ion mobility spectrometry (FAIMS) is composed of two electrodes that are perpendicular to the ion injection direction (Figure 1.25c).[81] The drift gas is flowing in the same direction with the ion injection to propel the ions. The dispersion voltage applied on one electrode has the waveform shown in the top of Figure 1.25c. The value of the positive voltage is greater than the absolute value of the voltage in the negative mode, but the absolute value of voltage multiplied by the time for the positive voltage and the negative voltage, respectively, are equivalent. The mobility of
ions at high electric field, $K_h$, is different from that at low electric field, $K$, which makes ions drift radially and eventually hit the electrodes. If the ions from two compounds have different ratios of $K_h$ to $K$, a compensation voltage can be applied on the electrode to specifically repel ions with a certain ion mobility. Thus, FAIMS is used as a filter on the basis of ion mobility, but CCS cannot be obtained from FAIMS.

1.4.2.2 Application

Because ion mobility is a gas phase separation technique, it can improve sensitivity and selectivity towards analytes of interest in a mixture. It has been shown by McLean and coworkers that different classes of biomolecules, such as lipids, peptides, carbohydrates and oligonucleotides from a complex biological sample occupy different conformational space as illustrate in Figure 1.26.[82] Cuyckens et al. found ion mobility spectrometry useful in extracting peptide ions from plasma matrix and reducing background signals, thus facilitating peptide metabolite identification.[83] The interrogation of a given species in a mixture can benefit from the further dimension of separation provided by ion mobility.
In our group, we perform SID/IM/MS experiments and the instrument schematic is shown in Figure 1.18. Protein complexes are composed of subcomplexes, which have some independent stability.\[33\] A variety of subcomplexes produced by SID provides inter-subunit contact information for a complex. Ion mobility after the SID cell assists us to confidently identify different product species with overlapping m/z. \[46\] For example, SID MS of a homo-11mer (Figure 1.27) does not clearly distinguish peaks from different species in m/z space alone, especially for those peaks in the high m/z range which cannot be resolved. However, with the aid of IM, oligomer (1mer to 10 mer) products can be separated and distinguished. Each of the oligomers appear to reside on an individual trend line in the SID/IM plot (mobiligram), because each type of oligomer has similar CCS for different charges states, their drift times are scaled with charge states.
Figure 1.27. SID (ΔV=100 V) spectrum of +15 TRAP 11mer with overlapping peaks in m/z space (up) and SID/IM plot with m/z overlapping peaks separated in the drift time space (down). Reproduced from Ref. [46]. Copyright © 2014 American Chemical Society

Observing ion drift time shift or converting drift time to CCS data allows numerous structural studies of biomolecules, such as conformation dynamics[84, 85], aggregation[86, 87], collision-induced unfolding[88] and protein complex architecture determination[89, 90].

In many cases, protein structures in the gas phase can be related to their structures in solution phase, as demonstrated by the fact that experimental CCSs from IM measurements are in good agreement with theoretical CCSs calculated from the
structures obtained from X-ray crystallography or NMR structures.[91] Traveling wave ion mobility is used in this dissertation. Due to the dynamic nature of the field of traveling wave IM, standards with published CCS are used as references to help measure the molecule with unknown CCS experimentally. The data processing steps are described in a published protocol[79] and the standards with published CCS have been summarized by the Bush group on the website: http://depts.washington.edu/bushlab/ccsdatabase/. [92, 93] For proteins with reported X-ray crystallography or NMR results, there are four common ways to calculate theoretical CCS from the atomic coordinates: projection approximation (PA)[94], exact hard sphere scattering (EHSS)[95], trajectory method (TM)[96] and projection superposition approximation (PSA)[97, 98, 99, 100]. The PA method is the least computationally expensive method. The atoms of a protein are treated as hard spheres with radii equal to hard sphere collision distances. The CCS of the PA method is determined by averaging the projection areas over all possible orientations. It takes neither the long-range interactions between the drift gas and the ion, nor the scattering in the drift gas into account. Thus, a scaled PA CCS \((1.14 \times \text{CCS}_{\text{PA}} \times \left(\frac{m_{\text{measured mass}}}{m_{\text{sequence mass}}}\right)^2\)) , which includes an empirically determined factor of 1.14 to estimate CCS, is used for comparing with experimentally measured CCS of proteins. [101] The more computationally expensive PSA methods can achieve very close CCS value to the scaled PA CCS, because the shape factor considered in PSA has an average number of 1.17. [97] In this dissertation, all the theoretical CCS for proteins with atomic coordinate files available are calculated through the scaled PA method.
In addition to using CCS comparative studies with structural information from X-ray crystallography or NMR, IM-MS can be a stand-alone technique to investigate unknown structures. For example, IM-MS was utilized to probe Amyloid-β (Aβ) protein oligomerization, which is an important process related to the etiology of Alzheimer’s disease.[102] As shown in Figure 1.28, different coarse-grained models for Aβ42 tetramer, hexamer and dodecamer are listed and compared with the measured CCSs from IM-MS experiments. The models that fit with the experimental CCS are chosen to represent general structures of specific states of oligomerization, and are further used for proposing an assembly pathway. Another example is that for full-length OmpA which is resistant to crystallization, Marcoux et al. proposed a low-resolution model for the full-length OmpA dimer by combining knowledge of partially known structures, site-directed mutagenesis, molecular modeling and ion mobility.[103] The models are screened by crosslinking data and experimental CCS data from IM-MS experiments.
Figure 1.28. The CCS of different models and experimental values of Aβ42 oligomers (left); A plausible mechanism for aggregation (right)
Reproduced from Ref. [102] with permission. Copyright © 2009, Rights Managed by Nature Publishing Group

### 1.4.3 Top-down Mass Spectrometry

In sequence analysis, also known as primary structure analysis, the protein of interest needs to be segmented into fragments to be analyzed. The segmentation process can be done in the solution phase, for example by proteolysis or in the gas phase, by tandem mass spectrometry. The “bottom-up” method refers to performing the digestion of protein in the solution phase and analyzing these peptides by mass spectrometry. The sequence information is then pieced together by the information from different peptides.

In contrast, when the segmentation process primarily occurs in the mass spectrometer, it is called a “down” type method. To be more specific, there are “top-
down” and “middle-down” methods. When the intact protein is subjected to ionization and subsequent gas-phase fragmentation, this method is called “top-down”. If the protein is cleaved to several large fragments in the solution phase and each of them is ionized and further fragmented to obtain structural information, it is called “middle-down” method.

The top-down method is widely used for determining the primary sequence of proteins because it can precisely differentiate protein isoforms, localize amino acid substitutions and identify post-translational modifications (PTMs).[104, 105] Besides the use for primary structure determination, top-down coupled with ECD can be performed on native protein complexes. Experiments on tetrameric yeast ADH, trimeric FMO antenna and tetrameric plant lectin conA, has revealed that fragmentation intensity correlates with the B-factor from X-ray crystallography, providing structural insights towards the flexible regions of protein complexes. [106] Further experiments on aldolase tetramer have shown that not only flexible regions of protein complexes are prone to ECD cleavage, but also the surface exposed regions.[107]

1.5 Bottom-up Mass Spectrometry based protein structural study

As introduced earlier in 1.4.3, bottom-up experiments involve digestion of proteins to smaller peptide fragments and peptide sequence analysis by liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). In order to reveal 3D structural information of the entire protein in bottom-up experiments, the information needs to be encoded into the peptides. The commonly used methods involve labeling of solvent-exposed amino acids of a protein or a protein complex in native
conditions or utilize reagents with two reactive groups to crosslink amino acids within a certain distance.

1.5.1 Covalent Labeling

Covalent labeling of proteins results from the reaction of chemical reagents with solvent-exposed functional group(s) of proteins or protein complexes. Identification of these modification sites is based on the mass increase as a result of the chemical reaction. The amino acid residues that are buried in the interior of the protein or involved in inter-subunit interaction of the protein complex are shielded from the labeling reagent, thus remain largely unmodified. This covalent labeling process is also called “footprinting” or “surface mapping”. The reaction can be highly specific or nonspecific towards side chains of different amino acids. Table 1.1 summarizes some of the reagents, their specificity towards amino acids and suitable conditions for reaction.
Table 1.1. Reagents used in covalent labeling of proteins

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Structure</th>
<th>Specificity</th>
<th>pH</th>
<th>Temp</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenylglyoxal And its derivatives</td>
<td><img src="image1" alt="Phenylglyoxal" /></td>
<td>R</td>
<td>7-10</td>
<td>RT-37</td>
<td>[108, 109, 110]</td>
</tr>
<tr>
<td>EDC+ Glycineamide</td>
<td><img src="image2" alt="EDC+ Glycineamide" /></td>
<td>D, E</td>
<td>4.5-6</td>
<td>RT</td>
<td>[111, 112]</td>
</tr>
<tr>
<td>Iodoacetamide</td>
<td><img src="image3" alt="Iodoacetamide" /></td>
<td>C</td>
<td>6.9-8.7</td>
<td>0-37 ,dark</td>
<td>[113, 114]</td>
</tr>
<tr>
<td>DEPC</td>
<td><img src="image4" alt="DEPC" /></td>
<td>H (K,Y,C,S,T)</td>
<td>5.5-7.5</td>
<td>RT-37</td>
<td>[115, 116]</td>
</tr>
<tr>
<td>Acetic anhydride</td>
<td><img src="image5" alt="Acetic anhydride" /></td>
<td>K, N-term</td>
<td>5-9</td>
<td>20-37</td>
<td>[117, 118]</td>
</tr>
<tr>
<td>N-hydroxysuccinimide derivatives</td>
<td><img src="image6" alt="N-hydroxysuccinimide" /></td>
<td>K, N-term</td>
<td>5-9</td>
<td>20-37</td>
<td>[119, 120]</td>
</tr>
<tr>
<td>2-hydroxy-5-nitrobenzyl bromide (Koshland’s reagent)</td>
<td><img src="image7" alt="2-hydroxy-5-nitrobenzyl bromide" /></td>
<td>W</td>
<td>4-6</td>
<td>RT</td>
<td>[121]</td>
</tr>
<tr>
<td>N-Acetylimidazole</td>
<td><img src="image8" alt="N-Acetylimidazole" /></td>
<td>Y (K,S)</td>
<td>6.5-9</td>
<td>RT-37</td>
<td>[122]</td>
</tr>
<tr>
<td>Hydroxyl radicals</td>
<td>·OH</td>
<td>14 out of 20</td>
<td>~7</td>
<td>RT</td>
<td>[123, 124]</td>
</tr>
</tbody>
</table>

EDC is 1-ethyl-3(3-dimethylaminopropyl)-carbobimide
RT is room temp
DEPC is diethylpyrocarbonate

The nonspecific covalent labeling, hydroxyl radical footprinting, takes advantage of highly reactive and short lived hydroxyl radicals. It can target side chains of 14 out of 20 natural amino acids. There are many ways to generate hydroxyl radicals, such as synchrotron radiolysis of water, Fenton reactions and electrical discharge. The method developed in the Gross group, Fast Photochemical Oxidation of Proteins (FPOP), generates hydroxyl radicals by laser photolysis. The reaction is accomplished in
approximately 1μs [123, 124], which can be used to study protein folding in hundreds of μs[125].

1.5.2 Non-covalent Labeling

Covalent labeling as described in the last section provides information about the surface accessibility on the amino acid residue level. Hydrogen/deuterium exchange (HDX) probes the protein amide backbone hydrogen exchange with deuterium upon exposure to deuterated water. The exchange rate is dependent on the protein conformation: amide hydrogens involved in secondary structure undergo slow exchange due to the hydrogen bonding, while amide hydrogens in disordered regions undergo fast exchange. Hydrogens in N-H, O-H or S-H side chain functional groups can undergo fast exchange with deuterium upon exposure to deuterated water. However, the fast back exchange from D to H during the digestion or separation steps in typical bottom-up experiments make the exchanges difficult to detect by mass spectrometry. Therefore, the measured mass shift does not contain these active hydrogens (hydrogens in N-H, O-H or S-H) on the amino acid side chains. [126] One exception is the slow exchanging of His imidazole C2-H.[127]

The most commonly used mass spectrometry-based method to probe the location and extent of HDX is the bottom-up approach and the workflow is shown by the solid arrows in Figure 1.29. First, the protein of interest is incubated in native-like buffer with D2O. Aliquots with different incubation times are taken out and the H/D exchange is quenched by adding acid to pH 2.5 and cooling to 0°C. Each aliquot is digested to peptides by passing through a column immobilized with pepsin, an enzyme that is active
at low pH. Then the peptides are rapidly separated in an analytical column kept at a low temperature and the peptides are subsequently analyzed by mass spectrometry. The low temperature and short time are necessary for preventing back exchange from D to H, because back exchange results in loss of coded structural information. Tandem MS with CID is useful for peptide sequence identification. Instrument automation, commercialization [128] and software development [129] has made the HDX-MS technique more and more popular.

Figure 1.29. Workflow of hydrogen/deuterium exchange monitored by mass spectrometry
The bottom-up approach is shown in the solid arrow, including H/D exchange, quenching, pepsin digestion, peptide chromatography, peptide tandem MS by ETD or ECD.
Reproduced from Ref. [130] with permission. Copyright © 2014, American Chemical Society
In order to obtain higher spatial resolution for localization of H and D, tandem mass spectrometry on HDX peptides by ETD or ECD is performed. Conventional CID on the HDX peptides suffers from H/D scrambling, which randomize the solution H/D exchange results.[131] One thing to pay attention to is that for the bottom-up based approach, each HDX peptide loses two amide hydrogen information due to the proteolysis process. One hydrogen loss is the lost upon enzymatic hydrolysis of an amide bond to produce amine, which undergoes complete back exchange, the other is the amide hydrogen of the second residue, which is exchanged back to H in most cases.[132] To circumvent this disadvantage, fragmentation of the HDX protein can be performed directly. But one needs to keep in mind that fragmentation efficiency decreases with an increase of protein size. The use of ETD and ECD in the HDX-MS top-down approach (shown in the dash arrow in Figure 1.29) has been demonstrated to provide an average resolution of 2 residues for carbonic anhydrase II, a nearly 30kDa protein.[133]

1.5.3 Protein Digestion

Protein digestion involves amide bond hydrolysis. Small peptides are easier to handle for mass spectrometric analysis, since they have higher ionization efficiency and the modified sites in labeling experiments are easier to identify. Table 1.2 summarizes some enzymes used for protein digestion and their digestion conditions.
Table 1.2. Enzymes used for protein digestion

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Cleavage sites</th>
<th>pH</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProTEV</td>
<td>ENLYFQ(G/S)</td>
<td>5.5–8.5</td>
<td>4–30 °C</td>
<td>1-6 hr</td>
</tr>
<tr>
<td>IdeS</td>
<td>F(ab’)_2</td>
<td>Fc</td>
<td>6-8</td>
<td>37 °C</td>
</tr>
<tr>
<td>Proteinase K</td>
<td>Aliphatic</td>
<td>Aromatic</td>
<td>4.3–12.0</td>
<td>37–56°C</td>
</tr>
<tr>
<td>Factor Xa</td>
<td>IEGR</td>
<td>X</td>
<td>7-8</td>
<td>23 °C</td>
</tr>
<tr>
<td>Chymotrypsin</td>
<td>Y, F, W</td>
<td>X (L, M, A, D, E</td>
<td>X)</td>
<td>7-8</td>
</tr>
<tr>
<td>Trypsin</td>
<td>K,R</td>
<td>X(not P)</td>
<td>7-9</td>
<td>37 °C</td>
</tr>
<tr>
<td>Lys-C</td>
<td>K</td>
<td>X</td>
<td>7-9</td>
<td>37 °C</td>
</tr>
<tr>
<td>Arg-C</td>
<td>R</td>
<td>X (K</td>
<td>X)</td>
<td>7.6–7.9</td>
</tr>
<tr>
<td>Asp-N</td>
<td>X</td>
<td>D, C</td>
<td>4-9</td>
<td>37 °C</td>
</tr>
<tr>
<td>Glu-C</td>
<td>E, D</td>
<td>X</td>
<td>4-9</td>
<td>37 °C</td>
</tr>
<tr>
<td>Pepsin</td>
<td>F, L, Y, W</td>
<td>X</td>
<td>1-3</td>
<td>37 °C</td>
</tr>
<tr>
<td>Thermolysin</td>
<td>X</td>
<td>L, I, F, V, A, M</td>
<td>5.0-8.5</td>
<td>70-95 °C</td>
</tr>
<tr>
<td>Elastase</td>
<td>A, V, S, G, L, I</td>
<td>X</td>
<td>9</td>
<td>37 °C</td>
</tr>
</tbody>
</table>

*X stands for any amino acid
*Cleavages in parentheses are less preferred.
*Most of the information is obtained from http://www.promega.com/

When performing the digestion, the general steps involve the target protein solubilization/denaturation, disulfide bond reduction and alkylation, and incubation of the sample with the specific digestion enzyme at the appropriate conditions. The first two steps make sure the target protein is unfolded and amino acids are exposed to the solvent for access of the enzyme. Any protease inhibitors that may impair the proteolytic ability of the enzyme needs to be removed as much as possible. Concentrations of denaturant, such as urea, guanidine·HCl, SDS, and acetonitrile, need to be lowered prior to adding the enzymes.

1.5.4 Structure modeling

The information obtained from bottom-up structural proteomics needs to be translated to the 3D structure of proteins/protein complexes. To our knowledge, there is
no software that can directly take the data from bottom-up structural proteomics as input and provide protein structure as output. In most of the studies, the data serves as structural constraints to filter the models generated by structure modeling.

Protein structure modeling is mainly divided into three categories based on the strategies used: homology modeling, threading modeling, and *ab initio* modeling. [134] Homology modeling utilizes templates with solved atomic coordinates that have high sequence identity with the target protein to build model candidates. Threading modeling also utilizes templates but the alignment relies on structural and biochemical similarities, thus allowing relaxed sequence similarity. *Ab initio* modeling does not depend on templates. It utilizes force field to fold protein to achieve satisfying energy terms. Due to the computational expenses, this method is currently limited to proteins with less than 200 amino acids. Some commonly used tools are listed: SWISS-MODEL [135, 136] ([swissmodel.expasy.org](http://swissmodel.expasy.org)), HHpred [137] ([toolkit.tuebingen.mpg.de/hhpred](http://toolkit.tuebingen.mpg.de/hhpred)), Rosetta [138] ([www.rosettacommons.org](http://www.rosettacommons.org)), I-TASSER [139] ([http://zhanglab.ccmb.med.umich.edu/I-TASSER/](http://zhanglab.ccmb.med.umich.edu/I-TASSER/)). The one used in prediction of complex structure from sequence information for the research described in this dissertation is Protinfo PPC ([http://ram.org/compbio/protinfo/ppc/](http://ram.org/compbio/protinfo/ppc/))

1.6 **Introduction Summary**

This introduction aims to provide basic terms and knowledge for readers to understand the research conducted for this dissertation. The methods described above are used throughout the work described in this dissertation.
Velos Pro ion trap mass spectrometer is used in bottom-up approaches applied in Chapter 2, 6 and 7. Orbitrap Elite mass spectrometer is used in top-down sequencing approaches mainly applied in Chapter 2. Quadrupole/ion mobility/TOF mass spectrometer is used throughout Chapter 2 to Chapter 8, for measuring intact mass of protein chains under denaturing conditions, transferring intact non-covalent protein complexes to the gas phase, performing SID dissociation studies and ion mobility measurement. Covalent labeling and structure modeling are applied in Chapter 6 and 7.
Chapter 2. Primary structure determination: Bottom-up with a customized database assisted Top-down approach for avian hemoglobin sequencing


2.1 Introduction

Hemoglobin (Hb), a tetrameric protein with four polypeptide chains, typically two α and two β, is the most abundant protein in erythrocytes. It plays an important role in oxygen transportation.[3] Modifications and mutations on the amino acids of the hemoglobin sequences may reveal health states of individuals. For example, glycated hemoglobin is commonly used for diagnosis of diabetes.[140, 141] Severe hemoglobinopathies including sickle cell disease may involve only one amino acid substitution or deletion on hemoglobin chains.[142] Mutations together with thalassemia can lead to a more severe anemia.[143] Also, the fully sequenced hemoglobin can be utilized in sequence evolution analysis and ancestral sequence reconstruction.[144] The
ability to identify small variations in hemoglobin sequences unambiguously and quickly is desirable.

Tick larvae often have their first blood meal on small mammals and birds and hence the possibility exists for the transmission of pathogenic microorganisms from those initial hosts to tick larvae. After feeding, the larvae leave their host and molt to nymphs. Later on, nymphs need to obtain a second blood meal from a larger host to molt to adult. As a consequence, disease acquired from the first host, if any exists, can be transmitted to a second host.[145] There are many bird species known to be host reservoirs for tick-borne diseases.[146, 147] The ability to identify major reservoirs in one geographical region is useful for disease prevention and intervention. It has been shown that host blood proteins still persist in ticks months after feeding and molting.[148] Because hemoglobin is the most abundant protein in erythrocytes and its sequence varies in different species,[149] Hb remnants from host in ticks can be identified to reveal host identities.[150] We have obtained blood samples from 33 bird species (142 individuals) that are common in the U.S. as hosts for ticks but that have unreported Hb sequences. Hb from most bird species contains one major component, HbA (α^A_2β_2) and one minor component, HbD (α^D_2β_2).[151, 152, 153, 154] Therefore, most bird species have three types of Hb chains: α^A, α^D and β with α^A and β chains being the most abundant Hb chains.

Bottom-up mass spectrometry, in which proteins are digested to form peptides that are fragmented in the mass spectrometer, is a very powerful tool in Hb identification. Gatlin and co-workers identified six variants of human hemoglobin by bottom-up
proteomics and SEQUEST-SNP algorithm, which dynamically generates all possible single-nucleotide polymorphisms (SNPs).[155] Laskay and co-workers utilized a bottom-up mass spectrometry approach to sequence hemoglobin from nine mammalian species.[150] Guo and co-workers combined bottom-up mass spectrometry and X-ray diffraction to achieve complete sequence coverage of hemoglobin from two feline species.[156] Onder et al. employed spectral matching in a proteome profiling technique to identify vertebrate host of ticks without a priori knowledge of protein sequence information.[157, 158] The method which Onder et al. adopted is similar with a spectral library searching approach for peptide identification in bottom-up mass spectrometry, which involves matching query spectra to library spectra.[159, 160, 161] The most common method for identifying digested peptide ions from bottom-up experiments is database searching. During this process, the mass and fragmentation pattern of the ions are matched with in silico generated peptide ions from known protein sequences in a database.[162] Obviously, one drawback of this database searching strategy is that the existing databases may not necessarily cover the protein/peptide under study. To overcome this disadvantage, several de novo sequencing algorithms have been developed.[163, 164] However, de novo sequencing results are lower throughput and the output can be hard to organize into interconnected peptides, because each tandem mass spectrum will return a peptide result and a typical run usually result in tens of thousands tandem mass spectra. In this study, a different bottom-up data processing method is needed.
Top-down mass spectrometry, which accurately measures the intact molecular weight of a protein and directly fragments the protein ions [104] without prior digestion to peptides is becoming increasingly popular due to advances in mass spectrometry instrumentation, such as higher resolution and accuracy [104, 165, 166], and improvements in top-down sequencing software[167, 168]. No digestion is performed in the top-down approach, thus the connectivity of the peptides is not lost. The use of multiple fragmentation methods, such as collision-induced dissociation (CID) and electron transfer dissociation (ETD), facilitates better fragmentation coverage.[169] The advantages of top-down MS include precise determination of protein isoforms and localization of amino acid substitutions and post-translational modifications (PTMs). Edwards and coworkers have successfully applied a top-down approach to screen neonatal dried blood spot samples for hemoglobin variants.[170, 171]

Here, we developed a top-down-assisted bottom-up mass spectrometry approach with a customized search database to successfully sequence the $\alpha^A$ and $\beta$ hemoglobin chains from the 33 bird species under investigation. The customized database was produced by aligning the known avian hemoglobin sequences from National Center for Biotechnology Information (NCBI), keeping sequences that are invariant in those constant and generating peptide sequences where variable amino acids were successively replaced with each canonical amino acid. The final resulting sequences produced by using this customized database to sequence the Hb of birds with unknown sequence and further confirmed with the top-down approach will be available to append to existing
hemoglobin databases so that LC-MS/MS of protein digests or ESI or MALDI MS of intact proteins can be searched against this database to reveal host identities.

2.2 Experimental

2.2.1 Full scan mass spectrometry analysis of bird blood

We have blood samples from 33 bird species. Within those, 5 species were represented by only one bird per species. The other 28 species were represented by multiple birds per species serving as biological replicates. To determine the difference of the intact mass of Hb that may exist between individuals within the same species, the blood samples were lysed, diluted and analyzed by a SYNAPT G2S Q-IM-TOF mass spectrometer (Waters Corporation, Manchester, UK). To prepare the spray solution, the blood aliquot was spun at 14,000 g for 10 min to separate plasma (supernatant) and packed red blood cells. Two microliters of red blood cells were incubated with 25 μL of water (Optima LC/MS water, Thermo Fisher Scientific, Waltham, MA) for 5 min for cell lysis. Cell lysate was centrifuged at 14,000 g for 5 min. The supernatant from the red blood cell lysate was diluted 200 fold with 0.1% formic acid (Thermo Fisher Scientific) in water in preparation for analysis. Multiply charged hemoglobin subunits were observed and deconvolved to zero charge mass with the SYNAPT’s embedded MaxEnt 1 algorithm. Some settings for MaxEnt 1 were: mass range 5000-20000, resolution 1 Da/channel, damage model uniform Gaussian width at half height 0.5 Da, minimum intensity ratios left 33% right 33%.
2.2.2 Bottom-up experiments

All the chemicals were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise specified. The bottom-up sequencing approach was performed after in-gel trypptic digestion of bird hemoglobin, with LC-MS/MS analysis on an LTQ mass spectrometer (Thermo Fisher Scientific) coupled to a capillary-flow LC (Surveyor, ThermoFinnigan, San Jose, CA). 15 µL of red blood cell lysate was mixed with 5 µL 4×Laemmli protein sample buffer (250 mM Tris-HCl, pH 6.8, 8% SDS, 30% glycerol, 0.02% bromophenol blue, 1% β-mercaptoethanol) followed by heating at 95°C for 5 min. The sample was loaded on a 4-20% precast protein gel (Mini-PROTEAN® TGX™, Bio-Rad, Hercules, CA) together with 10 µL Precision Plus Protein™ standard (Bio-Rad) in the adjacent lane. The gel was run at 90 V constantly for approximately 1.5 hrs until the dye front reached the reference line. The gel was rinsed three times in pure water (Millipore, Billerica, MA), and stained with Bio-safe Coomassie (Bio-Rad) for 1 hr, then background destained with pure water overnight on a shaker (Boekel, Feasterville, PA). The hemoglobin band (around 15kDa) was excised, diced into 1 mm³ cubes and transferred to a 1.5 mL Eppendorf tube. The gel pieces were washed with 200 µL pure water for 30 s and destained twice with 200 µL of methanol: 50 mM ammonium bicarbonate (1:1 v/v) for 1 min with vortex mixing. The gel pieces were then incubated for 5 min in 200 µL of acetonitrile: 50 mM ammonium bicarbonate (1:1 v/v) with vortex mixing. The gel pieces were further dehydrated in 200 µL acetonitrile then dried in SpeedVac (Savant, Irvine, CA) for 15 min. The dried gel pieces were reduced with 200 µL freshly prepared 25 mM dithiothreitol for 20 min at 56°C then alkylated with 200 µL
55 mM iodoacetamide at room temperature for 20 min in the dark. The gel pieces were washed twice with 500 μL of pure water by vortex mixing briefly. Dehydration and drying as described before were repeated. In-gel digestion was performed by incubating in 50 μL of 20 ng/μL Trypsin Gold (Promega, Madison, WI) in 0.01% ProteaseMAX (Promega) at 50°C for 1 hr. The digestion was stopped by adding in 0.5% (final v/v) trifluoroacetic acid to inactivate trypsin. The solution was transferred to a new 1.5 mL Eppendorf tube and centrifuged at 14,000 g for 10 min. Supernatant was subjected to LC-MS/MS analysis. The LC system was modified from a microbore HPLC (Surveyor) to achieve capillary flow. The column (8 cm x 100 μm ID) used was packed with 5 μm Xorbax C18 resin (Agilent, Santa Clara, CA). Peptides were separated with a gradient using mobile phase A (0.1% formic acid in water) and B (0.1% formic acid in acetonitrile). The initial flow was 5 min at 1.25uL/min at 5% B, followed by 100 min of 0.5uL/min with 5% B linearly ramping up to 45% B. Five min at 1.25uL/min at 95% B was used as a final wash step. An LTQ was used, with the full MS range set to 400 to 1600, and the three most intense ions automatically selected by Xcalibur software for fragmentation.

In addition to the above approach which was used for approximately 16 bird species, the bottom-up sequencing approach for 17 bird species was performed by in-solution tryptic digestion of bird blood, with LC-MS/MS analysis on a dual linear ion trap Velos Pro mass spectrometer (Thermo Fisher Scientific) coupled to a nanoACQUITY UPLC (Waters). 15 μL blood cell lysate obtained as described above was mixed with 28.5 μL 100 mM ammonium bicarbonate, 1.5 μL 1% ProteaseMAX
(Promega) and 5 μg Trypsin Gold (Promega). Digestion was performed at 37 °C for 3 hrs and stopped by adding in 0.5% (final v/v) trifluoroacetic acid. After centrifuging the digestion solution at 14,000 g for 10 min, supernatant was subjected to LC-MS/MS analysis. A Thermo EASY-Spray Column (Particle Size: 3 μm; 150 x 0.075 mm I.D.) was used. Flow rate was kept at 0.4 μL/min and the gradient started at 1% B, ramping to 35% B in 84 min. The full MS range was set to 400 to 2000, and the ten most intense ions were automatically selected using Xcalibur software for fragmentation.

2.2.3 Bottom-up data analysis

Because most of the bird species examined here have unreported hemoglobin sequences, we decided that a statistical analysis of known avian Hb sequences should first be performed to identify conserved regions in avian hemoglobin. The conserved regions could then be used in a customized database. We chose to use our in-house protein identification algorithm SQID[172] for peptide identification because it showed an improved number of identified peptides, due to incorporating pairwise fragmentation intensity information. The work flow of data analysis is shown in Figure 2.1.
A statistical analysis of avian Hb available in the NCBI bank (06/12/12) was performed. It indicates that arginine and lysine sites are mostly conserved, so that the Hb chains can be segmented into tryptic peptides ending with K or R. Peptides with enumerated mutations at variable sites were then generated to assemble hypothetical position-labeled peptide groups in the customized database for SQID searching. A general sequence of a Hb chain can be obtained by combining the SQID top hits from each hypothetical position-labeled peptide group in numerical order. The peptide group name (e.g. A1 - A9; B1 - B12) represents the position of the peptide in the Hb chain.
The statistical analysis of avian Hb in the National Center for Biotechnology Information (NCBI, 06/12/12) bank indicates that basic residue (arginine and lysine) sites are 98.3% and 99.8% conserved among 1045 avian Hb α chains and 760 avian Hb β chains. This gave us confidence in segmenting the Hb chain into tryptic peptides ending with K or R. Peptides with enumerated mutations at variable sites were then generated to assemble hypothetical position-labeled peptide groups. The customization step is elaborated in the Results section. A general sequence of a Hb chain can be obtained by combining the SQID top hits from each hypothetical position-labeled peptide group in numerical order. The peptide group name (e.g. A1 - A9; B1 - B12) represents the position of the peptide in the Hb chain.

2.2.4 Top-down experiments and data analysis

Top-down data acquisition was performed on an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific) with ESI infusion of bird blood lysate diluted in water/acetonitrile (1:1 v/v) containing 0.1% formic acid. Collision induced dissociation (CID) and electron transfer dissociation (ETD) were performed on the multiply charged (mostly, between +14 to +18) ions of the intact proteins. The isolation window was 2 Da. The CID energy was 20%-30% and ETD reaction time was 10-15 ms depending on precursor ion charge state and mass. Top-down data were charge-deconvolved and deisotoped by Thermo Xtract software. The Xtract settings were: generate MH+ Masses Mode, resolution 240000 @400, S/N Threshold 3, fit factor 44%, remainder 25%, AveragineLowSulfur model, max charge 30. After the fragments with S/N over 3 are deconvoluted and deisotoped, 10 ppm mass accuracy was used for manual matching
based on the results from bottom-up. To show the abundance of products produced in top-down sequencing, product maps were generated by an in-house built Matlab script with input of the final sequences and the charge-deconvoluted MS2 peak lists. The Matlab script is available in Online Resource 3.

Phylogenetic analysis of the Hb $\alpha^A$ chains was done via phylogeny.fr\cite{173, 174}. Two sequenced bird species in the NCBI database were subjected to analysis together with 35 sequenced $\alpha^A$ chains (two out of 33 species have two $\alpha^A$ isoforms) in this work.

2.3 Results and Discussions

2.3.1 Intact mass of hemoglobin chains

Blood samples from 28 out of the 33 bird species under study were collected from more than one individual bird. To check the heterogeneity of Hb within a species, the mass of Hb chains were measured on the SYNAPT G2S mass spectrometer. Figure 2.2 and Online Resource 1 “ESM_1. Intact average mass of hemoglobin chains” shows the intact average mass of $\alpha^A$, $\alpha^D$ and $\beta$ chains for the 28 bird species that contain more than one individual bird.
Figure 2.2. Intact mass of hemoglobin chains
a) $\beta$ b) $\alpha^D$ c) $\alpha^A$ from 28 bird species containing multiple bird individuals. Individual bird masses within a species are represented by “X” symbols with offset. Most bird species have consensus mass among different individuals. Several species, namely 7, 12, 13, 14, 15, 18, 19 contain outliers.

For most of the species, the same masses were present for the individual sampled birds in a given species, while some species, namely No.7 Catharus ustulatus, No.12 Icterus bullockii, No.13 Junco hyemalis, No.14 Melospiza melodia, No.15 Myiarchus cinerascens, No.18 Pheucticus melanocephalus, and No.19 Picoides nuttallii showed one or two outlier(s). We assume hemoglobin chains within a specific species have the same amino acid sequence if they have the same mass. The different masses observed in the outliers may due to neutral mutation [3]. The sequences reported for a given species in
this paper are from the individual bird samples that all provided the same masses. Outlier m/z values are included in Online Resource 1.

### 2.3.2 Customizing a database for searching

Bottom-up proteomics typically utilizes sequence database searching[162], a popular approach to identify sequence of the digested peptides, although spectral library searching[159, 160, 161] is also used. In sequence database searching, the intact mass and fragmentation pattern of the experimental peptide ions are matched to fragmentation spectra or m/z lists of peptide fragment ions produced for peptide sequences generated by *in silico* digestion of protein sequences in a database. This approach is a reasonably fast and high-throughput method to identify proteins. Because the Hb chains from the bird species in this study haven’t been sequenced before, a customized database was created in order to adopt the database searching approach.

The concept of customizing a sequence database has been applied to many studies to increase peptide detection sensitivity. [175, 176] Khatun et al. generated a database by 3-frame translation of a transcript database and 6-frame translation of genomic sequences.[177] Wang and coworkers a created protein sequence database from abundant RNA-seq data of a specific sample and further incorporated high-quality single nucleotide variations into the customized protein database.[178, 179] Park et al. considered single nucleotide variants, insertions, deletions, and stop-codon mutations derived from Exome-seq and RNA-seq to build a database for a specific sample.[180]
In our study, the steps of generating a customized database are shown in Figure 2.1. First, a statistical analysis of amino acids in already sequenced avian Hb chains was performed. From the NCBI database, 760 entries of avian Hb β chains were available; these were aligned and all were 146 amino acids in length. Arginine and lysine sites were 99.8% conserved among the 760 entries, so the β chain can be segmented into 12 peptides ending with K or R. For example, for the first peptide VHWTAEEK, with variants noted at H2 and T4, and no R, K, or I included (because I and L were treated as one amino acid, labeled L), there are 17 (20 minus 3) potential amino acid substitutions on the 2\textsuperscript{nd} and the 4\textsuperscript{th} positions. The 2\textsuperscript{nd} site mutation and the 4\textsuperscript{th} site mutation were considered independently. Therefore, a total of 289 (17 multiplied by 17) VXWXAEKEK peptide segments are generated as hypothetical position-labeled peptide group “B1”, in which X represents amino acids except for K, R and I. In a similar way, 12 (B1 to B12) position-labeled peptide groups were constructed in the database. Each peptide group B1-B12 exists as 2 to 83,521 sequences in the database. Hb α chains are more varied due to the fact that there are two copies of the α chain, one major component α\textsuperscript{A} and one minor component α\textsuperscript{D}. Both are 141 amino acids long. In order to compromise between the computational expense and the numerous variations of the peptides, some variable sites were limited to a certain number of substitutes that have over 1% occurrence in the sequenced Hb chains in the NCBI database. This compromise was justified by the results: More than 60% sequence coverage in bottom-up alone for α\textsuperscript{A} from any species was achieved. Top-down experiments performed later on enabled us to fully differentiate α\textsuperscript{A} from α\textsuperscript{D}. Nine (A1 to A9) position-labeled peptide groups for Hb α chain were generated for the database along with the twelve for the Hb β chain.
The constitution of each position-labeled peptide group is listed in Table 2.1. In this table, variable site X represents all amino acids except for K, R and I. Variable site $X^k$ represents all amino acids except for K and I (due to occurrence of K at these specific sites in some of the birds of known sequence). Sites with limited substitution are in parentheses with all the possible amino acids listed. The limited substitution is determined by the fact that in 99% of known sequences of avian hemoglobin, those residues are occupied at those sites. Peptides with enumerated mutations at variable sites (highlighted in yellow) were generated to assemble hypothetical position-labeled peptide groups (A1-A9, B1-B12). The number of sequences in each peptide group is listed in the third column of Table 1. An experimental peptide mass and MS/MS data were searched by the SQID proteomics search engine against each peptide in the hypothetical database. Because the peptide name represents the position of the peptide in the chain, a general sequence of a Hb chain can be obtained by combining the most confident peptide from each hypothetical position-labeled peptide group in the numerical order (A1-A9; B1-B12).
Table 2.1. Peptide components in the customized database (Variable sites are highlighted)

<table>
<thead>
<tr>
<th>Position</th>
<th>Sequence Variation</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>VLSXXDKE</td>
<td>2,312</td>
</tr>
<tr>
<td>A2</td>
<td>IGGHAVEEGGANSLEQT</td>
<td>31,104</td>
</tr>
<tr>
<td>A3</td>
<td>MLFTYPXTK</td>
<td>1734</td>
</tr>
<tr>
<td>A4</td>
<td>TYPFHDLSXKGASQDIKVRF</td>
<td>9792</td>
</tr>
<tr>
<td>A5</td>
<td>VAVGALVENVANINHDNADGISQTLSK</td>
<td>15,552</td>
</tr>
<tr>
<td>A6</td>
<td>LSDLHAKV</td>
<td>4</td>
</tr>
<tr>
<td>A7</td>
<td>LRVDPNF</td>
<td>2</td>
</tr>
<tr>
<td>A8</td>
<td>FLSGCFLIVVALAVATRHAALPSTVEVHSASAFDK</td>
<td>2,654,208</td>
</tr>
<tr>
<td>A9</td>
<td>FMCASVLTAK</td>
<td>192</td>
</tr>
<tr>
<td>B1</td>
<td>VXWXAEK</td>
<td>289</td>
</tr>
<tr>
<td>B2</td>
<td>QLIWXWKK</td>
<td>83,521</td>
</tr>
<tr>
<td>B3</td>
<td>VNVXCGXEALR</td>
<td>4,913</td>
</tr>
<tr>
<td>B4</td>
<td>LXXYPWTQR</td>
<td>289</td>
</tr>
<tr>
<td>B5</td>
<td>FSFTSGNLSSAAGPTAVIXGPK</td>
<td>544</td>
</tr>
<tr>
<td>B6</td>
<td>VLXSFGXAVK</td>
<td>289</td>
</tr>
<tr>
<td>B7</td>
<td>NXDIK</td>
<td>289</td>
</tr>
<tr>
<td>B8</td>
<td>XTXLSELHCDK</td>
<td>4,913</td>
</tr>
<tr>
<td>B9</td>
<td>LHVDPENFK</td>
<td>2</td>
</tr>
<tr>
<td>B10</td>
<td>LLGDVLXXVVLAXHFXXK</td>
<td>4,624</td>
</tr>
<tr>
<td>B11</td>
<td>DFSPXXQAAWQK</td>
<td>578</td>
</tr>
<tr>
<td>B12</td>
<td>XVAHALAR</td>
<td>17</td>
</tr>
</tbody>
</table>

X represents amino acids except for K, R, I; X^k represents amino acids except for R, I

2.3.3 Bottom-up approach

The software SQID [172] first indexed the database into peptides ending with K or R. The allowed peptides had mass ranges from 700 to 4500 Da and up to one missed
cleavage. Variable modifications included oxidation of methionine and carboxymethylation on cysteine. Thermo Discoverer was used to convert the LC-MS/MS raw data to .dta files. Each .dta file contained the precursor mass (calculated based on measured m/z), charge, fragments’ m/z values, and fragments’ intensities. In order to determine the peptide sequence of each experimental tandem mass spectrum or .dta file, the precursor mass was compared with those of the *in silico* digested peptides and sequence candidates were chosen within 1.5 Da mass error of the precursor mass. A series of b and y ions and their neutral loss ions were then calculated for each of the candidates. To evaluate the accuracy of the match between the candidates and the experimental tandem mass spectrum, a scoring function considering the number of matched ions (within 0.8 Da mass error), the number of consecutively matched ions, and a bonus for strong pairwise fragmentation intensity was utilized. The sequence candidate with the highest score was assigned to the experimental tandem spectrum. The output of SQID for each sample was a spreadsheet that included information such as spectral name, corresponding sequence, the group number that contains the peptide, etc. For each peptide group, which is a collection of all the possible sequences at one tryptic cleavage region of the hemoglobin chains, the sequence with the highest score is more likely to be the true sequence of the region. To be more specific, if more than one sequence was identified for a position-labeled peptide group, a minimum SQID score that corresponds to an FDR of 5% and a maximum delta SQID score of 0.3 were set to choose the sequence candidates. The FDR was established by searching against reverse decoys appended to our customized database, and the SQID cutoff was determined to be 16-21 for different species. The delta SQID score is represented by the normalized difference
between the SQID score of the top ranked sequence for the position-labeled peptide group and current sequence as shown in the following equation:

\[
\text{Delta SQID score} = \frac{(\text{SQID score}_{\text{rank1}} - \text{SQID score})}{\text{SQID score}_{\text{rank1}}}
\]

Sequences with small, but not 0, delta SQID scores are potential sequences for a specific Hb region, thus should be considered when constructing the general sequences of Hb from the bottom-up approach.

The Online Resource 2 “ESM_2. Peptide MS/MS Spectra Identified by SQID with Customized Database” contains all the successfully identified hemoglobin peptides for each of the bird species by this method. The confidence of assigning these sequences is strengthened by top-down analysis. Information on each page of the PDF file includes species number, species’ common name, peptide sequence, hypothetical protein name (which represents the location of the peptide in the hemoglobin chain), raw data file identifier, SQID score, the peptide spectrum and a three row sequence cell. In each spectrum, b and y ion peaks are labeled and plotted in blue and red respectively. The green peaks represent neutral losses from the b/y ions. The three row sequence cell facilitates counting the b and y ions. An example is shown below.
Figure 2.3. Tandem MS of the top peptide hit for A2 region identified by SQID with Customized Database for the species *Zonotrichia atricapilla*.

### 2.3.4 Top-down completion of partial bottom-up sequences

Top-down data analysis was used to fill in missing peptides and to confirm the sequences identified in the bottom-up approach. Both collision-induced dissociation
(CID) and electron transfer dissociation (ETD) methods were performed to improve the sequence coverage. All the precursor masses and fragment masses are within 10 ppm error. Figure 2.4 shows the product ions of intact hemoglobin αA and β chains from Zonotrichia atricapilla species generated by CID (b, y ions; blue mark) and ETD (c, z ions; red arrow).

![MS/MS product ion map of Hb αA chain (a) and β chain (b) from Zonotrichia atricapilla. (c) Legend for fragmentation symbols.](image)

The abundance of fragment ion information generated from the top-down approach provided sequence information that connected peptides identified in the bottom-up approach. For example, Figure 2.5 compares the result obtained from the bottom-up approach only to the result from the top-down-assisted bottom-up approach. The result from the top-down-assisted bottom-up approach is shown on the rows with amino acid number labels (1, 31, 61, 91, and 121) and the amino acids codes are in black or gray. A1 to A9 and B1 to B12 correspond to regions of hypothetical position-labeled peptide groups. A1, B7, B9 and B11 don’t have confident hits returned from bottom-up database.
searching because all hits had SQID scores lower than 16 (FDR 5%). B2 has a confident hit “QLIPSGWGK” returned from the bottom-up approach (SQID score 29), but the sequence doesn’t explain the fragment ions in the top-down experiments. A2 through A9, B1, B3, B4, B5, B6, B8, B10, and B12 are partially or completely sequenced correctly from bottom-up approach. For the Hb αA and β chains of Zonotrichia atricapilla, 67% and 66% sequence coverages were achieved without the assistance of the top-down method, respectively. However, the top-down-assisted bottom-up method achieved 100% confidence sequence coverage with ambiguity only at I and L. Moreover, the top-down method can provide post-translational modification information. For example, top-down data showed that K16 of the Corvus brachyrhynchos Hb αA chain was methylated. This modification was unexpected, and thus was not included in the database searching of the bottom-up data. The bottom-up approach is also necessary, because there are some regions in the middle of the hemoglobin sequences that don’t have much fragmentation in the top-down approach, for example as shown in Error! Reference source not found., not much fragmentation has been observed in the β67-76 (VLTSFGEAVK) region, but this region was successfully detected in the bottom-up approach as a top hit for the hypothetical position-labeled peptide group B6. All the product ion maps of αA and β from the 33 bird species are in Online Resource 3. The resulting sequences and deconvoluted top-down mass spectra were further subjected to ProSight Lite [181] matching and the P-score and the protein characterization score (PCS) were obtained. For all the Hb sequences, the median P-score is 2.4E-71 and the median PCS is 822.68.
Figure 2.5. Peptides aligned from a SQID search with a customized database in the bottom-up method for *Zonotrichia atricapilla* α\(^A\) chain (a) and β chain (b).

The amino acid letters showing the sequencing results from top-down-assisted bottom-up method. The boxes with the blue A1-A9 or B1-B12 show the Hb sequence segmentation in the bottom-up method. Peptides in black are correctly identified by bottom-up. The gray regions are only identified in the top-down approach.

Phylogenetic analysis (Figure 2.6) on the 35 sequenced Hb \(\alpha^A\) chains (two species has two \(\alpha^A\) isoforms) together with 2 known sequences was performed by uploading the sequences to the server phylogeny.fr [173, 174] in “One click” phylogeny analysis mode and the result is consistent with the conventional classification of the avian organisms. In Figure 2.6, the bird name is in the format of “Order_Family_Genus_Species”, which is the conventional classification. The phylogenetic tree is indicative of how similar the sequences are: dis-similar sequences diverge early on the tree while similar sequences stay in the same branch. For example, the three species *Colaptes auratus*, *Melanerpes formicivorus*, and *Picoides nuttallii* are known to come from the same order.
PICIFORMES. They are distant from all the other species that belong to the PASSERIFORMES order or the GALLIFORMES order. This is confirmed by the phylogenetic tree produced from the protein sequences of the 37 Hb $\alpha^A$ chains. Hb sequences from species *Vireo gilvus*, *Vireo cassinii*, and *Vireo huttoni* are from the same genus Vireo, and they grouped together. Similarly, Hb sequences from species *Catharus guttatus* and *Catharus ustulatus* are from the same genus Catharus, and they grouped into the same branch. *Turdus migratorius* from the same family as the two Catharus (TURDIDAE) but not the same genus diverged a little earlier on the tree. *Thryomanes bewickii* and *Troglodytes aedon* from TROGLODYTIDAE family, *Haemorhous mexicanus* and *Haemorhous purpureus* from the FRINGILLIDAE family and Haemorhous genus, *Euphagus cyanocephalus* and *Icterus bullockii* from the ICTERIDAE family, *Junco hyemalis*, *Melospiza melodia*, *Melozone crissalis*, *Passerella iliaca*, *Pipilo maculatus*, *Zonotrichia atricapilla* and *Zonotrichia leucophrys* from the EMBERIZIDAE family were all grouped separately based on their family. Aberrant from the classical classification are the species from the CARDINALIDAE family and the CORVIDAE family. They were grouped closely but not as close as exactly the same branch. The already sequenced *Callus gallus*, which belongs to GALLIFORMES order, was branched together with *Callipepla Californica* from our study, adding confidence to our sequencing results.
Figure 2.6. Phylogenetic tree of birds based on sequences of their Hb α chains and generated by using phylogeny.fr.

The tree matches very well with the conventional classification (Reflected by the name “Order_Family_Genus_Species”). The two known sequences from NCBI, and which were inserted as controls, are highlighted.

2.4 Conclusions

We have presented a top-down-assisted bottom-up sequencing approach (Figure 2.7). Building a customized database based on homologous and variable regions of already sequenced avian hemoglobin for the bottom-up approach expedited the alignment
of the peptide data, thus providing a reliable model for top-down data analysis. The top-down approach played a role of quality control in confirming the peptides alignment in the customized-database bottom-up approach but also provided sequence in regions where bottom-up failed. The combined approach has been successfully applied to sequence α^A and β hemoglobin chains from 33 avian species. The sequences obtained in this study will be appended to current hemoglobin databases for tick/host studies. In the future, the top-down-assisted bottom-up sequencing approach can be applied to facilitate fast and complete sequencing of other homologous proteins.

Figure 2.7. Top-down-assisted bottom-up approach
Chapter 3. Effect of Conformational Flexibility on Gas-Phase Unfolding of Noncovalent Protein Homodimers Probed by CID and SID

3.1 Introduction

Mass spectrometry has become a powerful tool in the structural analysis of noncovalent protein complexes.[182, 183, 184] Tandem mass spectrometry aids in the determination of the topology by releasing substructures following dissociation. However, collision induced dissociation (CID), the most commonly used dissociation method results in ejection of highly charged monomers for many protein complexes, thus providing limited direct interaction information.[38] Disproportional charge partitioning has been observed in CID of protein dimers with reduced disulfide bonds in the monomers, while not in their counterparts with intact intra-protein disulfide bonds.[36] Thus, one of the reasons why asymmetric charge partitioning occurs has been ascribed to the conformational flexibility of the monomers. Previous work from the Wysocki group has observed more symmetrical charge partitioning between subunits of protein complexes by surface induced dissociation (SID).[43, 47, 185, 186, 187, 188] Protein homo-dimers are the simplest protein complexes, with a single interface. Activation and dissociation of homo-dimers yields monomers, which always follows mass symmetric, but not necessarily charge symmetric dissociation. Thus, charge partitioning can be studied with fewer complicating factors. Herein, we first report the dissociation of non-disulfide containing enolase dimer by SID, because the dimer has been reported to have
asymmetric charge-to-mass partitioning in CID.[38] Then we report the effect of structural rigidity on SID behavior, by dissociating two intra-disulfide containing dimers, \( \alpha \)-lactalbumin and \( \beta \)-lactoglobulin and their disulfide-reduced dimers.

### 3.2 Experimental

Enolase (E6126, Sigma-Aldrich) from baker’s yeast was dissolved in 100 mM ammonium acetate (AmAc) to reach 100 \( \mu \)M monomer concentration. Supercharging was performed by adding 2.5\% 3-Nitrobenzyl alcohol (m-NBA), and charge reduction was performed by adding 30 mM triethylamine acetate (TEAA). Two disulfide containing proteins including \( \alpha \)-lactalbumin (type I, L5385, Sigma-Aldrich) and \( \beta \)-lactoglobulin (L3908, Sigma-Aldrich) were also examined. The protein samples were dissolved in 100 mM AmAc solution at a 150 \( \mu \)M monomer concentration (in equilibrium with its oligomers). In order to reduce the disulfide bonds, the protein solutions were mixed with 0.5 M dithiothreitol (DTT) solution at a volume ratio of 10:1 (protein solution: DTT) and reacted at 70\(^\circ\)C for 5 minutes. For the controls, DTT solution was replaced with 100mM AmAc solution. No heating was performed. Buffer exchange to 100 mM AmAc with 5 mM DTT and 100 mM AmAc was performed afterwards for reduced sample and control sample, respectively. Minor charge reduction is performed by adding 20\% volume of 100mM ethylenediamine diacetate (EDDA) to the protein complex solution.

Samples were sprayed via a nano-electrospray source into a Waters SYNAPT G2S mass spectrometer (Manchester, UK) with a customized SID device incorporated before the ion mobility cell.[50]
3.3 Results and Discussions

3.3.1 Drastically different dissociation behavior of enolase dimer by CID and SID

Studies initially focused on enolase, a dimer known to undergo asymmetric charge distribution by CID [38] and therefore an interesting precursor for SID. Studies were performed under “normal”, reduced, and supercharging conditions to enable assessment of the effect of charge on both the SID and CID behavior. The mass spectrum for enolase, prepared in 100 mM AmAc, shows the dimer presents over a range of charge states from +18 to +21. (Figure 3.1b) Monomer is also observed in the spectrum. Super charging was performed by adding 2.5% m-NBA to the solution, shifting the dimer charge state distribution to +24 to +22. (Figure 3.1a) Furthermore, charge reduction was also performed by adding 30 mM TEAA to the enolase in 100 mM AmAc. (Figure 3.1c) The observed charge state distribution for the charge reduced enolase dimer is +15 to +12.
To assess the effect of charge on the enolase dimer dissociation, representative charge states, namely +23 (Figure 3.2a, d), +21 (Figure 3.2b, e), and +13 (Figure 3.2c, f) from each solution condition were subjected to either CID or SID for dissociation study. It is clear in Figure 3.2a that the +23 dimer of enolase dissociates predominantly to +18 monomer and its complementary partner by CID. The transmission discrimination of the high m/z partner has been attributed to the increased difficulty in focusing large, low charged ions and the decreased response on the detector.[36, 189] Similar asymmetric dissociation was observed by following CID of +21 enolase dimer, as shown in Figure 3.2b. The 21 charges have been partitioned predominantly to +16 and its complementary partner, which is asymmetric considering the two monomers have exactly same sequence.
and mass. The further charge reduced enolase dimer resists dissociation even with the highest accessible CID voltage in this instrument (Figure 3.2c).

SID experiments on the aforementioned charge states, however, show drastically different fragmentation patterns (Figure 3.2d, e and f). Interestingly, the most abundant products share roughly equal charges. In summary, the dominant dissociation pathways are: +23 dimer dissociates to +12/+11 monomer pair, +21 dimer dissociates to +11/+10 monomer pair and +13 dimer dissociates to +7/+6 monomer pair. Other charge state complementary pairs are also present, but they have decreasing intensity as the charge state distribution becomes more asymmetric.

Figure 3.2. Dissociation of enolase dimer at different charge states by CID and SID. a) CID 3450 eV of +23 enolase dimer, b) CID 3570 eV of +21 enolase dimer, c) CID 2600 eV of +13 enolase dimer, d) SID 1840 eV of +23 enolase dimer, e) SID 2100 eV of +21 enolase dimer, f) SID 2340 eV of +13 enolase dimer. The precursors are shaded in pink and major monomer products are labeled with charge states. Contaminants peaks are labeled with “*”
To further compare the dissociation products of CID and SID, the conformations, and hence collisional cross-section (CCS) of the monomer products generated from the +21 enolase dimer by both methods were then considered (Figure 3.3). The +9 to +12 symmetric charge-to-mass partitioned monomer products were not observed in CID, thus the CCS could not be determined for these charge states following CID. Monomer products with the same charge state resulting from CID (black) and SID (red) are quite similar. Monomers with low charge states (+6 to +10) have CCS in good agreement with the CCS calculated from the monomer crystal structure (PDB ID 1EBH). Monomers with higher charge states (+11 and above) show higher CCS than the CCS calculated from folded structure, which indicates that these products undergo unfolding. One thing to note is that the CCS profile intensity of each product is normalized to its highest intensity to ease observation of the less abundant species clearly.

Figure 3.3. CCS of monomer products from +21 enolase dimer precursor by either CID or SID. SID and CID data are offset slightly for easier comparison. CCS profile intensity for each charge is normalized to the highest intensity.
Previously, we have ascribed the observation of more symmetric charge-to-mass partitioning of cytochrome C dimer following SID to subunits unfolding to a lesser extent in the single step energy deposition provided by SID.[43] Nowadays, we have ion mobility which can allow us to further explore the hypothesis that SID products are compact/folded, not unfolded. The CCS profiles of the remaining enolase dimer precursor upon CID and SID were also examined (Figure 3.4). Unfolding of the precursor can be clearly seen in both CID and SID. At the lowest SID energy (Figure 3.4b), enolase dimer has already unfolded (folded dimer has CCS around 48.2 nm$^2$), the decreased intensity was due to charge reduction to +20, +19 and +18 dimer. Monomer production starts from 1890eV in CID and 840eV SID, which correspond to +21 enolase dimer CCS unfolding to around 85nm$^2$ and 70nm$^2$, in CID and SID, respectively (Figure 3.4). Therefore, compared to CID, the enolase dimer unfolds to a lesser extent to generate monomers in SID. The less extensive dimer unfolding explains why there are more symmetric charge-to-mass dissociation products in SID. As shown in Figure 3.3, the monomer that has more than half of the charges on the precursor (>+10) expands more as the charge state increases. Less unfolding of the precursor correlates with the less unfolded products, which then present with lower charges state. So the products in SID are centered around precursor in m/z space. The fact that we observed unfolding of dimer precursor from the lowest energy in SID can be attributed to the strong interaction between the enolase dimer. The interface ratio which is obtained by dividing the interfacial area by the dimer complex solvent accessible surface area (SASA) is listed in Table 3.1. Compared with α-lactalbumin and β-lactoglobulin dimer we studied, the
enolase dimer has a significantly higher interface ratio (0.6 versus 0.2~0.3), which reflects stronger interactions between the monomers.

Figure 3.4. CCS profiles of remaining +21 enolase dimer upon CID (a) and SID (b). The yellow straight lines show the onset energy for production of monomers.
Table 3.1. Dimer complexes studied

<table>
<thead>
<tr>
<th>Dimer Complexes</th>
<th>Ligand</th>
<th>MW (kDa)</th>
<th>PDB ID</th>
<th>SASA (Å²)</th>
<th>#SB</th>
<th>#H B</th>
<th>Inter-facial area (Å²)</th>
<th>Intra-disulfide / 1mer</th>
<th>Interface ratio</th>
<th>CID</th>
<th>SID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enolase</td>
<td>Mg</td>
<td>93</td>
<td>1EBH</td>
<td>29540</td>
<td>4</td>
<td>16</td>
<td>1782</td>
<td>0</td>
<td>0.06</td>
<td>A</td>
<td>S</td>
</tr>
<tr>
<td>α-lactalbumin</td>
<td>Ca</td>
<td>28</td>
<td>1HFY</td>
<td>12465</td>
<td>0</td>
<td>3</td>
<td>275</td>
<td>4</td>
<td>0.02</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Reduced α-lactalbumin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-lactoglobulin</td>
<td>36</td>
<td>2Q2M</td>
<td>16010</td>
<td>2</td>
<td>6</td>
<td>468</td>
<td>2</td>
<td>0</td>
<td>0.03</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Reduced β-lactoglobulin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. #SB is the number of salt bridges between the two monomers in the complex  
b. #HB is the number of hydrogen bonds between the two monomers in the complex  
c. Interface ratio is obtained by dividing the interface area by SASA of the dimer complex  
d. CID and SID dissociation behavior; “A” means asymmetric charge-to-mass dissociation and “S” means symmetric charge-to-mass dissociation. In the case of dimer complex, symmetric dissociation produces monomers with about half of the charges of the dimer precursor.

3.3.2 Dissociation of disulfide intact and disulfide reduced α-lactalbumin dimer

Next the effect of structural rigidity on SID behavior was considered, by dissociating two intra-disulfide containing dimers, α-lactalbumin, β-lactoglobulin and their disulfide-reduced dimers. CID, previously reported by Williams and coworkers [36], was repeated here for direct comparison with SID. There are 123 amino acids and 4 intramolecular disulfide bonds in each protein monomer of bovine α-lactalbumin. Dimer charge states from +12 to +10 were observed when directly sprayed from AmAc solution. Monomers were also observed. The +11 charged homodimer (D11) was chosen from the full scan mass spectrum for further study. As shown in Figure 3.5a and c, both lower energy (330 eV) and higher energy (660 eV) CID of α-lactalbumin D11 results in symmetric charge partitioned monomers, i.e. the major monomer products are the +6 and +5 charge states. While the +7 and +4 charged monomer products exist, they are present
in a much lower abundance. In contrast to the disulfide-intact dimer, charge partitioning in the CID of α-lactalbumin D11 without disulfide bonds is asymmetric, with +7 and +4 charged monomers present as the major dissociation products as shown in Figure 3.5b (330 eV) and d (660 eV).

As illustrated in Figure 3.6a and c, results from both lower energy SID (440 eV) and higher energy SID (990 eV) of disulfide-intact α-lactalbumin D11 shows a symmetric charge distribution, but with a lower average charge state (+5 dominant) when
compared with the average charge state of monomer products in CID (Figure 3.5a and c). Interestingly, when the disulfide-reduced α-lactalbumin D11 is subjected to SID, a transition from asymmetric dissociation to symmetric dissociation is observed with an increase of SID energy applied. The mass spectrum obtained at low SID energy (440 eV) is shown in Figure 3.6b, the major dissociation products are +7 monomer and its complementary partner +4 monomer, which is a trait of asymmetric dissociation. In contrast, the spectrum at high SID energy (990 eV), shown in Figure 3.6d, shows that the major products correspond to +6 monomer and +5 monomer, representing symmetric dissociation.

Figure 3.6 SID spectra of +11 α-lactalbumin dimer
Native dimer (a,c) and disulfide-reduced dimer (b,d) with different collision energy applied: 440eV for a and b, 990eV for c and d. Symmetric dissociation was observed for native dimer. Asymmetric dissociation was observed for disulfide-reduced dimer at low energy SID. As more SID energy input, dissociation become more symmetric.

In order to better reflect the symmetry change of charge-to-mass partitioning as the dissociation energy changes, the percentage of remaining dimer precursor and dissociation product pairs (+8 and +3 monomer, +7 and +4 monomer as asymmetric dissociation pairs; +6 and +5 monomer as symmetric dissociation pair) as a function of energy is shown in Figure 3.7. For +11 disulfide-intact α-lactalbumin, in either CID (Figure 3.7a) or SID (Figure 3.7c), the symmetric dissociation products (M6/M5, red down triangle) are more abundant than any other products. While for +11 disulfide-reduced α-lactalbumin dimer in CID, the asymmetric products (M7/M4, green up triangle) are the most abundant products along the energy scale. In SID of +11 disulfide-reduced α-lactalbumin dimer, the symmetric pair (M6/M5) and asymmetric pair (M7/M4) are of approximately equal abundance at low energy (< 600 eV). As the SID energy increase over 600eV, however, the symmetric product pair (M6/M5) becomes the most abundant products. The increased flexibility of α-lactalbumin obtained by reducing intra-disulfide bonds completely changes the dissociation symmetry of the dimer in CID. The increased flexibility also influences the SID behavior at low energy, as more symmetric dissociation has been observed in the less flexible dimer. But the difference becomes smaller as the SID energy increases. It is highly possible that direct dissociation overcomes unfolding at high SID energy. What’s more interesting is that the remaining flexible dimer is more abundant than the less flexible dimer upon SID. It leads us to
consider rigidity is a factor that can influence the precursor survival yield in SID: complexes with more rigid subunits are easier to be ruptured in SID.

Figure 3.7 Energy resolved CID curves (a,b) and SID curves (c,d) for +11 charged native α-lactalbumin dimer (a,c) and disulfide-reduced α-lactalbumin dimer (b,d).
Symmetric product pair (total of +6 monomer and +5 monomer) is represented by red pointing down triangles. Asymmetric product pairs, namely total of +7 and +4 monomer or total of +8 and +3 monomer are represented by green triangles and pink dots respectively.

To better preserve the protein complexes structure in their native-like confromations, charge reduction was performed. The +9 charged homodimer (D9) was chosen from spraying solution with charge reducing additives for further study. Similar to D11, CID shows symmetric dissociation of the disulfide intact D9 and
asymmetric dissociation of the disulfide reduced D9. SID at high energy shows symmetric dissociation for both disulfide intact and reduced D9. (Figure 3.8)

Figure 3.8. Tandem mass spectra of charge-reduced α-lactalbumin +9 native dimer (a,c) and disulfide-reduced dimer (b,d) with different dissociation methods CID 900 eV for a and b, SID 1260 eV for c and d. By CID, symmetric dissociation was observed for native dimer while asymmetric dissociation was observed for the disulfide-reduced dimer. By SID, both are symmetric dissociation.

An observation of the change in collision cross section (CCS) of surviving D9 with increasing CID energy (Figure 3.9) reveals that unfolding occurs as CID energy increases. Furthermore, the disulfide-reduced D9 has a more flexible structure than the disulfide-intact D9 (Figure 3.9a) with CCS revealing a more extended structure for the reduced species (27.5 nm$^2$ versus 26.6 nm$^2$ for disulfide-intact D9).
Figure 3.9. Collision induced unfolding of disulfide-intract (a) and disulfide-reduced (b) +9 α-lactalbumin dimer. The CCS profiles of the dimer at distinct dissociation energies are shown in different colors with y offset for ease of viewing. The corresponding energies are labeled on the lines.

The protein α-lactalbumin naturally binds one Ca$^{2+}$ per monomer. [191] Examination of the $m/z$ of the +5 and +4 monomer released from +9 native α-lactalbumin dimer shows that the unbound monomer peak is almost at the same abundance as the Ca$^{2+}$ bound peak in the CID tandem spectra (Figure 3.10a and c). In contrast, by SID, monomers mainly retain Ca$^{2+}$ (Figure 3.10b and d). This further highlights that CID is an unfolding process, and this unfolding leads to the loss of Ca$^{2+}$. In contrast, SID barely unfolds the products, and the major product is the Ca$^{2+}$-bound species. The CID
unfolding and ligand loss may not necessarily induce a CCS change, or the CCS change is not evident under the instrument resolution.

Figure 3.10. α-lactalbumin monomer calcium binding information. +5 monomer dissociated from the +9 disulfide-intact α-lactalbumin dimer by (a) CID 900 eV and (b) SID 1260 eV; +4 monomer dissociated from the +9 disulfide-intact α-lactalbumin dimer by (c) CID 900 eV and (d) SID 1260 eV. The blue line and purple line indicate Ca\(^{2+}\) loss and Ca\(^{2+}\) bound species, respectively.

Considering the CCS of the monomer products M3, M4, M5 and M6 from D9, a general trend of increased CCS is observed as the charge increases on the monomer (Figure 3.11d to a). Collision energies of 900 eV (CID) and 1260 eV (SID) were chosen because dimers were close to fully dissociation at those energies. The CCS of +6 charged disulfide-reduced monomer is slightly larger than its disulfide-intact counterparts (Figure 3.11f and e, 18.4 nm\(^2\) versus 17.6 nm\(^2\)), showing a more extended structure without the constraint of disulfide bonds. The CCS of disulfide-intact α-lactalbumin M6 has a second
more extended conformation in CID (black solid trace in Figure 3.11e) compared to SID (red solid trace). The CCS of disulfide reduced α-lactalbumin M6 has two distributions in both CID and SID (dash traces in Figure 3.11e). The extended conformation is more abundant in CID while the two conformations are of similar abundance in SID. Thus, less monomer products unfolding by SID was observed for higher charged α-lactalbumin monomer products, especially for disulfide-intact α-lactalbumin.

Figure 3.11. CCS profiles of the monomers of different charge states released from disulfide-intact or disulfide-reduced α-lactalbumin dimer by either CID 900 eV or SID 1260 eV a) +6, b) +5, c) +4 and d) +3. e) and f) are zoomed-in spectra of a) with disulfide-intact and disulfide-reduced monomer plotted separately in two panels.
3.3.3 Dissociation of disulfide intact and disulfide reduced β-lactoglobulin dimer

To further confirm that the symmetric charge partitioning of SID is independent of conformational flexibility, further, similar experiments were performed on β-lactoglobulin homodimers. There are 162 amino acids and two disulfide bonds in each of the β-lactoglobulin monomers. In contrast to α-lactalbumin, which is composed of α helices, the major elements of secondary structure are β sheets for β-lactoglobulin. When directly sprayed from AmAc solution, the charge states of β-lactoglobulin dimer range from +13 to +11. In order to increase the abundance of +11 dimer and perform tandem MS study, EDDA was added in to slightly shift the charge distribution. Disulfide-intact D11 shows symmetric dissociation upon either CID or SID (Figure 3.12a and c). Disulfide-reduced D11, however, shows asymmetric charge-to-mass partitioning upon CID with M7 and M4 as dominant dissociation product pair (Figure 3.12b). In contrast, SID shows symmetric dissociation with M6 and M5 as dominant dissociation product pair (Figure 3.12d). The dissociation behaviors of +11 β-lactoglobulin dimer are similar with “normal” (+11) and reduced (+9) charged α-lactalbumin dimer system. Only minor differences in CCS were observed for monomer products with the same charge state generated by CID and SID. Furthermore, as expected monomer products with higher charge states have larger CCS then monomer products with lower charge state.(Figure 3.13)
Figure 3.12 Tandem mass spectra of β-lactoglobulin +11 native dimer (a,c) and disulfide-reduced dimer (b,d) with different dissociation methods. CID 990eV for a and b, SID 440eV for c and d. By CID, symmetric dissociation was observed for native dimer while asymmetric dissociation was observed for the disulfide-reduced dimer. By SID, both are symmetric dissociation.
Conclusions

In this study, CID and SID show dramatically different dissociation patterns for enolase dimer, with asymmetric charge partitioning between the two monomers in CID, but symmetric in SID. It is evident that the dimer unfolding profile in SID is smaller than in CID at the onset energy of monomer production. The fact that dissociating pathway overcomes the unfolding pathway by the high energy-sudden activation provided by SID makes the symmetric pathway dominant in SID.
For intra disulfide-intact α-lactalbumin and β-lactoglobulin homodimer, both CID and SID show symmetric charge partitioning between the two monomers. Previous reports have illustrated that intra disulfide-reduced α-lactalbumin homodimer undergoes asymmetric charge partitioning upon CID.[36, 37] and our results agree. We also observed that intra disulfide-reduced β-lactoglobulin homodimer also undergoes asymmetric charge partitioning upon CID. Thus, the ability to unfold by having higher conformational flexibility plays an important role in the asymmetric charge partitioning pathway in CID. However, we show here that a symmetric charge-to-mass partitioning in the homodimer systems of intra disulfide-reduced α-lactalbumin and β-lactoglobulin is observed following SID at higher energy up till the highest energy available in our instrument.

Examination of the CCS of the monomers produced from dimers reveals that charge plays the dominant role in monomer products’ compaction and unfolding. Less unfolding of monomer products by SID was observed for higher charged disulfide-intact α-lactalbumin monomer products. Higher conformational flexibility, which causes the asymmetric charge partitioning in CID, does not cause asymmetric charge partitioning in high energy SID products of the dimer studied here. Experiments performed on protein homo-dimers illustrate that SID is suitable for protein complex interface analysis, because SID minimizes conformational disruptions (unfolding) of subunits in the dissociation process.
Chapter 4. Top-down mass spectrometric study of monoclonal IgG by CID and SID

4.1 Introduction

Immunoglobin (Ig), or antibody, is utilized to identify and neutralize foreign objects by specific antigen-antibody binding.[192] Human immunoglobin is classified to five types: IgA, IgD, IgE, IgG and IgM, within which, IgD, IgE and IgG are Y-shaped proteins, IgA is a covalent dimer of Y-shaped proteins and IgM is a covalent pentamer of Y-shaped proteins. As of January 2015, 48 therapeutic monoclonal antibodies (mAbs) have been approved, or are in review, in the United States and the European Union (http://www.antibodysociety.org/news/approved_mabs.php). All of them are based on IgG. IgG can be further classified to IgG1, IgG2, IgG3 and IgG4, based on the number and location of disulfide bonds and the sequence of the heavy chains. Due to the short half-life of IgG3, it is rarely used as therapeutic mAbs.[193] The generic models for commonly used therapeutic monoclonal antibodies, IgG1, IgG2 and IgG4, are shown in Figure 4.1. IgG is composed of two identical heavy chains and two identical light chains. There are covalent disulfide bonds between the heavy and light chains, and also the heavy and heavy chains. The four chains are attached by both the covalent disulfide bonds and non-covalent interactions. The heavy chain is composed of one variable domain (VH) and three constant domains (CH1, CH2, CH3). There is a hinge region between the CH1 and CH2, where the inter heavy chain disulfide bonds are located.
Glycosylation is found in CH2. [194] The light chain is composed of one variable domain (VL) and one constant domain (CL). There is disulfide bond bridging the CL and CH1 to link the light chain to the heavy chain. The VH and VL function together to bind antigen. The complementarity determining regions (CDRs) within the VH and VL have highly variable amino acid sequences, thus providing specific binding towards diverse antigens. Fragment antigen-binding (Fab) refers to the light chain (VL and CL) and the upstream of the heavy chain (VH and CH1). Fragment crystallizable (Fc) refers to the downstream of the heavy chain (CH2 and CH3). Fc is responsible for serum half-life and effector functions, such as antibody-dependent cellular cytotoxicity, antibody-dependent cell phagocytosis and complement-dependent cytotoxicity.[195]

Figure 4.1. Structures of commonly used therapeutic monoclonal antibodies: IgG1, IgG2 and IgG4
Reproduced from Ref [194] with permission. Copyright © 2008 Wiley Periodicals, Inc.

For the studies reported here, a monoclonal antibody Refimab, was provided by the National Institute of Standards and Technology (NIST). The sequence and disulfide bond information for heavy chain and light chain is provided as in Figure 4.2 and Figure 4.3, respectively. The goal of the studies is to explore the instrument capability, especially surface induced dissociation (SID) of characterizing monoclonal antibody.
Figure 4.2. Sequence information for the heavy chain of Refimab
Variable Fab—Constant Fab—Hinge—Constant Fe—secretory tail

* cysteine involved in intrachain disulfide bond in constant region
* cysteine involved in interchain heavy-light disulfide bond
(asterisks are found above noted residue)

> sp|NLGT|NL_ NIST RefimAb light chain

<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>*</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIQMTQSPST LSASVGDRTV ITCSASSRVG YMHWYQQKFG KAPKLIIYDT SKLASGVP5R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>70</th>
<th>80</th>
<th>*</th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSGSGSTGTEF TLTSSLQFD DFATYYCFQG SGYPFTFGGG TKVEIKRTVA APSVFIEPPS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>130</th>
<th>*</th>
<th>140</th>
<th>150</th>
<th>160</th>
<th>170</th>
<th>180</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEQLKSSTAS VVCLNNFYP REAVQWKKVD NALQSGNSQE SVTEQSDKDS TYSLSSTTL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>190</th>
<th>*</th>
<th>200</th>
<th>210</th>
<th>*</th>
</tr>
</thead>
<tbody>
<tr>
<td>SKADYKEKHKV YACEVTHQGL SSPVTKSENK GEC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.3. Sequence information for the light chain of Refimab

4.2 Experimental

10 mg/mL Refimab in 12.5 mM L-Histidine HCl, 12.5 mM L-histidine (pH 6.0) was provided by NIST. 10 μL Refimab solution was buffer exchanged into 100 mM ammonium acetate (AmAc, Sigma-Adrich, St. Louis, MO) via Micro Bio-Spin™ P-6 Gel Columns (Bio-Rad, Hercules, CA) and diluted by adding 45 μL 100 mM AmAc. The final concentration of the Refimab was approximately 10 μM. Gentle disulfide bond reduction was performed by mixing buffer exchanged RefimAb and 50 mM dithiothreitol (DTT) at 10:1 (v:v) ratio, followed by heating on a 56°C, 300 rpm thermo mixer for 20 mins. Deglycosylation was performed by incubating buffer exchanged RefimAb: 10 U/uL PNGase F (Promega, Madison, WI) (6:1, v:v ratio) at room temperature for 18 hrs. Part of the reaction solution was buffer exchanged to 100 mM AmAc for spraying to
study proteins. Study of the cleaved glycan was done by spraying the reaction solution directly.

Samples were sprayed via a nano-electrospray source into a Waters SYNAPT G2S mass spectrometer (Manchester, UK) with a customized SID device incorporated before the ion mobility cell.[50]

4.3 Results and Discussion

4.3.1 Release of glycans

Peptide-N4-(N-acetyl-beta-glucosaminyl) asparagine amidase (PNGase F) is one of the most popular enzyme for releasing N-glycans. It cleaves the glycan as glycosylamine (structure 2 in Figure 4.4), which is readily converted to regular glycan under acidic conditions (structure 3 in Figure 4.4), but relatively stable under basic conditions.[196]

![Structure 1](image1.png) ![Structure 2](image2.png) ![Structure 3](image3.png)

Figure 4.4. Release of N-glycan from protein by PNGase F
Reproduced from Reference [196] with permission. Copyright © 2009, Oxford University Press
Direct nano-ESI spectrum of the Refimab/ PNGase F reaction solution is shown in Figure 4.5. The most abundant species are cleaved glycosylamine/glycan. The monoisotopic mass of the protonated peak in Figure 4.5b and c were determined to be 1461.54 Da and 1623.60 Da, respectively. Together with the reported glycan information of antibody [194, 197], the major glycans in Refimab were identified as G0F and G1F. The G0F in the glycosylamine form has structure as shown in Figure 4.6, and the G1F has one extra galactose.

Figure 4.5. Mass spectrum full scan of Refimab reacted with PNGase F.
a) Full range \( m/z \) 400-10000, the most abundant peaks respond to b) G0F and c) G1F glycosylamine. Zoomed-in spectra are shown in b) and c), sodiated peaks are labeled.
Chemical formula: C$_{56}$H$_{95}$N$_{5}$O$_{39}$; monoisotopic mass 1461.560430. Yellow square, red triangle and green circle represent N-acetylgalactosamine (GalNac), fucose and mannose, respectively. Further addition of a galactose residue to the left of GalNac yields G1F, which corresponds with a mass increase of 162 Da.

The reaction solution of Refimab and PNGase F was then buffer exchanged using a spin column with a 6 kDa molecular weight cut-off, to remove small molecule such as the glycan that been cleaved off from the protein following treatment with PNGase F. The buffer exchanged reaction solution was then reanalyzed by native MS, the spectrum is shown in Figure 4.7a. To reduce some of the non-covalent adducts such as buffer molecules, glycan, water, a high cone voltage 200 V was used. In addition to the peaks corresponding to the antibody (m/z 5000-8000), the PNGase F peaks can be seen clearly (m/z 3000-4000). For the purpose of comparison, mass spectrum of Refimab that did not go through deglycosylation is show in Figure 4.7b. Refimab has measured mass of
148,379±3 Da and deglycosylated Refimab has measured mass of 145,214±9 Da. The mass of the deglycosylated Refimab is around 3165 Da smaller than the intact Refimab, which is consistent with a combination of G0F and G1F glycan, which have a combined molecular weight of 3084 Da.

Figure 4.7.a) Mass spectrum full scan of Refimab reacted with PNGase F and small molecules removed by buffer exchange. b) Mass spectrum full scan of Refimab buffer exchanged to 100 mM AmAc. Cone voltage 200V was used to clean up adducts on the protein molecules.

4.3.2 Top-down +29 Refimab by CID and SID mainly releases light chain and b ions

Top-down experiments were performed on +29 Refimab utilizing either SID or CID. In CID, the products with the lowest energy onset (~4350 eV) correspond to the light chain, of which the mass matches with the loss of thiol group on a cysteine (-SH2, 34Da). With the increase of CID energy, large b ions from both the heavy chain and the
light chain are observed (Figure 4.8b). Fragment ions correspond to b114 to b117 ions from the light chain and b109 to b116 ions from the heavy chain. Interestingly, the cleavage sites to generate these ions are around linker region of the variable domains and the constant domains. These observations are consistent with a previous study on one IgG1 molecule and six IgG2 molecules performed by Zhang and Shah.[198] As shown in Figure 4.8a, high abundance charge reduced precursors were observed in SID. Besides those charge reduced precursors, there are covalent fragments in the smaller m/z range. These SID covalent fragments resemble the CID covalent fragments.

![Figure 4.8. a) SID 140 V and b) CID 180 V tandem mass spectra performed on +29 Refimab](image)

The main products are labeled. L means the light chain with loss of sulfur on cysteine that forms disulfide bond with the heavy chain. Large b ions from light chain (Lb) or heavy chain (Hb) were observed. b ions are labelled following the classical nomenclature [31].
4.3.3 Gently reduced Refimab has light chain and complementary partner non-covalent interacting together

Gentle disulfide bond reduction was performed on Refimab by using less reducing reagent DTT with the aim to preserve intra-chain disulfide bonds. The peaks corresponding to intact Refimab still existed as the most abundant species in the gently reduced Refimab solution. Top-down experiments were performed by CID on the same charge states (+29) of the reduced Refimab and the spectra were compared with the intact Refimab (Figure 4.9). Considering the much lower relative abundance of the +29 reduced Refimab compared with +29 intact Refimab at the same acceleration voltage, it is obvious that the reduced Refimab is easier to dissociate, consistent with some reduction of the inter-disulfide bonds. The major dissociation products of the reduced Refimab (Figure 4.9a) are multiply charged, intact, light chain with a mass of 23123±1 Da, which indicates the disulfide linkage of the light chain to the heavy chain is reduced to free thiol, since the theoretical mass the intact light chain is 23124 to 23129 depending on whether the intra-disulfide bonds are reduced or not. The major products resulting from dissociation of the +29 intact Refimab (Figure 4.9b) are multiply charged light chain with mass of 23089±1 Da, which is consistent with the intact light chain with the loss of SH2 (34Da), a result of how CID cleaves disulfide bonds.[199]
The major dissociation products in a) is multiply charged intact light chain (L*) The major products in b) is multiply charged light chain (L) with the loss of SH2 (34Da)

To further prove that the light chain to heavy chain disulfide bond breaks upon partial reduction, and intact complex observed in the mass spectrum is due to the light chain non-covalently interacting with the rest of the antibody, the CID energy was increased. Increasing the CID energy results in covalent fragmentation producing y ion fragments from the light chain of the reduced Refimab, however, these are absent in high energy CID fragmentation of the intact Refimab. Figure 4.10 shows the zoomed-in spectra (m/z=2070-2170) comparing high energy CID (ΔV= 190 V) performed on +29 reduced Refimab and intact Refimab. Production of y96 and y97 from light chain is obvious in the spectrum corresponding to the reduced Refimab (Figure 4.10a). The absence of light chain y ions following fragmentation of the intact Refimab is due to the disulfide linkage of the light chain C-terminus to the heavy chain.
4.3.4 Gently reduced Refimab releases light chain and heavy chain together in SID

Top-down fragmentation using SID was then performed on the +29 disulfide-reduced Refimab and +29 intact Refimab for comparison. Figure 4.11 shows the SID/IM 2D plots at the same energy (ΔV= 140 V) for the two systems. The high abundance of charge reduced precursor in the intact Refimab sample, in comparison to the reduced Refimab sample is consistent with the reduced Refimab being more fragile, and hence more easily dissociated by SID, due to chains being held together sole by non-covalent interactions as opposed to disulfide bonds present in the intact species. As shown in the mass spectrum Figure 4.8a and the mobiligram Figure 4.11b, the major products from dissociating the +29 intact Refimab is the light chain and b ions from the light chain, as
well as b ions from the heavy chain. Whereas, the major products following dissociating the +29 disulfide-reduced Refimab are the individual light and heavy chains as well as a dimer of the light and heavy chain with no obvious signs of covalent fragmentation (Figure 4.11a) The extracted mass spectra of the different product species are shown in Figure 4.12. Splitting of the peaks in the heavy chain (Figure 4.12b) and the dimer of the heavy chain and light chain (Figure 4.12c) was observed, which correspond to a mass different around 160 Da. The mass difference matches with the residue mass of galactose (162 Da), suggesting that one heavy chain of Refimab has G0F glycan modification and the other heavy chain has G1F glycan modification.

In comparison to CID of the reduced Refimab, which has the preferential pathway of ejecting the light chain (Figure 4.13), SID of the reduced Refimab has been shown to contain a variety of fragments, including light chain, heavy chain and half antibody (Figure 4.11a). It illustrates that SID is a symmetric dissociation method in regards to both mass and charge. Thus, SID can provide more substructural information.
Figure 4.11. Ion mobility - mass spectrum of all the products released from isolated +29 reduced Refimab a) and intact Refimab b) by SID (ΔV = 140 V)
Figure 4.12. Extracted mass spectra from SID/IM 2D plot of +29 reduced Refimab a) light chain and heavy chain, b) heavy chain, c) light chain, each spectra is normalized to its highest abundance peak and the most abundant peak in each spectrum is labeled with the charge state.

Figure 4.13. Mobiligram of +29 disulfide-reduced Refimab CID at an acceleration voltage of 150 V.
Same data with mass spectrum in Figure 4.9a, but plot in 2D mobiligram.
4.4 Conclusions

This project aims to explore the instrumental capability of characterizing IgG through the study of a standard monoclonal antibody (Refimab from NIST). We were able to deglycosylate the mAb in native-like conditions by PNGase F enzyme and identify the released glycan (or glycosylamine) as G0F and G1F. Top-down MS by either CID or SID was utilized to dissociate the intact mAb with the aim of generating sequence informative ions. The major products are light chain with the loss of disulfide bond between the light chain and the heavy chain by both dissociation methods. Increasing the acceleration voltage in CID or SID leads to the generation of some large b ions corresponding to the variable region of the light chain and the heavy chain, as reported by others using CID. Furthermore, the gentle reduction of disulfide bonds by dithiothreitol was performed on mAb. The reduced mAb has light chain and complementary partner non-covalently interacting together as shown by increased dissociation and release of intact light chain in Top-down MS experiments by CID, in comparison to the non-reduced sample. What is more interesting is that Top-down MS experiments performed on the disulfide-reduced mAb by SID can release light chain, heavy chain and dimer of the light chain and the heavy chain (half mAb). Future works will focus on determining how the heavy chain and half mAb are generated in SID but not in CID Top-down MS. One possible explanation is that the SID dissociates the inter-disulfide bonds between the heavy chains, while CID is not energetic enough. Another possible explanation is that some of the covalent disulfide bonds between the heavy chains have been reduced in the gentle reduction. However in CID, the preferential pathway is the ejection of the light
chain, which makes the heavy chain undiscernible. Since SID is a more symmetric pathways in regards to both mass and charges, thus heavy chain can be observed, as well as the heavy and light chains, being held together by noncovalent interactions.
Chapter 5. **SID behavior matches with *in silico* disassembly for several homo-hexamer protein complexes with known structures**

5.1 Introduction

The characterization of the assembly of protein subunits into functional complexes is a key focus of interest in structural biology. [200] Native mass spectrometry (native MS) has proven powerful in structural biology, being successfully applied to a number of intricate protein complex studies. [183, 184, 201, 202, 203, 204] In particular, solution phase destabilization experiments prior to native MS have been designed to study the building blocks of protein complexes. With careful adjustment of organic solvent percentage or ionic strength, sub-complexes can be generated from the complexes in solution and a contact map can be built based on the full MS scan of the solution.[205, 206] Systematic studies [207, 208] have shown that it is possible to predict the solution disruption pathways for protein complexes based on the relative interface strengths: smaller interfaces tend to be broken first yielding subcomplexes with larger interfaces being preserved in solution disruption.

In order to probe the constituents and architecture of the complexes, especially in a mixture, gas phase disassembly methods are often necessary, enabling mass selected products to be studied individually.[20] The most widely used gas-phase dissociation method is collision induced dissociation (CID), which involves an analyte going through tens of thousands collisions with inert gas. [33] A complication with CID is that the
products do not always reflect the real contacts in the structure of the complexes present in solution. To be more specific, the “typical CID process” dominates with one subunit unfolding and taking approximately half of the charges from the precursor complex and leaving the other half of charges behind on its complementary partner.[36, 209, 210, 211] This asymmetric dissociation pathway produced by CID, therefore, typically does not reflect the solution phase complex and the level of information that can be obtained in this way is limited. An alternative gas-phase disassembly method, surface induced dissociation (SID), which involves an analyte colliding with a surface, has been shown in several publications to reveal substructure information, rather than ejection of a monomer, the “typical CID process”. [43, 44, 45, 46, 47, 48, 186, 187, 188]

Homo-hexamers that have D₃ symmetry may be a dimer of trimers or a trimer of dimers. These different arrangements are thus expected to generate different subcomplexes that reflect the structure of the hexamer upon gas-phase activation and dissociation, in particular using SID, as CID wouldn’t be expected to reproduce different products. Here, we consider four homo-hexamer protein complexes with known crystal structures or NMR results. Their dissociation pathways are predicted in silico based on the interfacial areas which presumably track with interfacial strengths. Their gas-phase dissociation behaviors are experimentally studied by both CID and SID to see whether they are consistent with the predicted in silico dissociation pathways. Thus, the aim of this study is to discern the level of information on substructure that can obtained from homohexamers using gas-phase methods.
5.2 Experimental

Glutamate dehydrogenase from bovine liver (Bovine GDH), L-Glutamic Dehydrogenase (NADP) from Proteus sp. (Proteus GDH) and human insulin were purchased from Sigma-Aldrich (St. Louis, MO). Putative *Escherichia Coli* GTP cyclohydrolase 1 type 2 (GCH) was a host protein obtained from collaborator Dr. Bradley Tebo at the Oregon Health & Science University. Proteus GDH was buffer exchanged to 100 mM ammonium acetate (AmAc, Sigma-Aldrich, St. Louis, MO) via Micro Bio-Spin™ P-6 Gel Columns (Bio-Rad, Hercules, CA) and the concentration of the complex was adjusted to 10 μM. Human insulin hexamer was prepared by dissolving to 80 uM (as monomer) in 300 mM AmAc together with 5 mM phenol and 2 uM ZnCl₂. GCH, as an *E. Coli* host protein, was buffer exchanged to 100 mM AmAc together with the recombinant protein of interest in Chapter 8.

The interface area, number of salt bridges and hydrogen bonds broken *in silico* are calculated based on the subunit-subunit interface area, number of salt bridges and hydrogen bonds, which are obtained from PDBePISA[212] by entering the PDB ID. Additional constraints were applied to the PISA analysis, namely, the bonding distance was required to be within 3.5 Å, an amino acid residue could only be counted once in salt bridges, and duplicates of salt bridges in hydrogen bonds were removed.
5.3 Results and Discussions

A D₃ hexameric protein complex can be represented by a ball-and-stick model with the balls representing monomers and sticks representing interfaces (Figure 5.1). Due to the symmetry of the structure, there are generally three different kinds of interfaces (a, b, and c as shown in Figure 5.1). The direct dissociation of this D₃ hexamer can be predicted through breaking different combinations of interfaces yielding different products. [207, 208] The production of monomer and complementary pentamer involves one a, two b and one c interface being broken (M pathway in Figure 5.1). The production of dimer and complementary tetramer can originate from two pathways: the first one involves breaking four b and two c interfaces (D_1 pathway in Figure 5.1). The other one involves breaking two of each interface a, b, and c (D_2 pathway in Figure 5.1). Trimers can also be produced by two pathways: the first one involves breaking three a and three c interfaces (T_1 pathway in Figure 5.1). The second trimer pathway involves breaking one a, four b and one c interface (T_2 pathway in Figure 5.1).

The assessment of strength of interface for a known structure is normally done by interface area, number of salt bridges and number of hydrogen bonds. Thus, those values can be added up to assess the easiness for different disassembly pathways (Table 5.1).
Figure 5.1. Ball-and-stick representation for a D3 hexamer and its *in silico* disassembly pathways

Balls represent monomers and sticks represent interfaces. There are three kinds of interfaces in D3 hexamer denoted by a, b, and c. The dissociation to monomer and pentamer goes through “M” pathway. There are two pathways lead to dimer and complementary tetramer. Dissociation to two trimers can also go through two pathways

Table 5.1. Interface area, number of salt bridges and hydrogen bonds of different interfaces and those associated with different pathways

<table>
<thead>
<tr>
<th>PDB</th>
<th>Interface Area</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>M</th>
<th>D_1</th>
<th>D_2</th>
<th>T_1</th>
<th>T_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>3MVO Bovine GDH</td>
<td>991</td>
<td>1691</td>
<td>298</td>
<td></td>
<td>4671</td>
<td>7360</td>
<td>5960</td>
<td>3867</td>
<td>8053</td>
</tr>
<tr>
<td>SB</td>
<td>2</td>
<td>0.3</td>
<td>0</td>
<td></td>
<td>2.6</td>
<td>1.2</td>
<td>4.6</td>
<td>6</td>
<td>3.2</td>
</tr>
<tr>
<td>HB</td>
<td>9</td>
<td>10</td>
<td>0</td>
<td></td>
<td>29</td>
<td>40</td>
<td>38</td>
<td>27</td>
<td>49</td>
</tr>
<tr>
<td>2YFQ Bacterial GDH</td>
<td>1101</td>
<td>906</td>
<td>0</td>
<td></td>
<td>2913</td>
<td>3624</td>
<td>4014</td>
<td>3303</td>
<td>4725</td>
</tr>
<tr>
<td>SB</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td></td>
<td>10</td>
<td>12</td>
<td>14</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>HB</td>
<td>16</td>
<td>5</td>
<td>0</td>
<td></td>
<td>26</td>
<td>20</td>
<td>42</td>
<td>48</td>
<td>36</td>
</tr>
<tr>
<td>1MNO GCH</td>
<td>1077</td>
<td>280</td>
<td>949</td>
<td></td>
<td>2586</td>
<td>3018</td>
<td>4612</td>
<td>6078</td>
<td>3146</td>
</tr>
<tr>
<td>SB</td>
<td>2</td>
<td>0.5</td>
<td>5</td>
<td></td>
<td>8</td>
<td>12</td>
<td>15</td>
<td>21</td>
<td>9</td>
</tr>
<tr>
<td>HB</td>
<td>15</td>
<td>0</td>
<td>8</td>
<td></td>
<td>23</td>
<td>16</td>
<td>46</td>
<td>69</td>
<td>23</td>
</tr>
<tr>
<td>2AIY Insulin</td>
<td>787</td>
<td>337</td>
<td>103</td>
<td></td>
<td>1564</td>
<td>1554</td>
<td>2454</td>
<td>2670</td>
<td>2238</td>
</tr>
<tr>
<td>SB</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td></td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>HB</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td></td>
<td>12</td>
<td>8</td>
<td>20</td>
<td>24</td>
<td>16</td>
</tr>
</tbody>
</table>

* Salt Bridges (SB) and Hydrogen Bonds (HB) bond lengths are defined as <3.5Å
* The pathway(s) that break the least interface area is in bold
3MVO is the PDB code for the crystal structure of bovine glutamate dehydrogenase (GDH) complexed with NDP and Eu$^{3+}$. There does, however, exist a structural file for apo GDH (PDB code 1NR7), but the sequence does not match well to experimentally used GDH and, therefore, was not used here. The in silico analysis of 3MVO shows that the trimer pathway T_1 breaks the lowest interface area, which matches with the SID behavior of normal (+39) and reduced (+27) charged bovine GDH reported by Ma et al [51] and shown in Figure 5.2c and d. As shown in Figure 5.2a, +39 GDH dissociated exclusively through the typical CID dissociation pathway, with the ejection of highly charged monomer. In contrast, SID of +39 GDH (Figure 5.2c) delivers predominantly trimers. Tandem MS experiments of charge-reduced +27 GDH by CID 200 V, which is the highest acceleration voltage allowed in our instrument, does not result in dissociation of the complex. However as discussed earlier, SID of +27 GDH still results in generation of primarily trimers. Therefore, SID behavior of bovine GDH is consistent with the in silico disassembly analysis.
Figure 5.2. CID and SID tandem mass spectra of normal charged and reduced charged bovine GDH
(a) CID of +39 bovine GDH generates highly charged monomer products. (b) CID of +27 bovine GDH does not generate products even at the highest voltage that the instrument allows. SID of (c) +39 and (d) +27 bovine GDH primarily generates trimer
Reproduced from Reference [51] with permission. Copyright © 2014, American Society for Mass Spectrometry

We next considered a glutamate dehydrogenase from a different organism, namely *Proteus sp.*, which is a genus of gram-negative proteobacteria. This bacterial GDH also exists as a hexamer. No crystal structure exists for this hexamer and instead
comparisons were made with 2YFQ, the PDB ID for the crystal structure of glutamate dehydrogenase from *Peptoniphilus asaccharolyticus*. According to the interface analysis (Table 5.1), the monomer-pentamer dissociation pathway breaks the lowest interfacial area, making it theoretically the most accessible route to dissociate. When normal charged *Proteus* bacterial GDH (+36) is subjected to CID, monomer ejection is observed (Figure 5.3). The most abundant monomer peak has +23 charges, which is a highly disproportionate amount of charge considering 6 identical monomer sharing 36 charges on the hexamer. Once subjected to SID analysis at low energy (~2160 eV), monomer and pentamer products can be observed as predicted by interface analysis in Table 5.1. At medium SID energy (~3600 eV), monomer/ pentamer and dimer/ tetramer can all be observed, with monomer as the most abundance product (Figure 5.4). In addition, there are high abundances of charge-reduced precursors. There are two distributions of monomer: one corresponds to extended monomers with charge centered on +17. The other one corresponds to compact monomers with charge centered on +7. The extended monomers may be products from gas collisions, while +7 monomers, which is more charge-to-mass symmetric, are products from SID. The dimer products have charges centered on +11 while tetramers are centered on +15. At even higher SID energy (~4320 eV), the trimer products start to appear. The *in silico* disassembly analysis has shown that trimer production breaks a slightly lower interfacial area (8%) than dimer-tetramer pathway. The reason why the dimer-tetramer pathway is more accessible at lower energies than the trimer pathway, despite involving breaking a slightly larger interfacial area, may be due to the significantly lower number of potential hydrogen bonds (20 vs. 48). PISA analysis suggests that the dimer-tetramer pathway would break 20 hydrogen
bonds while the trimer pathway would break 48 hydrogen bonds. Another reason might be that the PDB ID 2YFQ closely, but not exactly, resembles the structure of the experimental GDH from Proteus sp. It is, therefore, possible that D_1 pathway involves breaking a lower interfacial area than the T_1 pathway for the Proteus sp. GDH. Overall, the predicted monomer-pentamer as the most dominant disassembly pathway by analysis of the crystal structure 2YFQ matches well with the SID behavior of Proteus sp. GDH and is distinctly different to the behavior observed previously for bovine GDH which exists as a dimer of trimers.[51]

Figure 5.3. CID (AV=125 V) mobiligram of +36 Proteus GDH
(Intensity is plotted in log color scale)
Considering the notable differences between bovine GDH and bacterial GDH, we expanded our study to include a third hexameric protein complex, Putative GTP cyclohydrolase 1 type 2 (GCH). In the hexamer, three dimers assemble to a ring-like structure surrounding a central part. [213] The corresponding PDB ID for GCH is 1MNO. In silico disassembly analysis of 1MNO has shown that the monomer-pentamer dissociation (M pathway) involves breaking the lowest interfacial area (Table 5.1), therefore, is the most accessible route to dissociate in theory. Considering the CID of +27 ions (normal charged) for GCH (Figure 5.5), the most abundant peaks correspond to monomer products carrying charges around +13. Unusually, dimers with +12 to +17 charges were also observed in CID process. SID experiment of the +27 ions of GCH (Figure 5.6) also resulted in monomer products, however in this case, the charge centered around +6, in addition to dimer products centered around +8, trimer products centered around +10, tetramer products centered around +11, and pentamer products
centered around +12 were observed. Closer examination of the abundance of different fragments (Figure 5.7) revealed that trimer products are much less abundant than the other products. The low abundance of pentamer observed here is due to the decreased ability to transmit ions of high m/z and detector bias common in this instrumentation. The experimental SID dissociation, therefore, shows an excellent correlation with the *in silico* disassembly analysis, which predicts that the trimer dissociation pathway is the least favored, because of the largest interfacial area that would have to be cleaved to produce these products.

Figure 5.5. CID (ΔV=125 V) mobiligram of +27 GTP cyclohydrolase (Intensity is plotted in log color scale)
Figure 5.6. SID (ΔV=140 V) mobiligram of +27 GTP cyclohydrolase (Intensity is plotted in log color scale)

Figure 5.7. Extracted tandem mass spectra of products with different oligometric states from SID (ΔV=140 V) mobiligram of +27 GTP cyclohydrolase (The intensity is normalized to +6 monomer, the highest intensity product)
In addition to large homo-hexamers GDH and GCH, which are over 150 kDa, we also examined the dissociation of a small homo-hexamer, insulin (~35 kDa). 2AIY is the PDB ID for the NMR results of R6 human insulin hexamer. Analysis of the interface information obtained from PISA for 2AIY and the corresponding dissociation pathway analysis reveals that monomer-pentamer and dimer-tetramer are the favored dissociation pathways as determined by considering the pathways which involves dissociating lowest interfacial area (1564Å² and 1564Å² respectively, Table 5.1). Experimentally, the products of the +11 ion for the human insulin hexamer range from monomer to pentamer by either CID (Figure 5.8) or SID (Figure 5.9). Comparison of the abundance of pentamer, tetramer and trimer products in either CID (Figure 5.10) or SID (Figure 5.11) reveals that the trimer products are the least abundant products with both dissociation methods. The reason of comparing the abundance of pentamer, tetramer and trimer is because they are more likely to be the direct dissociating products of the hexamer precursor via M, D, and T pathways as shown in Figure 5.1. Monomer and dimer may be generated via secondary fragmentation, rather than direct fragmentation from the hexamer precursor. Therefore, for human insulin hexamer, both the CID and SID dissociation behaviors match with the in silico disassembly analysis, in contrast to the other systems studied here in which SID was generally found to be more consistent with the in silico disassembly pathway analysis. The reason why several dissociation pathways have been observed in CID for insulin hexamer can be attributed to the small size of the complex (~35 kDa) allowing CID energy to overcome energy barriers for different dissociation pathways. While for larger systems, CID may only overcome the least energetic pathway, which involves monomer unfolding and ejection.
Figure 5.8. Insulin hexamer +11 CID (ΔV=70 V)
(Intensity is plotted in log color scale)

Figure 5.9. Insulin hexamer +11 SID (ΔV=60 V)
(Intensity is plotted in log color scale)
Figure 5.10. Extracted tandem mass spectra for trimer, tetramer and pentamer from Insulin hexamer +11 CID (ΔV=70 V)
(The intensity is normalized to +8 pentamer)

Figure 5.11. Extracted tandem mass spectra for trimer, tetramer and pentamer from Insulin hexamer +1 SID (ΔV=60 V)
(The intensity is normalized to +6 tetramer)
5.4 Conclusions

In this chapter, we performed *in silico* disassembly analysis of homo-hexamer protein complexes with $D_3$ symmetry. In general, the homo-hexamer can be, theoretically, directly dissociated through monomer – pentamer, dimer – tetramer and trimer – trimer pathways. In addition, each of the dimer – tetramer and trimer – trimer pathways can proceed *via* two pathways that involve breaking different amounts of interfacial area and different numbers of potential interactions in the form of salt bridges and hydrogen bonds. Based on the amount of interfacial area being broken in each of the pathways, the accessibility of different pathways can be assessed with the theoretically most favorable pathways having the least interfacial area to break. Four different hexamers bovine GDH, bacterial GDH, GCH and insulin are analyzed and the most favorable pathway(s) are obtained in *silico*.

Tandem mass spectrometry experiments were performed with these homo-hexamer protein complexes of interest. Except for the CID experiments of human insulin (2AIY), CID experiments almost exclusively proceed with ejection of highly charged monomer and its complementary pentamer providing limited information on subunit architecture. In contrast, the observed SID fragments are consistent with the *in silico* disassembly analysis for all the hexamers studied here. The trimer pathway is predicted to be the most accessible pathway for bovine GDH which matches with the fact that predominantly trimer products are observed in SID experiments. In contrast, the monomer pathway is predicted to be the most accessible pathway for bacterial GDH. In SID of bacterial GDH, monomer with charges that are proportional to its mass fraction
are observed at low energy SID, followed by dimer and trimer products at higher SID energy, showing that the monomer pathway is the least energy consuming pathway experimentally. The fact that all pathways are accessible under these experiments, through the variation of energy, is entirely consistent with the in silico dissociation. The monomer pathway is predicted to be the most favorable pathway for GCH by the analysis of the crystal structure, followed by dimer pathway and the least favored is the trimer pathway. Considering of the product abundance of monomer> dimer > trimer observed in SID, a consensus exists for in silico analysis and SID experiments. Based on these results, SID could be a useful tool in the study of unknown structures as it gives fragmentation reflective of starting structure. The next chapter demonstrates the power of SID in postulating the subunit connectivity of a hetero hexamer protein complex.
Chapter 6. Quaternary structure determination: Structure Refinement of a Hetero-hexamer protein by multiple mass spectrometry approaches

This chapter is adapted from “Refining the Structural Model of a Hetero-hexameric Protein Complex Using Multiple Mass Spectrometry Approaches.” Yang Song, Micah Nelp, Vahe Bandarian, and Vicki H. Wysocki, ACS Central Science, submitted

6.1 Introduction

Mass spectrometry has become an indispensable tool for characterizing proteins. The most well-known use of it lies in the field of proteomics, which mainly involves protein identification by digestion to peptides, followed by analysis by liquid chromatography coupled tandem mass spectrometry (LC-MS/MS).[214, 215, 216] More recently, mass spectrometry has begun to play an influential role in structural biology,[217] including a number of sophisticated studies considering protein conformations and dynamics.[184, 201, 202, 203, 204] Many proteins exist and function as multimeric complexes, in which subunits non-covalently interact with each other. In fact, more than 60% of entries in the Protein Data Bank are dimers or larger assemblies.[2] The characterization of protein complexes, especially those which do not yet have atomic coordinate information, including their composition and subunit interaction is crucial in structural biology.
Information provided through different mass spectrometry-based experiments can be complementary to each other, thus once combined, can contribute to a better understanding towards the structure of a specific protein complex.[218] In native mass spectrometry experiments protein samples, prepared in neutral aqueous buffers, are ionized, desolvated and introduced into the gas phase using nano-electrospray ionization (nano-ESI). In this way, non-covalent interactions can be preserved and the mass of the intact complex can be obtained. More likely the mass will be higher than the complex mass by sequence, since water, salts, and/or buffer molecules can be attached to the complex. In order to probe the constituents confidently and also gain information about architecture of the complexes, gas phase disassembly methods are applied. The most widely applied dissociation method is collision induced dissociation (CID). In CID the dominant dissociation pathway is via ejection of a single highly charged subunit, and therefore CID cannot provide substructural information. Several cases have been reported in which CID can provide dissociation products reflective of the initial complex, however, this remains the atypical pathway.[38] In contrast, surface induced dissociation (SID) has been shown to generate subcomplexes that truly reflect the structure of the complex.[43, 44, 46, 49, 50, 187]

Given the purity of protein complex sample, once the complex can be detected by native mass spectrometry, it can be destabilized by adding organic solvent or increasing ionic strength. The generated subcomplexes by solution destabilization can be analyzed by mass spectrometry, and provide subunit connectivity in addition to information on the relative binding strength of different subunits.[34] In 13 out of 16 protein homooligomers
Levy and coworkers studied,[207] subcomplexes containing the larger interface upon either dilution of the organic solvent or ionic strength change were observed. In an extensive study by Marsh and coworkers[208], 23 out of 27 hetero protein complexes showed excellent agreement between solution disassembly and interface sizes.

Ion mobility (IM) coupled to mass spectrometry (IM-MS) provides an additional dimension of information, enabling determination of the size and shape of the analytes.[219, 220, 221] Based on the time analytes spend travelling through an ion mobility cell, driven by electric gradient and retarded by collisions with a bath gas, size and shape information in the form of collisional cross-section (CCS) can be derived.[79, 92] The subunits/subcomplexes generated by solution destabilization can be assembled back into the intact complex based on CCS constrains from IM/MS experiments on the intact complex and the subunits/subcomplexes themselves.[222, 223]

Another powerful mass spectrometry-based approach in structural biology is covalent labeling.[217] This approach involves changing the mass of different parts of a protein in a conformational-dependent manner.[224] Covalent labeling is often performed either specifically or non-specifically towards amino acids. Non-specific covalent labeling is usually carried out with hydroxyl radicals.[225, 226] Laser is normally required in generating the hydroxyl radicals. Amino acid-specific labeling, in contrast, has the advantages of readily accessible reagent and easiness to use.[116] The reactions usually take place in solution with more solvent-exposed amino acids get labeled and the mass shift is detected by mass spectrometry in the form of digested peptides. Covalent modification on amino acids is usually stable during analysis, unlike back-exchange or
scrambling happened in H/D exchange. Using covalent labelling approaches consideration must be given to whether the introduction of modifications can alter the real structure of analytes,[116] therefore, checking the structural integrity of the protein subjected to covalent labeling is essential.

Toyocamycin nitrile hydratase (TNH) is a protein complex that catalyzes the hydration of toyocamycin to produce sangivamycin.[227] To-date, however, there have been no solved structures reported by either NMR or x-ray crystallography on this important enzymatic complex. It’s structure has, so far, only been characterized by mass spectrometry.[45] Mass spectrometry first confirmed that this enzyme is composed of two α, two β and two γ subunits (MW α 21190 Da, β 9974 Da, γ 11444 Da). CID applied on the +19 charge state of the hexamer resulted mainly in the ejection of α and β subunits. In contrast, the SID spectra are dominated by the αβγ trimer.[45] In consideration of the match of SID behavior with crystal structure for urease and GDH hexamer[51], it is likely that TNH is a dimer of αβγ trimers. There exists high resolution structural information for members of nitrile hydratase family (47 PDB entries), however, the majority are hetero-tetraromers comprising only two different subunits. There is a three-subunit-containing enzyme, thiocyanate hydrolase (SCNase) from Thiobacillus thioparus, but only the γ subunit of the SCNase is homologous to the α subunit of TNH. The other two subunits of the SCNase share little similarity with β and γ subunits of TNH.[227] Therefore, only limited information can be obtained from homology models. Here, we applied multiple mass spectrometry based methods to refine the structure of TNH.
6.2 Experimental

6.2.1 TNH expression and purification

TNH from *Streptomyces rimosus* was prepared by Micah Nelp in Dr. Bandarian’s lab at the University of Arizona. The genes encoding the three subunits of toyoacamycin nitrile hydratase were cloned as follows for heterologous co-expression in *E. coli* BL21 (DE3): the alpha subunit was cloned into the NdeI and XhoI sites of pACYCDuet-1, the gamma subunit was cloned into the NcoI and HindIII sites of pACYCDuet-1, and the beta subunit was cloned into the NdeI and XhoI sites of pET-29a. During initial purifications, the yield of the gamma subunit relative to the other two was found to decrease over the course of the multiple purification steps. To increase the yield of the gamma subunit and thus the replete heterohexamer at the beginning of the purification, a second copy of the gene encoding the gamma subunit was cloned in the NcoI and HindIII site of pETDUET-1.

Expression plasmids were introduced into *E. coli* with electroporation and plated on lennox broth agar containing 34 µg/mL of kanamycin and chloramphenicol and 100 µg/mL of ampicillin, which all further cultures shared. A single colony was used to inoculate a 0.1 L overnight starter culture which was spread evenly among six 2.5 L fernbach flasks containing 1 L of lennox broth. These were grown at 37 °C to an OD600 nm of 0.5 at which point expression was induced with the addition of isopropyl β-D-thiogalactopyranoside to 0.1 mM and CoCl$_2$ to a final concentration of 0.05 mM. After eight hours, these were spun down at 4,000xg, and cell paste was frozen with liquid N$_2$ and stored at -80 °C.
All purification steps were carried out at 4 °C. The cell paste was resuspended in 20 mM Tris HCl pH 8.0 and 1 mM phenylmethylsulfonyl fluoride and disrupted on ice with sonication at 50% for 10 min with 30 sec bursts and 30 sec rests. Insoluble material was removed with centrifugation at 18,000xg for 30 min. Clarified lysate was loaded onto a Q-sepharose column (2.6 x 12.5 cm) pre-equilibrated with 20 mM Tris HCl pH 8.0 and eluted over a 0.7 L gradient to the same buffer with 0.5 M NaCl. Fractions containing TNH as determined by SDS-PAGE and an orange color were pooled and brought to 1 M (NH₄)₂SO₄ with the addition of solid (NH₄)₂SO₄ over 20 min on ice with gentle stirring. This was loaded onto a butyl-sepharose column (2.6 x 12.5 cm) pre-equilibrated in 20 mM Tris HCl pH 8.0 and 1 M (NH₄)₂SO₄ and eluted with a 0.7 L gradient to the same buffer lacking (NH₄)₂SO₄ and with 10% ethylene glycol (v/v). Fractions containing TNH were pooled and concentrated by addition of solid (NH₄)₂SO₄ to 60% saturation over 20 min on ice with gentle stirring followed by centrifugation at 18,000xg for 30 min. The precipitated pellet was resuspended in 1 mL of 25 mM HEPES•NaOH pH 7.5 and loaded onto a HiPrep 16/60 Sephacryl S200 column (GE Healthcare) pre-equilibrated in the same buffer. Fractions containing TNH were pooled and concentrated with Amicon Vivaspin Turbo 10K (polyethersulfone membrane) centrifugal concentrators and flash frozen with liquid N₂ and stored at -80 °C.

6.2.2 Complex analysis by native mass spectrometry

To perform solution disruption and SID/IM experiments, the sample was first buffer exchanged into 100mM ammonium acetate (AmAc, Sigma-Adrich, St. Louis, MO) via Micro Bio-Spin™ P-6 Gel Columns (Bio-Rad, Hercules, CA). The final
concentration of the protein complex was 16 μM. Solution disruption was performed by adding 10% 20% 30% 40% and 50% methanol (Fisher Scientific, Pittsburgh, PA) to buffer exchanged protein but keeping the ionic strength (100mM AmAc) and protein complex concentration (8μM) constant. Charge reduction was performed by adding 30mM triethylamine acetate (TEAA) to sample solution.

Samples were nano electrosprayed to a SYNAPT G2-S mass spectrometer (Waters Corporation, Manchester, UK) with an SID device incorporated before the ion mobility cell (SID/IM experiments).[48, 49] The trap traveling wave ion guide (TWIG) was truncated to accommodate the SID cell. Under transfer mode, the voltages applied to the ten lenses of the SID cell were adjusted to guide the ions to fly through without colliding with the surface. While in SID mode, the ions were steered towards the surface by the voltages on the lenses. All the lenses are tuned to achieve better precursor transmission and products collection. For voltages applied to lenses at specific conditions, see Table 6.1. Glass capillaries for nano-electrospray ionization were pulled on a P-97 micropipet puller (Sutter Instruments, Hercules, CA), filled with sample solution and a 0.8-1.1kV ionization voltage was applied to a platinum wire inserted into the capillaries. Typical instrument settings for SID/IM experiments were: sampling cone 20 V, source offset 10 V, source temperature 25°C, trap gas flow 2 mL/min, helium cell gas flow 180 mL/min, IMS gas flow 60 mL/min; trap DC entrance 0 V, trap DC bias varies from 83V to 233V (corresponding to SID acceleration voltage 30V-180V), trap DC -2 V, trap DC exit 0 V, trap wave velocity 160 m/s, trap wave height 4 V; IMS DC entrance 10 V, helium cell DC 25 V, helium exit -5 V, IMS DC bias 5 V, IMS DC exit 0
V, IMS wave velocity 300 m/s, IMS wave height 20 V; transfer DC entrance 2 V, transfer DC exit 15 V, transfer wave velocity 50 m/s, transfer wave height 4V.

Table 6.1. SID/IM experiment, trap DC bias and trap SID lenses voltage at different SID acceleration voltages. (Unit: V)

<table>
<thead>
<tr>
<th>Trap SID ΔV</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
<th>130</th>
<th>140</th>
<th>150</th>
<th>160</th>
<th>170</th>
<th>180</th>
</tr>
</thead>
<tbody>
<tr>
<td>trap DC bias</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Exist 2</td>
<td>115</td>
<td>115</td>
<td>115</td>
<td>115</td>
<td>115</td>
<td>115</td>
<td>115</td>
<td>115</td>
<td>115</td>
<td>115</td>
<td>115</td>
<td>115</td>
<td>115</td>
<td>115</td>
<td>115</td>
<td>115</td>
</tr>
<tr>
<td>2. Exit 1</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
<td>130</td>
</tr>
<tr>
<td>3. Front Top</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
</tr>
<tr>
<td>4. Front Bottom</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
<td>-60</td>
</tr>
<tr>
<td>5. Mid Bottom</td>
<td>125</td>
<td>125</td>
<td>125</td>
<td>125</td>
<td>125</td>
<td>125</td>
<td>125</td>
<td>125</td>
<td>125</td>
<td>125</td>
<td>125</td>
<td>125</td>
<td>125</td>
<td>125</td>
<td>125</td>
<td>125</td>
</tr>
<tr>
<td>9. Entrance 2</td>
<td>105</td>
<td>105</td>
<td>105</td>
<td>105</td>
<td>105</td>
<td>105</td>
<td>105</td>
<td>105</td>
<td>105</td>
<td>105</td>
<td>105</td>
<td>105</td>
<td>105</td>
<td>105</td>
<td>105</td>
<td>105</td>
</tr>
</tbody>
</table>

6.2.3 Collisional cross-section measurement and spherical subunit modeling

The CCS was measured following a published protocol.[79] The calibrants used for α, β, γ monomer and αβ dimer CCS measurement were +3, +4 melittin from honey bee venom, +6, +7 cytochrome c from equine heart, +8, +9 β-lactoglobulin monomer and +11 to +13 β-lactoglobulin dimer from bovine milk. The calibrants used for αβγ trimer and (αβγ)₂ hexamer were +14 to +16 transthyretin from human plasma, +15 to +18 avidin.
from egg white, +20 to +23 concanavalin A from canavalina ensiformis, +23 to +26 alcohol dehydrogenase from yeast. All the protein standards were purchased from Sigma-Adrich (St. Louis, MO) and prepared as 10uM in 100mM AmAc.

CCS of each subunit was used to generate the spherical models. As described by Zoe Hall, et al.[223], the radius of a spherical model \( r = (\text{CCS}/\pi)^{0.5} - r_{\text{He}} \), in which \( r_{\text{He}} = 1.4 \text{Å} \). Projection approximation (PA) in Mobcal program[94, 95] and a scaled PA method[228] was used to perform theoretical CCS calculation. The mobcal.f (downloaded from http://www.indiana.edu/~nano/software.html) was modified to calculate the CCS of a complex of spherical models as explained in a published protocol.[79] Briefly speaking, it was done by adding subunit as an atom in the file but specifying the mass and radius derived from IM-MS analysis. The specific lines added are shown below:

Remove the trajectory method (TJM) by placing a ‘c’ in front of the Line#338.

Insert at Line#671 and Line#2682 the following lines describing the radius of coarse-grained sphere \( \alpha, \beta \) and \( \gamma \) subunits.

```
c     alpha (in TNH)

c
if(imass(iatom).eq.100) then
  itest=1
  xmass(iatom)=21190.3d0
  eolj(iatom)=1.35d-3*xe
```
rolj(iatom)=23d0*1.0d-10

rhs(iatom)=23d0*1.0d-10

endif

c

c beta (in TNH)

c

if(imass(iatom).eq.101) then

ittest=1

xmass(iatom)=9974.4d0
eolj(iatom)=1.35d-3*xe
rolj(iatom)=17.4d0*1.0d-10
rhs(iatom)=17.4d0*1.0d-10

endif

c
c gamma (in TNH)

c

if(imass(iatom).eq.102) then

ittest=1

xmass(iatom)=11444.1d0
eolj(iatom)=1.35d-3*xe
rolj(iatom)=18.3d0*1.0d-10

rhs(iatom)=18.3d0*1.0d-10

endif

The coordinate of the centers of the spheres (in .mfj file) determine how the spheres are stacked together. An example of .mfj file for αβγ trimer is shown below:

```
1
3
ang

none

1.0000

  0  0  0  100   .10000
  0  26.5  0  101   .10000
  0  20.7  21  102   .10000

3
```

"
6.2.4 Bottom-up approach for surface mapping modification localization

A flow chart showing experimental procedure of surface mapping experiments is shown in Figure 6.1A and the reaction of labeling amino acids side chain is shown in Figure 6.1B.

Figure 6.1. A) The flow chart of the surface mapping experiment. B) Labeling reagent PGO and its reaction with arginine rendering a mass increase of 134.0368 or 116.0262.
The surface mapping of arginine in TNH or glutamate dehydrogenase (GDH, Sigma-Adrich, St. Louis, MO) was performed at room temperature for 14 hrs. 1mM, 2mM, 4mM and 7mM of phenylglyoxal (PGO, Sigma-Adrich, St. Louis, MO) were chosen to react with 8μM of TNH in 25 mM HEPES (pH 7.5), yeilding 125:1, 250:1, 500:1, and 875:1 molar ratios of phenylglyxal to TNH. The ratios of PGO to arginine targets in TNH were 1.5:1, 3:1, 6:1, and 11:1, respectively. To keep approximately the same PGO to arginine ratios for GDH, the concentration used was 4μM GDH. The reaction was stopped by depleting PGO through buffer exchanging to 100mM AmAc via Micro Bio-Spin™ P-6 Gel Columns. A part of the sample was subjected to native MS analysis to check hexamer integrity and the rest was subjected to bottom-up experiments to localize the modified arginine.

The bottom-up approach to identify arginine modifications was performed by in-solution pepsin digestion of PGO labeled samples with LC-MS/MS analysis on a linear ion trap mass spectrometer (Velos Pro, ThermoFisher Scientific Inc., Waltham, MA) coupled to a UPLC (nanoACQUITY, Waters Corporation, Manchester, UK). Pepsin digestion was performed first by adjusting pH of the labeled sample solution with 1M HCl (Sigma-Adrich, St. Louis, MO) to below 4, followed by adding pepsin (Promega, Madison, WI) at 1:10 enzyme: protein ratio. The digestion was performed at 37 °C, on a 150 rpm thermomixer for 7 hrs. The reaction was stopped by heating on a 95 °C block for 5min. Prior to LC-MS/MS injection, the sample was centrifuged at 10,000 g for 10 min. The amount of injection was 0.4 μg.
LC-MS/MS data analysis was performed by SEQUEST HT in Proteome Discoverer software (ThermoFisher Scientific Inc., Waltham, MA). Some searching parameters are listed here: Precursor mass ranged from 400 to 5000 Da. Peptide length ranged from 4 to 144. Precursor mass tolerance was 1.5 Da and fragment mass tolerance was 0.8 Da. Database was the sequence of α, β, γ and pig pepsin. No enzymes were specified for in silico digestion. Dynamic modifications were methionine oxidiation (M+15.995) and phenylglyoxal reacted arginine (R+116.026 or 134.0368). Peculator was used to validate the identities of peptides with 1% false discovery rate (FDR).

6.2.5 Molecular modeling

“Protinfo PPC” web server (protinfo.compbio.washington.edu/ppc/) was used to predict atomic level structure information for protein complexes from their amino-acid sequences based on homology. [229] The structural model candidates for the TNH αβγ trimer are generated based on four templates (PDB ID: 3hht, 1ugq, 1ugs and 1ahj) with αβ and αγ dimer models predicted by protinfo ppc server with sequence information as input. A 10 ns all-atom molecular dynamics simulation with periodic water box was conducted using NAMD 2.9 [230] with the CHARMM force field [231, 232]. Periodic water box were set up as there was a layer of 10 Å water from the atom with the largest coordinate. The charges were neutralized by 0.15 M NaCl. Long-range forces in the periodic system were calculated using the Particle Mesh Ewald (PME) [233] with 1.0 grid spacing. Linear bonds involving hydrogens are treated as rigid bonds to allow the simulation to be performed at 2 fs steps. Different trimer candidates were first subjected
to 1000 steps of energy minimization, followed by equilibrium simulation in constant temperature, 310K and constant pressure, 1 atm using Langevin piston method [234].

Also, a 1000-step energy minimization and 10 ns all-atom molecular dynamics simulation in vacuum was conducted using NAMD 2.9 [230] with the CHARMM force field [231, 232]. In this case, none of periodic boundary conditions, PME, and constant pressure was used.

6.3 Results and Discussions

6.3.1 Solution-disruption experiments reveal TNH subcomplexes \( \alpha \beta \gamma \) and \( \alpha \beta \)

Our previous energy resolved-SID on a QTOF platform showed that TNH hexamer is a dimer composed of two \( \alpha \beta \gamma \) trimers, with subunits \( \alpha \), \( \beta \) interacting strongly. [45] Solution disruption was used here as an alternative tool to generate subcomplexes for comparison with our previous SID results. As shown in Figure 6.2A, increasing the percentage of methanol in the sample solution leads to the dissociation of TNH hexamer (dark stripped area decreases). The subcomplexes generated from solution disruption studies can provide insight into the arrangement of subunits in the intact complex. For TNH, the \( \alpha \beta \gamma \) trimer and \( \alpha \beta \) dimer (black full and purple crosshatch areas, respectively) are observed as the methanol content increases. In addition, the single subunits of \( \alpha \), \( \beta \) and \( \gamma \) can also be observed with increasing methanol percentage (red, blue and green areas, respectively).
Figure 6.2. Methanol-induced solution-disruption of TNH hexamer
A). The detection of the TNH complex and its subcomplexes are by mass spectrometry full scan. 
B) A cartoon showing the connectivity of subunits in TNH by methanol disruption results. The hexamer is composed of two αβγ trimers and within the trimer, α-β interaction is the strongest.

Solution disruption studies have successfully predicted the subcomplexes in several protein complex systems, for example Hernández et al.[206] were able to derive a three-dimensional interaction map of yeast exosome, which contains ten different subunits, by generating subcomplexes from the highly pure native protein complex via methanol and DMSO solution disruption. Besides adding organic solvent to perturb the native complexes, ionic strength can also be manipulated to generate protein subcomplexes. Zhou et al.[235] observed a series of subcomplexes from eIF3, a complex
containing 13 subunits, with increasing AmAc concentration from 250mM to 500mM. It was suggested that electrostatic interactions play an important role in sustaining the whole complex, thus can be disrupted by high ionic strength. Increasing ionic strength for TNH protein complex by raising the AmAc concentration up to 1M did not induce the dissociation of TNH hexamer (Spectra shown in Figure 6.3). This suggests that the strongest interactions between subunits are most likely to be hydrophobic interactions, which would not be disrupted with high ionic strength solutions.[236] Based on the discovery of the αβγ trimer and αβ dimer subcomplexes, an interaction map can be generated (Figure 6.2B). The hexamer is clearly composed of two αβγ trimers. Within the trimer, α-β interaction is the strongest, enabling these subcomplexes to be preserved. However, the solution disruption experiments reveal neither how γ interact with α and β within the trimer, nor how the two trimers interact.

Figure 6.3. Increasing ionic strength from F) 0.1M ammonium acetate to A) 1M ammonium acetate doesn’t induce the dissociation of TNH hexamer.
6.3.2 SID/IM and SID/IM/SID reveal increased information on subunit arrangement

SID is a gas phase dissociation method which involves an analyte ion colliding with a surface. The dissociation occurs over a very fast time scale (picoseconds), in one step and has the advantage that greater energy is transferred in a single collision than in conventional gas collision induced dissociation. Thus SID can access more dissociation pathways and provide more information about the contact within a protein complex.[43, 44, 46, 49, 50, 187] Furthermore, Zhou et al.[48] have shown that dissociation by SID of reduced charge states better preserves subunit contacts of products. It has been shown that products from TNH hexamer by CID dissociation are dominated by monomer ejection.[45] However, upon SID dissociation, for +14 TNH hexamer (charge-reduced precursor), several subcomplexes could be generated. The major SID products observed in low energy SID coupled with ion mobility (SID/IM) (Figure 6.4A), shown with a square root intensity scale, are αβγ trimer. At higher SID energies, αβ dimer and αβγ2 tetramer become more abundant (Figure 6.4B). The simplest interpretation of this observation is that the two γ subunits are in contact and may provide a significant interface between the two trimers. The interaction between α and β is the strongest within the αβγ trimer. With the separation provided by IM and a square root intensity scale, some less abundant subcomplexes can also be observed clearly in high energy SID/IM (Figure 6.4B), namely the γ2 dimer, βγ dimer, αγ dimer, αγ2 trimer, αβ2 trimer, α2β2 tetramer, and αβ2γ tetramer. These products appear only at the higher SID energies, suggesting that their dissociation from the complex may involve breaking a greater
number of, or stronger, interfaces. The presence of αβ, αγ and βγ suggests that the αβγ trimer may adopt trigonal arrangement, with any two of the three subunits in contact. The presence of γ2 dimer and αγ2 trimer suggests that two γ subunits are in contact. γ2 dimer and α2β2 tetramer may come from dissociation of the precursor with multiple interfaces broken at the same time. αγ dimer, αγ2 trimer may be generated from secondary fragmentation from αβγ trimer and αβγ2 tetramer respectively. Alternatively, αγ dimer, αγ2 trimer might be generated directly from the hexamer, yielding complementary αβ2γ tetramer and αβ2 trimer. These complementary oligomers might then undergo secondary dissociation quickly to αβγ trimer and αβ dimer, explaining their extremely small absence in the spectrum. The observation of αβ2 trimer and αβ2γ, α2β2 tetramer, makes it highly likely that the two β subunits are in weak contact. The information obtained on the contact between subunits from SID experiments enables the composition of this complex to be visualized, as shown in the cartoon representation given in Figure 6.4C. The stronger non-covalent interaction of subunits, based on intensities of subcomplexes, is represented by thicker lines connecting spheres. The strongest interaction is α-β, followed by α-γ and γ-γ while the weakest interactions are between β-β and β-γ with strength of interaction inferred from ion intensities in the SID/IM data of Figure 6.4B.
Figure 6.4. A) SID/IM/MS plot (700 eV) and B) SID/IM/MS plot (1680 eV) for +14 TNH hexamer. The intensity of spots shows the normalized abundance of the ion species in square root scale. C) A cartoon showing the connectivity of subunits in TNH by SID/IM/MS results.

In A), the most abundant species are αβγ trimer (average charge ~7) and remaining precursors with charge reduction. In B), the most abundant species in circled area are α monomer (average charge ~5), αβ dimer (average charge ~6), αβγ trimer (average charge ~7), αβγ2 tetramer (average charge ~8) and αβγ2 pentamer respectively. There are also species with lower abundance in circled regions which correspond to γ2 dimer, βγ dimer, αγ dimer, αγ2 trimer, αβ2 trimer, αβ2 and αβ γ tetramer. In C), the lines represent contacts between pairs of subunits and the thicknesses denote the strength of non-covalent interactions: 1≈1’<2≈2’<3
The $\alpha\beta\gamma$ trimer is the most abundant product from the hexamer upon SID at the lowest energy onset. In order to understand the relative interface strengths within this trimer, a pseudo MS3 experiment can be performed with two SID processes. The first SID occurs after the selection of the precursor (+17 hexamer), following which the products are separated in the IM cell. The second SID event then occurs after the ion mobility separation. If the ions are spread out in the drift time space during analysis the IM cell can be used to identify the origin of fragments. Manual extraction of a specific drift time of a trimer from the 2D mobiligram can identify dissociation products of this trimer following increasing secondary SID energies. The energy resolved SID results of either the +8 or +9 $\alpha\beta\gamma$ highlights that the trimer decreases with increasing SID acceleration voltage, as expected (Figure 6.5). The most abundant products are $\gamma$ and its complementary dimer $\alpha\beta$ across the different SID acceleration voltages. This directly shows that within $\alpha\beta\gamma$ trimer, the interaction between $\alpha$ and $\beta$ is much stronger than $\alpha-\gamma$ or $\beta-\gamma$, as the earlier SID/IM experiments inferred.

![Figure 6.5](image)

Figure 6.5 The energy resolved SID results for dissociation of +8 (A) and +9 (B) TNH $\alpha\beta\gamma$ trimer generated in the trap-SID and separated in the ion mobility cell.
6.3.3 Predicting the TNH structure via collisional cross-section constraints on coarse-grained models

Good agreement has been reported between experimental CCS of subcomplexes generated from solution perturbation and theoretical CCS calculated from structural information in PDB file.[223] It has been shown that the CCS of a set of protein complexes represented by overlap spheres (subunits) is in very close agreement to the CCS calculated from high-resolution atomic structural information.[223]. Pukala and coworkers[222] applied CCS constraints to predict the topology of two 3-unit subcomplexes, f:h:m and e:l:k from human eukaryotic initiation factor 3. The former of which adopted a trigonal geometry while the latter one was linearly arranged. Bernstein et al.[102] utilized CCS of oligomers of amyloid-β proteins to determine its qualitative structure of each of the aggregates and their results showed that two isoforms of amyloid-β proteins aggregate differently. The previous success of such methods to predict complex structures clearly demonstrates that this approach can be very useful in structural characterization of unknown systems, and hence is applied here to the hexameric TNH complex.

The α and β monomers can be generated from methanol disruption of the complexes without significant unfolding of the subunits. The CCS of the α and β subunits was, therefore, measured directly from the MS spectra obtained from the 50% methanol perturbed solution. Only low abundance γ subunit can be generated by methanol disruption with poor reproducibility. In high energy SID, however, the γ subunit can be readily observed. Thus the CCS of the γ subunit could be obtained from tandem MS. The
experimental CCS of the three subunits measured either from methanol disruption or SID is shown in Table 6.2.

<table>
<thead>
<tr>
<th></th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>αβ</th>
<th>αβγ</th>
<th>6mer</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCS (Å)</td>
<td>1870±40</td>
<td>1110±30</td>
<td>NA</td>
<td>2450±70</td>
<td>2890±70</td>
<td>4860±70*</td>
</tr>
<tr>
<td>MeOH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCS(Å) SID</td>
<td>1880±40</td>
<td>1110±30</td>
<td>1210±30</td>
<td>2440±60</td>
<td>2730±160</td>
<td></td>
</tr>
<tr>
<td>r(Å)</td>
<td>23.0</td>
<td>17.4</td>
<td>18.3</td>
<td>Linear</td>
<td></td>
<td>?</td>
</tr>
</tbody>
</table>

* The CCS of hexamer is measured in 100mM AmAc aqueous solution. Standard deviations were calculated from four different day repeats.

There is no significant difference between the CCS measured by methanol disruption and SID fragmentation for the subunits α and β, thus it is valid to use the CCS measured in SID for the γ subunit. The CCS of the individual subunit was used to simulate the radius of the corresponding sphere models as described in the methods section. The radii of the spheres representing different subunits are shown in Table 6.2. The CCS of the αβ dimer and the αβγ trimer were used as constraints to limit the possible arrangement of the subunit spheres within the subcomplexes. The experimental CCS of the two subcomplexes was measured from 50% methanol perturbation and full MS scan. There is no significant difference of CCS measured for αβ dimer by methanol perturbation full MS and SID tandem MS (Table 6.2). However, the CCS measured for αβγ by SID tandem MS is slightly smaller than the CCS measured by methanol perturbation. This may be due to the subcomplexes collapsing following SID, which has been reported by Zhou et al.[48] previously. The simulation of αβ dimer is based on optimizing the distance d between the two centers to achieve the experimental CCS.
When $d$ equals $0.66(r_\alpha + r_\beta)$, or the $\alpha\beta$ linear overlap is 34%, a simulated model reaches the same CCS as experimental value. After determination of the $\alpha\beta$ dimer structure, the location of $\gamma$ is explored to achieve a CCS that matches with the experimental trimer CCS. As Figure 6.6A illustrates, in the plane defined by the center of $\alpha\beta\gamma$ with the coordinates $\alpha$ and $\beta$ fixed, the location of $\gamma$ can be represented by center $(x_\gamma, y_\gamma)$ and a known radius. At each $(x_\gamma, y_\gamma)$, a CCS of the trimer was calculated and its difference from experimental value is show as the color of $(x_\gamma, y_\gamma)$ spot, with the more intense green demonstrating a closer representation of the CCS of experimental value. It was observed that the average linear overlap derived from models with crystallographic positions was $29\pm10\%$.\[223\] Thus, together with the linear overlap exceed 20% and deceed 40%, the center of $\gamma$ was determined to be around the red arrow area. It shows that $\alpha\beta\gamma$ adopts a trigonal topology, rather than a linear arrangement.
Figure 6.6. Coarse-grained model of TNH trimer and hexamer

A) Possible location of the center of a coarse-grained sphere γ subunit in the plane defined by the centers of coarse-grained spheres α (0,0), β (0,26.5) and γ (Xγ, Yγ). (Xγ, Yγ) is varied to obtain relative difference of modeled CCS from the experimentally measured αβγ CCS, as shown by the intensity of the green. More intense green shows closer approximation to the measured CCS as shown by the color bar. The white area represents a CCS error over 7.3%. The dash curve shows the 40% and 20% linear overlap limits of the γ center from α and β. Considering both the CCS constraint and overlap constraint, the most likely location of γ center is highlighted by the red arrow. B) A possible arrangement of the hexamer by docking the two trimers represented by sphere models. The two variables are the angle of two trimer plane and the distance between the two trimers.

After assembly of the αβγ trimer, the next step is to associate the two trimers into a hexamer and use the hexmer CCS to constrain the possible structure. Two γ subunits are in close contact and also two β subunits are in contact was derived from SID/IM experiments. Most protein complexes adopt symmetry to some extent,[207] here we assume the sphere-modeled hexamer will follow C2 symmetry. If the first trimer is fixed in the yz plane and the center of βγ is the origin. The second trimer triangle can be initiated by the mirror image of the first triangle along the xy plane. Then it need to be
rotated along z axis (θ degree) and moved along z axis (z angstrom) to accommodate spherical units overlapping. A 2D color map shows a combination of θ and z and their corresponding absolute relative CCS error from the measurement (Figure 6.6B). The experimental CCS can be achieved, confirm that the ββ and γγ interaction between the two trimers. A possible arrangement of TNH hexamer represented by stacked spheres is also shown in Figure 6.6B.

6.3.4 Phenylglyoxal surface mapping of TNH helps determining models of αβγ trimer

Surface mapping involves covalently labelling solvent-exposed amino acids with specific functional groups. Thus, the structural information of the protein in native solution can be preserved post-digestion, in the form of modified peptides. Lysine is a target heavily used for covalent labeling in protein chemistry due to its reactive free amine group in the side chain.[224] There are, however, only four lysines in the whole TNH complex which makes lysine a less favorable target in our analysis. In TNH, arginine takes 95% percent of the basic residues (K and R) and hence a reagent that targets arginine is more favorable for our analysis. Phenylglyoxal (PGO) is reported to react with guanidinyl group.[237] To demonstrate its sensitivity and selectivity of exposed arginine versus buried arginine, glutamate dehydrogenase (GDH) hexamer with a known crystal structure was reacted with PGO. GDH is a homo-hexamer with D3 symmetry. Arginines on different chains but with the same residue number share a similar chemical environment. Among the 30 arginine on one chain, 20 of them have a percentage SASA (calculated solvent accessible surface area compared to SASA of Gly-
X-Gly) over 30% by GETAREA[238] software, which is a way of assessing a residue is exposed.[116] Table 6.3 shows all the SASA ratios calculated based on PDB 3MVO hexamer or monomer and the labeling results at different concentration of PGO (Y means labeled peptides have been identified, N means that no labeled peptides have been identified. NC means the arginine is not covered by any identified peptides).
Table 6.3. GDH arginine exposed information and labeling results.

<table>
<thead>
<tr>
<th>Res. No.</th>
<th>Hexamer SASA ratio (%)</th>
<th>Monomer Exp. %</th>
<th>PGO (7 mM) label results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>3</td>
<td>86</td>
<td>100</td>
<td>76</td>
</tr>
<tr>
<td>19</td>
<td>48</td>
<td>50</td>
<td>51</td>
</tr>
<tr>
<td>35</td>
<td>66</td>
<td>69</td>
<td>62</td>
</tr>
<tr>
<td>42</td>
<td>34</td>
<td>37</td>
<td>35</td>
</tr>
<tr>
<td>44</td>
<td>33</td>
<td>33</td>
<td>30</td>
</tr>
<tr>
<td>46</td>
<td>62</td>
<td>67</td>
<td>66</td>
</tr>
<tr>
<td>50</td>
<td>52</td>
<td>54</td>
<td>54</td>
</tr>
<tr>
<td>66</td>
<td>45</td>
<td>46</td>
<td>42</td>
</tr>
<tr>
<td>67</td>
<td>26</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>79</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>86</td>
<td>28</td>
<td>29</td>
<td>28</td>
</tr>
<tr>
<td>94</td>
<td>3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>146</td>
<td>28</td>
<td>32</td>
<td>29</td>
</tr>
<tr>
<td>147</td>
<td>35</td>
<td>38</td>
<td>38</td>
</tr>
<tr>
<td>174</td>
<td>42</td>
<td>41</td>
<td>37</td>
</tr>
<tr>
<td>211</td>
<td>11</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>217</td>
<td>19</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>261</td>
<td>28</td>
<td>26</td>
<td>25</td>
</tr>
<tr>
<td>265</td>
<td>54</td>
<td>49</td>
<td>52</td>
</tr>
<tr>
<td>338</td>
<td>61</td>
<td>64</td>
<td>60</td>
</tr>
<tr>
<td>363</td>
<td>33</td>
<td>35</td>
<td>37</td>
</tr>
<tr>
<td>396</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>403</td>
<td>35</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>419</td>
<td>59</td>
<td>46</td>
<td>64</td>
</tr>
<tr>
<td>439</td>
<td>15</td>
<td>35</td>
<td>25</td>
</tr>
<tr>
<td>459</td>
<td>48</td>
<td>47</td>
<td>47</td>
</tr>
<tr>
<td>462</td>
<td>44</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>466</td>
<td>56</td>
<td>52</td>
<td>51</td>
</tr>
<tr>
<td>478</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>491</td>
<td>50</td>
<td>37</td>
<td>56</td>
</tr>
</tbody>
</table>

| No. of labeled R that >30% SASA | 8   | 12  | 13  | 13  | 14  | 8   | 13  | 14  | 14  | 15 |
| No. of labeled R that <30% SASA | 0   | 0   | 0   | 1   | 0   | 0   | 0   | 0   | 1   | 2  |
| No. of R that are not labeled   | 19  | 15  | 14  | 13  | 10  | 19  | 14  | 13  | 12  | 10 |
Figure 6.7 plots the concentration-dependent number of labeled arginine as a percentage of the 17 covered (red dot) arginine that are considered to be solvent exposed by having over 30% relative SASA criteria. The results for GDH hexamer demonstrate a good sensitivity of the PGO reagent for labelling surface exposed arginine. Also, buried arginine can be modified by PGO at higher concentration of the reagent. The blue triangle curve in Figure 6.7 shows the falsely labeled rate curve: number of buried but labeled arginine divided by itself plus the number of unlabeled arginine. It shows PGO within 2mM has a good specificity towards exposed arginine on GDH.

![Graph](image)

Figure 6.7. Sensitivity and specificity of phenylglyoxal surface labeling on glutamate dehydrogenase hexamer.

A consensus result between the classical crystallographic approach and the PGO covalent labeling strategy was obtained for protein complex GDH. For example R396, which becomes buried upon the formation of the GDH complex (shows over 30% SASA in monomer but less than 30% SASA in hexamer), did not get labeled at all levels of
PGO concentrations studied. R86, however, also becomes buried upon the formation of the GDH complex and was not labeled by PGO at low concentration but got labeled in the most concentrated PGO sample. The fact shows that low concentration of PGO cannot get into the interface of subunits, but high concentration of PGO may alter the structure of protein complexes and thus can result in the labeling of interfacial arginine.

PGO was further used to probe the exposed arginine on TNH hexamer. One way to check the structural integrity of protein complex after labeling is to check whether the complexes have dissociated after labeling using native MS. As shown in Figure 6.8A, as the concentration of labeling reagent PGO increases, the hexamer abundance decreases and αβγ trimer and αβ dimer increases slightly. This clearly shows that covalent labeling has the drawback of perturbing the structure of analytes. Once subject the hexamer to dissociation, different labeling pattern for different subunits can be observed. The average number of modifications is shown in Figure 6.8B. Increasing concentration of PGO leads to more modifications on all the three subunits, as expected.
Figure 6.8. Higher concentration of the labeling reagent disrupts the structure of TNH hexamer to subcomplexes A). As the concentration of the labeling reagent increases, the percentage of the hexamer decreases (dash line), the percentage of the αβγ trimer (solid dark line) and αβ dimer (solid purple line) increase. By fragmenting the hexamer via either CID or SID, the monomer can be easily observed. The average modification number on the monomers is plotted versus the labeling reagent concentration in B).

Covalent labeling results were further used in screening structural data generated by computer modeling. “Protinfo PPC” web server was used to predict atomic level structure information for protein complexes from their amino-acid sequences based on
homology.\[229\] The input of all the three sequences – α, β, γ did not return any results. Since αβ dimer and αγ dimer has been observed as stable subcomplexes in either solution destabilization experiments or SID/IM experiments. The two dimers were modeled separately. The sequences of α and β subunits were submitted to perform modeling. 5 results were returned based on PDB ID 1ugq, 3hht, 1ugs, 2d0q and 1ahj. The submission of α and γ sequences returned 6 results based on structures with PDB ID 3hht, 2cz6, 1ahj, 1ugp, 1ugq and 1ugs. So the modeled structure for the trimer can be aligned by αβ and αγ from the same PDB ID structure: 3hht, 1ugq, 1ugs and 1ahj.

The four αβγ trimer models that were common to ProtinfoPPC searches conducted with both αβ and αγ (1ugq, 3hht, 1ahj and 1ugs) were each subjected to a 1000-step energy minimization and 10 ns equilibrium in water box or vacuum by NAMD software\[230\] to reduce any steric clashes and also test model stability. The calculated root-mean-square deviation (RMSD) of backbone atoms (relative to the initial trimer PDB) over the simulation time frame for each candidate is shown in Figure 6.9. Models based on 3hht and 1ahj maintain lower RMSD throughout the course of the simulation as compared to 1ugq and 1ugs, suggesting that they are more stable constructs.
RMSD of backbone atoms over simulation time for trimer model candidates

A). Simulation in a periodic water box (a layer of water 10 Å on edge and charge balanced by 0.15M NaCl); B) Simulation in vacuum

Figure 6.9. RMSD of backbone atoms over simulation time for trimer model candidates

Furthermore, the CCS of candidates in dynamic simulation were calculated by scaled projection approximation (PA) method and compared with the experimental αβγ trimer CCS from solution disruption. The relative CCS deviations of atomic coordinates of every 2 ns simulation in either periodic water box or vacuum were plotted in Figure 6.10. For all the candidate structures, the CCS in water equilibrium is greater than the
measured αβγ trimer CCS and the CCS in vacuum equilibrium is smaller than the measured αβγ trimer CCS. The candidate structure based on 3hht has the closest CCS to the experimental value, which is 7% higher in water equilibrium and 3% lower in vacuum equilibrium.

Figure 6.10. Collisional cross-section for trimer model candidates during simulation compared with the experimental value
Atomic coordinates are subjected to CCS calculation every 2 ns. A). Simulation in a periodic water box (a layer of water 10 Å on edge and charge balanced by 0.15M NaCl); B) Simulation in vacuum
Covalent labeling results from surface mapping experiments are further used to eliminate models that possess buried Arg residues that are modified even at the lowest concentration of PGO (Table 6.4). Since the PGO labeling experiments are performed in solution phase, the NAMD simulation employing a periodic water box were used for comparison with the experimental results. The model based on 3hht is most consistent with the experimental labeling results (Figure 6.11). For each arginine, the greatest %SASA among 2, 4, 6, 8, 10, 12 and 14 ns simulation frames of model based on 3hht is shown in Table 6.4. Arg residues that have %SASA over 30% are considered to be exposed. The table indicates that none of the buried Arg residues in the model based on 3hht is labeled with PGO at low concentration. There are some arginines that are predicted to be exposed by modeling and simulation that are not labeled by PGO at any concentration level, namely α134, β2 and γ26 (Table 6.4, Figure 6.11). We cannot exclude the possibility that the local environment around these residues, though accessible to solvent, occludes the larger, hydrophobic PGO label. Similar protection patterns are also observed in the GDH experiment. Another possibility is that Arg residues that are exposed on the surface of trimer but not labeled are involved in trimer-trimer interactions. The location of β2 and γ26 is highly likely to be involved in the trimer-trimer interface predicted in the SID/IM experiments.
Table 6.4. Table of PGO labeling results and extent of solvent exposer by trimer model based on 3HHT.

<table>
<thead>
<tr>
<th>Res. No.</th>
<th>SASA %&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Expo?</th>
<th>PGO(mM)</th>
<th></th>
<th>Res. No.</th>
<th>SASA %&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Expo?</th>
<th>PGO(mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 2 4 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1 2 4 7</td>
</tr>
<tr>
<td>a12</td>
<td>49</td>
<td>Y</td>
<td>NC NC NC NC</td>
<td>β2</td>
<td>69</td>
<td>Y</td>
<td>N</td>
<td>N N N N</td>
</tr>
<tr>
<td>a14</td>
<td>79</td>
<td>Y</td>
<td>NC NC NC NC</td>
<td>β3</td>
<td>64</td>
<td>Y</td>
<td>Y Y Y Y</td>
<td></td>
</tr>
<tr>
<td>a15</td>
<td>64</td>
<td>Y</td>
<td>NC NC NC NC</td>
<td>β5</td>
<td>37</td>
<td>Y</td>
<td>NC NC NC NC</td>
<td></td>
</tr>
<tr>
<td>a19</td>
<td>38</td>
<td>Y</td>
<td>NC NC N N</td>
<td>β5</td>
<td>7</td>
<td>N</td>
<td>N N N Y</td>
<td></td>
</tr>
<tr>
<td>a41</td>
<td>58</td>
<td>Y Y Y Y Y Y</td>
<td>β5</td>
<td>7</td>
<td>Y</td>
<td>N</td>
<td>N N N N</td>
<td></td>
</tr>
<tr>
<td>a49</td>
<td>35</td>
<td>Y N N N Y</td>
<td>β72</td>
<td>24</td>
<td>Y</td>
<td>N N N N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a53</td>
<td>40</td>
<td>Y Y Y Y Y</td>
<td>β89</td>
<td>86</td>
<td>Y</td>
<td>Y Y Y Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a61</td>
<td>50</td>
<td>Y N N Y Y</td>
<td>γ3</td>
<td>62</td>
<td>Y</td>
<td>Y Y Y Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a73</td>
<td>80</td>
<td>Y Y Y Y Y</td>
<td>γ4&lt;sup&gt;b&lt;/sup&gt;</td>
<td>61</td>
<td>Y</td>
<td>N N N N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a88</td>
<td>42</td>
<td>Y Y Y Y Y Y</td>
<td>γ12</td>
<td>41</td>
<td>Y</td>
<td>N N Y Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a126</td>
<td>2</td>
<td>N</td>
<td>N N N N N</td>
<td>γ14</td>
<td>48</td>
<td>Y</td>
<td>Y Y Y Y</td>
<td></td>
</tr>
<tr>
<td>a128</td>
<td>14</td>
<td>N N N N N</td>
<td>γ23</td>
<td>9</td>
<td>N</td>
<td>N N N N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a131</td>
<td>30</td>
<td>N N N N N</td>
<td>γ26</td>
<td>78</td>
<td>Y</td>
<td>N N N N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a134</td>
<td>61</td>
<td>Y N N N N</td>
<td>γ29</td>
<td>30</td>
<td>N</td>
<td>N N N N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a151</td>
<td>38</td>
<td>Y Y Y Y Y Y</td>
<td>γ50&lt;sup&gt;b&lt;/sup&gt;</td>
<td>83</td>
<td>Y</td>
<td>N N N N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a160</td>
<td>6</td>
<td>N NC N N N</td>
<td>γ51</td>
<td>74</td>
<td>Y</td>
<td>N N N N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a166</td>
<td>60</td>
<td>Y Y Y Y Y Y</td>
<td>γ52</td>
<td>81</td>
<td>Y</td>
<td>Y Y Y Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a167</td>
<td>41</td>
<td>Y Y Y Y Y Y</td>
<td>γ56</td>
<td>62</td>
<td>Y</td>
<td>Y Y Y Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a182</td>
<td>58</td>
<td>Y Y Y Y Y Y</td>
<td>γ64</td>
<td>11</td>
<td>N</td>
<td>NC NC NC NC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a185</td>
<td>27</td>
<td>N N N N N</td>
<td>γ68</td>
<td>76</td>
<td>Y</td>
<td>Y Y Y Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>γ95</td>
<td>45</td>
<td>Y Y Y Y</td>
<td></td>
</tr>
</tbody>
</table>

a. Due to the dynamic nature of molecules in solution, the % SASA is the biggest number chosen from 2 ns, 4 ns, 6 ns, 8 ns, 10 ns, 12 ns and 14 ns simulation results.

b. Arginines on γ4 and γ50 are not labeled, may due to steric effect when the PGO labeling of the adjacent arginines on γ3 and γ52, respectively.
Figure 6.11. Modeled TNH αβγ trimer structure based on homolog PDB 3hht and equilibrated in periodic water box for 10ns. Subunits α, β and γ are in red, blue and green respectively. The exposed arginines are in stick representation. Labeled arginines are in cyan color. Arginine β2 and γ26 in gray are not labeled and thus postulated to be involved in trimer-trimer interaction to form hexamer.

6.3.5 ZDOCK helps determining models of TNH hexamer $\alpha_2\beta_2\gamma_2$

After predicting the αβγ structure and the trimer-trimer interfacial residues β2 and γ26, efforts were put into assembling trimers to TNH hexamer $\alpha_2\beta_2\gamma_2$. Two identical modeled TNH αβγ trimer structures (Figure 6.11) based on homolog PDB 3hht and equilibrated in periodic water box for 10ns were subjected to docking using the program ZDOCK [239] (http://zdock.umassmed.edu/) with β2 and γ26 Arg residues selected as contacting residues. Ten hexamer models were returned and further subjected to a 1000-step energy minimization, followed by 10 ns water equilibrium. Hexamer 1, 2, and 3 are the most stable structures as evidenced by their low RMSD (< 3.5 Å, Figure 6.12)
throughout the course of the simulation. The calculated CCS of the energy-minimized (0 ns) of hexamers 1, 2, and 3 structures are within 5% error compared with the experimental hexamer CCS (Figure 6.13). Furthermore, the %SASA of γ26 Arg is below 30% among all the three structures, indicating that it is buried upon hexamer formation. Therefore, the surface mapping experiments together with homology modeling predicts three possible TNH hexameric structures (Figure 6.14).

![RMSD of backbone atoms over simulation time for TNH hexamer model candidates](image)

Figure 6.12. RMSD of backbone atoms over simulation time for TNH hexamer model candidates Simulation in a periodic water box (a layer of water 10 Å on edge and charge balanced by 0.15M NaCl); Hexamer 7 and 8 failed simulation due to instability
Atomic coordinates are subjected to CCS calculation every 2 ns for simulation in a periodic water box (a layer of water 10 Å on edge and charge balanced by 0.15M NaCl). Hexamer 7 and 8 failed simulation due to instability. Hexamer 5 and Hexamer 10 do not have β-β and γ-γ interaction between the two trimers, thus eliminated.

Figure 6.13. Collisional cross-section for TNH hexamer model candidates during simulation compared with the experimental value.
Figure 6.14. Possible TNH hexamer structures (Hexamer 1 in green, 2 in magenta and 3 in blue) with docking of αβγ trimer structure based on homolog PDB 3hht and equilibrated in periodic water box for 10ns.

6.4 Conclusions

As shown in Figure 6.15, multiple mass spectrometry based approaches have been applied to refine the structure of hexameric TNH. Solution-phase organic solvent disruption coupled with native MS experiments highly suggest that TNH is a dimer
composed of $\alpha \beta \gamma$ trimers, and that $\alpha$-$\beta$ interact strongly. More subcomplexes containing contact information of TNH were generated in surface-induced dissociation coupled with ion mobility mass spectrometry (SID/IM) experiments. The results reveal that $\gamma$-$\gamma$ subunits are primarily in contact between the two trimers and weak interactions between two $\beta$ may also exist. But no clear evidence has shown that the two $\alpha$ subunits can interact. Collisional cross-section measured from IM experiments were used as constraints for postulating arrangement of the subunits represented by coarse-grained spheres. Covalent labeling together with protein complex homology modeling is utilized to propose possible atomic level structures of TNH hexamer. The combined results from different mass spectrometry approaches can achieve a more thorough structural analysis towards protein complexes.

Figure 6.15. A summary of TNH structure elucidation by multiple mass spectrometric techniques
Chapter 7. The Effect of Charge and His-tag on the Gas Phase Dissociation of Toyocamycin Nitrile Hydratase Hexamer Probed by CID and SID

7.1 Introduction

Numerous proteins exist and function as complexes, in which protein subunits assemble via non-covalent interactions.[240] The characterization of these protein complexes including their composition, subunit connectivity and arrangement is very important in structure biology. Mass spectrometry (MS) is a recognized approach for characterizing proteins and their complexes.[182, 183, 184, 241] The native MS strategy makes it possible to transfer proteins in neutral aqueous buffers to the gas phase.[242, 243] The transmission of these large ions in the mass spectrometer is drastically improved by increasing the pressure in the early pumping stages.[20, 244, 245] The analysis of these large ions benefits from the ability to measure high m/z on the time-of-flight (TOF) [246], FT-ICR [247], and Orbitrap [28] mass analyzers and the coupling with ion mobility (IM) further extends the analysis to the conformation of these large ions.[91, 202]

In order to study the building blocks of protein complexes, solution phase destabilization has been employed. With careful adjustment of organic percentage or ionic strength sub-complexes can be generated and measured in a mass spectrometer.[205, 206] In order to probe the constituents and architecture of the complexes, especially in a mixture, gas phase disassembly methods are necessary due to its mass selection.
The most widely used gas-phase dissociation method is collision induced dissociation (CID), which involves an analyte going through multiple collisions with a low mass inert gas. A complication with CID is that the products do not always reflect the contacts in the structure of the complexes. To be more specific, the “typical CID process” dominates with one subunit unfolding and taking away about half of the charges from the precursor complexes and leaving the other half of charges behind on its complementary partner.[36, 209, 210, 211] An alternative gas-phase disassembly method, surface induced dissociation (SID), which involves an analyte colliding with a surface, has been shown from previous work in the Wysocki group to generate product ions that reflect the native quaternary structure of the protein complexes, rather than ubiquitous ejection of a monomer as “typical CID process”. [43, 44, 45, 46, 47, 48, 186, 187, 188]

The effect of charge has been extensively studied for dissociating protein complexes. Pagel and coworkers discovered that C-terminal fragments were released from low charge state TTR during the CID process.[190] Zoe Hall and coworkers have shown that increased charges on some protein complexes can lead to an atypical CID dissociation pathway by CID, which can provide subcomplex information.[38] By IM analysis, they found the lower charge states of complexes that have cavities experience more compaction.[228] Zhou et.al. have shown that reduced charge states preserve of integrity of the protein complexes in SID, thus leading to better characterization of the protein complexes.[48] Most of the charge manipulation experiments have been done with homo-oligomers. Studying a hetero-oligomer dissociation behavior contributes to a better understanding of the dissociation process, because different subunits possess
different mass, and the change in ejection order as precursor charge state changes can be readily revealed.

His-tagging has been popular for decades as a protein purification tool.[248] It involves expressing the recombinant proteins with six histidines and a cleavable linker to one terminus of the protein. The histidine sequence has strong affinity toward nickel beads and thus achieves enrichment of the recombinant protein. The tag may be or may not be removed before the proteins are studied. It has been shown that His-tagging typically has no significant influence on the structure of native proteins.[249] Some simulation work has been done on the influence of the charged site on controlling the dissociation channels. Sarah Fegan and Mark Thachuk have shown that the N-terminus leads the unraveling of TTR tetramer by molecular dynamics (MD) simulations.[250] It suggests that experimentally the modification of the N-terminus of one protein subunit in a protein complex may change the dissociation behavior. Adding a His-tag is also an easy way to change proton affinity of protein N-terminus.

In the previous chapters, efforts have been put into elucidation the structure of toyocamycin nitrile hydratase (TNH), an enzyme complex that is comprised of two α, two β and two γ subunits (MW α 21190 Da, β 9974 Da, γ 11444 Da). We also have a His-tagged TNH hexamer with a His-tag on the γ subunit (MW His-γ 13608 Da). In addition, we can manipulate the hexamer charging by solution additives. Thus, the influence of charge and His-tag on dissociation of the TNH hexamer is studied here, with the aim of better understanding gas phase dissociation of protein complexes in general.
7.2 Experimental

7.2.1 Reagents and surface mapping reactions

Untagged or His-tagged TNH, composed of two α, two β and two γ subunits, was prepared by Micah Nelp in Dr. Bandarian’s lab at the University of Arizona. The preparation steps are elaborated in the previous chapter. His-tagged TNH has His-tag on the N-terminus of the γ subunit. The sequences of the subunits are shown in Table 7.1.

Table 7.1. Sequence information of TNH α, β, γ subunit

<table>
<thead>
<tr>
<th>Subunit</th>
<th>Sequence Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>ToyJ (α) (Average molecular weight: 21085)</td>
<td>STEHVLPIAARVRLEERVIAAGLVTDEQLDTILEHNLSRATPFNGARLVARAW TSPDFRDLLLGEPAALREMGLDGDLADDEHLRVVANTPTVHNVVVCTLSCYP VLLLPSPWSWYKSDAYRARVREPRAVLAEFTVLPAPAEDVVRVWASAEARYM VLPRRPAGTEGLDEELAARVTRAGLGTAPV</td>
</tr>
<tr>
<td>ToyL (β) (Average molecular weight: 9974)</td>
<td>ARLNDGGTGFGYSIPMDGAEPPHPWRHDWEARVFAVLAGAMGVTASELRTDEERVPPPNDYLAASYYERVLMGMELLDKEDGIPSRT</td>
</tr>
<tr>
<td>ToyK (γ) (Average molecular weight: 11444)</td>
<td>GERRFSVGDPVRVRPAPPHTRVPRYVRGHLGHVTVQPPCPLDPDDVARRRDPFRVLPYTVRFARAARLWGSGLWLIDLWECYLEPAAETARGLGETS</td>
</tr>
<tr>
<td>His-tagged ToyK (His-γ) (Average molecular weight: 13607)</td>
<td>GSSHHHHHHSSGLVPRGSHMGERRFSVGDPVRVPADPPHHTRVPRYVRGHLGHVTVQPPCPLDPDDVARRRDPFRVLPYTVRFARAARLWGSGLWLIDLWECYLEPAAETARGLGETS</td>
</tr>
</tbody>
</table>

The sample was buffer exchanged into 100 mM ammonium acetate (AmAc) via Micro Bio-Spin™ P-6 Gel Columns (Bio-Rad, Hercules, CA). The final concentration of the protein complex was adjusted to 20 μM. Ammonium acetate solid and triethylammonium acetate (TEAA, ~1.0 M in H₂O) were purchased from Sigma-Adrich (St. Louis, MO) and dissolved in ultrapure water (Sartorius Corporation, Bohemia, NY) to a concentration of 100 mM. Charge reduction is performed by adding 20% volume of 100 mM TEAA to the protein complex solution. 100mM ethylenediammonium diacetate
(EDDA) was prepared by mixing ethylenediammonia (Sigma-Adrich, St. Louis, MO) with glacial acetic acid (Sigma-Adrich, St. Louis, MO) in ultrapure water and adjust the pH to 7.5. Minor charge reduction is performed by adding 20% volume of 100 mM EDDA to the protein complex solution.

The surface mapping reaction of arginines in TNH was performed at room temperature for 14 hrs. Several concentrations (1 mM-7 mM) of phenylglyoxal (PGO, Sigma-Adrich, St. Louis, MO) were chosen to react with 8uM of TNH hexamer in 25 mM HEPES (pH 7.5). The reaction was stopped by depleting PGO through buffer exchanging to 100 mM AmAc via Micro Bio-Spin™ P-6 Gel Columns. Diethylpyrocarbonate (DEPC, Sigma-Adrich, St. Louis, MO) surface mapping, or the addition of a carboxyethyl group to H, K, C, S, T, Y was performed at 37 °C, 150 rpm in a thermomixer for 1 min. DEPC was diluted to 100 mM stock in acetonitrile and further diluted by 100 mM AmAc and applied quickly to 8 uM of TNH hexamer in 25 mM HEPES (pH 7.5). The concentration of DEPC used was from 0.2 mM to 7 mM. The reaction was quenched by adding 5 mM imidazole. The labeled protein was purified by 10k MWCO Amicon® Ultra 0.5 mL filters (EMD Millipore, Billerica, MA) with 100 mM AmAc prior to native MS and bottom-up experiments.

7.2.2 Instrumentation for native mass spectrometry analysis

Samples were nanoelectro-sprayed into a SYNAPT G2s mass spectrometer (Waters Corporation, Manchester, UK) with a custom SID device incorporated before the ion mobility cell, the detail of which has been described previously.[50] The trap traveling wave ion guide (TWIG) was truncated to accommodate the SID cell. Under full
scan mode or CID mode, the voltages applied to the ten lenses of the SID cell were adjusted to guide the ions to fly through without colliding with the surface. While in SID mode, the ions were steered towards the surface by the voltages on the lenses. All the lenses are tuned to achieve better precursor transmission and products collection. Proteins solutions were filled in glass capillaries pulled on a P-97 micropipet puller (Sutter Instruments, Hercules, CA) and ionized with 1.1kV voltage applied on a platinum wire inserted to the capillaries. Typical instrument settings were: sampling cone 20 V, source offset 10 V, source temperature 25 ºC, trap gas flow 4 mL/min, helium cell gas flow 180 mL/min, IMS gas flow 60 mL/min; trap DC entrance 0 V, trap DC bias 45 V, trap DC -2 V, trap DC exit 0 V, trap wave velocity 160 m/s, trap wave height 4 V; IMS DC entrance 10 V, helium cell DC 25 V, helium exit -5 V, IMS bias 5 V, IMS DC exit 0 V, IMS wave velocity 300 m/s, IMS wave height 20 V; transfer DC entrance 2 V, transfer DC exit 15 V, transfer wave velocity 50 m/s, transfer wave height 4 V.

7.2.3 Bottom-up approach for surface mapping modification localization

The bottom-up approach to identify arginine modifications was performed by in-solution pepsin digestion of PGO labeled samples and in-solution trypsin digestion of DEPC labeled samples, with LC-MS/MS analysis on a linear ion trap mass spectrometer (Velos Pro, ThermoFisher Scientific Inc., Waltham, MA) coupled to a UPLC (nanoACQUITY, Waters Corporation, Manchester, UK). Pepsin digestion was performed first by adjusting the pH of the labeled sample solution with 1 N HCl (Sigma-Adrich, St. Louis, MO) to below 4, followed by adding pepsin (Promega, Madison, WI) at 1:10.
enzyme: protein ratio. The digestion was performed on a 37 °C, 150rpm in a thermomixer for 7 hrs. The reaction was stopped by heating on a 95 °C block for 5min. Trypsin digestion was performed first by unfolding the protein in 0.03% ProteaseMAX surfactant (Promega, Madison, WI), followed by adding trypsin gold (Promega, Madison, WI) at 1:10 enzyme: protein ratio. The digestion was performed on a 37 °C, 150rpm thermomixer for 3 hrs. To quench the digestion, 0.5% TFA (Fisher Scientific, Pittsburgh, PA) was added to the digestion. Prior to LC-MS/MS injection, the sample was centrifuged at 10 kg for 10 min. The amount of injection was 0.4 μg.

LC-MS/MS data analysis was performed by SEQUEST HT in Proteome Discoverer software (ThermoFisher Scientific Inc., Waltham, MA). Some searching parameters are listed here: precursor mass ranges from 400 to 5000 Da, peptide length ranges from 4 to 144, precursor mass tolerance is 1.5 Da, and fragment mass tolerance is 0.8 Da. For PGO labeled TNH data, no enzymes are specified. Dynamic modifications are methionine oxidation (M+15.995), phenylglyoxal reacted arginine (R+116.026 or 134.0368). For DEPC labeled TNH data, trypsin is specified as enzyme. Dynamic modifications are methionine oxidation (M+15.995), carboxyethylation on histidine, lysine, cysteine, serine, threonine and tyrosin (H, K, C, S, T, Y+ 72.021129). Percolator with 1% FDR was used to validate the identities.
7.3 Results and Discussions

7.3.1 Surface mapping suggests that His-tag is on the protein surface and does not significantly alter the quaternary structure of TNH

Covalent labeling probes amino acids that are exposed to solvent, thus protein structural information in the solution phase can be preserved in the form of modified peptides in the gas phase, which is readily measured by mass spectrometry. It is reported that diethylpyrocarbonate (DEPC) is a promising reagent to probe the surfaces of proteins. It can react with histidine, lysine, cysteine, serine, threonine, and tyrosine, which make up to 30% of the residues in a protein on average.[116, 251] For the particular purpose of mapping the surface of the TNH protein, the target residues of DEPC make up 17%, 14%, 19%, and 25% of subunit α, β, γ, and His-γ residues respectively. To add more confidence to the surface mapping experiments, another labeling reagent, phenylglyoxal (PGO) was also used. PGO reacts with arginine side chains.[237] Arginine was chosen for modification because 95% of the basic residues (R or K) in TNH are arginine. As percentage of all the residues, arginine makes up 10%, 8%, 14% and 12% for α, β, γ, and His-γ subunits respectively.

To check the structural integrity of the protein complexes after labeling, native MS analyses of the labeled samples were performed. At the lowest concentration of labeling reagent, only hexamer peaks can be observed. Increased concentrations of DEPC and PGO gradually lead to dissociation of the hexamer. When the hexamer is subjected to CID or SID dissociations, subunit peaks with different extents of modification could be observed. The relationship of the average number of labels added on each subunit to the
concentration of the labeling reagent is shown in Figure 7.1. The average DEPC label number of His-tagged γ is much higher than wild type γ. The labeled samples are subjected to bottom-up MS analysis to locate the labeled sites.

Figure 7.1. Dose-dependent average number of labels of different subunits. a) wild type TNH by PGO, b) wild type TNH by DEPC, c) His-tagged TNH by PGO, d) His-tagged TNH by DEPC. The fact that His-tag was heavily labeled as shown in d) indicates that His-tag is on the surface of the protein complex, it did not influence the interfaces between subunits.

Figure 7.2 shows the labeling results on the sequence of the wild type TNH (left) and His-tagged TNH (right). Blue shaded residues are targets of DEPC labeling and yellow shaded residues are targets of PGO labeling. The experimentally DEPC labeled residues are squared out in black and PGO labeled residues squared out in red. The dash
square means the labeling occurred only at higher concentration of the labeling reagents.

Wild type TNH and His-tagged TNH have similar surface mapping results (except for the His-tag region) indicating a similar structure between the two. The fact that the His-tag was heavily DEPC labeled suggests the His-tag may flank outside of the TNH protein complex, thus the tag does not disrupt the formation of the hexamer.

Figure 7.2. Surface mapping results. Sequence information of wild type TNH is shown on the left and His-tagged TNH on the right.

Blue shaded residues are targets of DEPC and yellow shaded residues are targets of PGO. The experimentally DEPC labeled residues are squared out in black and PGO labeled residues squared out in red. The dash square means the labeling occurred only at higher concentration of the labeling reagents. Similar labeled sites between wild type TNH and His-tagged TNH (except for the His-tag) can be observed.
7.3.2 His-tagging or charge states did not influence SID behavior of TNH significantly

+20~+18 charged TNH hexamer was observed when directly sprayed from AmAc solution. By adding EDDA, the charge states reduced to +17~+15. By adding TEAA, the charge states further reduced to +15~+13. The same charge state distribution and charge shift by solution additives can be observed for His-tagged TNH hexamer. Hexamer precursor with +19 to +14 were each individually isolated and fragmented by SID.

As shown in energy resolved SID plots of Figure 7.3, different charge states (a,d: +19; b,e: +17; c,f: +14) of the wild type TNH hexamer (a, b, c) and the His-tagged TNH hexamer (d, e, f) break down to trimer products (black square) containing one α subunit, one β subunit and one γ subunit or its His-tagged counterpart at lower SID energies. At higher SID energies, the trimer further dissociates to αβ dimer (purple up triangle) and γ or His-γ monomer (Olive green down triangle). Also, the hexamer can access other dissociation pathways to form αβ dimer and αβγ2 tetramer (pink hollow up triangle), α monomer (red dot) and αβγ2 pentamer, β monomer (blue square) and α2βγ2 pentamer, γ and α2βγ pentamer at higher SID energies. (Due to the low intensity of the pentamers, that result from instrument discrimination, less effective transmission, and/ or detection of high m/z products, pentamers are not shown in the graphs)
Figure 7.3. Energy-resolved SID results of (a)+19, (b)+17, (c)+14 TNH wild type hexamer; (d)+19, (e) +17, (f)+14 His-tagged TNH hexamer.

Why is SID not significantly affected by His-tagging? SID is a single-step and more energetic activation method that can access dissociation pathways that are more authentic to the structure of the non-covalent protein complex. In a lot of cases, when the structures of non-covalent protein complex are known by either X-ray crystal structure or NMR results, SID has shown to break the weakest interface first to reveal the subunit interaction faithfully.[43, 44, 45, 46, 47, 48, 186, 187, 188] Although the high resolution structural information of wild type TNH or His-tagged TNH os not available, the labeling experiments shown in the previous section indicate that there are no significant differences between the contacts within the complex, thus the SID behaviors are similar.
for the wild type TNH and His-tagged TNH. In the charge states studied (+19-+14), SID behavior is not affected by the change of charges on the hexamer precursor. We ascribe this to the fact that there are no significant structural changes among the charge states studied for TNH. It is also manifested by no major collision cross section (CCS) changes among different charge states in collision-induced unfolding experiments. (Figure 7.4a-f: compact wild type TNH CCS 48.7±0.3 nm², Figure 7.4g-l: compact His-tagged TNH CCS 50.3±0.4 nm² with the smallest activation energy)
Figure 7.4. Unfolding of wild type (a-f) and His-tagged (g-l) THN hexamer with different charges by CID.
7.3.3 His-tagging or charge states influence CID behavior of TNH significantly

+19 to +14 wild type TNH hexamer or His-tagged TNH hexamer was isolated and fragmented by CID. The spectra at low collision energy and high collision energy are shown in purple and black, respectively, in Figure 7.5. Comparison of the monomer ejection of wild type TNH (left) and His-tagged TNH (right) reveals enhanced ejection of the His-tagged gamma subunit versus that of its untagged counterpart. Further, the change of the monomer products abundance with an increase of collision energy can be plotted. The energy-resolved CID results are shown in Figure 7.6.
Figure 7.5. CID spectra at low collision energy (purple spectra) and high collision energy (black spectra) of (a)+19, (b)+18, (c)+17, (d)+16, (e)+15, (f)+14 TNH wild type hexamer; (g)+19, (h) +18, (i) +17, (j) +16, (k)+15, and (l)+14 His-tagged TNH hexamer.
Figure 7.6. Energy-resolved CID results of (a)+19, (b)+18, (c)+17, (d)+16, (e)+15, (f)+14 TNH wild type hexamer; (g)+19, (h)+18, (i)+17, (j)+16, (k)+15, (l)+14 His-tagged TNH hexamer.
As Figure 7.6a, b and Figure 7.7a and d illustrate, for +19, +18 wild type TNH hexamer, at low energy CID, the β subunit is most readily ejected. By increasing the CID energy, the α subunit is the most abundant product. The γ subunit is greatly suppressed at any CID energy. As shown in the right of Figure 7.6, His-tagged hexamer has a different monomer percentage and ejection order from wild type hexamer. For +19 and +18 high charge states (Figure 7.6g and h, Figure 7.7b and e), the three monomers have about same energy onset. At higher CID energy, the α subunit is dominant, followed by His-tagged γ subunit.

Figure 7.7. Tandem mass spectrum of +19 wild type(a,d), His-tagged (b,e) and chimera (c,f) hexamer by various collision energy. Zoom in monomers (m/z=1000-2500, yellow shadow) are shown in d, e and f. His-tagged γ subunit is preferentially ejected versus wild type γ subunit.
For +17 and +16 wild type TNH hexamer (Figure 7.6c and d, Figure 7.8a and d), the α subunit is ejected preferentially as the CID energy increase. The γ subunit is still suppressed at any CID energy. The β subunit is slightly more abundant than the γ subunit. In contrast, for +17 and +16 His-tagged hexamer (Figure 7.6i and j, Figure 7.8b and e), the His-γ subunit is ejected first at low CID energy, as shown in Figure 7.8e. At medium CID energy, His-γ is the most abundant monomer product, followed by α subunit. At higher CID energy, the abundance are in the order of α> His-γ> β.

![Tandem mass spectrum of +17 wild type(a,d), His-tagged (b,e) and chimera (c,f) hexamer by various collision energy. Zoom in monomers (m/z=1200-2700, yellow shadow) are shown in d, e and f. His-tagged γ subunit is preferentially ejected versus wild type γ subunit](image)

For +15 wild type TNH hexamer (Figure 7.6e), the β subunit is preferentially ejected at low CID energy (<2.2 keV) while the α subunit and the γ subunit have similar
abundance. At higher CID energy (>2.2 keV), the α subunit takes over the high percentage while the β subunit and the γ subunit have similar abundance. For +15 His-tagged hexamer (Figure 7.6k), the β subunit eject first at low CID energy. At high CID energy, these three monomers are of the same abundance.

For +14 wild type TNH hexamer (Figure 7.6f, Figure 7.9a and d), the ejection order for the three subunits becomes β, γ, α as CID energy increasing. Charge states affect the dissociation pattern of wild type TNH drastically. For the lowest charge state, +14 His-tagged hexamer (Figure 7.6f, Figure 7.9b and e), the β subunit ejects first and dominates along with CID increasing. Both the α and the His-γ subunit are suppressed.

![Figure 7.9. Tandem mass spectrum of +14 wild type (a,d), His-tagged (b,e) and chimera (c,f) hexamer by various collision energy. Zoom in monomers (m/z=1100-4000, yellow shadow) are shown in d, e and f.](image)
It is widely observed that CID always results in highly charged monomer ejection. A lot of experiments and theoretical calculations have demonstrated this asymmetric charge-to-mass charge partitioning involves charge enrichment on one subunit and Coulomb repulsion-induced unfolding. Jurchen and Williams[36] have provided a convincing evidence for monomer unfolding by the experiments involving cytochrome c dimer with surface crosslinking on each subunit and α-lactalbumin dimer with 4 disulfide bonds on each monomer (hard to unfold) which showed symmetric dissociation while uncross-linked cytochrome c dimer and disulfide-reduced α-lactalbumin dimer (easy to unfold) showed asymmetric dissociation. A follow-up experiment[37] on α-lactalbumin dimer with one monomer disulfide-intact and another monomer disulfide-reduced resulted in the reduced monomer carried away more charges. This further proved that one subunit unfolds upon CID activation. As for multimeric protein complexes, Benesch and coworkers[209] established a correlation between the relative surface area of the unfolded monomer to the complementary (n-1)-mers and the relative charge on the products of CID. Hall et al.[38] summarized data for several non-covalent protein oligomers to show that high subunit flexibility and stronger subunit interaction promotes unfolding and thus asymmetric charge-to-mass partitioning. Fegan and Thachuck[211] developed a charge moving algorithm to perform MD simulation of the gas-phase dissociation of transthyretin tetramer. They suggested that charge migration and energy activation are necessary for unfolding to occur. Asymmetric dissociation has also been observed by blackbody infrared radiative dissociation, Klassen’s group [210, 252] proposed that Coulomb repulsion contributed to dissociation and quantified its influence on thermodynamic coefficients.
The major CID difference of wild type TNH and His-tagged TNH is the ejection of His-tagged $\gamma$. The His-tagged $\gamma$ is more abundant than the wild type $\gamma$ as the monomer is ejected from the respective hexamers with higher charge states (>16). The charges on the monomer products are closely examined (Figure 7.10).

![Figure 7.10](image)

Figure 7.10. Average charges on monomer products by a) CID of different charge state hexamer precursor; b) SID of different charge state hexamer precursor. The solid symbols represent monomer products from wild type TNH while the open symbols represent monomer products from His-tagged TNH hexamer. (Other oligomer products in SID are not shown here)

The monomers generally have fewer charges by SID than by CID, consistent with previous results showing that SID is a more charge-to-mass symmetric dissociation method.[44] In CID, the charge on the $\alpha$ subunit (red dots) almost increase linearly with slope 1 with the increase of charges on the precursor. The charge on the $\beta$ subunit (blue square) remains between 6 and 7 as the precursor charge change. The charge on the $\gamma$ subunit (green down triangle) increases slightly as the charge on the precursor increases. The charging situation remains the same for the $\alpha$ and $\beta$ subunits coming from His-tagged hexamer with 0.4%-7% charge decrease and -2%-4% variation for $\alpha$ and $\beta$ respectively.
depending on the precursor charge states. But the His-tagged γ subunit holds more charges than its unHis-tagged counterpart, with a 6%-32% charge increase depending on the precursor charge states. The His-tag may help the γ subunit take more charges and unravel in the CID process.

Another explanation for the easier ejection of His-tagged γ subunit is that the His-tag lies on the surface of the complex, as shown by the results of the surface mapping experiments. It can be peeled away easily but does not introduce much disruption to the rest of the complex. Erba and coworkers investigated the CID of two dodecamer protein complex: HSP16.9 and SP-1, with the former one having “typical CID behavior” and latter one having atypical CID behavior.[253] One reason why HSP16.9 has monomer ejection in CID, is that for all the subunits, the C-terminus lies on the surface, which more easily initiates unfolding by charge migration.[209] The N-terminal His-tag on the surface of the γ subunit can serve in the same role: a terminus that can take on charge which also increases the flexibility to unfold.

To further investigate the role of the His-tag, a chimera hexamer has been made by mixing wild type hexamer and His-tagged hexamer. This chimera hexamer contains two α subunits, two β subunits, one γ subunit and one His-γ subunit. The Figure 7.11 shows an SID mobiligram of +19 chimera hexamer. The existence of αβγHis-γ tetramer confirmed the correct isolation of the chimera hexamer. Two different trimers, αβγ and αβHis-γ, can be clearly seen as the major products.
As shown in Figure 7.12a and b (spectra shown in Figure 7.7c and f, Figure 7.8c and f), upon CID, at precursor charge states +19 and +17, His-γ is more readily ejected at low energy and more abundant than γ thought the entire CID energy scale. But when the charge is limited, as shown in Figure 7.12c (spectra shown in Figure 7.9c and f), where the chimera precursor has +14 charges, the ejection of the γ subunit is more preferential than His-γ subunit. This shows that the His-tag can grasp protons or add in flexibility to help the γ subunit unraveling when the precursor complex has excessive charges on it.
Figure 7.12. Energy-resolved CID results of (a)+19, (b)+17, (c)+14 TNH chimera hexamer, which is composed of two α, two β, one γ and one His-tagged γ subunits. The very left stacked-column graphs represent each monomer percentage takes of the total monomers abundance as CID energy increases. Red is α, blue is β, olive green solid is γ and olive green mesh is His-γ. The collision cross section of the (d)+19, (e)+17, (f)+14 chimera hexamer changes as CID energy change is shown in the right-side heat maps. Red color shows high survival yield and blue color shows low survival yield. Transitional colors show different level of survival yield in reference to the color bar.

The drastic monomer ejection patterns alteration as the precursor charge changes may due to the change of relative energy barrier of compaction, unfolding and different dissociation pathways. [48, 190, 228] CCS profile for wild type TNH (Figure 7.4a-f) and His-tagged TNH (Figure 7.4g-l) upon CID has the same trend as chimeric TNH (Figure 7.12d-f). The +19 hexamer exhibits extensive CCS enlarging upon CID activation (Figure 7.12d). Both the +17 and +14 experience the compaction before unfolding. The
compaction of +14 hexamer continues over a larger CID energy span than for +17 hexamer, which almost complete dissociation then taking over. (Figure 7.12e and f)

7.4 Conclusions

Toyocamycin nitrile hydratas (TNH) is a non-covalent hexameric protein complex composed of three different subunits – α, β, and γ. His-tagged TNH has the His-tag on the γ subunit, but performs the same biological function as the wild type TNH. Here, wild type TNH and His-tagged TNH are shown to have similar subunit interaction by surface mapping using diethylpyrocarbonate and phenylglyoxal with bottom-up mass spectrometry analysis. Further, wild type TNH and His-tagged TNH were subjected to native mass spectrometry analysis in a modified quadrupole/ ion mobility/ time-of-flight mass spectrometer. By surface induced dissociation (SID), the major dissociation product of the hexamer is trimer composed of one copy each of the three subunits, regardless of the precursor charge state and His-tagging. SID results matches with the fact that no major structure changes with different charge states or His-tagging. In contrast, upon dissociation by collision induced dissociation (CID), the dissociation is dominated by ejection of monomers. Manipulating the charges on the complex changes the order of different subunit ejection. Together with collisional cross-section examination, the results suggest that charge states influence energy barriers for unfolding, compaction and different dissociation pathways. Also the His-tag helps the tagged subunit with increased ejection. Two processes may contribute to the ejection of the His-tagged subunit. The
first is that His-tag helps to sequester charge. The second is that the His-tag flanks on the surface of the subunit and facilitates the unfolding and charge migration.
Chapter 8. SID/IM/MS as a Tool for Analysis of Composition and Subunit Organization of a Bacterial Multicopper Oxidase

8.1 Introduction

Manganese (Mn) is the second most abundant redox active transition metal next to iron in the crust of the Earth.[254] Mn(IV) oxide is used widely in the dry cell industry. The most prevalent oxidation states of Mn are II, III and IV. It is believed that conversion of Mn(II) to higher oxidation states are largely performed by bacteria in nature. Genetic analysis including protoplast transformation and mutagenesis, was applied to identify chromosomal regions (MnxA-MnxB) involved in Bacillus species manganese oxidation.[255] Follow-up studies showed that products of these genes were exosporium proteins of Bacillus spores and MnxB has an amino acid sequence similar to the family of multicopper oxidases (MCO), a group of proteins that adopt multiple copper ions to oxidize different substrates, which was responsible for Mn(II) oxidation. [256, 257] Later studies directly identified via bottom-up proteomics that MnxB protein (or a MnxF/G protein complex) from marine Bacillus spores catalyzes formation of Mn oxide. The oxidation involves two sequential one-electron oxidations from Mn(II) to Mn(III) and from Mn(III) to Mn(IV).[258, 259] The most recent study was able to purify The MCO complex in a large quantity by expressing the mnxDEFG construct (mnxD can be omitted from the construct) in E. coli. The complex was characterized to contain one full-length
MnxG and 6-8 MnxE and MnxF subunits of unknown stoichiometry. [260] Here we use native MS and surface-induced dissociation to propose a structure of the Mnx complex.

8.2 Experimental

The purified Mnx protein complex obtained by expressing the mnxDEFG construct was provided by the Tebo group at the Oregon Health and Science University and further processed by Mowei Zhou in the Pacific Northwest National Laboratory. The sample was buffer exchanged immediately before MS analysis to 100 mM ammonium acetate (AmAc, Sigma-Adrich, St. Louis, MO) via Micro Bio-Spin™ P-6 Gel Columns (Bio-Rad, Hercules, CA). The protein was further diluted with 100 mM AmAc and the concentration was measured by a BCA protein assay (Thermo Scientific Pierce, Rockford, IL) to be 1 μg/μL. Samples were sprayed via a nano-electrospray source into a Waters SYNAPT G2 or G2S mass spectrometer (Manchester, UK) with a customized SID device incorporated before the ion mobility cell.[50]

The bottom-up experiment was performed by in-solution trypsin digestion of the Mnx sample, with LC-MS/MS analysis on a linear ion trap mass spectrometer (Velos Pro, ThermoFisher Scientific Inc., Waltham, MA) coupled to a UPLC (nanoACQUITY, Waters Corporation, Manchester, UK). Trypsin digestion was performed first by unfolding the protein in 0.03% ProteaseMAX surfactant (Promega, Madison, WI), followed by adding trypsin gold (Promega, Madison, WI) at 1:10 enzyme: protein ratio. The digestion was performed on a 37 °C, 150rpm thermomixer for 3 hrs. To quench the digestion, 0.5% TFA (Fisher Scientific, Pittsburgh, PA) was added to the digestion. Prior to LC-MS/MS injection, the sample was centrifuged at 10 kg for 10 min. The amount of
injection was 0.4 \( \mu \text{g} \). LC-MS/MS data analysis was performed by SEQUEST HT in Proteome Discoverer software (ThermoFisher Scientific Inc., Waltham, MA). The database contains protein sequences of Mnx, E.coli proteome and human keratin. Some searching parameters are listed here: precursor mass ranges from 400 to 5000 Da, peptide length ranges from 4 to 144, precursor mass tolerance is 1.5 Da, and fragment mass tolerance is 0.8 Da. Percolator with 1% FDR was used to validate the identities.

8.3 Results and Discussion

Based on the sequence information of the E and F subunits, which is further confirmed by denaturing ESI top down LCMS experiments performed at PNNL, their individual accurate mass can be obtained (Listed in Table 8.1). The mass of G subunit is calculated from the gene sequence and 70% sequence coverage was obtained in bottom-up proteomics experiment. The full mass spectrum of the Mnx sample is shown in Figure 8.1. The complex has mass of 211,010±285 Da. Often the measured mass of protein complexes is higher than the mass by sequence, because water, salts, and/or buffer molecules may remain attached to the complex. In addition, MnxE and MnxF are similar in mass, and is much lower in mass than the mass of MnxG. The composition of the Mnx complex was thus estimated to be one MnxG with six MnxE and/or MnxF, with the uncertainty in the MnxE and MnxF stoichiometry. The CCS of the complex was determined to be 8500±40 \( \text{Å}^2 \) by ion mobility analysis. Assuming the complex is globular, the diameter of the complex can be estimated to around 10 nm, which is
calculated by \((\frac{8500}{\pi})^{0.5-1.4})\times2\). The dimension results from ion mobility experiments match well with estimates from transmission electron microscopy.

Figure 8.1. Native full mass spectrum of Mnx, the major complex is \(E_{x}F_{(6-x)}G\).
### Table 8.1 Sequence information of subunits MnxE, MnxF, MnxG

<table>
<thead>
<tr>
<th>Subunit</th>
<th>Sequence Information from Uniprot</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MnxE</strong>&lt;br&gt;–S-S–&lt;br&gt;12,169 Da</td>
<td>tr</td>
</tr>
<tr>
<td><strong>MnxF</strong>&lt;br&gt;N-&lt;br&gt;11,193 Da</td>
<td>tr</td>
</tr>
<tr>
<td><strong>MnxG</strong>&lt;br&gt;138,250 Da</td>
<td>tr</td>
</tr>
</tbody>
</table>

The +29 peak was further isolated for analysis by tandem MS. A representative CID-IM-MS mobilogram is shown in Figure 8.2. It has products MnxE monomer (E), MnxF monomer (F), and their complementary partners Mnx-E, Mnx-F and Mnx-EF. CID analysis indicates that the Mnx complex contains both MnxE and MnxF subunits. However, the stoichiometry of MnxE and MnxF is still unknown.
Figure 8.2. Ion mobility - mass spectrum of all the products released from isolated +29 of Mnx complex by collision induced dissociation (ΔV=150 V)

The relative abundance of products and precursor are shown in log scale color scheme

The +29 peak was further isolated for fragmentation by SID. Figure 8.3 shows the low energy SID (ΔV=40 V) results. The complex is primarily fragmented into two products: one at m/z 5000-8000 with a mass of 70,760±70 Da, and the other one at m/z 9000-14000 with a mass of 141,400±500 Da. Also there are monomer products MnxE and MnxF. The masses of monomers are 61 higher than the masses calculated from sequences, which may be caused by the addition of one copper ion. The hexamer made of different copies of Cu bound MnxE and MnxF can then be calculated and the masses are shown in Table 8.2. By comparison with the theoretical masses, the 70,760 Da product was determined to be E3F3 hexamer. The 141,400 Da product was determined to be MnxG, which has a mass calculated from sequence of 138,250 Da.
Figure 8.3. Ion mobility - mass spectrum of all the products released from isolated +29 of Mnx complex by surface induced dissociation with low energy (ΔV= 40 V).
The relative abundance of products and precursor are shown in log scale color scheme

Table 8.2. Theoretical masses of hexamers composed of different combination of MnxE and F

<table>
<thead>
<tr>
<th>Composition</th>
<th>E+Cu: 12230 Da</th>
<th>F+Cu: 11254 Da</th>
<th>Mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>E6</td>
<td>x 6</td>
<td>x 0</td>
<td>73380</td>
</tr>
<tr>
<td>E5F1</td>
<td>x 5</td>
<td>x 1</td>
<td>72404</td>
</tr>
<tr>
<td>E4F2</td>
<td>x 4</td>
<td>x 2</td>
<td>71428</td>
</tr>
<tr>
<td>E3F3</td>
<td>x 3</td>
<td>x 3</td>
<td>70452</td>
</tr>
<tr>
<td>E2F4</td>
<td>x 2</td>
<td>x 4</td>
<td>69476</td>
</tr>
<tr>
<td>EF5</td>
<td>x 1</td>
<td>x 5</td>
<td>68500</td>
</tr>
<tr>
<td>F6</td>
<td>x 0</td>
<td>x 6</td>
<td>67524</td>
</tr>
</tbody>
</table>

Figure 8.4 shows the high energy SID-IM results (ΔV=100 V). As the SID energy goes higher, the E3F3 hexamer further dissociates to smaller oligomers. With the separation of ion mobility, the oligomers with overlapping m/z could be separated in drift time space thus clearly being distinguished. In the Figure 8.4, only the dominant
composition is labeled for each oligomeric state. Besides EF as dimer products, there are also F2 dimer products. Dominant trimer product and tetramer product are EF2 and E2F2, respectively, but E2F and EF3 also exist. The pentamer product is predominantly E2F3.

Figure 8.4. Ion mobility - mass spectrum of all the products released from isolated +29 of Mnx complex by surface induced dissociation with high energy (ΔV= 100 V). The relative abundance of products and precursor are shown in log scale color scheme

The variety of products in the SID spectra not only leads to accurate determination of the composition of the Mnx complex, but also provides more structural information than products produced in CID. The break down to E3F3 hexamer and MnxG at the lowest energy onset (1160 eV) indicates that the interaction between the hexamer and MnxG is the weakest. The breakdown of the E3F3 hexamer and generation of different smaller oligomers or subunits as the SID acceleration voltage increases is shown in Figure 8.5. Excluding MnxG species and only considering MnxE and/or MnxF containing species, the relative intensity of each species was taken by normalizing against
the sum of intensities of these species, and plotted over the SID energy increment. The higher relative abundance of a species indicates that the pathway forming the species is favored. In this sense, the monomer dissociation pathway E+E2F3 is the most preferred pathway, followed by the dimer pathway EF+E2F2 and the trimer pathway EF2+E2F. Some oligomer fragments may be generated from secondary fragmentation, such as F, F2, EF3. Based on the disassembly pathways, a structure with intersubunit contacts for E3F3 is proposed in the left of Figure 8.6, with three types of interfaces labeled as a, b, and c. The interface strength of E-F (a), E-E (b), and F-F (c) are proposed to be a>c>b. If one follows this model to disassemble the complex, the E+E2F3 is the easiest to dissociate, which involves breaking 2a+2b. The EF+E2F2 pathway involves breaking 3a+2b+c, which is smaller than dimer pathway F2+E3F (breaking 5a+c), explaining the EF as the dominant dimer products. The trimer pathway is relatively energy consuming, which involves breaking 4a+b+c. The higher abundance for EF2 over E2F may be attributed to the fact that the interaction of F-F (c) is higher than that of E-E (b).

Figure 8.5. a) Energy-resolved SID results for fragmentation of MnxE3F3G, focusing on E3F3 hexamer, and normalized to total EF oligomer intensity. b) Zoom in low intensity oligomers.
The zoomed-in spectra of +4 MnxF and MnxE from CID of Mnx complex are shown in Figure 8.7 (a) and (b). The MnxF subunit is almost exclusively in apo form, without any metal binding, while the MnxE subunit almost exclusively binds to Cu. A close look at the +4 monomers in SID reveals that approximately 60% of MnxF binds to one or more Cu and almost 100% of MnxE binds to one or more Cu (Figure 8.7 (c) and (d)). Comparing the +4 MnxF (Figure 8.7 (a) and (c)) in CID and SID, the Cu holo peaks are more abundant in SID. Also the dominant MnxF charge state +6 in CID barely has any Cu bound species. This is due to the fact that CID is an unfolding process, so the Cu may easily be lost in CID process.
Close scrutiny of the Mnx complex at harsher ionization condition enable us to see the heterogeneity of the complex, because more efficient desolvation allows us to resolve the peaks in the proximity of the total complex m/z. As shown in Figure 8.8 (a), the gray trace shows the full scan of Mnx complex with +29 to +27 charges under increased cone voltage. The peaks split into two envelopes of different masses, indicating the heterogeneity of the sample. Isolation and fragmentation of the +28 peaks that split to lower m/z (left) peak and higher m/z (right) peak was further performed to investigate the peak identity. The isolation of +28 left and right peaks was shown in Figure 8.8 (a) top and bottom, respectively. The two species were fragmented with SID and a similar
pattern of fragments as in Figure 8.4 could be observed. Different product species were extracted to mass spectra for the convenience of comparison between fragmentation of left and right peaks. Extracted MnxE, MnxF and MnG monomer peaks from surface induced dissociation of the left and right peaks were fixed in the same \( m/z \) scale and aligned vertically for comparison (Figure 8.8 (b), (c)). It shows that the mass of the individual subunit MnxE, MnxF and MnxG doesn’t change. However, differences between the left and right peaks are observed in oligomer fragments comprise of MnxE and MnxF. The left peak primarily contains E2F4, while the right peak mainly comprises E3F3 (Figure 8.8 (h)). For the hexamers losing one subunit to pentamer, E2F4 tend to lose a E or F equally to EF4 and E2F3 in similar ratio, while E3F3 primarily loses E to E2F3 (Figure 8.8 (g)). Then almost all the pentamers EF4 and E2F3 lose one F to tetramers EF3 and E2F2, respectively (Figure 8.8 (f)). The dimer product is mainly EF dimer and the trimer product is mainly EF2 in both left and right peaks. The experiments with harsher ionization condition prove that the Mnx complex has heterogeneity with the most abundant E3F3G configuration and the second most abundant E2F4G configuration.
Figure 8.8. The heterogeneity of the Mnx complex.
(a) Isolation of the left and right +28 Mnx. When the left and right slices of the peak are subjected to SID (ΔV=120V) analysis, the products are extracted and categorized as b) MnxE, MnxF monomer, c) MnxG monomer, d) Dimer, e) Trimer, f) Tetramer, g) Pentamer and h) Hexamer that are composed of E and F. It indicates that the left precursor is E2F4G and the right precursor is E3F3G.

8.4 Conclusions

Surface induced dissociation coupled with Ion mobility mass spectrometry has confidently determined the composition of Mnx, a bacterial multicopper oxidase enzyme complex. Mnx is primarily composed of one MnxG, three MnxE and three MnxF. The assembly pathway is more likely to be the initial formation of highly symmetric hexamer with alternating MnxE and MnxF, followed by the association of MnxG with the hexamer. Moreover, it has been discovered by SID that the MnxE subunit almost
exclusively binds with Cu and the MnxF subunit loosely binds Cu. The binding of MnxF to Cu is greatly diminished in CID due to the unfolding in CID process. Also CID of Mnx is dominated by MnxE and MnxF monomer ejection, and thus provides limited subcomplexes for structural elucidation. This protein has not yet been successfully crystalized for structural determination, probably due to the heterogeneity. The mass spectrometry and SID data presented here are the first experiments to determine its quaternary structure. The successful application of SID/IM in the analysis of the Mnx structure definitely makes it a valuable tool in structural biology.
Chapter 9. Conclusions and Future Directions

This dissertation has shown that mass spectrometry is a powerful tool for characterization of protein structures, from primary to quaternary structures.

Chapter 2 describes the application of mass spectrometry to elucidate primary structures of Hemoglobin (Hb) from different bird species in an effort to advance tick-borne disease prevention and intervention. Blood samples were collected from 33 bird species that are common in the U.S. as hosts for ticks but that have unknown Hb sequences. Ticks are vectors for disease transmission because they are indiscriminant in their feeding on multiple vertebrate hosts, transmitting pathogens between their hosts. Identifying the hosts on which ticks have fed is important for disease prevention and intervention. We have previously shown that Hb remnants from a host on which a tick fed can be used to reveal the host’s identity. A top-down-assisted bottom-up mass spectrometry approach with a customized searching database, based on variability in known bird hemoglobin sequences, has been devised to facilitate fast and complete sequencing of Hb from birds with unknown sequences. The top-down approach played a role of quality control in confirming the peptides alignment in the bottom-up approach. The bottom-up approach with a customized searching database resulted in a general sequence quickly, and also provided sequence in middle sequence regions where the top-down approach failed. These hemoglobin sequences will be added to a hemoglobin
database and used for tick host identification. The general approach has the potential to sequence any set of homologous proteins completely in a rapid manner.

Tandem mass spectrometry not only can be applied to dissociating covalent bonds of peptides or protein chains, yielding sequence information. It can also be applied to non-covalent protein complexes, probing their quaternary structures by generating structurally informative subcomplexes. The most widely used gas-phase dissociation method is collision induced dissociation (CID), which involves an analyte going through tens of thousands of collisions with inert gas atoms/molecules. However, the “typical CID process” proceeds with both mass and charge asymmetric partitioning, and the level of structural information that can be obtained in this way is limited. An alternative gas-phase disassembly method, surface induced dissociation (SID), which involves an analyte colliding with a surface, has been shown in several publications to reveal substructure information ever since it was realized in the Wysocki lab about a decade ago. Especially in recent years, with the incorporation of an SID device to Waters SYNAPT G2 and G2-S mass spectrometers with ion mobility capability, the conformation of SID products, as well as remaining precursors were studied to improve our understanding of the SID process. The dissociation behavior of the simplest protein complexes, homodimers, has been studied by several groups to explore the origin of asymmetric charge partitioning in CID. Factors including lower charge states, higher activation energy, and higher conformational flexibility have been shown to promote asymmetric charge partitioning in CID. We hypothesized that these factors would influence the symmetric charge partitioning in SID to a lesser extent. In Chapter 3, fundamental studies of gas-phase
dissociation behavior of homodimers, enolase, α-lactalbumin, and β-lactoglobulin were performed by CID and SID. Drastically different charge-to-mass partitioning has been observed for enolase dimer with SID versus conventional CID. The difference can be ascribed to subunits unfolding to a lesser extent in the single step, high energy deposition provided by SID. It is evident that the dimer unfolding profile in SID is smaller than in CID at the onset energy of enolase monomer production. The intra-disulfide intact α-lactalbumin homodimer and β-lactoglobulin homodimer both show symmetric charge partitioning upon CID. Increasing subunit flexibility by reducing the intra-disulfide bonds causes asymmetric charge partitioning by CID throughout different collision energy levels. In contrast, both intra-disulfide intact and reduced α-lactalbumin homodimer and β-lactoglobulin homodimer show symmetric charge partitioning upon higher energy SID, indicating that increased flexibility plays a less important role in affecting the SID dissociation behavior of protein complexes. The experiments performed on protein homo-dimers indicate that SID is beneficial for protein complex interface analysis, because SID minimizes conformational disruptions (unfolding) of subunits in the dissociation process.

Chapter 4 describes another model system studied in order to expand the SID capability, monoclonal antibodies. Light chains with free thiol groups were observed as products by dissociating gently disulfide-reduced mAb by either CID or SID. Interestingly, heavy chain and half mAb are generated in SID but not in CID by top-down MS. It is highly possible that some of the covalent disulfide bonds between the heavy chains have been reduced in the gentle reduction. However in CID, the preferential
pathway is the ejection of the light chain, which makes the heavy chain fragments unobservable. Because SID produces more symmetric dissociation with respect to both mass and charges, the heavy chain can be observed, as well as the heavy and light chains, being held together by noncovalent interactions.

Chapter 5 describes research performed, in order to further explore the level of information on substructure provided by SID, larger and more complicated protein complexes with multiple different interfaces were under investigation. Four homo-hexamer protein complexes with known crystal structures or NMR results were dissociated by SID. Conventional CID generated limited fragment species, mainly monomers and complementary pentamers, thus provided little substructural information. For all the four homo-hexamers, the in silico dissociation analysis based on the relative strengths of interfaces is consistent with the SID result. The excellent correlation of the SID behaviors and the in silico analysis demonstrates that SID can be very useful in structural characterization of unknown systems, especially in predicting the interfaces and relative interfacial strength.

SID characterization studies with protein complex standards or relatively well-studied protein complex systems described in Chapter 3, 4, and 5 provide a sound basis encouraging the tackling of challenging tasks, like predicting unknown quaternary structures based on the SID products pattern. Also, we would like to have other complementary methods to cross-validate our prediction. One example highlighting the elucidation of quaternary structure is shown in Chapter 6. Toyocamycin nitrile hydratase (TNH) is a protein hexamer that catalyzes the hydration of toyocamycin to produce
sangivamycin. The structure of hexameric TNH and arrangement of subunits within the complex is, however, unknown. Native mass spectrometry (MS) clearly shows that TNH is composed of two copies each of α, β, and γ subunits. Elucidating the arrangement of the subunits within this complex is of interest. Previous surface-induced dissociation (SID) tandem mass spectrometry on a QTOF platform highly suggests that TNH hexamer is a dimer composed of two αβγ trimers, furthermore the results suggest that the subunit α, β interact most strongly (Blackwell, 2011). Here, multiple mass spectrometry based approaches have been applied to refine the structure of TNH. Solution-phase organic solvent disruption coupled with native MS agrees with the previous SID results. By coupling surface induced dissociation with ion mobility mass spectrometry (SID/IM), further information on the inter-subunit contacts can be obtained. The results confirm that TNH is a dimer of αβγ trimers and that within the trimer the α, β subunits bind most strongly and that the primary contact between the two trimers is through a γ-γ interface. Collisional cross-section (CCS) measured from IM experiments are used as constraints for postulating the arrangement of the subunits represented by coarse-grained spheres. Covalent labeling together with protein complex homology modeling is utilized to propose likely quaternary structure of TNH. The combined results from different mass spectrometry approaches and computational modeling provide a robust framework for analysis of protein complexes.

As a continuation of structural elucidation work of TNH, we found that the His-tagged γ subunit can still form a hexamer with subunits α, β to perform the same biological function with native TNH. (Chapter 7) Because His-tagging on the terminus of
proteins is a popular tool for protein purification, we were trying to understand how His-tagging may affect the gas phase dissociation behavior of TNH, and hope the results can be extended to understanding many other protein complexes that contain His-tags. Native TNH and His-tagged TNH (His-tag on γ) have been shown to have similar structure by surface mapping using diethylpyrocarbonate (DEPC) and phenylglyoxal (PGO) with bottom-up mass spectrometry analysis. Further, native TNH and His-tagged TNH were subjected to native mass spectrometry analysis in a modified quadrupole/ ion mobility/ time-of-flight mass spectrometer. By SID, the major dissociation product of the hexamer is trimer composed of one copy of each of the three subunits, regardless of the precursor charge state and His-tagging. Upon CID, spectra are dominated by ejection of monomers, whether a His-tag is present or not. Interestingly, the monomer ejection order is influenced by the precursor charge state and the His-tagging. The phenomena suggest that in CID, the His-tag can help unravel the corresponding subunit in a protein complex, but enough charge on the protein complex is one requisite.

Another example showing our effort in elucidation of quaternary structure is shown in Chapter 8. Native MS and SID/IM have been applied to elucidate the structure of Mnx, a bacterial multicopper oxidase enzyme complex. This complex has not yet been successfully crystalized for structural determination by X-ray crystallography and NMR has not yielded meaningful results, probably due to its heterogeneity and large size (over 200 kDa). By SID/IM, Mnx is determined to be primarily composed of one MnxG, three MnxE and three MnxF. The assembly pathway is more likely to be the initial formation of highly symmetric hexamer with alternating MnxE and MnxF, followed by the
association of MnxG with the hexamer. A less abundant component was mass selected and fragmented. The fragmentation pattern shows it is composed of one MnxG, two MnxE and four MnxF. The successful application of SID/IM in the analysis of the Mnx structure definitely makes it a valuable tool in structural biology. Multiple MS-based techniques, including CCS constrained coarse-grained modeling, surface exposed lysine labeling by NHS-biotin or acetic anhydride, ab initio subunit modeling, homology modeling and docking haven been attempted to propose high resolution structural data for Mnx complex, but not yet included in this dissertation. In the future, the experiments need to be repeated for further refinement of the Mnx model.

Overall, this dissertation has demonstrated that mass spectrometry is an indispensable tool for characterization of protein structures, from primary to quaternary structures.
References

52. Ma, X.; Loo, J. A.; Wysocki, V. H., Surface induced dissociation yields substructure of Methanosarcina thermophila 20S proteasome complexes. *Int. J. Mass spectrom.* 2015; 377 (0), 201-204.


82. Fenn, L.; Kliman, M.; Mahsut, A.; Zhao, S.; McLean, J., Characterizing ion mobility-mass spectrometry conformation space for the analysis of complex biological samples. *Analytical and Bioanalytical Chemistry* 2009; 394 (1), 235-244.


86. Smith, D. P.; Radford, S. E.; Ashcroft, A. E., Elongated oligomers in β2-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry. *Proceedings of the National Academy of Sciences* 2010; 107 (15), 6794-6798.


