Study of Proteoforms, DNA and Complexes using Trapped Ion Mobility Spectrometry-Mass Spectrometry

Alyssa Lynn Garabedian
Florida International University, agara016@fiu.edu

DOI: 10.25148/etd.FIDC004097
Follow this and additional works at: https://digitalcommons.fiu.edu/etd
Part of the Analytical Chemistry Commons

Recommended Citation
https://digitalcommons.fiu.edu/etd/3567

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.
FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

STUDY OF PROTEOFORMS, DNA AND COMPLEXES USING TRAPPED ION MOBILITY SPECTROMETRY – MASS SPECTROMETRY

A dissertation submitted in partial fulfillment of
the requirements for the degree of
DOCTOR OF PHILOSOPHY
in
CHEMISTRY

by

Alyssa L. Garabedian

2018
To: Dean Michael R. Heithaus  
College of Arts, Sciences and Education

This dissertation, written by Alyssa L. Garabedian, and entitled Study of Proteoforms, DNA and Complexes using Trapped Ion Mobility Spectrometry-Mass Spectrometry, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

_______________________________________  
Yuan Liu

_______________________________________  
Jessica Siltberg-Liberles

_______________________________________  
Fenfei Leng

_______________________________________  
Yuk-Ching Tse-Dinh, Co-Major Professor

_______________________________________  
Francisco Fernandez-Lima, Co-Major Professor

Date of Defense: March, 26, 2018

The dissertation of Alyssa L. Garabedian is approved.

_______________________________________  
Dean Michael R. Heithaus  
College of Arts, Sciences and Education

_______________________________________  
Andrés G. Gil  
Vice President for Research and Economic Development  
and Dean of the University Graduate School

Florida International University, 2018
© Copyright 2018 by Alyssa L. Garabedian

All rights reserved.
DEDICATION

To my parents, brother, sister and my husband, for helping me see this through and for being there every step of the way.
ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Francisco Fernández-Lima, for introducing me to the world of mass spectrometry and for providing numerous opportunities to attend conferences and to travel to various labs for research. His continuous involvement means that he has always dedicated immeasurable amounts of time, energy and support to my research projects, publications and presentations. I would like to express my gratitude to my co-mentor, Dr. Yuk-Ching Tse-Dinh, for her patience, knowledge, guidance and for always being a reliable source of assistance.

I am thoroughly grateful for all the support I have received from numerous people. In particular, my Lima Lab ‘family’ (both past and present members) have continuously proven that they are the absolute best in terms of support, advice, assistance and friendship - thank you Kendra Adams, Paolo Benigni, Anthony Castellanos, Kevin Jeanne Dit Fouque, Camilo Molano-Arevalo, Alan McKenzie Coe, Jacob Porter, Emily Schenk and Lilian Tose.

Thank you to Dr. Cesar Ramirez and Dr. Mario Gomez for being my second mentors. In any scenario, they have always been eager and willing to help and advise.

Thank you to Dr. Arasu Annamalai and Dr. Jeremy Chambers for being a constant presence throughout my graduate studies and for being my go-to source for any biological question. Your encouragement and support has been invaluable.

I am very appreciative of the administrative help provided by Stephanie Rosas, Jackelyn Marcos and Maggie Autie - thank you for your time and patience.

My committee members, Dr. Fenfei Leng, Dr. Jessica Siltberg-Liberlies and Dr. Yuan Liu, have been an instrumental part of my Ph.D. training; thank you for providing
me with research direction, professional guidance, feedback, and encouragement. I appreciate the thought behind every question and the time spent helping to advance the quality of my dissertation research.

Finally, I would like to acknowledge the FIU Dissertation Year Fellowship which funded my final year and allowed me to focus on my research and writing.
ABSTRACT OF THE DISSERTATION

STUDY OF PROTEOFORMS, DNA AND COMPLEXES USING TRAPPED ION MOBILITY SPECTROMETRY – MASS SPECTROMETRY

by

Alyssa L. Garabedian

Florida International University, 2018

Miami, Florida

Professor Francisco Fernández-Lima, Co-Major Professor
Professor Yuk-Ching Tse-Dinh, Co-Major Professor

The characterization of biomolecules and biomolecular complexes represents an area of significant research activity because of the link between structure and function. Drug development relies on structural information in order to target certain domains. Many traditional biochemical techniques, however, are limited by their ability to characterize only certain stable forms of a molecule. As a result, multidimensional approaches, such as ion mobility mass spectrometry coupled to mass spectrometry (IMS-MS), are becoming very attractive tools as they provide fast separation, detection and identification of molecules, in addition to providing three-dimensional shape for structural elucidation. The present work expands the use and application of trapped ion mobility spectrometry-coupled to mass spectrometry (TIMS-MS) by analyzing a range of biomolecules (including proteoforms, intrinsically disordered peptides, DNA and molecular complexes). The aim is to i) evaluate the TIMS platform measuring sensitivity, selectivity, and separation of targeted compounds, ii) pioneer new applications of TIMS for a more efficient and higher
throughput methodologies for identification and characterization of biomolecular ions, and
iii) characterize the dynamics of selected biomolecules for insight into the folding
pathways and the intra-or intermolecular interactions that define their conformational
space.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAPTER 1: INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Overview</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Analytical Methods for Molecular Ionization, Separation and Characterization</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Dissertation Structure</td>
<td>7</td>
</tr>
<tr>
<td>1.4 References</td>
<td>9</td>
</tr>
<tr>
<td>CHAPTER 2: TOWARDS DISCOVERY AND TARGETED PEPTIDE BIOMARKER DETECTION USING NANOESI-TIMS-TOF MS</td>
<td>16</td>
</tr>
<tr>
<td>2.1 Abstract</td>
<td>17</td>
</tr>
<tr>
<td>2.2 Introduction</td>
<td>17</td>
</tr>
<tr>
<td>2.3 Experimental</td>
<td>20</td>
</tr>
<tr>
<td>2.3.1 Tumor protein extraction and tryptic digestion</td>
<td>20</td>
</tr>
<tr>
<td>2.3.2 Tryptic peptide fractionation</td>
<td>21</td>
</tr>
<tr>
<td>2.3.3 Trapped Ion Mobility Spectrometry – Mass Spectrometry Analysis</td>
<td>22</td>
</tr>
<tr>
<td>2.3.5 LC-ESI-MS/MS Analysis</td>
<td>24</td>
</tr>
<tr>
<td>2.4 Results and Discussion</td>
<td>25</td>
</tr>
<tr>
<td>2.5 Conclusions</td>
<td>35</td>
</tr>
<tr>
<td>2.6 Acknowledgements</td>
<td>36</td>
</tr>
<tr>
<td>2.7 References</td>
<td>37</td>
</tr>
<tr>
<td>CHAPTER 3: LINEAR AND DIFFERENTIAL ION MOBILITY SEPARATIONS OF MIDDLE-DOWN PROTEOFORMS</td>
<td>44</td>
</tr>
<tr>
<td>3.1 Abstract</td>
<td>45</td>
</tr>
<tr>
<td>3.2 Introduction</td>
<td>45</td>
</tr>
<tr>
<td>3.1 ESI-TWIMS-MS Instrumentation</td>
<td>48</td>
</tr>
<tr>
<td>3.2 nESI-TIMS-MS Instrumentation</td>
<td>51</td>
</tr>
<tr>
<td>3.3 Results and Discussion</td>
<td>52</td>
</tr>
<tr>
<td>3.3.1 TWIMS Separations</td>
<td>52</td>
</tr>
<tr>
<td>3.3.2 TIMS Separations</td>
<td>56</td>
</tr>
<tr>
<td>3.3.3 Correlations between Separation Dimensions</td>
<td>60</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>6.3.1 Materials and reagents</td>
<td>129</td>
</tr>
<tr>
<td>6.3.2 DNA UV melting studies</td>
<td>130</td>
</tr>
<tr>
<td>6.3.3 Fluorescence Measurements</td>
<td>130</td>
</tr>
<tr>
<td>6.3.4 Isothermal Titration Calorimetry (ITC)</td>
<td>131</td>
</tr>
<tr>
<td>6.3.5 Trapped Ion Mobility Spectrometry – Mass Spectrometry Analysis (TIMS-MS)</td>
<td>131</td>
</tr>
<tr>
<td>6.3.6 Theoretical Calculations</td>
<td>133</td>
</tr>
<tr>
<td>6.4 Results and Discussion</td>
<td>134</td>
</tr>
<tr>
<td>6.5 Conclusions</td>
<td>142</td>
</tr>
<tr>
<td>6.6 Acknowledgements</td>
<td>143</td>
</tr>
<tr>
<td>6.7 References</td>
<td>143</td>
</tr>
</tbody>
</table>

CHAPTER 7: PEPTIDE SEQUENCE INFLUENCE ON THE CONFORMATIONAL DYNAMICS AND DNA BINDING OF THE INTRINSICALLY DISORDERED AT-HOOK 3 PEPTIDE ................................................................. 150

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1 Abstract</td>
<td>151</td>
</tr>
<tr>
<td>7.2 Results and Discussion</td>
<td>153</td>
</tr>
<tr>
<td>7.2 Experimental Methods</td>
<td>163</td>
</tr>
<tr>
<td>7.2.1 Sample preparation</td>
<td>163</td>
</tr>
<tr>
<td>7.2.2 Peptide nomenclature</td>
<td>164</td>
</tr>
<tr>
<td>7.2.3 Ion Mobility Spectrometry-Mass Spectrometry</td>
<td>164</td>
</tr>
<tr>
<td>7.2.4 Correction of amino acid substituted collision cross sections</td>
<td>166</td>
</tr>
<tr>
<td>7.3 Acknowledgements</td>
<td>167</td>
</tr>
<tr>
<td>7.4 References</td>
<td>167</td>
</tr>
</tbody>
</table>

CHAPTER 8: MICRO HETEROGENEITY OF TOPOISOMERASE IA/IB AND THEIR DNA-BOUND STATES ................................................................. 173

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1. Introduction</td>
<td>174</td>
</tr>
<tr>
<td>8.2 Experimental Methods</td>
<td>175</td>
</tr>
<tr>
<td>8.2.1 Materials and reagents</td>
<td>175</td>
</tr>
<tr>
<td>8.2.2 Trapped Ion Mobility Spectrometry – Mass Spectrometry Analysis (TIMS-MS)</td>
<td>176</td>
</tr>
<tr>
<td>8.3 Results and Discussion</td>
<td>177</td>
</tr>
<tr>
<td>8.4 References</td>
<td>186</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1. Peptide sequence, m/z of light and heavy peptides, ion-neutral collision cross section (CCS) and total in-fraction peptide concentration of the five targeted peptides in the fractions. LC-ESI-MS/MS concentrations using the 535.2 → 187.0 Da (DVVICPDASLEDAKK), 585.7 → 201.1 Da (VFDFGDNGYISAAELR) and 474.9 → 130.1 Da (TTILQSTGK) channels are denoted with an asterisk (*).</td>
<td>35</td>
</tr>
<tr>
<td>Table 3.1. PTM localizations in H3 tail (ART3K4Q_T6ARK9S10_TGGK14A_PRK18QL_ATK23AA_RK27S28AP_ATGGV_K36KPHR_Y41RPGT_VALRE)</td>
<td>48</td>
</tr>
<tr>
<td>Table 3.2. Summary of Linear Correlations between Separations (averaged over all PTMs and charge states): r² values with Standard Errors of Mean</td>
<td>61</td>
</tr>
<tr>
<td>Table 5.1. Experimental and theoretical collision cross sections for i-motif DNA bands observed in positive and negative ion mode in nanoESI-TIMS-MS and proposed candidate structures, respectively.</td>
<td>109</td>
</tr>
<tr>
<td>Table 7.1. Summary of charge location and cis/trans proline configuration of the Native ATHP3 per IMS band (A-J)</td>
<td>157</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1. Typical mass spectra and IMS projection plots of the five targeted peptides of interest. Notice the high mobility resolution obtained using nanoESI-TIMS-MS for singly and doubly charged molecular ions.</td>
<td>26</td>
</tr>
<tr>
<td>Figure 2.2. a) TIMS-CID-MS of isolated isobaric precursor ions for the DFTPAELR and TTILQSTGK peptides and the corresponding ladder fragmentation pattern. B) Mobility selected MS showing the separation of the precursor ions. C) CID spectra of the fragments corresponding to DFTPAELR and TTILQSTGK.</td>
<td>28</td>
</tr>
<tr>
<td>Figure 2.3. a) Typical 2D-IMS-MS contour plot using nanoESI-TIMS-MS for a fraction containing the target peptide TTILQSTGK. The 2D-IMS-MS profile highlights the complexity of each fraction and shows the charge state specific trend lines. b) The 2D IMS-MS at the level of nominal mass depicts the isomeric interferences in the region of the targeted peptide and heavy analog. c) IMS projection plots for the targeted and corresponding heavy peptide using 5mDa window show identical CCS values.</td>
<td>30</td>
</tr>
<tr>
<td>Figure 2.4. a) Typical TIMS profiles of a pure heavy standard and in-fraction light peptide standard (TTILQSTGK [M+H]+) as a function of the concentration. b) Trapping efficiency of TIMS analysis illustrating that trapping time does not impact the calculated concentration or linear response and c) Measured light peptide concentrations for each targeted peptide, via TIMS-MS, in relation to the heavy analog displaying a linear response.</td>
<td>32</td>
</tr>
<tr>
<td>Figure 2.5. a) Response curves for the LSASTASELSPK [M+2H]+2 peptide standards analyzed in fraction (matrix effect) and in the blank via nanoESI-TIMS-MS and LC-QqQ-MS and b) comparisons of targeted peptide concentrations measured in fractions by nanoESI-TIMS-MS and LC-QqQ-MS.</td>
<td>34</td>
</tr>
<tr>
<td>Figure 3.1. TWIMS spectra for histone tail variants with z = 6 - 11, measured with solvent (i) using s = 650 m/s.</td>
<td>50</td>
</tr>
<tr>
<td>Figure 3.4. TIMS analysis of histone tail variants: spectra (cross section scale) for z= 6–11 (with solvent (v), tramp = 500 ms).</td>
<td>57</td>
</tr>
<tr>
<td>Figure 3.5. TIMS spectra for K23me3 8+ measured at (a) tramp = 100 and 500 ms from solvent (v) and (b) tramp = 500 ms from solvents (v) and (vi). Results for other tramp values, variants, and charge states are given in Appendix 3.8.</td>
<td>58</td>
</tr>
</tbody>
</table>
Figure 3.6. Linear correlations between FAIMS and TIMS separations for ac variants (r2 marked). The plots involving TWIMS and for p variants are in Appendix 3.14. 63

Figure 4.1. Scheme utilized for the nESI-TIMS-MS experiments with organic gas modifiers. Notice that the gas velocity in the TIMS analyzer is kept constant. 76

Figure 4.2. a) Typical mass spectra and b) native IMS spectra of ATHP3 as a function of c) starting solvent (methanol:H2O or acetone:H2O) or d) dopant bath gas (methanol or acetone). 78

Figure 4.3. The relative abundances of ATHP 3 [M+2H]+2 conformers as a function of the trapping time, stating solvent conditions and bath gas composition. Starting solvent and bath gas are listed to the left of the graphs. 79

Figure 5.1. Circular dichroism spectra of i-motif DNA (30 µM in deionized water) as a function of the solution pH =4.0 – 9.0 at 20 °C. Notice that no changes in the spectra are observed in the pH = 7.0 – 9.0 range. 98

Figure 5.2. Typical positive (red) and negative (blue) mass spectra (top) and mobility profiles for i-motif DNA as a function of the solution pH (4.0 – 9.0) using low energy injection nanoESI-TIMS-MS. Notice the change in the charge distribution as a function of the solution pH. Changes in individual charge states mobility profiles were not observed as a function of the solution pH conditions. 100

Figure 5.3. Plot of the fast (ϕ1) and slow (ϕ2) phases as a function of Cpp/β in the 16 – 26 °C temperature range. In the inset, PAC traces for 2-NBA photo-dissociation in the presence of i-motif DNA sample (blue) and 4SP reference compound (red) at 16 °C. 102

Figure 5.4. Typical CIA-TIMS-MS profiles of i-motif DNA as a function of the deflector voltage Vdeflector 60V (top) and 140V (bottom). Notice the change in the mobility distribution for the z= −5 and z= −6 as a function of the molecular ion activation prior the TIMS-MS analysis. 106

Figure 5.5. Candidate structures proposed for the i-motif DNA mobility bands observed. Main intramolecular interactions are denoted. CCSs of the candidate structures are reported in Table 1. 108

Figure 5.6. Typical mass spectra and mobility profiles for different i-motif DNA molecular ions observed at the z= −7 and z= −8 charge state. Notice the change in the mobility distribution in the presence of the ammonium and sodium adducts. 111

Figure 6.1. Typical mobility profiles for the [M+4H]4+ and [M+5H]5+ charge states of ATHP 1, 2, or 3 in complex with the DNA hairpin (FL876). The isotopic distribution for each complex is shown (inset) and candidate structures for each complex are proposed. 136
Figure 6.2. Typical mobility and isotopic distribution of DNA complexes. Single mobility peaks were observed for all complexes and charge states. Specifically, the DNA and the minor groove binding (MGB) Hoescht 33258 compound produced a [M+4H]+4 charge state. ATHP 1 binding to the DNA-Hoescht complex produced a [M+4H]+4 and [M+5H]+5 charge states. Proposed structures are also shown. .................... 138

Figure 6.3. Typical mobility profiles of ATHP 1, 2, or 3 in complex with the DNA hairpin (FL876). Stoichiometry of 2:1 was observed at a [M+5H]+5 charge state. The isotopic distribution for each complex is shown (inset) and candidate structures shown for the FL876 hairpin in complex with ATHP1, ATHP2 and ATHP3. .................... 139

Figure 6.4. a) DNA UV melting curves for uncomplexed FL876 and complexed FL876 with ATHP 1, 2 or 3 and b) ITC results determining the binding stoichiometry and affinity of ATHP 1, 2 and 3 to both the major and minor groove of FL876 DNA and c) Collision induced dissociation for FL876 complexed with ATHP1. ATHP2 or ATHP 3 (charge state and degrees of freedom were considered) with the curve derivative shown as an inset. ................................................................. 141

Figure 7.1. Typical, normalized ATHP3 and corrected variant (*) mobility profiles for the [M+H]+, [M+2H]2+ and [M+3H]3+ for intrinsic size parameters upon residue substitution to alanine (a). Peptide IMS fingerprint (b) and correlation matrix (c) are used for assessment of the primary sequence effect on the secondary structure. .............................. 155

Figure 7.2. Typical mass spectra (a) and ‘corrected’ ion mobility profiles (b) of the native and substituted ATHP3 : DNA complexes. The IMS profile fingerprint (c) of the complexes was used to generate a correlation matrix (d). Values reported in (a) correspond to peptide : DNA : complex ratios. ................................................................. 159

Figure 7.3. Typical CID profiles (a) showing the decrease in intensity of the [M+4H]4+ peptide : DNA complexes as a function of collision energy (eV), charge (z) and degrees of freedom (DoF). The dissociation threshold (inset) shows the absolute derivative of each CID profiles as a function of the collision energy. The binding affinity (b) and distance matrix (c) illustrates the influence of the single amino acid substitution on the peptide-DNA binding affinity. ................................................................. 161

Figure 8.1. a) Typical mass spectra of all charge states observed under native starting solvent conditions, and (b) the mobility profiles for the native (+10 to +13) and partially folded forms (+13 to +21) of unbound vTopIB. The structure presented is a modified version pdb 3IGC,10 without the DNA substrate. ................................................................. 179

Figure 8.2. a) Typical mass spectra of all charge states observed under native starting solvent conditions, and (b) the mobility profiles for the native (+16 to +21) state of unbound EcTopI. The structure presented is a modified version pdb 4RUL, without the DNA substrate. ................................................................. 181
Figure 8.3. a) Typical mass spectra, and (b) mobility profiles for the native states (+12 to +14) of vTopIB in complex with a 59 base stem loop DNA substrates (MBLong2), as a function of native starting solvent conditions.

Figure 8.4. a) Typical mass spectra, and (b) mobility profiles for the native states (+17 to +20) of EcTopI in complex with two ssDNA substrates (2x sub 1), as a function of native starting solvent conditions.

Figure 8.5. a) Typical mass spectra, and (b) mobility profiles for the native states (+17 to +22) of EcTopIB in complex with a 59 base stem loop DNA substrates (MBLong2), as a function of native starting solvent conditions.
ABBREVIATIONS AND ACRONYMS

°C          Celsius
2-NBA       2-nitrobenzaldehyde
ATHP        AT-hook peptide
AT-rich     Adenosine Thymine-rich
CCS         Collision cross section
CIA         Collision induced activation
CID         Collision Induced Dissociation
CD          Circular dichroism spectroscopy
CE          Collision energy
C₃H₆O      Acetone
CH₃OH      Methanol
CM          Center of mass
CPTAC      Clinical Proteomic Tumor Analysis Consortium
C-rich     Cytosine-rich
CSD        Charge state distribution
CXP        Collision cell exit potential
DE         Delustering potential
DNA        Deoxyribonucleic acid
DT         Drift tube
E. coli    Escherichia coli
ESI        Electrospray ionization
ETD        Electron transfer dissociation
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAIMS</td>
<td>Field asymmetric ion mobility spectrometry</td>
</tr>
<tr>
<td>FT-ICR</td>
<td>Fourier transform-ion cyclotron resonance</td>
</tr>
<tr>
<td>HDX</td>
<td>Hydrogen/deuterium exchange</td>
</tr>
<tr>
<td>HMGA2</td>
<td>High mobility group protein A2</td>
</tr>
<tr>
<td>HPLC</td>
<td>High-performance liquid chromatography</td>
</tr>
<tr>
<td>IDP</td>
<td>Intrinsically disordered protein</td>
</tr>
<tr>
<td>i-motif DNA</td>
<td>Intercalated-motif DNA</td>
</tr>
<tr>
<td>IMS</td>
<td>Ion mobility spectrometry</td>
</tr>
<tr>
<td>IMS-MS</td>
<td>Ion mobility spectrometry-mass spectrometry</td>
</tr>
<tr>
<td>IRMPD</td>
<td>Infrared multiphoton dissociation</td>
</tr>
<tr>
<td>ITC</td>
<td>Isothermal titration calorimetry</td>
</tr>
<tr>
<td>LC</td>
<td>Liquid chromatography</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular dynamics</td>
</tr>
<tr>
<td>MGB</td>
<td>Minor groove binding compound</td>
</tr>
<tr>
<td>MS</td>
<td>Mass Spectrometry</td>
</tr>
<tr>
<td>NH₄Ac</td>
<td>Ammonium Acetate</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance spectroscopy</td>
</tr>
<tr>
<td>PAC</td>
<td>Photoacoustic calorimetry</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
</tr>
<tr>
<td>PNNL</td>
<td>Pacific Northwest National Laboratory</td>
</tr>
<tr>
<td>PTM</td>
<td>Post-translational modification</td>
</tr>
<tr>
<td>QQQ</td>
<td>triple quadrupole mass spectrometer</td>
</tr>
<tr>
<td>ssDNA</td>
<td>single-stranded DNA</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>TIMS</td>
<td>Trapped ion mobility spectrometry</td>
</tr>
<tr>
<td>TOF</td>
<td>Time of flight</td>
</tr>
<tr>
<td>TWIMS</td>
<td>Travelling wave ion mobility spectrometry</td>
</tr>
<tr>
<td>UV-vis</td>
<td>Ultraviolet and visible</td>
</tr>
<tr>
<td>WHIM</td>
<td>Washington University human in mouse</td>
</tr>
</tbody>
</table>
CHAPTER 1: INTRODUCTION
1.1 Overview

Interactions between proteins and DNA control several fundamental biological processes, including gene regulation, replication, repair, transcription, translation and recombination\(^1\). Studying complex formation and conformational changes as a function of binding is a central goal in biochemical and biophysical research. The interaction between such interfaces is crucial to understanding their potential as therapeutic targets and for disease prevention.\(^2\) The basis of these interactions (e.g., association and dissociation) are rooted in the kinetics and thermodynamics of the systems.\(^3\) Data from thermodynamic studies (i.e., molecular modeling) have been successful in providing information on binding free energies. However, because of the highly complex dynamics involved in protein-DNA interactions there is still much ambiguity regarding the kinetic nature of the folding process. As a result, extensive research has been dedicated to instrumental and method development to further probe protein-DNA complex folding and allow detailed insight into structural assembly.

Through the use of various techniques, structural elucidation of biomolecular complexes have grown substantially;\(^4\) however, some significant limitations still exist. Specifically, the heterogeneous nature and dynamic character of most assemblies and their low relative concentrations within physiologically relevant conditions has provided challenges for traditional structural biology tools. That is, high-resolution methodologies (e.g., x-ray crystallography and NMR) dominate the field of structural biology because they excel at revealing atomic level information. Unfortunately, their major limitation is the requirement for large quantities of highly purified and homogenous samples. Even then, such approaches can prove futile because of low sensitivity, lack of quality crystals, or the
size of the molecular assembly. Additionally, in the case of x-ray crystallography, flexible or disordered regions are often excluded and removed so that crystals can form for analysis. Since between 30% and 50% of eukaryotic proteins have unstructured parts, the majority of structural characterizations of eukaryotic proteins is incomplete. Even when these efforts are successful, only a snapshot of the most stable biomolecule or biomolecular complex conformation is described, which may not necessarily represent *in vivo* structures. Thus, the inability to access information regarding kinetic intermediates at native conditions means that a comprehensive view and significant portion of potential knowledge is left inaccessible.

It is important, during the structural characterization of biomolecules, to preserve the biologically relevant concentrations and conditions. Proteins can be heavily influenced by their cellular environment, often being altered to some degree by post-translational modifications (PTMs) to carry out specific functions. Many of these modified or unmodified biomolecules are expressed at very low concentrations (<picomolar), therefore, most analytical and biochemical techniques are not effective at these detection limits or are restricted to one-dimensional type of analysis. For example, methods that can detect compounds will often lack the ability to determine structure. To further advance the field of structural analysis there is a push for multidimensional approaches that can detect low abundant biomolecules followed by identification, quantification and tertiary structure characterization. Comprehensive cellular measurements alone can potentially allow progress in protein biomarker discovery, where distinct molecular signatures can lend insight into the presence of specific disease. Early and effective disease detection provides a greater potential for successful therapeutic applications and such research is well
complemented by structural analysis as protein folding (i.e., the arrangement of amino acids in space) is imperative in the search for drug targets.  

1.2 Analytical Methods for Molecular Ionization, Separation and Characterization

Recent innovations in speed, accuracy, and sensitivity have established mass spectrometry (MS) methods as a key technology within the field of structural biology and complex mixture analysis. Specifically, native MS techniques, which have been developed over the last two decades, permit the structural interrogation of intact biomolecules and biomolecular complexes at biologically relevant conditions, which are not accessible by other methods. For example, electrospray ionization (ESI) is a soft ionization technique that has become a standard method in clinical laboratories. The successful application of ESI by John Fenn’s group in 1988 showed that large proteins could successfully and very efficiently be transferred intact from solution to the gas phase. Fenn’s work opened the door for significantly low detection limits (femtomolar) and larger mass range of analytes (kDa). Also significant was the fact that noncovalent interactions are preserved throughout the ionization process, allowing complexes to be interrogated for stoichiometry studies.

Over the past decade ESI has been combined with ion mobility spectrometry (IMS), coupled to mass spectrometry (IMS-MS) to become an analytical research tool employed in the study of the structure of gas-phase ions, as it provides a more dynamic view and native-like folding information. Used as a gas-phase separation technique, traditional IMS relies on the separation of ions as they drift in a bath of inert neutral molecules (e.g., N₂ or He) under the influence of a weak electric field. The ion’s mobility gives information on their size and shape via the momentum transfer ion-neutral collision cross...
section (CCS) and allows the identification of different ion conformations based on the variation in size. Many different forms of IMS have been used in the analysis of biological molecules, the most common being FAIMS, DT-IMS and TWIMS. While different in the mode of operation, resolution and sensitivity, their implementation has allowed for the study of isotopomers, proteins, protein complexes, folding pathways, unstructured/intrinsically disordered proteins, as well as collisionally activated states of peptides and proteins. In addition to biomolecule activation, variations in starting solvent conditions (e.g., pH, organic, and salt content) have become a traditional approach for studying conformational folding via IMS following the fact that disruption of the hydrogen bonding networks can provide complimentary information on protein composition, connectivity and architecture. In addition to altering starting solvent conditions, post translational modifications (PTMs) or single amino acid mutations can impact the function and folding of many proteins or complexes. The influence of chemical or environmental changes on biomolecule structure have gone relatively undocumented because most methods are insensitive to residue mutations at critical junctions. However, with mass spectrometry coupled experiments, sensitive structural information can be elucidated.

Ion mobility spectrometry has also been extensively employed in areas of trace analysis for the detection of explosives and chemical warfare agents and has proven itself to be an invaluable tool in field of proteomics, genomics, and metabolomics as a result of the ability to differentiate and identify isomeric species in complex samples. Ultimately, in comparison to traditional LC-MS approaches, IMS has the ability to provide
more sensitive detection and higher throughput analysis of biological standards and complex biological matrices for biomarker identification.

In 2011, Dr. Francisco Fernandez-Lima, Dr. Desmond Kaplan and Dr. Melvin A. Park pioneered the development of a new IMS analyzer - Trapped Ion Mobility Spectrometer (TIMS). Since then, collaborative efforts between Dr. Fernandez-Lima research group, Dr. Melvin A. Park at Bruker Daltonics Inc. and others have allowed for major breakthroughs in the field of analytical chemistry. The possibility of decoupling the time domain from the IMS separation allows for the study of conformationally trapped molecular ions in the gas-phase as a function of the desolvation time, temperature and bath gas composition. In the case of structural biology, the gas-phase studies take advantage of the desolvation process to effectively reduce sample complexity, permitting molecular characterization in the absence of bulk solvent. Trapped ion mobility spectrometry-mass spectrometry has proven to consistently achieve high resolution and high mobility separation; it boasts unparalleled resolving power up to ~400,\textsuperscript{54} compared to a typical TWIMS and DT-IMS analyzer which will achieve ~30-60 and ~60-80. The new capabilities for fast separation, identification, and quantification of complex mixtures, and the ability to study biomolecular conformational space for structural elucidation makes TIMS an exceptional platform. In particular, multiple molecules and their conformations can be measured simultaneously, allowing it to be applied to a variety of analytical challenges and questions. Kinetic studies can be performed to elucidate conformational interconversion/intermediates of proteins and their complexes.\textsuperscript{55-62} When complemented with molecular dynamics (MD), or other solution- and condensed-phase techniques, TIMS-
MS has proven itself to be a very powerful instrument for biomolecule structural characterization, and a leading technique in terms of speed and selectivity.

1.3 Dissertation Structure

This dissertation contains some research that has been published in peer-reviewed journals. Specifically, material from chapter 2 was published in the Journal of the American Society for Mass Spectrometry and published on their website (DOI: 10.1007/s13361-017-1787-8). Chapter 3 was published in Analytical Chemistry (Anal. Chem., 2018, 90 (4), pp 2918-2925). Chapter 4 was accepted for publication in the International Journal for Ion Mobility Spectrometry. Chapter 5 was published in Physical Chemistry Chemical Physics (Phys. Chem. Chem. Phys., 2016, 18 (38), pp 26691-26702). We aim to quickly submit chapter 6 to the Journal of the American Chemical Society. Chapter 7 has been submitted for review for publication in Scientific Reports. Chapter 8 is being prepared for submission in Biophysics.

Throughout this dissertation we combine several objectives that aim to understand and develop new TIMS-MS capabilities, including complex mixture separation for discovery and targeted molecular monitoring and quantitation, structural characterization of intrinsically disordered peptides, DNA, proteins, and protein-DNA assemblies. Complemented by extensive solution- and gas-phase experimental and computational approaches, the goal was to pioneer new applications of TIMS for a more efficient and higher throughput methodologies for identification and characterization of biomolecular ions. In Chapter 2, we detail the first report on TIMS-CID-MS capabilities for offline discovery and targeted monitoring of potential peptide biomarkers using data independent mobility fragmentation and quantitation with internal heavy standards. The first analytical
separation of isomeric, middle down, histone tail proteoforms using TIMS-MS is shown in Chapter 3; this chapter also describes the first comparison across FAIMS, TWIMS and TIMS platforms and the analytical potential of using FAIMS and TIMS due to their substantial orthogonality. In Chapter 4, for the first time, TIMS-MS was applied to the study of intrinsically disordered AT-hook peptides (ATHPs 1, 2 and 3) as a function of the time after desolvation, organic content and bath gas composition; this study describes the influence of the molecular environment on the conformational populations of intrinsically disordered peptides. In Chapter 5, the first application of TIMS-MS for the study of DNA conformers, specifically the kinetically trapped intermediate structures of i-motif DNA in their folded and unfolded states is shown. This study provided the intramolecular interactions that are responsible for stabilizing the folding pathway and for the first time, the possibility to manipulate conformational populations prior to TIMS-MS using collisional induced activation (CIA-TIMS-MS). In Chapter 6, we describe for the first time, the interaction of intrinsically disordered peptides with DNA (ATHP with AT-rich DNA hairpin) and demonstrated that the ATHP peptides can effectively bind to both the major and minor DNA grooves. Chapter 7 provided for the first time an analytical workflow that combines single amino acid substitutions with TIMS-MS for direct assignment of intramolecular interactions and cis/trans locations in the ATHPs, as well as primary sequence influence on the binding dynamics to AT-rich DNA. In chapter 8, for the first time, the conformational space and binding dynamics to single stranded DNA and a stem loop DNA of two topoisomerases – E.coli Topoisomerase IA and Vaccinia Topoisomerase IB using TIMS-MS are described.
1.4 References


CHAPTER 2: TOWARDS DISCOVERY AND TARGETED PEPTIDE BIOMARKER DETECTION USING NANOESI-TIMS-TOF MS

This chapter was published in Journal of The American Society for Mass Spectrometry and reproduced with permission.

2.1 Abstract

In the present work, the potential of Trapped Ion Mobility Spectrometry coupled to TOF Mass Spectrometry (TIMS-TOF MS) for discovery and targeted monitoring of peptide biomarkers from human-in-mouse xenograft tumor tissue was evaluated. In particular, a TIMS-MS workflow was developed for the detection and quantification of peptide biomarkers using internal heavy analogs, taking advantage of the high mobility resolution \( (R = 150–250) \) prior to mass analysis. Five peptide biomarkers were separated, identified and quantified using offline nanoESI-TIMS-CID-TOF MS and results were in good agreement with measurements using a traditional LC-ESI-MS/MS proteomics workflow. The TIMS-TOF MS analysis permitted peptide biomarker detection based on accurate mobility, mass measurements, and high sequence coverage for concentrations in the 10-200 nM range, while simultaneously achieving discovery measurements.

2.2 Introduction

The level of chemical complexity during proteomic analysis and the large dynamic range of commonly studied and potential biomarkers represents an analytical challenge that requires the further development of high throughput, orthogonal, reproducible and robust analytical platforms. Nowadays, mass spectrometry based analysis offers an unparalleled, non-targeted, analysis tool for dissecting complex protein samples at the molecular level; however, prior to mass spectrometry analysis, pre-separation techniques, such as high performance liquid chromatography (HPLC) and nano-liquid chromatography (nanoLC), are often required to mitigate matrix effects and to enhance the peak capacity of the analysis. These pre-separation methods have the advantage of reducing problems associated with ion suppression during competitive ionization of complex samples, a
phenomenon that is more typically observed during direct infusion ionization analyses [1, 2]. However, traditional LC-based protocols require long separation times in order to analytically separate most of the compounds of interest, and high solvent volumes per sample; which for large scale profiling, represents a major obstacle in analysis due to added time and cost [3-5]. In addition, these techniques still suffer from poor separation of isobaric species, which significantly challenges protein sequencing and identification using bottom-up approaches. These challenges become major hindrances for the analysis of a complex biological system, such as cancer proteomic samples, which typically contain a myriad of molecular species. In addition, large-scale profiling in bottom-up proteomics is often limited by the sensitivity of the current mass spectrometry instruments to isolate and detect parent and fragment ions during tandem MS analysis of complex mixtures [6]. For example, current bottom-up proteomic strategies require the chemical treatment of samples (i.e., trituration, protein extraction, enzymatic digest) prior to analysis which result in highly complex mixtures which then require further separation and preparation prior to MS analysis [7].

An alternative approach is the use of gas-phase, post-ionization separations such as ion mobility spectrometry coupled to mass spectrometry (IMS-MS), which promises further gains in the speed, sensitivity and selectivity for the analysis of complex biological mixtures [8, 9]. Specifically, the added mobility dimension of separation yields an increase in peak coverage [6, 10-12], a factor that has often inhibited the analysis of complex mixtures with MS-only detection. The IMS-MS coupling readily enhances peptide/protein coverage and identification by allowing more ions, specifically isomers, to be resolved while simultaneously reducing chemical noise [13, 14]. Previous studies have illustrated
the advantages of IMS-MS in terms of profiling mixtures [8, 15-19], making it one of the most powerful platforms for identification and characterization of proteins and peptides in biological samples. Our group has been working on the development of alternative, time-independent IMS approaches based on Trapped Ion Mobility Spectrometry coupled to MS (TIMS-TOF MS and TIMS-FT-ICR MS) for the study and manipulation of gas-phase molecular ions [10, 20-33]. Briefly, the concept behind TIMS is the use of an electric field to hold ions stationary against a moving gas, so that the drift force is compensated by the electric field and ion packets are separated based on their respective ion mobilities [20, 21, 27]. This concept follows the idea of a parallel flow ion mobility analyzer [34], with the main difference that ions are also confined radially using a quadrupolar field to guarantee higher ion transmission and sensitivity [20, 21]. Since the introduction of TIMS-MS in 2011 [20, 21], our group [10, 22-33, 35] and others [8, 36-44] have shown the potential of TIMS-MS for fast, gas-phase separation and for molecular structural elucidation. In particular, we have demonstrated the advantages of TIMS over traditional IMS analyzers for fast screening [22] and targeted [10, 35] analysis of molecular ions from complex chemical mixtures; the study of isomerization kinetics of small molecules [23, 24], peptides [25], DNA [33], proteins [28, 29], DNA-protein complexes and protein-protein complexes in their native and denatured states [32]. In a more recent report, we showed the isomer separation of polybrominated diphenyl ether metabolites using nanoESI-TIMS-TOF MS with mobility resolutions of up to 400 (the highest reported mobility resolution for singly charged species) [30].

Herein, we present for the first time a nanoESI-TIMS-CID-TOF MS workflow, developed for fast, gas-phase ion separation and accumulation, with efforts focused on
targeted quantitative analysis and discovery measurements of breast cancer markers. The ability of TIMS-CID-TOF MS to separate and sequence isobaric peptides in a complex mixture is illustrated. We address typical challenges and targeted and discovery monitoring strategies using isotopically-labeled internal standards for effective peptide identification and sequencing. While LC-TIMS-MS separations has been recently shown in the case of peptide markers [8], the presented workflow targets offline separations in order to shorten the MS analysis time while tailoring the TIMS analysis for high mobility separation and sensitivity.

2.3 Experimental

2.3.1 Tumor protein extraction and tryptic digestion

A patient-derived mouse xenograft model of luminal B human breast cancer - Washington University Human-in-Mouse (WHIM16) was used for all the studies [45]. The WHIM16 xenograft tumor pieces were transferred into pre-cooled Covaris Tissue-Tube 1 Extra (TT01xt) bags (Covaris no. 520007) and processed in a Covaris CP02 Cryoprep device using an impact setting of 3 (all tumor tissue wet weights were less than 100 mg). The tissue powder was then transferred into precooled cryovials (Corning no. 430487). All procedures were carried out on dry ice and liquid nitrogen to maintain tissue in a powdered, frozen state. Approximately 50 mg of WHIM16 tumor tissue was homogenized in 600 µL of lysis buffer (8 M urea, 100 mM NH₄HCO₃, pH 7.8, 0.1% NP-40, 0.5% sodium deoxycholate, 10 mM NaF, phosphatase inhibitor cocktails 2 and 3, 20 µM PUGNAlc). Protein concentrations of tissue lysates were determined by BCA assay (Pierce). Proteins were reduced with 5 mM dithiothreitol for 1 h at 37 °C, and subsequently alkylated with
10 mM iodoacetamide for 1 hour at room temperature in the dark. Samples were diluted 1:2 with Nanopure water, 1 mM CaCl$_2$ and digested with sequencing grade modified trypsin (Promega, V5113) at 1:50 enzyme-to-substrate ratio. After 4 h of digestion at 37 °C, samples were diluted 1:4 with the same buffers and another aliquot of the same amount of trypsin was added to the samples and further incubated at room temperature overnight (~16 h). The digested samples were then acidified with 10% trifluoroacetic acid to ~pH 3. Tryptic peptides were desalted on strong cation exchange (SCX) SPE (SUPELCO, Discovery-SCX, 52685-U) and reversed-phase C18 SPE columns (SUPELCO Discovery, 52601-U) and dried using Speed-Vac.

2.3.2 Tryptic peptide fractionation

The tryptic peptide sample was separated on a Waters reversed phase XBridge C18 column (250 × 4.6 mm, 5-μm and protected by a 4.6 mm × 20 mm guard column) using an Agilent 1200 HPLC System. After sample loading, the column was washed for 35 min with 10 mM triethylammonium bicarbonate, pH 7.5 (solvent A), before applying a 102-min LC gradient in combination with 10 mM triethylammonium bicarbonate, pH 7.5, 90% acetonitrile (solvent B). The LC gradient started with a linear increase to 10% B in 6 min, then to 30% B in 86 min, 42.5% B in 10 min, 55% B in 5 min and 100% solvent B in another 5 min. The flow rate was 0.5 mL/min. A total of 96 fractions were collected into a 96 well plate throughout the LC gradient. These fractions were concatenated into 48 fractions by combining 2 fractions that are 48 fractions apart (i.e., combining fractions #1 and #49; #2 and #50; and so on)[46]. The concatenated fractions were dried in a Speed-Vac and stored at −80°C. Fractions of various volumes were prepared at PNNL based upon BCA analyses to have total peptide concentrations of 0.5 µg/µL and shipped for analysis.
at FIU. Each fraction was selected based on LC-MS/MS analyses conducted at PNNL (to be reported separately) to contain specific peptides to target for identification in those analyses (see Table 1). Heavy standards of the target peptides were purchased from ThermoFisher and used as received. The last residue of the sequence (Arg or Lys) was modified with $^{13}$C6 and $^{15}$N4 or $^{13}$C6 and $^{15}$N2, respectively. Light (non-isotopically labelled) standards of the target peptides were also purchased from GenScript and used without further purification. All samples were diluted with Optima grade 0.1% Formic Acid.

2.3.3 Trapped Ion Mobility Spectrometry – Mass Spectrometry Analysis

Individual fractions, each spiked with the corresponding internal heavy peptide standard, were analyzed by directly infusing the sample via nanoESI into the TIMS-MS spectrometer. A detailed overview of the TIMS analyzer and its operation can be found elsewhere [20, 21, 27]. The nitrogen bath gas flow is defined by the pressure difference between entrance funnel $P_1 = 1.8$-$2.6$ mbar and the exit funnel $P_2 = 0.6$-$1.0$ mbar at ca. 300 K. The TIMS analyzer is comprised of three regions: an entrance funnel, analyzer tunnel (46 mm axial length), and exit funnel. A 880 kHz and 200 Vpp RF potential was applied to each section creating a dipolar field in the funnel regions and a quadrupolar field inside the tunnel. In TIMS operation, multiple ion species are trapped simultaneously at different E values resulting from a voltage gradient applied across the TIMS tunnel. After thermalization, species are eluted from the TIMS cell by decreasing the electric field in stepwise decrements (referred to as the “ramp”) and can be described by a characteristic voltage (i.e., $V_{\text{elution}} - V_{\text{out}}$). Eluted ions are then mass analyzed and detected by a maXis
impact Q-ToF MS (Bruker Daltonics Inc, Billerica, MA). In a TIMS device, the total analysis time can be described as:

\[
\text{total IMS time} = t_{\text{trap}} + (V_{\text{elution}}/V_{\text{ramp}}) t_{\text{ramp}} + \text{ToF} = t_0 + (V_{\text{elut}}/V_{\text{ramp}}) t_{\text{ramp}}
\]  

where, \(t_{\text{trap}}\) is the thermalization/trapping time, \(\text{ToF}\) is the time after the mobility separation, and \(V_{\text{ramp}}\) and \(t_{\text{ramp}}\) are the voltage range and time required to vary the electric field, respectively. The elution voltage was experimentally determined by varying the ramp time for a constant ramp voltage. This procedure also determines the time ions spend outside the separation region \(t_0\) (e.g., ion trapping and time-of-flight). The TIMS cell was operated using a fill/trap/ramp/wait sequence of 10/10/50-500/50 ms. The ToF analyzer was operated at 10 kHz (\(m/z\) 100-3500). The data was summed over 100 analysis cycles yielding an analysis time of ~50 s for the largest trapping times. Mobility calibration was performed using the Tuning Mix calibration standard (G24221A, Agilent Technologies, Santa Clara, CA) in positive ion mode (e.g., \(m/z\) 322, \(K_0 = 1.376 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}\) and \(m/z\) 622, \(K_0 = 1.013 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}\)) [27]. The TIMS operation was controlled using in-house software, written in National Instruments LabVIEW, and synchronized with the maXis Impact Q-ToF acquisition program [20]. A custom-built source using pulled capillary nanoESI emitters was utilized for all the experiments. Quartz glass capillaries (O.D.: 1.0 mm and I.D.: 0.70 mm) were pulled utilizing a P-2000 micropipette laser puller (Sutter Instruments, Novato, CA) and loaded with 10 µL aliquot of the 20x diluted sample solution. A typical nanoESI source voltage of +600-1200 V was applied between the pulled capillary tips and the TIMS-MS instrument inlet. Ions were introduced via a stainless steel inlet capillary (1/16 x 0.020”, IDEX Health
Science, Oak Harbor, WA) held at room temperature into the TIMS cell. Reduced mobility values ($K_0$) were correlated with Collisional cross section ($\Omega$) using the equation:

$$\Omega = \frac{(18\pi)^{1/2} z}{16 (k_B T)^{1/2}} \left[ \frac{1}{m_i} + \frac{1}{m_b} \right]^{1/2} \frac{1}{K_0 N^*}$$

(2)

where $z$ is the charge of the ion, $k_B$ is the Boltzmann constant, $N^*$ is the number density and $m_i$ and $m_b$ refer to the masses of the ion and bath gas, respectively [47]. All resolving power ($R = \Omega/\Delta\Omega$) values were determined from Gaussian peak fits using OriginPro (version 8.0).

**2.3.5 LC-ESI-MS/MS Analysis**

Confirmation studies using tandem mass spectrometry were performed by a QTRAP 5500 Triple-Quadrupole mass spectrometer (AB Sciex, Concord, Ontario, Canada) equipped with a Turbo V ion source (ESI) operated in the positive mode. Solutions of peptides and heavy analogs (5.0 µM) in 50% acetonitrile, 0.1% formic acid in water were directly infused (10 µL/min) into the TurboV ion source. Once suitable species (usually [M+2H]$^{+2}$) were detected in manual tuning mode, automatic optimization was performed of the collision energy (CE), declustering potential (DP) and collision cell exit potential (CXP) to obtain best parameters for MS/MS via collision-induced dissociation (CID). A multiple reaction monitoring (MRM) detection method was thus developed for each peptide and heavy analog, using the two most intense transitions observed for quantitative and confirmation purposes. HPLC separations (40 µL injections) used a reverse phase column (Dionex Acclaim 120 C18 Column, 250x2.1 mm, 5 µm) and a Shimadzu Prominence LC-20AD Ultra-Fast Liquid Chromatograph. Mobile phase
gradient was performed between 0.1% formic acid dissolved in water (mobile phase A) and 0.1% formic acid dissolved in acetonitrile (mobile phase B), all purchased commercially and of Optima LC-MS grade. The auto sampler was kept at 4° C. Analysis was performed at 35° C with a flow rate of 0.80 mL/min, according to the following 11.0 min program: hold 10% B for 0.25 min; ramp to 65% B in 4.5 min; ramp to 98% in 0.1 min; hold for 1.65 min; return to 10% B in 0.5 min; hold for 4 min until end.

2.4 Results and Discussion

Commonly used peptide biomarkers during detection of protein DJ-1 [48], calmodulin [49], parafibromin [50, 51], MAP7 domain-containing protein 1 [52-54] and membrane-associated progesterone receptor component 1 [55, 56] (sequences: DVVICPDASLEDAKK, VFDKDNGYISAAELR, TTILQSTGK, LSASTASELSPK, DFTPAELR, respectively) were used in this study (see Table 2.1). The selection of the targeted peptides was guided towards covering a diverse protein abundance range based on previous analyses of the patient-derived breast cancer mouse xenograft tissue sample (WHIM16) using LC-QQQ by the Clinical Proteomic Tumor Analysis Consortium (CPTAC) [57]. Single peptide standards and their respective heavy versions were analyzed using TIMS-MS in order to determine the charge state distribution (CSD) and collision cross section (CCS) when sprayed from the same starting solvent conditions as those of the WHIM16 tryptic digested fractions. Peptides DFTPAELR and TTILQSTGK showed similar CSDs with the [M+H]+ producing the most abundant signal, while LSASTASELSPK, VFDKDNGYISAAELR, and DVVICPDASLEDAKK showed larger abundance for the [M+2H]+2 charge state (Figure 2.1). In addition to targeting m/z peaks, based on their abundance as a function of the charge state, a second criteria utilized
Figure 2.1. Typical mass spectra and IMS projection plots of the five targeted peptides of interest. Notice the high mobility resolution obtained using nanoESI-TIMS-MS for singly and doubly charged molecular ions.
was the simplicity of the CCS profiles for the [M+H]+ and [M+2H]+² charge states in order to avoid potential interferences. A typical mobility resolving power of R ~ 200 was obtained for the [M+H]+ and [M+2H]+² charge states and CCS values correlate well with previously identified peptide mobility trend lines observed during IMS-MS analysis [58]. Comparison of the IMS profiles of the targeted peptides and the heavy analogs present the same distribution and CCS values (Table 2.1); moreover, an exception to this rule was observed for CAM modified heavy analogs customized to prevent disulfide association. For the latter, the confirmation and quantification was made based on the targeted and heavy analog m/z and CCS value. The power of the TIMS-MS for peptide characterization was further examined by sequencing two of the targeted peptides possessing the same nominal mass (e.g., m/z 948.479 and m/z 948.536 for DFTPAELR and TTILQSTGK, respectively). While traditional proteomics analysis is based on peptide identification using MS/MS strategies, for TIMS-CID-TOF MS the m/z and CCS characterization of the parent ion can be complemented with CID without the need for m/z preselection if separation in the CCS domain is achieved (Figure 2.2). Inspection of the experimental 2D IMS-MS contour plots of the isobaric peptide mixture shows the fragment ions of DFTPAELR and TTILQSTGK peptides falling directly in line with the mobilities of their respective parent ion (Figure 2.2a). The incorporation of IMS prior to CID holds multiple advantages for molecular identification since direct correlation of fragment ions with precursors ion can be performed in the 2D-IMS-MS [59-65]. Analysis using TIMS provided baseline separation of the targeted peptides, where a minimum \( R_{\text{IMS}} \sim 100 \) (i.e., CCS of 295 and 298 Å²) was required (Figure 2.2b). Closer inspection of the product ions (mostly b and y type fragments and some internal fragments) permitted the verification of the peptide sequences
Figure 2.2. a) TIMS-CID-MS of isolated isobaric precursor ions for the DFTPAELR and TTILQSTGK peptides and the corresponding ladder fragmentation pattern. B) Mobility selected MS showing the separation of the precursor ions. C) CID spectra of the fragments corresponding to DFTPAELR and TTILQSTGK.
The advantage of this approach when compared to traditional LC-MS/MS proteomics is that the CCS values (or profiles) of each parent and corresponding fragments are common parameters and can be used as additional identification confirmation. That is, the precursor and product ions will share the same CCS, while characteristic LC elution times may depend on several experimental conditions and may not be as reproducible, nor specific, to a given peptide. However, because the CCS is a property of the peptide parent ion, the possibility to uniquely trap the mobility range of interest in a TIMS analyzer significantly enhances the multiple reaction monitoring capabilities of the TIMS-MS analyzer by ultimately reducing chemical noise and increasing the TIMS selectivity of the parent and fragment ions.

Despite the LC pre-fractionation step, the samples of interest provided highly complex spectra with multiple peaks present at the nominal mass level in the 2D-IMS-MS domain (Figure 2.3a). The 2D IMS-MS contour plots showed that each fraction contained two main trend lines, corresponding to singly and doubly charged species [8]. Closer inspection confirmed that the TTILQSTGK [M+H]+ molecular ion (m/z 948.536) was accompanied by two other compounds within 5 mDa, which are not distinctly separated in the MS domain alone, despite the high resolution of the TOF analyzer (RMS ~20-40k). When combined with TIMS analysis, however, the three signals can be easily separated, distinguished, and identified (Figure 2.3b). Further comparison of the targeted peptide and the corresponding heavy analog IMS projections of TTILQSTGK [M+H]+ (m/z 956.555) confirmed the assignment in the 2D-IMS-MS contour plots (Figure 2.3c).

After verifying the TIMS-MS workflow for high reproducibility and accuracy in measuring and identifying the targeted peptides from the WHIM16 tryptic digested
Figure 2.3. a) Typical 2D-IMS-MS contour plot using nanoESI-TIMS-MS for a fraction containing the target peptide TTILQSTGK. The 2D-IMS-MS profile highlights the complexity of each fraction and shows the charge state specific trend lines. b) The 2D IMS-MS at the level of nominal mass depicts the isomeric interferences in the region of the targeted peptide and heavy analog. c) IMS projection plots for the targeted and corresponding heavy peptide using 5mDa window show identical CCS values.
fractions, the potential for quantitative analysis via TIMS-MS was evaluated. To mimic matrix effects, one of the fractions, with confirmed absence of the target peptide, was spiked with known concentrations of the light and heavy peptide standards. The use of internal heavy standards accounted for variations in the nanoESI spray between experiments and from sample to sample, as well as changes in the spraying conditions as a function of time. Figure 4 shows a linear dependence between the TIMS peak area and the sample concentration for the case of the TTILQSTGK [M+H]^+ peptide (Figure 2.4a), regardless of the TIMS trapping time (e.g., t_{ramp} = 100-500 ms) and analytical ramp slope (Figure 2.4b). The robustness of this TIMS-MS quantitation procedure was also confirmed for all the peptides of interest at the low concentration (e.g., 1, 5, 10 and 20 nM) and the use of internal heavy peptide analogs accounted for all the potential nanoESI spray variability (Figure 2.4c).

The analysis and quantitation of targeted compounds in complex mixtures using direct infusion ESI (and nanoESI) can be subject to ion suppression effects [66, 67]. To further evaluate this consequence, dilution (up to 20 times) of a WHIM16 tryptic digested fraction with known spiked concentrations of targeted and heavy standards showed less than 10% variability in the TIMS-MS quantification results (Figure 2.5a, top). Complementary LC-ESI-MS/MS based traditional proteomic analyses of the same sample showed similar ion suppression effects (Figure 2.5a, bottom).

The ultimate test for the TIMS-MS workflow consisted of blind analysis of WHIM16 tryptic digested fractions known to have the targeted peptides. Briefly, the highest concentration of the targeted peptide observed was for DFTPAELR at 182 ± 7.0 nM (364 fmol/µg), follow by TTILQSTGK peptide at 21.0 ± 5.0 nM (42 fmol/µg),
Figure 2.4. a) Typical TIMS profiles of a pure heavy standard and in-fraction light peptide standard (TTILQSTGK [M+H]+) as a function of the concentration. b) Trapping efficiency of TIMS analysis illustrating that trapping time does not impact the calculated concentration or linear response and c) Measured light peptide concentrations for each targeted peptide, via TIMS-MS, in relation to the heavy analog displaying a linear response.
VFDDKDGNGYISAAEELR peptide at 16.0 ± 4.5 nM (32 fmol/µg), DVVCPDASLEDAKK peptide at 18 ± 5.0 nM (35 fmol/µg) and LSASTASELSPK peptide at 17 ± 4.8 nM (33 fmol/µg). The results of the TIMS-MS quantification procedure and their comparison to LC-ESI-MS/MS based traditional proteomic analyses of the same sample are summarized in Figure 5b and Table 1. Overall, comparable results for targeted peptides per fraction were observed using nanoESI-TIMS-MS and LC-ESI-MS/MS; moreover, it is worth stating that while nanoESI-TIMS-MS was routinely done in five minutes, each LC-ESI-MS/MS analysis typically required 20-25 minutes. In addition, while TIMS-MS measurements were geared toward the separation, identification and quantitation of five targeted peptides, simultaneously, discovery TIMS-MS measurements were collected without compromising the targeted analysis. That is, the TIMS-MS analysis allowed for the identification of multiple tryptic peptides (both targeted and untargeted) from the above five specified proteins. For example, TIMS-MS data analysis revealed 15 to 60% protein sequence coverage, using peptide IDs that were not initially targeted, over the various fractions analyzed (see Appendix 3.2 and Appendix 3.1 in the Supporting Information).

While new biomarker detection is out of the scope the present study, it should be noted that a posteriori screening (or discovery) of potential biomarkers of interest is an inherent potential of the current TIMS-MS workflow. That is, the use of TIMS-MS workflow as a broad measurement technique, concurrently with the targeted quantitative approach, effectively reduces the problems associated with individual analyses and holds the advantage of increasing proteome coverage and new biomarker detection while maintaining speed, accuracy and sensitivity.
Figure 2.5. a) Response curves for the LSASTASELPK [M+2H]+2 peptide standards analyzed in fraction (matrix effect) and in the blank via nanoESI-TIMS-MS and LC-QqQ-MS and b) comparisons of targeted peptide concentrations measured in fractions by nanoESI-TIMS-MS and LC-QqQ-MS.
Table 2.1. Peptide sequence, m/z of light and heavy peptides, ion-neutral collision cross section (CCS) and total in-fraction peptide concentration of the five targeted peptides in the fractions. LC-ESI-MS/MS concentrations using the 535.2 → 187.0 Da (DVVICPDASLEDAKK), 585.7 → 201.1 Da (VFDKDGNGYISAAELR) and 474.9 → 130.1 Da (TTILQSTGK) channels are denoted with an asterisk (*).

<table>
<thead>
<tr>
<th>Peptide</th>
<th>m/z</th>
<th>CCS_{N2} (Å²)</th>
<th>Concentration (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DFTP AELR</td>
<td>948.483 [M+H]^+</td>
<td>298</td>
<td>182 ± 7.0</td>
</tr>
<tr>
<td>DFTP AELR</td>
<td>958.486 [M+H]^+</td>
<td>298</td>
<td></td>
</tr>
<tr>
<td>TTILQSTGK</td>
<td>948.521 [M+H]^+</td>
<td>295</td>
<td>21 ± 5.0</td>
</tr>
<tr>
<td>TTILQSTGK</td>
<td>956.555 [M+H]^+</td>
<td>295</td>
<td>(19 ± 3.0)</td>
</tr>
<tr>
<td>DVVICPDASLEDAKK</td>
<td>801.904 [M+2H]^2</td>
<td>430</td>
<td>18 ± 5.0</td>
</tr>
<tr>
<td>DVVICPDASLEDAKK</td>
<td>834.422 [M+2H]^2</td>
<td>436</td>
<td>(19.0 ± 3.0)*</td>
</tr>
<tr>
<td>LSASTASELPK</td>
<td>595.815 [M+2H]^2</td>
<td>380</td>
<td>17 ± 4.8</td>
</tr>
<tr>
<td>LSASTASELPK</td>
<td>599.824 [M+2H]^2</td>
<td>380</td>
<td></td>
</tr>
<tr>
<td>VFDKDGNGYISAAELR</td>
<td>877.925 [M+2H]^2</td>
<td>452</td>
<td>16 ± 4.5</td>
</tr>
<tr>
<td>VFDKDGNGYISAAELR</td>
<td>882.944 [M+2H]^2</td>
<td>452</td>
<td>(15 ± 3.0)*</td>
</tr>
</tbody>
</table>

2.5 Conclusions

The demand for fast, accurate, and sensitive analytical tools for the detection and quantification of biomolecules is increasing as a way to offset the challenge of drug discovery and biomarker identification. While several strategies have been developed, some current efforts are focused on reducing sample preparation and analysis time, while increasing detection limits and peak capacity by using complementary, orthogonal separation techniques. In the present work, the concepts of characterizing proteomes using offline nanoESI-TIMS-MS were evaluated by performing targeted and discovery analysis of cancer biomarkers from a human-in-mouse xenograft tumor tissue. Results showed that targeted peptide separation, identification and sequencing can be performed based on
accurate mobility, mass and fragmentation pattern measurements, and that peptide quantitation can be routinely achieved utilizing heavy peptide analogs as internal standards. The capacity of the TIMS analyzers for selective mobility trapping with high resolving power increases the selectivity and sensitivity of the analysis and provides unique advantages for offline targeted studies compared with traditional LC-ESI-MS-MS proteomic strategies. A good agreement was obtained between the quantitation using offline nanoESI-TIMS-MS and LC-ESI-MS/MS. This work serves as a stepping-stone and proof of concept for quantitative proteomics of targeted peptides without the need for online LC separation, an aspect that can significantly lower costs and lead to increased sample throughput during targeted biomarker detection.

2.6 Acknowledgements

This work was supported by the National Cancer Institute (NCI) through Leidos Biomedical Inc. (No. 14X2333 to FFL) and National Institute of General Medicine (Grant No. R00GM106414 to FFL). The PNNL efforts were performed in the Environmental Molecular Sciences Laboratory, a U.S. Department of Energy (DOE) national scientific user facility located at PNNL in Richland, Washington. The author would like to acknowledge the support from Kristin E. Burnum-Johnson and Marina A. Gritsenko for the liquid chromatography fractionation and peptide selection per fraction. PNNL is a multi-program national laboratory operated by Battelle Memorial Institute for the DOE under Contract DE-AC05-76RL01830.
2.7 References


34. Zeleny, J.: On the ratio of velocities of the two ions produced in gases by Röngten radiation, and on some related phenomena. Philos. Mag. 46, 120-154 (1898)


CHAPTER 3: LINEAR AND DIFFERENTIAL ION MOBILITY SEPARATIONS OF MIDDLE-DOWN PROTEOFORMS

This chapter was adapted with permission from Garabedian, A., et al. *Analytical Chemistry*, 2018, 90 (4), 2918-2925. Copyright 2018 American Chemical Society.
3.1 Abstract

Comprehensive characterization of proteomes comprising the same proteins with distinct post-translational modifications (PTMs) is a staggering challenge. Many such proteoforms are isomers (localization variants) that require separation followed by top-down or middle-down mass spectrometric analyses, but condensed-phase separations are ineffective in those size ranges. The variants for “middle-down” peptides were resolved by differential ion mobility spectrometry (FAIMS), relying on the mobility increment at high electric fields, but not previously by linear IMS on the basis of absolute mobility. We now use complete histone tails with diverse PTMs on alternative sites to demonstrate that high-resolution linear IMS, here trapped IMS (TIMS), broadly resolves the variants of ~50 residues in full or into binary mixtures quantifiable by tandem MS, largely thanks to orthogonal separations across charge states. Separations using traveling-wave (TWIMS) and/or involving various time scales and electrospray ionization source conditions are similar (with lower resolution for TWIMS), showing the transferability of results across linear IMS instruments. The linear IMS and FAIMS dimensions are substantially orthogonal, suggesting FAIMS/IMS/MS as a powerful platform for proteoform analyses.

3.2 Introduction

As the proteomics tools mature, the front line moves to characterizing proteoforms and revealing the activitymodulating impacts of post-translational modifications (PTMs). Many proteoforms feature different numbers or types of PTMs, detectable by mass spectrometry (MS) on the basis of the mass increment. Others are isomers with identical PTMs on different residues. Such “localization variants” are individually distinguishable by unique fragments in tandem MS, particularly employing electron
transfer dissociation (ETD) that severs the protein backbone while retaining weaker PTM links.\textsuperscript{3,7,9–11} The conundrum is that multiple variants frequently coexist in cells, but MS/MS cannot disentangle mixtures of more than two, as those with PTMs on internal sites yield no unique fragments.\textsuperscript{12,13} This calls for variant separation at least to binary mixtures before the MS/MS step.\textsuperscript{12–14} Liquid chromatography (LC) could resolve some variants for peptides in the “bottom-up” mass range (<2.5 kDa) usual for tryptic digests,\textsuperscript{15} but not “middle-down” peptides (2.5–10 kDa) or intact proteins. Unfortunately, splitting proteins into peptides using proteases precludes global PTM mapping by obliterating the proteoform-specific connectivity information between the modified peptides.\textsuperscript{9,16}

This problem is most prominent for histones that combine exceptional importance to life with great diversity of PTM types and sites.\textsuperscript{9,16–26} Histones (H2A, H2B, H3, and H4) consisting of ∼100–140 residues are nucleosome core particles—the spools that store the DNA in cell nuclei and regulate chromatin structure and function through dynamic reversible PTMs including methylation (me), trimethylation (me3), acetylation (ac), phosphorylation (p), and others.\textsuperscript{9,14,16–26} Permuting their order and modulating the site occupation levels in this ”histone code” drastically alters the activity of the whole genome, defined chromatin domains, genomic regions, and/or individual genes. Nearly all PTMs in histones are on the enzymatically cleavable N-terminal domains (“tails”) protruding from the nucleosome.\textsuperscript{16,24,25} The H3 tail of ∼50 residues is cleavable by the endoproteinase Glu-C, and its characterization approaches that of intact histone.\textsuperscript{23–25}

A growing alternative to LC is ion mobility spectrometry (IMS), which is based on the ion transport in gases driven by an electric field,\textsuperscript{27,28} with the key benefits of speed and distinct (often superior) selectivity. Linear IMS\textsuperscript{27} measures the absolute ion mobility ($K$)
at low field strength \((E)\), whereas differential or field asymmetric waveform IMS (FAIMS)\(^{28}\) relies on the difference between \(K\) at high and low \(E\) elicited by an asymmetric waveform. That \(\Delta K\) is less correlated\(^{29,30}\) to the ion mass \((m)\) than \(K\), rendering FAIMS more orthogonal to MS than linear IMS is—by about 4-fold for many biomolecular classes comprising peptides.\(^{31,32}\) Therefore, FAIMS commonly separates isomers better than linear IMS of the same resolving power \((R)\), including peptides with sequence inversions\(^{32}\) and localization variants with diverse PTMs.\(^{14,33-37}\) In particular, complete histone tails and their segments involving various PTMs and sites have been resolved.\(^{14,34,35}\)

Linear IMS separations of such variants were limited to phosphopeptides under \(\sim 1.5\) kDa.\(^{38,39}\) Expanding this capability to larger peptides and smaller PTMs is topical, as linear IMS platforms can be more sensitive than high-definition FAIMS. They also determine the collision cross section \((\Omega)\) unavailable from FAIMS,\(^{27,28}\) which may help understanding and predicting the PTM-controlled differences in the stability of peptide folds with implications for activity in vivo.\(^{40}\) Here we deploy linear IMS in the commercial traveling wave (TWIMS)\(^{41-47}\) and trapped (TIMS)\(^{48-53}\) platforms to separate localization variants for complete histone tails. The instrumental resolving power of TIMS can exceed 300, far over \(\sim 50\) with TWIMS.\(^{42,53,54}\) However, \(R\) for proteins in linear IMS has been capped at \(\sim 30\) by peak broadening due to conformational multiplicity.\(^{55,56}\) A critical advantage of TIMS is achieving for some protein conformers the same peak width as that for small peptides, as in FAIMS.\(^{52,57}\)

We utilize the H3 variants investigated\(^{14}\) by FAIMS to compare performance and evaluate the orthogonality between two dimensions for middle-down proteoforms. We also
inspect the correlation between TWIMS and TIMS to gauge the transferability across linear IMS platforms.

### 3.1 ESI-TWIMS-MS Instrumentation

In TWIMS, ions “surf” along a stack of addressable electrodes that create an axial wave with spatial period $L$ and radially confining rf field. We employed the Synapt G2 system (Waters, Milford, MA), where exiting ions are injected into an orthogonal reflectron time-of-flight (ToF) stage (resolving power $R_{MS}$ of 20000) and registered. As isobaric ions have the same velocity under vacuum, their temporal separation at the detector equals the difference of transit times ($t_T$) through the IMS stage determined by mobility. Unlike the case with drift-tube (DT) IMS, the $t_T(K)$ function is not reducible to closed form. Hence, extracting $K$ (to deduce the ion geometries by matching calculations or preceding measurements) necessitates a multipoint calibration using standards and is especially challenging for macromolecules because variable source conditions and field heating prior to and during IMS separation affect the geometries of pertinent standards. Still, Synapt has become the prevalent IMS/MS platform in proteomics and structural biology.

<table>
<thead>
<tr>
<th>PTM</th>
<th>Positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>me</td>
<td>K4, K9, K23</td>
</tr>
<tr>
<td>me3</td>
<td>K4, K9, K23, K27, K36</td>
</tr>
<tr>
<td>ac</td>
<td>K9, K14, K18, K27, K36</td>
</tr>
<tr>
<td>p</td>
<td>T3, T6, S10, S28, Y41</td>
</tr>
</tbody>
</table>
biology. Here we look at the variant separations without assigning structures; thus, the \( t_T \) scale was not converted into \( \Omega \) terms. However, as in FAIMS, an internal calibrant—a peptide of similar mass (insulin, 5.8 kDa)—was spiked to validate consistency and accurate spectral comparisons. The spectra were linearly scaled to align the \( t_T \) for calibrant peaks.

The key parameters of TWIMS are peak voltage \((U)\), wave speed \((s)\), and the buffer gas identity, pressure \((P)\), and temperature \((T)\). Separations are mainly governed by the ion drift velocity at wavefront relative to its speed:

\[
c = \frac{KU}{sL} = \frac{K_0P_0TU}{(PT_0Ls)}
\]

(1)

where the subscript “0” denotes quantities at STP (including the reduced mobility \( K_0 \)). The resolution is maximized at some \( c \); therefore, the variants with unequal mobility (reflecting different geometries and/or charge states \( z \) involved) may separate best in differing regimes. However, the said maximum is near-flat over \( c \approx 0.3–0.8 \), allowing \( \sim 4 \)-fold variation of \( K \) with little resolution loss. The mobilities of large peptides with \( z > 3 \) depend on \( z \) weakly, as charging induces unfolding (elevating \( \Omega \)), and the mobility range for conformers at a given \( z \) is limited as well. Hence, peptides in different charge states can often be run together. Ions in TWIMS are materially field-heated, which may isomerize flexible macromolecules with mobility shifting over time. As reducing \( c \) slows the ion transit, that effect may influence the variant resolution for large peptides apart from its dependence on \( c \) for fixed geometries. Therefore, we have repeated analyses over the practical \( c \) range using \( s \) values of 650, 1000, and 1900 m/s at \( U = 40 \) V with \( \text{N}_2 \) gas at \( P = 2.2 \) Torr. The gas flows were 0.5 L/min \( \text{N}_2 \) to the source (at 100 °C), 0.09 L/min \( \text{N}_2 \) to the (unheated) cell, and 0.18 L/min \( \text{He} \) to the helium gate in front of it.
The ESI source with a 32-gauge steel emitter was run with the infusion flow rate of 20 μL/min, capillary at 2.8 kV, and sampling cone at 45 V. The geometries of protein and peptide ions from ESI may keep the memory of folding in solution and thus depend on the

Figure 3.1. TWIMS spectra for histone tail variants with z = 6 - 11, measured with solvent (i) using s = 650 m/s
solvent,\textsuperscript{61,62} modifying the variant resolution. To assess that, we tested 0.1 μM peptide solutions in (i) default 50/49/1 MeOH/H\textsubscript{2}O/acetic acid (pH = 3), (ii) predominantly organic 90/9/1 MeOH/H\textsubscript{2}O/acetic acid, (iii) extremely acidic 97/3 H\textsubscript{2}O/formic acid (pH 1.5), and (iv) 99/1 isopropyl alcohol/acetic acid.

The apparent TWIMS resolving power is $R = t_T/w$, where $w$ is the full peak width at half-maximum. The true $R$ is greater by the logarithmic derivative of $t_T(\Omega)$, which is $\sim 2$ over the practical $c$ range where $t_T(\Omega)$ is near-quadratic.\textsuperscript{42,54}

### 3.2 nESI-TIMS-MS Instrumentation

In TIMS,\textsuperscript{48-53} ions radially confined by rf field in a straight section of electrodynamic funnel are axially stratified by flowing gas (sucked by MS vacuum) and retarding longitudinal dc field $E$. As $E$ is ramped down, the flow pushes ions in order of decreasing mobility to the MS stage—here, an Impact Q-ToF (Bruker, Billerica, MA) with $R_{MS} = 30000$ (at 10 kHz frequency). Separations depend on the gas flow velocity ($v_g$), trapping voltage ($V_{ramp}$), base voltage ($V_{out}$), and ramp duration ($t_{ramp}$). Isomers emerge at elution voltages ($V_{elution}$) given by

$$K = v_g/E = A/(V_{elution} - V_{out})$$

where $A$ is a constant fit using internal calibrants\textsuperscript{52} (here the Agilent Tuning Mix components with $K_0$ values of 1.013, 0.835, and 0.740 cm$^2$/(V s) for respective $m/z$ values of 622, 922, and 1222) with $V_{elution}$ for each determined from the analysis time corrected for delay after elution (using varying ramp times).\textsuperscript{50} All electrode voltages were managed by custom software synchronized with the MS platform controls. The rf amplitude was 250 $V_{PP}$ at 880 kHz frequency. The typical dc voltages were: inlet capillary at 40 V, funnel entrance at 0 V, $V_{ramp} = -(50–200)$ V, and $V_{out} = 60$ V. Lower scan rates ($Sr = \Delta V_{ramp}/t_{ramp}$)
improve the resolving power; we generally adopted $Sr = 0.3$ V/ms. The overall fill/trap/ramp/wait sequence was 10/10/(100–500)/50 ms. With summation of 100 cycles, the longest acquisition took $\sim 1$ min.

The buffer gas was N$_2$, with $v_g$ set by the difference between pressures at the funnel entrance (2.6 Torr) and exit (1.0 Torr). Ions were generated by a pulled-tip nESI emitter (biased at 700–1200 V) from 10 µL sample aliquots (0.5 µM in (v) 50/50 MeOH/H$_2$O or (vi) H$_2$O) and introduced into the TIMS device via an orthogonal unheated metal capillary. More details on the nESI/TIMS hardware and mobility calibration are given in the Supporting Information.

The measured mobilities were turned into $\Omega$ using the Mason-Schamp formula:

$$\Omega = \frac{3}{16} \left[ \frac{2\pi}{k_BT} \left( \frac{1}{m} + \frac{1}{M} \right) \right]^{1/2} \frac{ze}{N_K}$$

where $z$ is the charge state, $e$ is the elementary charge, $k_B$ is the Boltzmann constant, and $N$ and $M$ are the gas number density and molecular mass, respectively. The resolving power is $R = \Omega/w$.

### 3.3 Results and Discussion

#### 3.3.1 TWIMS Separations

Using solvent (i), we observed all variants in $z = 5–11$. This range is lower than the $z = 8–12$ examined in nESI/FAIMS experiments with the same solvent, which reflects a different ion source and greater instrumental sensitivity that allows collecting IMS data for more states (although with low signal at $z = 5$).

Most IMS spectra were obtained using the default $s = 650$ m/s (Figure 3.1). Each variant exhibits one defined peak in $z = 10, 11$ but up to three (fully or partly resolved)
peaks in \( z = 6–9 \). This suggests a gradual transition from compact conformers at low \( z \) to unfolded conformers at high \( z \) over several charge states exhibiting rich structural heterogeneity, ubiquitous for proteins.\(^{56,59}\) As the scaling\(^ {42} \) of \( t_T \) as \( \sim \Omega^2 \) renders \( \Omega \) about proportional to \( z(t_T)^{1/2} \) over the practical \( t_T \) range, we can estimate relative \( \Omega \) with no scale anchoring (Figure 3.2 and Appendix 3.2). The S shape of these plots with a jump between two trend lines for all variants confirms unfolding at intermediate charge states. The apparent \( R \) is 29–33 for all PTMs (average over variants and charge states) and 30–34 in \( z = 7 \) and 9–11 (average over variants and PTMs). In \( z = 8 \), the slightly wider peaks and lower \( R = 27 \) likely reflect unresolved conformers broadening the peaks in unfolding region. Hence, the performance is consistent across PTMs, their locations, and charge states.
The spectra for variants in many charge states significantly differ, but rarely enough for satisfactory resolution. The greatest separation is for me3 tails, proven using the mixtures of two to five variants (Appendix 3.3a-d). The best resolution is in \( z = 6, 8, 9 \): at the peak apexes, K23me3 is largely resolved from all but K27me3 as 8+ ions and all but K36me3 as 9+, K27me3 is largely resolved from all but K23me3 or K36me3 as 8+, and K36me3 is baseline-resolved from others as 6+ and 9+. K9me3 is filtered from others in \( z = 10, 11 \) (not at the apex). As MS/MS can fully characterize binary variant mixtures, this partial resolution helps more than may seem: for example, one can use 10+ or 11+ to detect and reasonably quantify K9me3, 8+ for K27me3 (in K27me3/K36me3 mix), and 9+ for K23me3 (in K23me3/K36me3 mix), while the K4me3 and K36me3 variants with PTMs on bookend sites need no separation. This strategy demands no prior knowledge of the IMS spectra for each variant, although that would accelerate analyses by revealing the optimum drift times and charge states.

This successful separation was limited to the me3 case. For the isobaric acetylation, no variant is fully resolved in any state. The K9ac and K36ac are filtered in 10+ at the longest and shortest \( t_R \), respectively (with large signal loss), but separating those “bookend” variants is not crucial. K14ac is enriched at the lesser peak in 9+, but intense contamination by other variants makes that of little utility. The situation for phosphorylation is more promising. One can cleanly filter the Y41p variant at its peak apex in 7+ and T3p and S10p (away from apexes) in respectively 11+ and 10+, and T6p/S28p mix near the apex of S28p in 6+ (the S10p contribution there would not compromise the analysis for T6p and S28p with occupied external sites). For single methylation with just three variants here, the major task is separating K9me with PTM in the middle. That is feasible (a bit off apex) in 10+
and 11+, and the K4me variant can be filtered (away from the apex) in 10+. The profile for K23me differs from those for K4me and K9me in 8+ and 9+ substantially, but not enough for clean filtering. The separations for p and me variants are also verified using selected mixtures (Appendix 3.3e-f).

The peak pattern in Figure 3.1 is consistent over the practical wave speed range: raising $s$ from 650 to 1000 and 1900 m/s increases $t_T$ from 4–7 to 6–10 and 10–25 ms without significantly moving the relative peak positions (Figure 3.3 and Appendix 3.4). To quantify, the $t_T$ sets at $s$ values of 650 and 1000 m/s are correlated with $r^2$ (average over all charge states) of 0.95 for ac and 0.85 for me3, where the transitions between major conformers at some $z$ interfere with correlation (Appendix 3.5). The respective $r^2$ values for pairs at $s = 1000$ and 1900 m/s decrease to still high 0.90 and 0.79 (excluding one outlier). Hence, the ion geometries are largely conserved between $\sim$5 and $\sim$20 ms. The resolving power is unchanged at $s = 1000$ m/s (apparent $R$ of 29–35 in $z = 7$, 9–11 and $R = 25$ in $z = 8$ upon averaging over all me3 and ac variants), but drops at $s = 1900$ m/s (to $R = 17–28$ in $z = 7$, 9–11 and $R = 14$ in $z = 8$). Thus, the variant resolution at $s = 1000$ m/s is close to that at $s = 650$ m/s but deteriorates at $s = 1900$ m/s outside the optimum range.$^{42}$ Substitution of ESI solvent has minor effects on IMS spectra in any given charge state (Appendix 3.6). This agrees with the analyses$^{64}$ of unmodified histone tails using Synapt G2, where the mobilities at fixed $z$ were the same with solvent pHs of 2 and 6.5. More acidic or organic media favor higher $z$ as anticipated,$^{64,65}$ and solvents (ii) and (iii) produced me3 variants in $z = 12$ observed$^{14}$ in FAIMS. However, we saw no significant variant resolution for 12+ ions (Appendix 3.7). Hence, the variant separations by ESI-TWIMS are independent of the source and kinetic factors, likely reflecting the equilibrium
ion geometries formed in the desolvation region. Then overcoming insufficient variant resolution requires IMS of higher resolving power, such as TIMS.

### 3.3.2 TIMS Separations

We observed \( z = 6–11 \) for all PTMs (K4me3 and K27me3 were not studied because of sample shortage). The resolving power for base peaks at \( Sr = 0.3 \ V/\text{ms} \) is \( \sim 80–280 \), with a mean of \( \sim 150–170 \) for each PTM. The overall average \( (R = 167) \) is \( \times 5 \) that with TWIMS \( (R = 32) \), yielding multiple (up to \( \sim 10 \)) substantial peaks for all variants in each \( z \) except 6 and 10 (Figure 3.4 and Appendix 3.1). These metrics match those for multiply charged unmodified peptides.\(^6\) We now note no drop of \( R \) in \( z = 8 \): instead of peak broadening, multiple conformers produce rich spectra for all variants. The \( \Omega \) values increase at higher \( z \) due to unfolding, and relative \( \Omega \) values match those estimated from TWIMS data (Figure 3.2 and Appendix 3.2). This validates our approximation to obtain the relative \( \Omega \) from raw TWIMS spectra and points to similar ion geometries in the two separations.

---

**Figure 3.3.** TWIMS spectra for K27me3/K36me3 mix \( (z = 9) \) measured with solvent (i) depending on the waveform speed (solid black lines), with fits by scaled individual traces (colored lines) and their computed sum (dotted lines). Data for other speeds and mixtures are in Appendix 3.4.
Figure 3.4. TIMS analysis of histone tail variants: spectra (cross section scale) for z= 6–11 (with solvent (v), tramp = 500 ms).

With the TIMS residence time of \(\sim 40\text{–}400\) ms (depending on \(t_{\text{ramp}}\)), even the shortest is much beyond the longest in TWIMS. Gas-phase protein conformations may evolve over time, specifically on the \(\sim 5\text{–}500\) ms scale relevant here.\(^{67,68}\) Present TIMS experiments employed soft ion injection without activation. However, the IMS spectra for all variants and charge states do not significantly depend on \(t_{\text{ramp}}\) or solvent (v) versus (vi) (Figure 3.5 and Appendix 3.8). Therefore, we focus on the data obtained at maximum resolution (\(t_{\text{ramp}} = 500\) ms) using solvent (v), which provides a higher and more stable ion signal.
The three \textit{me}3 variants can be largely separated using $z = 6$–9, 11 (Figure 3.4). One can filter K36me3 from K9me3 and K23me3 best at the major peak \textit{c} in 6+ and lesser \textit{a} in 9+, largely K23me3 from others at the major peaks \textit{e} in 8+ and \textit{b} in 9+, and readily K9me3 from K36me3 in $z = 6, 8, 9, 11$. Resolving K9me3 from K23me3 is difficult: the best outcome is a $\sim 3\times$ enhancement in 8+ at the major peak \textit{d} or \textit{e}. However, separation to the binary mixtures (by resolving the K9me3/K23me3 mix and K36me3) is trivial. As seen in DTIMS and FAIMS analyses,\textsuperscript{14,38} the spectra are “quantized”: most variants exhibit features at discrete $\Omega$ bands (labeled in Figure 3.4) in different proportions. This suggests a set of energetically competitive folds persisting across variants, with relative energies and thus populations dependent on the PTM position.

Despite many more features, these separations track the order and often the relative spread of cross sections found in TWIMS (Figure 3.1): K9me3 $\leq$ K23me3 $<$ K36me3 in 6+, similar $\Omega$ values for leftmost peaks with features \textit{c}, \textit{d} for K36me3 and (with higher $\Omega$) \textit{d} for K23me3 in 7+, K36me3 $<$ K23me3 $<$ K9me3 for major peaks in 8+, and K9me3
< K23me3 ≤ K36me3 for those in 11+. The starkest similarity is in 9+: here K9me3 has one major peak d with feet b and c, K23me3 has three peaks (largest b, smallest c, and medium d), K36me3 has two intense peaks (a and larger c), and the overall order is K36a < K23b ~ K9b < K36c ≤ K23c < K9d < K23d. The only difference is that in 10+ all variants coincide in Figure 3.4 but K9me3 lies to the left of others in Figure 3.1.

The results for other PTMs are similar. With acetylation (Figure 3.4), there is modest separation in 6+, but K9ac and K18ac are well-resolved from K14ac and K27ac (and vice versa) at the peak apexes in 7+. The blow-up of conformational multiplicity in 8+ obstructs separations, but K27ac is filtered from others at f. The 9+ state permits excellent resolution of K14ac from others at the major peak d and intense e (and vice versa at the major peaks for others a, b, c) and of K9ac at b from K14ac and K27ac. Each variant exhibits one major peak in 10+ as with the me3 case, but here those are dispersed enough to resolve K9ac and K36ac from others at the apexes. In 11+, all variants are similar except K36ac filtered at the major peak a. These properties permit multiple protocols to quantify all variants in a mixture. The optimum may be to isolate K9ac in 10+, K14ac in 9+, K27ac in 8+, and K36ac in 10+ or 11+ (not truly necessary for the bookends K9ac and K36ac). K18ac is not resolved in any state individually but is resolved to binary mixtures (K9ac/K18ac at the peak apex in 7+ and K18ac/K27ac right of the c apex in 9+), allowing redundant quantification by ETD. The order of peaks across charge states also correlates with TWIMS data. For example, that in 10+ is K36ac < K18ac ≤ K27ac < K14ac < K9ac in TIMS and similar K36ac < K27ac ≤ K18ac = K14ac < K9ac in TWIMS (Figure 3.1).

With phosphorylation (Figure 3.4), one can pull out (at apexes) S28p and Y41p in 6+, T3p in 10+ and 11+, and S10p in 10+. As with ac variants, here one (T6p) is not cleanly
resolved in any $z$ but is filtered in T6p/S28p mix at the apex in 6+ and T6p/S10p mix at the apex in 10+ (best) and peak i in 8+. Hence, all variants are quantifiable employing ETD. The correlation with TWIMS data is clear: e.g., the peak order (Figure 3.1) is consistently Y41p < S10p < T3p < T6p < S28p in 6+ and T3p < T6p < Y41p < S28p < S10p in 10+. As with TWIMS, the separations projected from individual spectra were confirmed using binary mixtures (Appendix 3.9).

With me variants, the spectra in $z = 6$–8 provide only a limited separation (Figure 3.4). We can filter K4me at the major peak apex in 10+ and (less cleanly) K23me at peak a in 9+. The K9me is filtered from K4me right of the apex in $z = 10$ and (not cleanly) from K23me on the left of the major peaks in 6+ or 11+. Thus, each variant can be filtered as an individual or as a dominant component of binary mixtures. The correlation with TWIMS data is seen from the peak order K9me < K4me < K23me in 11+ or intense peaks on the left for only K23me in 8+ and 9+ (Figure 3.1).

### 3.3.3 Correlations between Separation Dimensions

The analyses of the same peptide set in FAIMS\textsuperscript{14} and two linear IMS systems allow exploration of pairwise correlations between separations within and between those dimensions: across charge states in TWIMS and TIMS and for the same species in the TWIMS/TIMS/FAIMS space.

Separations of all variants in TWIMS notably differ across charge states. This may be quantified via pairwise linear correlation between separation parameter sets\textsuperscript{14,34} Here, the mean $r^2$ values for $t_r$ correlations over $z = 5$–11 (Appendix 3.10) equal 0.23, 0.24, and 0.25 for me3, ac, and p variants, respectively (with 21 pairs each). The values for $\Omega$ in
TIMS are the same: 0.23 (ac variants) and 0.24 (p variants) for \( z = 6, 7, 10, 11 \) with single dominant peaks (Appendix 3.11) and 0.26 and 0.18, respectively, if we add \( z = 8, 9 \) using Table 3.2. Summary of Linear Correlations between Separations (averaged over all PTMs and charge states): \( r^2 \) values with Standard Errors of Mean.

<table>
<thead>
<tr>
<th></th>
<th>TWIMS (( z_1 ))</th>
<th>TIMS (( z_1 ))</th>
<th>FAIMS (( z_1 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>TWIMS (( z_1 ))</td>
<td>0.91 ± 0.03^a</td>
<td>0.52 ± 0.10^c</td>
<td></td>
</tr>
<tr>
<td>TWIMS (( z_2 ))</td>
<td>0.24 ± 0.04^b</td>
<td></td>
<td>0.22 ± 0.05^f</td>
</tr>
<tr>
<td>TIMS (( z_1 ))</td>
<td>0.86 ± 0.05^c</td>
<td></td>
<td>0.52 ± 0.11^g</td>
</tr>
<tr>
<td>TIMS (( z_2 ))</td>
<td>0.22 ± 0.03^d</td>
<td>0.22 ± 0.04^b</td>
<td></td>
</tr>
<tr>
<td>FAIMS (( z_2 ))</td>
<td></td>
<td>0.25 ± 0.05^l</td>
<td></td>
</tr>
</tbody>
</table>

a. In TWIMS at \( s = 650 \) vs. 1000 m/s  
b. In TWIMS for same peptides in different \( z \)  
c. For same ion species in TWIMS vs. TIMS  
d. In TIMS for same variants in different \( z \)  
e. For same ion species in TWIMS vs. FAIMS (8 pairs)  
f. For variants in TWIMS vs. same variants with different \( z \) in FAIMS (24 pairs)  
g. For same ion species in TIMS vs. FAIMS (8 pairs)  
h. For variants in TIMS vs. same variants with different \( z \) in FAIMS (24 pairs)  
i. In FAIMS for same variants in different \( z \) (30 pairs).^{14}

The aggregate \( r^2 \) over all PTMs is 0.24 ± 0.04 standard error (for 63 pairs) with Synapt and likewise 0.22 ± 0.03 with TIMS, also equal to 0.25 ± 0.05 (for 30 pairs with \( z = 8–12 \) for me3, ac, and p variants) with^{14} FAIMS (Table 3.2). This manifests an essentially perfect orthogonality across charge states, previously demonstrated in FAIMS^{14,34} but not linear IMS separations of any PTM localization variants.

We can also quantify the correlation between TWIMS and TIMS seen in comparisons of cross sections (Figure 3.2) and spectra (Figure 3.1 and 3.4) best for ac and p variants with five \( t_\text{f} \) and \( \Omega \) points. Calculations for \( z = 8, 9 \) are complicated by multiple intense features in both data sets that need integration; therefore, we restricted the comparison to \( z = 6, 7, 10, 11 \) with at most two major peaks. The resulting \( r^2 \) values
(Appendix 3.12) are 0.7–1.0 (mean 0.76) for ac and 0.9–1.0 (mean 0.95) for p variants (higher $r^2$ values for the latter reflect a greater variant separation diminishing the relative random error of peak spacings). These values with aggregate $r^2 = 0.86 \pm 0.05$ (Table 3.2) show strong correlation, especially as we ignored the smaller features in TIMS spectra and $t_T$ is not proportional to $\Omega$. The accord between TWIMS and TIMS data despite dissimilar ESI and ion heating regimes and $\sim 50\times$ longer separation in TIMS shows the ion geometries conserved over $\sim 5$–300 ms and supports the formation of equilibrium conformers in the source. The present similarity between TWIMS and TIMS separations mirrors that for peptides with D/L residue swaps,\textsuperscript{66} though just two epimers per peptide there allowed no $r^2$ values.

This orthogonality of separations across charge states, their number generated by ESI, and impressive resolving power enable TIMS to disentangle all variants tried to at least the binary mixtures. That said, separation to individual variants would be beneficial. Also, the histone stoichiometries have up to $\sim 50$ known variants,\textsuperscript{69,70} with further less abundant variants likely to be discovered. Fully characterizing such complex endogenous samples involving spectral congestion requires yet greater peak capacity (pc) that could come from 2-D FAIMS/IMS separations, depending on the orthogonality between dimensions.

The complementarity of FAIMS and linear IMS separations of histone tails is evident from different loci of variant resolution across charge states. For example, that for me3 variants maximizes for $z = 8, 9$ in TWIMS (Figure 3.1) and TIMS (Figure 3.4) vs 10 and 11 in FAIMS.\textsuperscript{14} Within a given state, some variants resolved by FAIMS may coelute in TIMS and vice versa. For instance, in $z = 10$, the K18ac and K27ac merged in TIMS are
separated by the FAIMS baseline,\textsuperscript{14} whereas TIMS partly resolves K14ac and K27ac merged in FAIMS.\textsuperscript{14} Broadly, the FAIMS dimension is correlated to TWIMS/TIMS with mean $r^2$ (over $z = 8–11$) of 0.51/0.42 for ac and 0.53/0.60 for p variants (Figure 3.6 and Appendix 3.13), with the aggregate of 0.52 ± 0.07 for 16 pairs (Table 3.2). Proteomic findings are often validated by negative testing of a priori false suppositions using decoy databases.\textsuperscript{71} Inspired by that, we computed the “decoy correlations” of FAIMS to TWIMS/TIMS separations for same variants in all wrong charge states (48 pairs, Appendix 3.14). The associated mean $r^2$ value of 0.22 ± 0.05 (with TWIMS or TIMS) is apart from the above for correct states but matches the $r^2$ for correlations across those in TWIMS or

![Figure 3.6. Linear correlations between FAIMS and TIMS separations for ac variants ($r^2$ marked). The plots involving TWIMS and for p variants are in Appendix 3.14.](image-url)
TIMS that apparently make the random baseline (Table 3.2). Therefore, the correlation between linear IMS and FAIMS is real but is below 50% upon baseline subtraction. Accordingly, the 2-D pc of FAIMS/IMS separations for middle-down peptides must be over half of the product of pc for each stage (defined as the occupied separation space, $d$, over mean $w$ of peaks). Here in TIMS, the typical $d \approx 100 \, \text{Å}$ and $w \approx 10 \, \text{Å}$ in a “good” charge state yield pc $\sim 10$ (e.g., 8 for p variants in 6+ and 10+, or 14 and 11 for me3 variants in 8+ and 9+). In FAIMS, $^{14}$ the typical pc in one state was $\sim 25$ (with $d \approx 30 \, \text{V/cm}$ and $w \approx 1.2 \, \text{V/cm}$). Hence, the pc of FAIMS/IMS would be $>125$ in one state and easily $>500$ in all (near-orthogonal) states. The values would be greater for more complex samples (as the separation space statistically widens), and the number of available charge states can be augmented (e.g., via supercharging). $^{72-74}$ Despite much of this pc taken up by the conformers of each variant, $^{14}$ it should still suffice to largely fractionate the known isomeric proteoform sets at least into binary mixtures.

### 3.4 Conclusions

Linear IMS with resolving power $>100$ (specifically TIMS) can broadly separate the PTM localization variants of “middle-down” peptides, here histone tails with $\sim 50$ residues comprising common PTMs: methylation(s), acetylation, or phosphorylation. Although only some variants (at best) are resolved in each charge state generated by ESI, the separations are orthogonal across states and all variants were filtered in some to at least binary mixtures quantifiable by ETD MS/MS. The serial Bruker timsToF Pro system featuring another funnel trap prior to the TIMS cell would deliver similar separations with improved sensitivity due to a higher duty cycle. The much lower resolving power of
TWIMS limits separation to a few variants, but all relative mobilities reproduce those in TIMS despite dissimilar ESI and IMS conditions. Separations are also independent of the ESI solvent or IMS residence time (from ∼5 to ∼300 ms), though less denaturing solvents and/or conditions may change that. This suggests that we deal with stable conformers thermalized prior to separation, wherein results transfer to other IMS systems including DTIMS. This indicates cataloging the Ω values for all histone proteoforms. However, ETD (with a normal time scale of ∼10–100 ms) is harder to add after time-dispersive separations that output transient ion packets (such as DTIMS and TWIMS) in comparison to TIMS, where the ramp can be arbitrarily slow. These findings agree with those for D/L peptides but extend beyond ∼3 kDa considered there.

The linear IMS and FAIMS separations for same set of variants are ∼50% orthogonal (as for tryptic peptides). Hence, online FAIMS/IMS based on existing technology ought to provide a 2-D peak capacity of several hundred across charge states, enabling separation of most complex known proteoform mixtures.

3.5 References


<table>
<thead>
<tr>
<th>Reference</th>
<th>Details</th>
</tr>
</thead>
</table>
CHAPTER 4: TAILORING PEPTIDE CONFORMATIONAL SPACE WITH
ORGANIC GAS MODIFIERS IN TIMS-MS

This chapter was published in *International Journal of Ion Mobility Spectrometry* and reproduced with permission.

4.1 Abstract

Recently, we showed the advantages of Trapped Ion Mobility Spectrometry for the study of kinetic intermediates of biomolecules as a function of the starting solvent composition (e.g., organic content and pH) and collisional induced activation. In the present work, we further characterize the influence of the bath composition (e.g., organic content) on the conformational space of an intrinsically disordered, DNA binding peptide: AT-hook 3 (Lys-Arg-Pro-Arg-Gly-Arg-Pro-Arg-Lys-Trp). Results show the dependence of the charge state distribution and mobility profiles by doping the solution and the bath gas with organic modifiers (e.g., methanol and acetone). The high resolving power of the TIMS analyzer allowed the separation of multiple IMS band per charge state, and their relative abundances are described as a function of the experimental conditions. The use of gas modifiers resulted in larger ion-neutral collision cross sections, with a direct correlation between the size of the modifier and the CCS differences. Conformational isomer inter-conversion rates were observed as a function of the trapping time. Different from solution experiments, a larger variety of organic gas modifiers can be used to tailor the peptide conformational space, since peptide precipitation is not a problem.

4.2 Introduction

Mass spectrometry-based methods have increasingly become a complementary or alternative research tool for investigating the conformational space of biomolecules under a variety of conditions, including biologically relevant conditions.1-5 Specifically, ion mobility spectrometry combined with mass spectrometry (IMS-MS) has the capability to perform separation and selection of
gas-phase ions, from heterogeneous solutions. It provides insight into both stable and intermediate structures, allowing for a more dynamic view and native-like folding information, \(^6\) while resembling solution structures (memory effect). \(^7^-^1^5\) Previous studies from our group showed the advantages of ESI-TIMS-MS for the study of kinetically trapped intermediates of biomolecules. \(^1^4^-^2^6\) Relevant to this study, we presented the folding pathways between local, free energy minima of AT-hook peptide 3 (ATHP3) leading to multiple, stabilized conformations. \(^2^7\) Protonation site, backbone relaxation and side-chain orientations were implicated in defining each structure. We have shown that the conformational space can be altered by introducing dopants into the TIMS cell for the case of flavin adenine dinucleotide. \(^2^6\) Different from other experiments where gas modifiers are used to increase the analytical power of IMS by increasing the size of the collision partner or inducing higher order multi-pole interactions, \(^2^8^-^3^3\) in this project we focused on the influence of the microenvironment on the stabilization of the conformational space of biomolecules.

In the present work, a ten amino acid intrinsically disordered peptide, Lys-Arg-Pro-Arg-Gly-Arg-Pro-Arg-Lys-Trp, was studied using nanoESI-TIMS-MS as a function of starting solvent (e.g. organic content and pH), bath gas collision partner and time after desolvation. This study is the first to report on the use of TIMS gas modifiers to tailor the peptide conformational space. 
4.2 Experimental Methods

4.2.1. Materials and reagents

AT-hook peptides 3 (Lys-Arg-Pro-Arg-Gly-Arg-Pro-Arg-Lys-Trp) was purchased from Advanced ChemTech Inc. (Louisville, KY) and used as received. Methanol and acetone solvents, and ammonium acetate salts utilized in this study were analytical grade or better and purchased from Fisher Scientific (Pittsburgh, PA). A Tuning Mix calibration standard (G24221A) was obtained from Agilent Technologies (Santa Clara, CA) and used as received.

4.2.2. Trapped Ion Mobility Spectrometry – Mass Spectrometry Analysis (TIMS-MS).

Details regarding the TIMS operation and specifics compared to traditional IMS can be found elsewhere. \(^{34-38}\) Briefly, mobility separation in TIMS is based on holding the ions stationary against a moving gas using an electric field. The separation in a TIMS device can be described in the center of the mass reference frame using the same principles as in a conventional IMS drift tube. \(^{39}\) Since mobility separation is related to the number of ion-neutral collisions (or drift time in traditional drift tube cells), the mobility separation in a TIMS device depends on the bath gas drift velocity, ion confinement and ion elution parameters. The reduced mobility, \(K\), of an ion in a TIMS cell is described by:

\[
K = \frac{V_g}{E} \approx \frac{A}{(V_{elution} - V_{out})}
\]
where \(v_g\) and \(E\) are the velocity of the gas and the applied electric field across the TIMS analyzer region. \(V_{\text{elution}}\) is the voltage when the ions elute in the \(V_{\text{ramp}}\) sweep and \(V_{\text{out}}\) is the voltage applied at the end of the TIMS analyzer region.

A custom-built, pulled capillary nanoESI source was utilized for all the experiments. Quartz glass capillaries (O.D.: 1.0 mm and I.D.: 0.70 mm) were pulled utilizing a P-2000 micropipette laser puller (Sutter Instruments, Novato, CA) and loaded with 10 \(\mu\)L aliquot of the sample solution. A typical nanoESI source voltage of +/- 600-1200 V was applied between the pulled capillary tips and the TIMS-MS instrument inlet. Ions were introduced via a stainless steel tube (1/16 x 0.020’’, IDEX Health Science, Oak Harbor, WA) held at room temperature into the TIMS cell. It should be noted that all solvent studies were performed with nitrogen as the bath gas, and that all dopant experiments were conducted with peptides sprayed from 10 mM NH\(_4\)AC.

Mobility calibration was performed using the Tuning Mix calibration standard (G24221A, Agilent Technologies, Santa Clara, CA) in positive ion mode (e.g., \(m/z = 322, K_0 = 1.376 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}\) and \(m/z = 622, K_0 = 1.013 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}\)).\(^{38}\) The TIMS operation was controlled using in-house software, written in National Instruments Lab VIEW, and synchronized with the maXis Impact Q-ToF acquisition program.\(^{34-35}\) Gas modifiers were introduced at the entrance of the TIMS cell via vaporization of the respective solvents (e.g., methanol or acetone) at a ratio of 2:1 air:air modified mix (scheme shown in Figure 4.1). For simplified mobility calibration, the gas velocity was kept constant in all experiments (P1 and P2 values).
4.3 Results and Discussion

The analysis of ATHP3 peptide using nES-TIMS-MS resulted in a charge state distribution of [M+H]$^+$ to [M+3H]$^{+3}$. The ATHP3 motif is mainly comprised of basic residues with seven potential locations for protonation (e.g., N-terminus, four arginines and two lysines), however, the most abundant ion under all experimental conditions was the [M+2H]$^{2+}$ charge state (Figure 4.2a). The mobility distributions observed from our previous study of ATHP 3 [M+2H]$^{+2}$ using ESI-TIMS-MS are consistent with the current analysis by nanoESI-TIMS-MS.27

At native conditions (Figure 4.2b, pink mobility bands) ATHP3 [M+2H]$^{2+}$ populates four conformers (A-D). The mobility bands are also conserved across the different organic solvent conditions (Figure 4.2c, pink mobility bands in blue panel). At higher trapping times (e.g., 500 ms), a kinetically trapped structure, and what we consider a more stable “desolvated” conformer, appears between bands B and C of the native. The presence of methanol in the TIMS cell did not alter the [M+2H]$^{2+}$
conformers, while acetone significantly changed the relative abundance and distribution of structures (Figure 4.2d, pink mobility band in pink panel). Three structures (A, B and C) were observed for ATHP3 [M+3H]^{3+} over the range of starting solvent solutions (Figure 4.2b and 4.2c, green mobility bands). Conformer A was the major structure present in all experiments, followed by conformer B and C. The abundance of conformer B, however, increased in the presence of acetone in the TIMS cell. The mobility profiles of ATHP3 [M+H]^+ showed the presence of two structures (A and B) which were observed over the range of experimental conditions (Figure 4.2b, 4.2c and 4.2d, blue mobility bands). The distribution of ATHP3 structures using acetone solvent cannot be recreated using acetone in the TIMS cell. One possible explanation is that the differences in conformational space is due to interaction with the ketone functional group of acetone. Unlike methanol’s alcohol group, acetone’s ketone group can form various interactions with the peptide, including 1) hydrogen bonding with the amide of the peptide backbone, 2) disruption of hydrogen bonding networks or, more likely, 3) dipole-dipole interactions with the charged residues of the peptide. Confirming this explanation will be the subject of future studies.

Changes in the conformational space as a function of the trapping showed stabilization towards more energetically favored structures as a function of the trapping time for [M+H]^{+2} charge state (Figure 4.3). While our measurements are only sensitive to the 50-500 ms time scale, potential rearrangements are possible in the first 50 ms after desolvation^{40}. In the case of varying the starting solution (10 mM ammonium acetate, and with methanol and acetone), a common trend in the gas
phase kinetics in nitrogen is the increase of the band E, which corresponds to the largest $1/K_0$ for this charge state. However, in the case of gas modifiers, band E is not observed, and the trends are best characterized by a decrease.

Figure 4.2. a) Typical mass spectra and b) native IMS spectra of ATHP3 as a function of c) starting solvent (methanol:H$_2$O or acetone:H$_2$O) or d) dopant bath gas (methanol or acetone).
Figure 4.3. The relative abundances of ATHP 3 [M+2H]$^{+2}$ conformers as a function of the trapping time, stating solvent conditions and bath gas composition. Starting solvent and bath gas are listed to the left of the graphs.
of band B and band A for methanol and acetone, respectively, which correspond to the smallest $1/K_0$. We interpret these results as the most stable gas-phase structures tending to have larger $1/K_0$ than those initially observed in solution. These effects may be a consequence of the absence of the solvent, since in the gas-phase the lack of water molecules promotes long range interactions. In the case of the $[\text{M+H}]^+$ and $[\text{M+H}]^{+3}$ charge states, similar trends were observed regarding the increase of larger $1/K_0$ bands as a function of the trapping time (see Figure 4.2).

4.4 Conclusions

The results presented here displayed the utility of gas modifiers in TIMS-MS for investigating and monitoring solution versus gas-phase microenvironment contribution to the peptide conformational space. When ionized from native conditions (10 mM NH$_4$Ac and nitrogen bath gas), the mobility profiles of ATHP 3 show an ensemble of conformers, which were preserved as a function of increasing organic content (methanol and acetone). Although the overall IMS profiles were maintained, changes in the relative abundance of conformers (e.g., conformational isomerization to the more stable gas-phase structure) were observed and recorded. The interconversion of structures, however, was small and often did not exceed growth or decay abundances of $\sim$10%. Comparison between starting solvent and bath gas with the same organic modifier showed that acetone as a dopant consistently changed the original IMS profiles. We attribute these rearrangements to the highly polar ketone functional group of acetone and its interaction with the peptide (i.e., disruption of the hydrogen bonding network or interaction with the charged residues). Overall, we find evidence for multiple stable conformers of these “disordered” motifs as a function of starting solvent (e.g. organic content), bath gas collision partner and time after
desolvation. The sensitivity of TIMS-MS allows for the observation of many low abundant conformers, separation of closely related structures and tracking of gas-phase stable structures via isomerization kinetics. This methodology opens new avenues for the study of biomolecules in the presence of gas modifiers that are not accessible during solution experiments, due to the typical precipitation of biomolecules during non-native conditions.

4.5 Acknowledgements

This work was supported by the National Science Foundation Division of Chemistry, under CAREER award CHE-1654274, with co-funding from the Division of Molecular and Cellular Biosciences to F.F.-L. The authors will also like to acknowledge the helpful discussions and technical support from Dr. Mark E. Ridgeway and Dr. Melvin A. Park from Bruker Daltonics Inc. during the development and installation of the custom-built TIMS-TOF MS instrument.

4.5 References


23. Ridgeway, M. E.; Silveira, J. A.; Meier, J. E.; Park, M. A., Microheterogeneity within conformational states of ubiquitin revealed by high resolution trapped ion mobility spectrometry. *Analyst* 2015, 140 (20), 6964-6972.


32. Fernández-Maestre, R.; Wu, C.; Hill, H. H., Buffer Gas Modifiers Effect Resolution in Ion Mobility Spectrometry through Selective Ion-Molecule Clustering


CHAPTER 5: STRUCTURES OF THE KINETICALLY TRAPPED I-MOTIF DNA INTERMEDIATES

This chapter was reproduced from Garabedian, A., et al. *Phys. Chem. Chem. Phys.*, 2016, 18 (38), 26691-26702 with permission from PCCP Owner Societies.
5.1 Abstract

In the present work, the conformational dynamics and folding pathways of i-motif DNA were studied in solution and in the gas-phase as a function of the solution pH conditions using circular dichroism (CD), photoacoustic calorimetry analysis (PAC), trapped ion mobility spectrometry - mass spectrometry (TIMS-MS), and molecular dynamics (MD). Solution studies showed at thermodynamic equilibrium the existence of a two-state folding mechanism, whereas during the pH = 7.0 \rightarrow 4.5 transition a fast and slow phase ($\Delta H_{\text{fast}} + \Delta H_{\text{slow}} = 43 \pm 7 \text{ kcal mol}^{-1}$) with a volume change associated with the formation of hemiprotonated cytosine base pairs and concomitant collapse of the i-motif oligonucleotide into a compact conformation were observed. TIMS-MS experiments showed that gas-phase, kinetically trapped i-motif DNA intermediates produced by nanoESI are preserved, with relative abundances depending on the solution pH conditions. In particular, a folded i-motif DNA structure was observed in nanoESI-TIMS-MS for low charge states in both positive and negative ion mode (e.g., $z = +/-3$ to $+/-5$) at low pH conditions. As solution pH increases, the cytosine deprotonation leads to the loss of cytosine-cytosine$^+$ (C•CH$^+$) base pairing in the CCC strands and in those conditions we observe partially unfolded i-motif DNA conformations in nanoESI-TIMS-MS for higher charge states (e.g., $z = -6$ to -9). Collisional induced activation prior TIMS-MS showed the existence of multiple local free energy minima, associated with the i-motif DNA unfolding at $z = -6$ charge state. For the first time, candidate gas-phase structures are proposed based on mobility measurements of the i-motif DNA unfolding pathway. Moreover, the inspection of partially unfolded i-motif DNA structures ($z = -7$ and $z = -8$ charge states) showed that the presence of inner cations may or may not induce
conformational changes in the gas-phase. For example, incorporation of ammonium adducts does not lead to major conformational changes while sodium adducts may lead to the formation of sodium mediated bonds between two negatively charged sides inducing the stabilization towards more compact structures in new local, free energy minima in the gas-phase.

5.2 Introduction

Recent innovations in speed, accuracy, and sensitivity have established Mass Spectrometry (MS) based methods as a key technology in the field of structural biology. Over the last two decades, native MS of intact biomolecules and biomolecular complexes has permitted structural interrogation at biologically relevant conditions that are not accessible by other methods. Most common gas-phase structural probes are based on (or a combination of) tandem mass spectrometry (ergodic and non-ergodic), gas-phase hydrogen-deuterium exchange, ion spectroscopy, and ion mobility spectrometry. In particular, ion mobility spectrometry combined with mass spectrometry (IMS-MS), can be a multifaceted approach that quickly provides insight into both the stable and intermediate structures of biomolecules in the gas-phase. Traditional ion mobility spectrometry (IMS) is based on the separation of ions under the influence of a weak electric field as they collide with a bath gas of inert neutral molecules. The mobility provides information on the its size and shape via the momentum transfer ion-neutral collision cross section (CCS). While this description holds true for most contemporary IMS analyzers (e.g., periodic focusing DC ion guide, segmented quadrupole drift cell, multistage IMS, transient wave ion guide, trapped ion mobility spectrometers and SLIM devices), a common feature is that gas-phase studies take advantage of the desolvation process.
to effectively reduce sample complexity, permitting molecular characterization in the absence of bulk solvent by studying gas-phase structures in kinetically trapped intermediates. More recently, the combination of IMS-MS measurements with infrared multi-photon dissociation (IRMPD) spectroscopy and infrared vibrational spectroscopy for structural characterization of gas-phase molecular ions have shown the unique advantage of gas-phase studies in order to better understand the intramolecular forces that define the conformational space of biomolecules. 25-29 With the recent introduction of a new IMS analyzer - Trapped Ion Mobility Spectrometer (TIMS)- the possibility to decouple the time domain from the IMS separation allows for the study of kinetically trapped intermediates of molecular ions in the gas-phase as a function of the desolvation time, temperature and bath gas composition. TIMS’ mode of operation and its advantages over traditional IMS are described in ref 19, 30-31. We have shown the use of TIMS for the study of isomerization kinetics of small molecules32-34, peptides35-36, and proteins37-42, the influence of the collision partner on the molecular structure43, and the factors that affect molecular-adduct complex lifetime and stability during TIMS measurements44.

Here we report on a proton-sensitive DNA molecule based on the i-motif conformation (i-motif DNA) that can reversibly fold in response to the solution pH conditions.45-47 The i-motif DNA is a four-stranded DNA structure that forms due to intramolecular noncanonical base pair interactions between a protonated and a neutral cytosine residue (i.e., a CH⁺:C base pair) under slightly acidic conditions. At higher pH values, the cytosines are neutral and the DNA strand adopts a random coil conformation.48-49 Regardless of its impending role in oncogene regulation and anti-cancer therapies,50-53 i-motif DNA has found multiple applications in the field of nanotechnology (e.g.,
nanomotors, environmental sensors and pH switches). Recent studies have shown that nucleic acids can maintain their solution phase topology (e.g., duplexes, triplexes or G-quadruplexes\textsuperscript{55-57}) when transferred to the gas phase via electrospray sources.\textsuperscript{58-59} A recent work using ESI-TWIMS-MS and ESI-IRMPD-MS concluded that zwitterionic i-motif DNA structures are preserved in negatively charged DNA ions.\textsuperscript{60} However, the detailed structural features of the folded i-motif DNA and the variations in structure that occur in response to changes in pH have received little attention.

In the present work, the conformational dynamics and folding/unfolding pathways of i-motif DNA are investigated in solution and in the gas-phase. That is, i-motif DNA was studied in solution for the pH range 4.0 – 9.0 at thermodynamic equilibrium using circular dichroism (CD) and during the pH 7.0 \to 4.5 transition using photoacoustic calorimetry analysis (PAC), and in the gas-phase using nano electrospray ionization trapped ion mobility spectrometry - mass spectrometry (nanoESI-TIMS-MS) with and without collisional induced activation prior the TIMS-MS analysis. The TIMS-MS measured ion-neutral collision cross sections are used to propose candidate structures based on molecular dynamics (MD) simulations as a way to better understand the i-motif DNA free energy landscape. In addition, the influence of inner cations on the stabilization of gas-phase structures is discussed for partially unfolded i-motif DNA.

5.3 Experimental

5.3.1 Materials and reagents

A 21 base, cytosine-rich DNA sequence mimicking the telomeric repeat sequence 5’-\{CCCTAA\}_3CCC-3’ (C\textsubscript{198}H\textsubscript{256}N\textsubscript{72}O\textsubscript{121}P\textsubscript{20}, MW 6200 Da) was obtained indistinctly from
Eurofins Scientific (Huntsville, AL) and from Integrated DNA Technologies (Coralville, IA) and used without further purification. Nitrobenzaldehyde (2-NBA) powder was obtained from Sigma-Aldrich (St. Louis, MO). Solvents, hydrochloric acid, sodium hydroxide, ammonium acetate, ammonium formate, and ammonium hydroxide salts utilized in this study were analytical grade or better and purchased from Fisher Scientific (Pittsburgh, PA). A Tuning Mix calibration standard (G24221A) was obtained from Agilent Technologies (Santa Clara, CA) and used as received.

5.3.2 Circular Dichroism Analysis (CD)

Circular dichroism spectra were collected using a Jasco J-815 CD spectrometer (Easton, MD) as a function of the solution pH. Spectra were collected from 1 mm quartz cuvettes containing 30 µM i-motif DNA solubilized in deionized water at pH = 4.0, 5.0, 6.0, 7.0, 8.0 and 9.0. Solution pH was adjusted using hydrochloric acid (HCl) and sodium hydroxide (NaOH).

5.3.3 Photo Acoustic Calorimetry Analysis (PAC)

A detailed description of PAC principles of operation and data analysis can be found elsewhere. Briefly, a solution of 30 µM i-motif DNA and 1 mM 2-NBA was prepared in deionized water and stored in the dark to prevent unwanted photolysis. Solution pH was adjusted using hydrochloric acid (HCl) and sodium hydroxide (NaOH). A 500 µL solution of i-motif DNA and 2-NBA solution was placed in a 1 cm x 0.5 cm quartz cuvette (Starna Cells) and the cuvette was placed in a temperature-controlled sample holder with magnetic stirring capability (TC 125, Quantum Northwest). The sample solution was stirred during measurements to prevent accumulation of the photoproduct and a fresh
aliquot was used for each temperature measurement; the PAC measurements were performed over 16-26 °C temperature range. 2-NBA photo-dissociation and concomitant proton release was triggered by a 355 nm laser pulse (500 µJ, Surelite I-10, Continuum). Fe(III)tetraphenylsulfonato porphyrin (4-SP) solubilized in deionized water at pH 7.0 was employed as a reference compound. The absorbance of the reference solution matched the absorbance of the sample at the excitation wavelength of 355 nm. A piezoelectric transducer (Panametrics RV103, 1 MHz) was adhered to the side of the cuvette and the signal was amplified using an ultrasonic preamplifier (Panametrics 5662). The PAC traces were recorded using a digital oscilloscope (Wave Surfer 42Xs, LeCroy) and were deconvoluted using Sound Analysis software (Quantum Northwest, WA).

Acoustic transducers used in PAC measurements are sensitive to the amplitude of an acoustic wave as well as to its temporal profile. For a two-step sequential decay process on the time scale of the transducer resolution (∼50 ns to ∼10 µs), the individual contributions of ΔV and ΔH for each kinetic process can be resolved. Changes in the phase and amplitude of the PAC traces can then be correlated with structural changes in solution as a consequence of the pH jump (reaction volume and enthalpy)61. For the initial process (τ < 50 ns), the ratio of the sample and reference PAC signal amplitudes (φ1) is related to the reaction volume (ΔV_{fast}) and reaction enthalpy (ΔH_{fast}) according to Eq. 1:

$$E_{h\nu}(\phi_1-1)/\Phi = \Delta V_{fast} \left[ \frac{C_p \rho}{\rho} \right] - \Delta H_{fast}$$

(1)

where $E_{h\nu} = 80.59$ kcal mol$^{-1}$, Φ is the quantum yield of the primary photo-reaction, and $C_p \rho / \beta$ is a temperature-dependent parameter made up of the heat capacity ($C_p$), density ($\rho$) and expansion coefficient (β) of the solvent.
For slower processes (50 ns < \( \tau \) < 10 \( \mu s \)), the ratio of the sample and reference PAC signal amplitudes (\( \phi_2 \)) is related to the reaction volume (\( \Delta V_{\text{slow}} \)) and enthalpy (\( \Delta H_{\text{slow}} \)) according to Eq. 2:

\[
E_{hv}(\phi_2)/\Phi = \Delta V_{\text{slow}}\left[\frac{C_p q}{\beta}\right] - \Delta H_{\text{slow}}
\]  \( (2) \)

### 5.3.4 Trapped Ion Mobility Spectrometry – Mass Spectrometry Analysis (TIMS-MS)

Details regarding the TIMS operation and specifics compared to traditional IMS can be found elsewhere. \(^{19-21, 62-63}\) Briefly, mobility separation in TIMS is based on holding the ions stationary using an electric field against a moving gas. The separation in a TIMS device can be described by the center of the mass frame using the same principles as in a conventional IMS drift tube. \(^{10}\) Since mobility separation is related to the number of ion-neutral collisions (or drift time in traditional drift tube cells), the mobility separation in a TIMS device depends on the bath gas drift velocity, ion confinement and ion elution parameters. The reduced mobility, \( K \), of an ion in a TIMS cell is described by:

\[
K = \frac{v_g}{E} = \frac{A}{(V_{\text{elution}} - V_{\text{out}})}
\]  \( (3) \)

where \( v_g \) and \( E \) are the velocity of the gas and the applied electric field across the TIMS analyzer region. \( V_{\text{elution}} \) is the voltage when the ions elute in the \( V_{\text{ramp}} \) sweep and \( V_{\text{out}} \) is the voltage applied at the end of the TIMS analyzer region.

A custom-built, pulled capillary nanoESI source was utilized for all the experiments. Quartz glass capillaries (O.D.: 1.0 mm and I.D.: 0.70 mm) were pulled utilizing a P-2000 micropipette laser puller (Sutter Instruments, Novato, CA) and loaded with 10 \( \mu L \) aliquot of the sample solution. Sample solutions consisted of 1 \( \mu M \) i-motif
DNA in 10 mM ammonium acetate solution at three pH values (pH= 4.5, 7.0 and 9.0). Solution pH was adjusted using ammonium formate and ammonium hydroxide salts. A typical nanoESI source voltage of +/- 600-1200 V was applied between the pulled capillary tips and the TIMS-MS instrument inlet. Ions were introduced via a stainless steel tube (1/16 x 0.020”, IDEX Health Science, Oak Harbor, WA) held at room temperature into the TIMS cell.

The TIMS cartridge is comprised of three regions: an entrance funnel, analyzer tunnel (46 mm axial length), and exit funnel. A 880 kHz and 200 Vpp RF potential was applied to each section creating a dipolar field in the funnel regions and a quadrupolar field inside the tunnel. For collisional induced activation experiments prior the TIMS separation (CIA-TIMS-MS), 2D-TIMS-MS spectra were collected as a function of the axial electric field across the deflector plate and the end of the entrance funnel in 10V intervals of the deflector voltage (V_{deflector} = 0-200 V); the voltages across the entrance and exit of the entrance funnel were adjusted relative to V_{deflector} to guarantee an homogenous axial electric field in all CIA measurements. TIMS separation was performed using nitrogen as a bath gas at ca. 300 K, at a constant gas flow velocity defined by the pressure difference between entrance funnel P_1 = 1.1-4.3 mbar, and the exit funnel P_2 = 0.6-3.0 mbar.\(^{33, 41}\)

In TIMS operation, multiple geometric isomers/conformers are trapped simultaneously at different E values resulting from a voltage gradient applied across the TIMS tunnel. After thermalization, geometrical isomers/conformers are eluted by decreasing the electric field in stepwise decrements (referred to as the “ramp”). Each isomer/conformer eluting from the TIMS cell can be described by a characteristic voltage
(i.e., \( V_{\text{elution}} - V_{\text{out}} \)). Eluted ions are then mass analyzed and detected by a maXis impact Q-ToF mass spectrometer (Bruker Daltonics Inc, Billerica, MA).

In a TIMS device, the total analysis time can be described as:

\[
\text{Total IMS time} = t_{\text{trap}} + (V_{\text{elution}}/V_{\text{ramp}}) \cdot t_{\text{ramp}} + \text{ToF} = t_0 + (V_{\text{elut}}/V_{\text{ramp}}) \cdot t_{\text{ramp}}
\]  

(4)

where, \( t_{\text{trap}} \) is the thermalization/trapping time, ToF is the time after the mobility separation, and \( V_{\text{ramp}} \) and \( t_{\text{ramp}} \) are the voltage range and time required to vary the electric field, respectively. The elution voltage was experimentally determined by varying the ramp time \( (t_{\text{ramp}} = 100, 200, 300, 400 \text{ and } 500 \text{ ms}) \) for a constant ramp voltage sweep. This procedure also determines the time ions spend outside the separation region \( t_0 \) (e.g., ion trapping and time-of-flight). The TIMS cell was operated using a fill/trap/ramp/wait sequence of 10/10/50-500/50 ms. The TOF analyzer was operated at 10 kHz \((m/z = 100-3500)\). The data was summed over 100 analysis cycles yielding an analysis time of \(~50 \text{ s for } t_{\text{ramp}} = 500 \text{ ms}\). Mobility calibration was performed using the Tuning Mix calibration standard (G24221A, Agilent Technologies, Santa Clara, CA) in positive ion mode (e.g., \( m/z = 322, K_0 = 1.376 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1} \) and \( m/z = 622, K_0 = 1.013 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1} \)).\(^{21}\) The TIMS operation was controlled using in-house software, written in National Instruments Lab VIEW, and synchronized with the maXis Impact Q-ToF acquisition program.\(^{19-20}\)

Reduced mobility values \((K_0)\) were correlated with CCS \((\Omega)\) using the equation:

\[
\Omega = \frac{(18\pi)^{1/2}z}{16} \frac{1}{(k_B T)^{1/2}} \left[ \frac{1}{m_i} + \frac{1}{m_b} \right]^{1/2} \frac{1}{K_0} \frac{1}{N^*} 
\]  

(5)
where $z$ is the charge of the ion, $k_B$ is the Boltzmann constant, $N^*$ is the number density and $m_1$ and $m_b$ refer to the masses of the ion and bath gas, respectively.\textsuperscript{10} All resolving power ($R$) values reported herein were determined from Gaussian peak fits of the features in the TIMS distributions ($R = \Omega/\Delta\Omega$) using OriginPro (version 8.0). The FWHM of the mobility band was used to calculate the $\Delta\Omega$.

\textbf{5.4 Theoretical}

Candidate structures search and assignment from TIMS-MS data was performed based on the workflow described in reference\textsuperscript{64}. Briefly, an initial guess was performed based on the folded i-motif DNA 1ELN structure from protein data bank \textsuperscript{65}. Molecular dynamic simulations were performed with NAMD (http://www.ks.uiuc.edu/Research/namd/)\textsuperscript{66} using CHARMM36 force field\textsuperscript{67}, which shows significant improvements over previous force-fields for simulating the nucleic acids.\textsuperscript{68} CHARMM-GUI\textsuperscript{69} web-server was used to initialize the MD simulations. The i-motif DNA was solvated in a 53 Å cubic box with TIP3P water model and the system was neutralized by adding eight Mg\textsuperscript{2+} and two Cl\textsuperscript{-} ions.

A 10,000-step energy minimization was performed using the conjugate gradient and line search algorithm, followed by a 100-ps NVT equilibration at 300 K and 1 atm pressure with periodic boundary conditions. The particle mesh Ewald method\textsuperscript{70} was used to treat the long-range interactions with a 12 Å nonbonded cutoff. The RATTLE algorithm was used to constrain protein bonds involving hydrogen, and SETTLE was used to maintain water geometry. A 400-ns NPT simulation was performed with a 2 fs integration time step at 300 K and 1 atm pressure. A total of 18,000 structures were submitted to the
CCS calculation and clustered based on RMSD using our custom-built R-code software. A minimum of five distinct center of mass structures (CM) were selected per experimental IMS band observed from the cluster analysis based on energy minimization and CCS calculation as a function of the charge state \([CM, CCS]_{\text{charge}}\). The MD force field energy minimization did not properly describe the hemiprotonated cytosine interactions leading to mostly \([CM, CCS]_{\text{charge}}\) of unfolded i-motif DNA. To overcome this limitation, the final set of candidate structures also included structures generated based on the trends observed during the i-motif DNA unfolding 400-ns trajectory and the previously reported i-motif DNA folded structure. CCSs were calculated using the IMoS (v1.04b) and PSA packages with nitrogen as a bath gas at ca. 300K. In the IMoS calculations, 100 total rotations were performed using the trajectory method with a Maxwell distribution. Molecular visualization was performed using Visual Molecular Dynamics software.

5.5 Results and Discussion

This work utilizes solution and gas-phase experiments for the study of equilibrium and kinetic intermediates of the i-motif DNA system. Specifically, CD and TIMS-MS provide information of i-motif DNA structures as a function of solution pH, whereas PAC and CIA-TIMS-MS interrogate kinetic intermediates during the folding and unfolding pathways, respectively. These complementary techniques allow different conformational states to be unveiled and provide a more detailed interpretation of the folding pathway and unfolding intermediates in the presence and absence of bulk solvent, respectively. The focus of this paper is towards a better description of the i-motif DNA free energy landscape.

At thermodynamic equilibrium in solution, the transition from basic to acidic pH solution conditions of the i-motif DNA can be characterized by an increase of the 287 nm
band and the concomitant decrease of the 272 nm band in the CD signal (see Figure 5.1), being more significant in the pH = 4.0 - 7.0 range. The observation of two isosbestic points at 245 nm and 276 nm suggests a two state folding mechanism (F↔U) for the i-motif DNA as a function of the solution pH. It should be noted that a two state folding mechanism does not limit the number of conformations to two; that is, multiple conformations may exist at equilibrium in the folded and unfolded states (e.g., the CD signal at pH = 6 contains mixed states).

Complementary gas-phase analysis was performed as a function of the solution pH and time after desolvation using nanoESI-TIMS-MS. This gas-phase analysis takes

![Circular dichroism spectra of i-motif DNA (30 µM in deionized water) as a function of the solution pH =4.0 – 9.0 at 20 °C. Notice that no changes in the spectra are observed in the pH = 7.0 – 9.0 range.](image.png)
advantage of the desolvation process to effectively reduce sample complexity and permit molecular characterization in the absence of bulk solvent. The rationale is to kinetically trap gas-phase structural populations (or ensemble) in local, free energy minima close to the initial, thermodynamic equilibrated solution-phase structural populations (or ensemble). Previous works have shown the presence of solution memory effects by looking at the changes on the IMS profile as a function of the desolvation time.62, 77-80.

The nanoESI-TIMS-MS analysis of i-motif DNA resulted in the observation of molecular ions in positive and negative ion mode (see Figure 5.2). In positive mode, a narrow charge state distribution (+3 - +5) is observed with no major variation as the solution pH. In negative mode, a broad distribution of charge states (-3 - -9) is observed with relative abundances changing with the solution pH. We interpret the changes in the MS relative abundances in negative mode with the solution pH increase as a consequence of the loosening of the C•CH+ base pairing that stabilizes the folded i-motif conformation leading to more deprotonation sites in the partially unfolded i-motif DNA structures. In the case of positive mode, as the pH decreases the cytocines become protonated leading to the i-motif folding, limiting the number of accessible protonation sites. Notice that this trend is unique for the i-motif DNA compare to other reported studies of biological molecules (e.g., some peptides and proteins) where denaturing at low pH conditions leads to the observation of higher charge states and native conditions (pH = 6 - 8) show a narrow low charge state distribution.77, 81-89 That is, the stabilization of the i-motif DNA with the reduction of the solution pH results in the observation of a narrow charge state distribution in positive mode. Despite the changes in the MS relative abundances of the charge states with the solution pH, closer inspection of the mobility profiles did not show any
dependence on the solution condition or the time after desolvation (details for the TIMS profiles as a function of the time after desolvation over the 50-500 ms range can be found in the supporting information). Moreover, major changes in the CCS values were observed as a function of the i-motif DNA charge state. Low charge states (e.g., +3 to +5 and -3 to -

Figure 5.2. Typical positive (red) and negative (blue) mass spectra (top) and mobility profiles for i-motif DNA as a function of the solution pH (4.0 – 9.0) using low energy injection nanoESI-TIMS-MS. Notice the change in the charge distribution as a function of the solution pH. Changes in individual charge states mobility profiles were not observed as a function of the solution pH conditions.
5) can be described by a single, broad mobility band\textsuperscript{60}, with slightly increasing CCS as a function of the charge state. Notice that the increase of CCS with the charge state is an expected trend and does not necessarily reflect conformational changes for $z = +3$ to $+5$ and $-3$ to $-5$. However, a major transition is observed at $z = -6$, with the observation of partially unfolded i-motif DNA structures in the $z = -6$ to $-9$ charge states. The CCS trend observed in negative mode as a function of the charge state is in good agreement with previous ESI-TWIMS-MS experiments\textsuperscript{90-93}; however, this study takes advantage of TIMS’s higher mobility resolution to resolve a higher number of IMS bands and to report accurate ion-neutral collision cross sections.

Solution phase analysis of i-motif DNA transition ($U \rightarrow F$) during the pH jump from 7.0 to 4.5 was performed by PAC and permitted the interrogation of the folding process (kinetic intermediates) on the nanosecond to microsecond timescale. The photolysis of 2-NBA leads to a prompt pH decrease ($\Delta p\text{H} \sim 2.5$ units) due to the 2-NBA photo-conversion to nitroso-benzoate (see details in Scheme 1). For the 2-NBA photo-dissociation, reaction enthalpy $\Delta H = -29.2 \pm 6.6$ kcal mol\textsuperscript{-1} and volume change $\Delta V = -5.1 \pm 1.6$ mL mol\textsuperscript{-1} were observed in agreement with previously published data.\textsuperscript{94} In the presence of i-motif DNA, the fast increase in the proton concentration (pH 7.0 $\rightarrow$ 4.5) initiates the folding of i-motif DNA. Acoustic traces for the photolysis of 2-NBA in the presence of i-motif DNA show a phase shift and amplitude change with respect with the reference acoustic traces, indicative of the presence of two processes, one occurring within the first 50 ns (fast phase, $\phi_1$), and a temperature-dependent process with a time constant of 140 ns (slow phase, $\phi_2$).
at 20 °C (see Figure 5.3). After subtraction of ΔV/ΔH determined for 2-NBA photolysis in the absence of i-motif DNA, the reaction parameters for the fast and slow processes are

$$\Delta H_{\text{fast}} = 8.5 \pm 7.0 \text{ kcal mol}^{-1} \quad \text{and} \quad \Delta V_{\text{fast}} = 10.4 \pm 1.6 \text{ mL mol}^{-1} \quad \text{and} \quad \Delta H_{\text{slow}} = -51.5 \pm 4.8 \text{ kcal mol}^{-1} \quad \text{and} \quad \Delta V_{\text{slow}} = -6.6 \pm 0.9 \text{ mL mol}^{-1},$$

respectively.

The fast process observed in PAC measurements ($\tau < 50 \text{ ns}$) can be attributed i) to the photo-dissociation of 2-NBA and concomitant release of protons (with a time constant of ~20 ns$^{95}$) and ii) to cytosine residue protonation (assuming the rate of pyridine protonation in solution ~ tens of picoseconds)$^{96}$. During the pH jump (pH 7.0 $\rightarrow$ 4.5), approximately six cytosine residues (pKa = 4.58)$^{97}$ will be protonated with an enthalpy

Figure 5.3. Plot of the fast ($\phi_1$) and slow ($\phi_2$) phases as a function of $C_{pp}/\beta$ in the 16 – 26 °C temperature range. In the inset, PAC traces for 2-NBA photo-dissociation in the presence of i-motif DNA sample (blue) and 4SP reference compound (red) at 16 °C.
change of -30 kcal mol$^{-1}$ (assuming a $\Delta H = -5.0$ kcal mol$^{-1}$ per cytosine residue$^{96}$). Since a positive enthalpy change of $8.5 \pm 7.0$ kcal mol$^{-1}$ was determined for the fast process, an additional endothermic process with $\Delta H \sim 38.5 \pm 7.0$ kcal mol$^{-1}$ must occur within 50 ns of 2-NBA dissociation. This endothermic enthalpy change can be attributed to the decrease in the solvent electrostriction due to reduction of the overall DNA charge.

The reaction parameters determined for the slow phase are consistent with i-motif DNA nucleation. Nucleation of the intramolecular i-motif DNA structure requires the formation of at least one hemiprotonated cytosine base pair, an interaction that consists of three hydrogen bonds between one protonated and one neutral cytosine residue within an oligonucleotide. Though empirical measurements of the enthalpy of this specific base-pairing interaction are not available, the enthalpy of inter base pair hydrogen bond formation was determined for the canonical GC Watson-Crick base to be -8.0 kcal mol$^{-1}$.\textsuperscript{99}

Within the intramolecular i-motif DNA with six protonated cytosines, up to six cytosine-cytosine base-pairing interactions can be formed for a contribution of -48 kcal mol$^{-1}$ to the observed enthalpy for the folding process. In addition, formation of the i-motif structure results in stacking of the cytosine bases, an interaction with an enthalpy of -5.0 kcal mol$^{-1}$ per base pair stack.\textsuperscript{100} Formation of the folded i-motif DNA structure results in the formation of 5 base pair stacking interactions with a contribution of -25 kcal mol$^{-1}$ to the observed enthalpy. Therefore, formation of the fully folded i-motif DNA should provide $\Delta H_{\text{obs}} = -73$ kcal mol$^{-1}$ or $\sim 12$ kcal mol$^{-1}$ per hemiprotonated cytosine base pair. The enthalpy change determined here for the slow process $\Delta H_{\text{slow}} = -51.5 \pm 4.8$ kcal mol$^{-1}$ reflects the formation of approximately 4.5 hemiprotonated cytosine base pairs. The observed volume change ($\Delta V_{\text{slow}} = -6.6 \pm 0.9$ mL mol$^{-1}$) is consistent with the formation of
hemiprotonated cytosine base pairs and concomitant collapse of the oligonucleotide into a more compact conformation with a smaller volume as observed previously using small angle X-ray scattering.\textsuperscript{101} The overall reaction enthalpy change ($\Delta H_{\text{fast}} + \Delta H_{\text{slow}} = 43 \pm 7$ kcal mol$^{-1}$) is similar to that previously determined by NMR for the intramolecular folding of a similar cytosine-rich strand (5'-'[TCC]$_4$-3') with fewer potential base pairing interactions (-43 kcal mol$^{-1}$)\textsuperscript{102} and for the reverse process of i-motif DNA unfolding using differential scanning calorimetry (66 kcal mol$^{-1}$).\textsuperscript{100} That is, the PAC experiments suggest that during the pH 7.0 $\rightarrow$ 4.5 transition, the i-motif DNA folds in two consecutive steps: i) protonation of cytosines and ii) nucleation and stacking towards the i-motif DNA structure. In addition, the time scale of the PAC measurement (50ns - 10$\mu$s) does not probe rearrangements occurring in timescales larger than a few microseconds; that is, these results suggest that due to the nature of the i-motif folding (C•CH$^+$ pairing and stacking) other kinetic intermediates are likely to be populated in the millisecond timescale. Notice that PAC measurements access kinetic intermediates, not accessible during CD measurements performed at thermodynamic equilibrium. That is, CD and PAC measurements provide complementary information on the solution conditions and intermediate transitions characteristic of the DNA i-motif folding, respectively.

In the gas-phase, collisional induced activation prior to the TIMS-MS analysis was used to access kinetic intermediates not accessible during the solution conditions (e.g., pH variation) and desolvation process. Previous reports have shown the advantage of this methodology for the study of biomolecules and biomolecular complexes using in-source activation and collision induced unfolding (CIU) prior to IMS-MS analysis.\textsuperscript{14, 80, 103-107}
Notice that in the case of CIA-TIMS-MS, the extent of activation is limited by the local pressure and the electric field range accessible in the TIMS cartridge (i.e., there is a higher limit in the electric field to avoid breakdown). We prefer to called it CIA instead of CIU since the cooling process that occurs during the TIMS measurement can also lead to refolding into a more compact structure not accessible during the molecular ion formation. In the case of i-motif DNA, the intramolecular forces (C•CH+ pairing) that define the folded states are overcome for the populations observed at \( z = -6 \). Our interpretation of the broad, bimodal mobility distribution is the existence of a transition point towards multiple local free energy minima: from folded to unfolded i-motif DNA (see Figure 5.2). Notice that the observation of a broad distribution can be justify with the co-existence of multiple minima reflecting the noncanonical hemiprotonated cytosine base pairing mechanism. A more detailed investigation using collisional induced activation of the i-motif-DNA prior to the TIMS mobility separation confirmed that most of the unfolding of the i-motif DNA structure occurs for \( z = -6 \) (see Figure 5.4). The \( z = -5 \) charge state also showed a movement toward unfolding, with the population of a new IMS band at larger CCS value. This result suggests that the conformational interconversion energy barriers for the \( z = -5 \) and \( z = -6 \) charge state structure can be easily overcome with collisional activation prior to the mobility separation and that the existence of multiple, local free energy minima can be visualized by the broad IMS profiles observed at \( z = -5, -6 \) and -7. The broad distribution of \( z = -5, -6 \) and -7 IMS bands does not allow to make a discrete annotation on the number of conformational states (at least with TIMS measurements performed at room temperature). Changes in the mobility profiles for the other charge states upon activation were not observed. The bimodal distribution and changes in the mobility profile upon
activation using CIA-TIMS-MS agree with previous reports of i-motif DNA activation in the entrance cone of a TWIMS-MS.  

The search for candidate structures resulted in the generation of an average structure for the major mobility trends observed in Figures 5.2 and 5.4. Inspection of the candidate structures suggest that the folded state observed for the low charge states (+3 to +5 and -3 to -5) shares the same main interactions: stabilization via cytosine-cytosine+ (C•CH+) and adenosine – thymidine (A-T) base pairing at weakly acidic pH values (see Figure 5.5). The most compact structure was observed for z = +3, with only small differences in the distances between the base pairs (vertical direction in Figure 5.5) as the
charge state increases from +3 to +5 and from -3 to -5. That is, major conformational differences were not observed between candidate structures proposed for positive and negative ions (e.g., +3 to +5 and -3 to -5); our interpretation is that differences in the CCS values are mostly due to the charge state and the internal base pairing stacking. Inspection of the candidate structures at the z = -6 charge state suggests that the broad, bimodal CCS distribution is probably due to cytosine neutralization. That is, the i-motif DNA structure begins transformation towards an overall larger CCS trend where a semi-open hairpin structure with a relaxed 5' end first appears. Specifically, this larger CCS trend has the common loss of the cytosine-cytosine\(^+\) (C\(\bullet\)CH\(^+\)) interactions in one of the CCC strands for the extended z = -6 and the z = -7 charge states, with a further loss of the second CCC strand interactions for the z = -8 and z = -9 charge states. The folding pattern proposed as a function of the charge state is in good agreement with the previous studies suggesting that the hairpin-like structures are the most stable conformations during the i-motif DNA deprotonation.\(^{108}\) While full unfolding of the i-motif DNA is never reached during CIA-TIMS-MS, the CCS profiles and candidate structures are in good agreement with previously reported DNA structural models\(^{101}\). Notice that the proposed candidate structures were generated based on a 400ns MD trajectory and the previously reported i-motif DNA folded structure\(^{65}\). The MD simulation described the unfolding of the i-motif DNA using a starting structure with deprotonated cytosines. Other MD attempts to simulate the folding of the i-motif- DNA using protonated cytosines were not successful; our interpretation is that further development of the force fields are required to better simulate DNA interactions and base pair stacking\(^{109}\). Nevertheless, for the first time candidate
Figure 5.5. Candidate structures proposed for the i-motif DNA mobility bands observed. Main intramolecular interactions are denoted. CCSs of the candidate structures are reported in Table 1.
Table 5.1. Experimental and theoretical collision cross sections for i-motif DNA bands observed in positive and negative ion mode in nanoESI-TIMS-MS and proposed candidate structures, respectively.

<table>
<thead>
<tr>
<th>z</th>
<th>$K_0$ (cm$^2$ V$^{-1}$ s$^{-1}$)</th>
<th>Experimental CCS$_{N_2}$ (Å$^2$)</th>
<th>Theoretical CCS$_{N_2}$ (Å$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+3</td>
<td>0.748</td>
<td>800</td>
<td>811</td>
</tr>
<tr>
<td>+4</td>
<td>0.944</td>
<td>846</td>
<td>846</td>
</tr>
<tr>
<td>+5</td>
<td>1.089</td>
<td>922</td>
<td>926</td>
</tr>
<tr>
<td>-3</td>
<td>0.729</td>
<td>821</td>
<td>832</td>
</tr>
<tr>
<td>-4</td>
<td>0.931</td>
<td>863</td>
<td>865</td>
</tr>
<tr>
<td>-5</td>
<td>1.096</td>
<td>917</td>
<td>903</td>
</tr>
<tr>
<td>-6</td>
<td>1.178</td>
<td>1023 (A)</td>
<td>1021</td>
</tr>
<tr>
<td></td>
<td>1.013</td>
<td>1190 (B)</td>
<td>1166</td>
</tr>
<tr>
<td>-7</td>
<td>1.267</td>
<td>1111 (A)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>1.140</td>
<td>1234 (B)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>1.071</td>
<td>1314 (C)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>1.034</td>
<td>1360 (D)</td>
<td>1378</td>
</tr>
<tr>
<td>-8</td>
<td>1.204</td>
<td>1335 (A)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>1.181</td>
<td>1362 (B)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>1.153</td>
<td>1394 (C)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>1.119</td>
<td>1436 (D)</td>
<td>1445</td>
</tr>
<tr>
<td></td>
<td>1.100</td>
<td>1462 (E)</td>
<td>1455</td>
</tr>
<tr>
<td></td>
<td>1.085</td>
<td>1482 (F)</td>
<td>1492</td>
</tr>
<tr>
<td>-9</td>
<td>1.168</td>
<td>1548</td>
<td>1546</td>
</tr>
</tbody>
</table>
structures are proposed for kinetically trapped intermediates of the i-motif DNA folding which allows for a better mechanistic understanding of the free energy landscape.

In addition to structural changes of the i-motif DNA as a consequence of the noncanonical hemiprotonated cytosine base pairing, the presence of inner cations may also induce preference for alternative kinetically trapped conformations in the gas-phase. For example, closer inspection of the higher charge state distributions showed the presence of multiple sodium and ammonium adducts for $z = -7$ and $z = -8$ (see Figure 5.6). These adducts can be attached to the i-motif-DNA structure and result in: i) no conformational changes (no shift in the IMS profile), or ii) can stabilize the i-motif DNA structure in other folding states (shift in the IMS profile). Inspection of the mobility profiles in the presence of adducts shows that while ammonium incorporation does not drive major changes in the mobility profile, sodium adducts act as a positively charge contact point between negatively charged strands (most likely phosphate groups) of the i-motif-DNA resulting in more compact structures. Since both $z = -7$ and $z = -8$ are partially unfolded, the appearance of new mobility bands suggests that sodium ions acts as bridges between the CCC strands, stabilizing the complex in other local free energy minima. That is, the introduction of the sodium cation may allow CCC strands to come together and create a sodium mediated bond, forming a tighter structure (smaller CCS). The interaction of inner cations on DNA folding will be further investigated utilizing this methodology in a following publication.
Figure 5.6. Typical mass spectra and mobility profiles for different i-motif DNA molecular ions observed at the $z=-7$ and $z=-8$ charge state. Notice the change in the mobility distribution in the presence of the ammonium and sodium adducts.
5.6 Conclusions

Folding dynamics of i-motif DNA in solution and in the gas-phase were studied using circular dichroism, photoacoustic calorimetry, trapped ion mobility spectrometry and molecular dynamics. Solution studies using CD showed a two state folding mechanism (F↔U) for the i-motif DNA as a function of the solution pH = 4.0 – 9.0 at thermodynamic equilibrium. The photolysis of 2-NBA in the presence of i-motif DNA displayed a phase shift indicating the presence of two processes, one occurring within 50 ns (fast phase) of the laser pulse and a temperature-dependent process with a time constant of 140 ns (slow phase) at 20 °C with an overall reaction enthalpy change (ΔH_{fast} + ΔH_{slow} = 43 ± 7 kcal mol\(^{-1}\)). The observed volume change during PAC experiments suggested the formation of hemiprotonated cytosine base pairs and concomitant collapse of the i-motif oligonucleotide into a more compact conformation with a smaller volume during the pH 7.0 → 4.5 transition at kinetic equilibrium. The nanoESI-TIMS-MS experiments showed that during the nanoESI process the i-motif DNA structures are kinetically trapped in intermediates that retain the main conformational motifs (C•CH\(^{+}\) base pairing) for low charge states. In particular, an i-motif DNA structure is observed for the low charge states in both positive and negative ion mode (e.g., +/- 3 to +/-5) at low pH conditions. With the pH increase the cytosines become neutralized leading to the loss of the cytosine-cytosine\(^{+}\) (C•CH\(^{+}\)) base pairing in the CCC stretches and a charge state distribution increase to higher charges is observed in negative mode. The loss of the cytosine-cytosine\(^{+}\) (C•CH\(^{+}\)) base pairing in the CCC stretches results in the observation of partially unfolded i-motif DNA conformations. CIA-TIMS-MS experiments showed a broad IMS distribution for z = -5, -6 and -7 and the existence of multiple, local free energy minima different from those obtained by varying
the pH solution conditions. Molecular dynamics simulations were used to sample the gas-phase conformations and to generate candidate structures during i-motif DNA unfolding. For the first time, an unfolding mechanism is supported by average candidate structures from mobility measurements as a function of the i-motif DNA unfolding pathway and the main intramolecular interactions that stabilized the process are described. Inspection of partially unfolded i-motif DNA structures, for intermediate charge states in negative ion mode, showed that sodium, but not ammonium, induces changes in the population of gas-phase kinetic intermediates.

5.7 Acknowledgements

This work was supported by the National Institute of Health (Grant No. R00GM106414) and a FFL Bruker Daltonics, Inc. fellowship. The authors would like to thank Dr. Carlos Larriba-Andaluz and Dr. Christian Bleiholder for helpful discussions during the theoretical CCS calculations. We will like to acknowledge the Instructional & Research Computing Center (IRCC) at Florida International University for providing computing resources.

5.8 References


40. Ridgeway, M. E.; Silveira, J. A.; Meier, J. E.; Park, M. A., Microheterogeneity within conformational states of ubiquitin revealed by high resolution trapped ion mobility spectrometry. *Analyst* 2015, 140 (20), 6964-6972.


102. Leroy, J. L.; Gueron, M., Solution structures of the i-motif tetramers of d(TCC), d(5methylCCT) and d(T5methylCC): novel NOE connections between amino protons and sugar protons. Structure 1995, 3 (1), 101-120.


CHAPTER 6: AT-HOOK PEPTIDES BIND THE MAJOR AND MINOR GROOVE OF AT-RICH DNA
6.1 Abstract

The mammalian high mobility group A2 protein (HMGA2) houses three motifs that preferentially bind short stretches of AT-rich DNA regions. These DNA binding domains, known as “AT-hooks”, are traditionally characterized as being unstructured. Upon binding to substrates, they form ordered assemblies. It is this disordered-to-ordered transition that has implicated HMGA2 as a protein actively involved in many biological processes, with abnormal HMGA expression linked to a variety of health problems including diabetes, obesity and oncogenesis. In the current work, gas-phase ion mobility spectrometry -mass spectrometry (IMS-MS) analysis was used to observe the conformations populated by each AT-hook peptide (ATHP) and DNA complex. The resulting mobility distributions showed that the three ATHPs bind DNA to form a single “key-locked” conformational ensemble. Together with molecular models, complex formation was determined to occur between ATHP and a pre-folded AT-rich DNA hairpin substrate. Interestingly, candidate structural analysis also revealed the ability of the peptides to bind the minor and major grooves of the DNA. Employing additional gas-phase IMS-MS studies and solution isothermal titration calorimetry (ITC) we observe for the first time results of a 2:1 stoichiometry of ATHP:DNA.

6.2 Introduction

HMGA2 is a 12 kDa member of the non-histone chromosomal HMG protein family. Three DNA binding motifs present in the protein, known as “AT-hook” domains, exhibit specificity for binding to AT-rich DNA sequences. A conserved set of tripeptide core residues (-Arg-Gly-Arg-), primarily surrounded by basic amino acids, are responsible for this specificity. Variants of HMGA2 have been linked to human height and levels.
of expression are associated with obesity,\textsuperscript{6-7} diabetes,\textsuperscript{8} human intelligence,\textsuperscript{9} stem cell youth\textsuperscript{10} and oncogenesis.\textsuperscript{11-14} Such involvement in significant biological processes, specifically the direct binding to transcription factors, has implicated HMGA proteins as potential therapeutic drug targets.\textsuperscript{15-17}

However, HMGA2 provides an additional level of interest because of its classification as an intrinsically disordered protein\textsuperscript{18-19}. The presence of ‘unstructured’ biomolecules in eukaryotic proteins comprises a surprisingly large number, with as many as 25\% predicted to be disordered.\textsuperscript{20} The important role these proteins have in maintaining cellular function most likely comes from their ability to conform to multiple binding partners. For example, in the absence of an AT-rich DNA substrate, the high mobility group AT-hook 2 (HMGA2) is highly flexible and adopts random coils.\textsuperscript{18} However, it is the conformational changes upon DNA binding, and subsequent transition from unstructured to structured, that allows HMGA2 to be involved in several fundamental biological processes, including DNA replication, translation, recombination and gene regulation.\textsuperscript{1, 9, 11, 21-22}

Through the use of various techniques, structural elucidation of protein-DNA complexes have grown substantially, as knowledge of complex interfaces have been crucial to the development of new therapeutic targets and better disease prevention strategies,\textsuperscript{23} yet the specifics still remain unclear regarding the conformational changes required by the protein and DNA to form complexes.\textsuperscript{24} Previous studies, using NMR\textsuperscript{18} and crystallography,\textsuperscript{15} provide insight into the intermolecular interactions that support the binding of the AT-hook motifs with DNA. The consensus is that the tripeptide core motif is positioned deep into the minor groove forming hydrogen bond contacts with the bases,
along with a variety of electrostatic and van der waals interactions. More recent experimental analysis, involving ATHP complexes, show the ability for RNA binding via an extended peptide motif\textsuperscript{25} and increase in DNA binding affinity upon conjugation to a minor groove binder.\textsuperscript{26} Unfortunately, information comparing the structural differences between the three AT-hook domains and their capacity to form DNA complexes is limited, confined only to select solution-phase and x-ray diffraction studies.

However, the combination of solution and gas-phase techniques, has proven to be very versatile and powerful in terms of structural analysis of biomolecules. In particular, intermediate states can be easily probed by IMS-MS technology, which when complemented by solution measurements provides a definitive picture of complex structural isomers, binding affinity and molecular stability.

In the present work, we report on the conformational changes and ATHP-DNA complex formation of the three AT-hook peptides of HMGA2 in solution and in the gas-phase using DNA UV melting studies, fluorescence spectroscopy, isothermal titration calorimetry (ITC), IMS-MS and molecular modeling. The main motifs that define the conformational space of the HMGA2 domains (ATHP 1-3) and the DNA-ATHP complex formation are described. Although prior studies note only minor groove ATHP interactions, here we present solution- and gas-phase experimental and theoretical evidence of both ATHP major and minor groove binding capabilities.

6.3 Experimental

6.3.1 Materials and reagents

An AT-rich DNA oligomer, denoted as FL876, sequence GGATATTGCCCCCGCAATATCC (C\textsubscript{212}H\textsubscript{270}N\textsubscript{79}O\textsubscript{130}P\textsubscript{21}, MW 6655.1561) was
purchased from Eurofins Scientific (Luxembourg, Luxembourg) and used as received. This 22 nucleotide DNA hairpin contains a 9 base pair stem including a 5 base pair AT DNA in the middle of the stem. AT-hook peptides 1 (Lys-Arg-Gly-Arg-Gly-Arg-Pro-Arg-Lys), 2 (Pro-Lys-Arg-Pro-Arg-Gly-Arg-Pro-Lys) and 3 (Lys-Arg-Pro-Arg-Gly-Arg-Pro-Arg-Lys-Trp) were purchased from Advanced ChemTech Inc. (Louisville, KY) and used as received. Solvents, BPE buffer, and ammonium acetate salts utilized in this study were analytical grade or better and purchased from Fisher Scientific (Pittsburgh, PA). A Tuning Mix calibration standard (G24221A) was obtained from Agilent Technologies (Santa Clara, CA) and used as received.

6.3.2 DNA UV melting studies

DNA UV melting curves were determined using a Cary 100 UV-Vis spectrophotometer (Agilent Technologies, Santa Clara, CA) equipped with a thermoelectric temperature-controller. DNA oligomer FL876 in 1×BPE buffer was used for melting studies. DNA-peptide samples were prepared to a final concentration of 2 µM by directly mixing at a molar ratio of 1:10 FL876 DNA with each ATHP peptide, followed by incubation for 60 min at room temperature to ensure equilibration. Samples were typically heated in the 20-100 °C range at a rate of 1 °C min⁻¹, while continuously monitoring the absorbance at 260 nm. Primary data were transferred to the graphic program Origin (MicroCal, Inc., Northampton, MA) for plotting and analysis.

6.3.3 Fluorescence Measurements

Fluorescence spectra of DNA-ATHP 3 were acquired using a Jobin Yvon Horiba FluoroMax-3 with excitation wavelength of 355 nm. In the titration experiments, 99 nM
FL876 and 50 nM Hoechst stain in 1xBPE was titrated by increasing concentrations of ATHP 3. The fluorescence spectra were recorded from 400 to 550 nm.

6.3.4 Isothermal Titration Calorimetry (ITC)

ITC experiments were performed using a VP-ITC titration calorimeter (MicroCal, Inc., Northampton, MA) interfaced to a personal computer. Origin 7.0 was used for data acquisition and processing. A typical ITC experiment was set up so that 10 uL of ATHP ligand, i.e. 300 uM ATHP1, 300 uM ATHP2 and 150 uM ATHP3, was injected every 120 seconds for 30 injections into a 5 uM DNA sample (1 x BPE buffer) in the sample cell.

6.3.5 Trapped Ion Mobility Spectrometry – Mass Spectrometry Analysis (TIMS-MS)

Details regarding the TIMS operation and specifics compared to traditional IMS can be found elsewhere. Briefly, mobility separation in TIMS is based on holding the ions stationary against a moving gas using an electric field. The separation in a TIMS device can be described in the center of the mass reference frame using the same principles as in a conventional IMS drift tube. Since mobility separation is related to the number of ion-neutral collisions (or drift time in traditional drift tube cells), the mobility separation in a TIMS device depends on the bath gas drift velocity, ion confinement and ion elution parameters. The reduced mobility, $K$, of an ion in a TIMS cell is described by:

$$K = \frac{V_g}{E} = \frac{A}{(V_{elution} - V_{out})}$$

where $v_g$ and $E$ are the velocity of the gas and the applied electric field across the TIMS analyzer region. $V_{elution}$ is the voltage when the ions elute in the $V_{ramp}$ sweep and $V_{out}$ is the voltage applied at the end of the TIMS analyzer region.
A custom-built, pulled capillary nanoESI source was utilized for all the experiments. Quartz glass capillaries (O.D.: 1.0 mm and I.D.: 0.70 mm) were pulled utilizing a P-2000 micropipette laser puller (Sutter Instruments, Novato, CA) and loaded with 10 µL aliquot of the sample solution. Sample solutions consisted of 1-10 µM ATHP or DNA in 10 mM ammonium acetate solution at physiological pH (pH = 6.7). For the observation of the DNA-ATHP complexes, a 1:1 and 1:2 ratio of 5 µM concentration of the DNA and AT-hook peptide (1, 2 or 3) was prepared in 10mM ammonium acetate immediately prior infusion. A typical nanoESI source voltage of +/- 600-1200 V was applied between the pulled capillary tips and the TIMS-MS instrument inlet. Ions were introduced via a stainless steel tube (1/16 x 0.020”, IDEX Health Science, Oak Harbor, WA) held at room temperature into the TIMS cell.

Mobility calibration was performed using the Tuning Mix calibration standard (G24221A, Agilent Technologies, Santa Clara, CA) in positive ion mode (e.g., \( m/z = 322, K_0 = 1.376 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1} \) and \( m/z = 622, K_0 = 1.013 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1} \)).\(^{31}\) The TIMS operation was controlled using in-house software, written in National Instruments LabVIEW, and synchronized with the maXis Impact Q-ToF acquisition program.\(^{27-28}\)

Reduced mobility values (\( K_0 \)) were correlated with CCS (\( \Omega \)) using the equation:

\[
\Omega = \frac{(18\pi)^{1/2}}{16} \frac{z}{(k_B T)^{1/2}} \left[ \frac{1}{m_i} + \frac{1}{m_b} \right]^{1/2} \frac{1}{K_0 N^*}
\]

where \( z \) is the charge of the ion, \( k_B \) is the Boltzmann constant, \( N^* \) is the number density and \( m_i \) and \( m_b \) refer to the masses of the ion and bath gas, respectively.\(^{32}\)
3.3.6 Theoretical Calculations

Candidate structures were proposed for the mobility bands observed in the TIMS-MS experiments. The protein data bank (PDB) entry 2EZD was used as a template for the peptide-DNA candidate structures. First, a 2-dimensional DNA hairpin structure was obtained for the DNA sequence using mFold server\textsuperscript{33}. Based on this structure, a 3-dimensional hairpin structure was then created by merging PDBs 2EZD (for the stem) with 2K71 (for the tetra-nucleotide loop) followed by mutations with appropriate bases in order

Scheme 1. Preparation of PDB template for modeling of DNA.
to achieve the desired structure (see Scheme 1). The modeled DNA hairpin structure was optimized by energy minimization with CHARMM36 force field\textsuperscript{34} using NAMD Molecular Dynamics package\textsuperscript{35}. Molecular docking was performed using AutoDock vina\textsuperscript{36} to generate the DNA-ATHP complexes. Theoretical CCS were calculated using the IMoS (v1.04b)\textsuperscript{37-39} and PSA\textsuperscript{40} packages with nitrogen as a bath gas at ca. 300K. In the IMoS calculations, 100 total rotations were performed using the trajectory method with a Maxwell distribution. Molecular visualization was performed using Visual Molecular Dynamics software.\textsuperscript{41}

### 6.4 Results and Discussion

Gas phase analysis of ATHP 1, 2 or 3 in complex with the FL876 oligonucleotide produced [M+4H]\textsuperscript{+4} and [M+5H]\textsuperscript{+5} ions (Figure 6.1). MS analysis of equal molar sample ratios shows that a 1:1 peptide with DNA stoichiometry is produced, accompanied by each species in free form (Appendix 6.1, left panel). Inspection of the corresponding IM complex profiles show a single mobility band recorded for each charge state. Structures were proposed based off these initial findings and suggest the likelihood of the ATHPs attaching to either the major or minor grooves of the DNA hairpin. This contrast from current reports in literature (e.g., ATHPs are specific minor groove binding molecules) subsequently motivated a series of experiments to evaluate the binding specificity of each motif.

We proceeded by using Hoescht 33528, a fluorescent dye known as a minor groove binding (MGB) compound that is highly specific to AT-rich DNA substrates. By blocking the minor groove, ATHP attachment was restricted to the major groove of FL876 so that the presence or absence of binding could be determined. Although displacement of the
Hoescht 33258 compound in the minor groove by ATHPs is possible, we hypothesized that the majority of ATHPs would attach to the major groove, if applicable, leaving the minor groove occupied by the dye. In fact, experimental IMS-MS results confirmed that all three ATHPs have the ability to bind to the FL876 : Hoescht complex, forming 1 : 1 : 1 molecular assemblies (Figure 6.2). Additional theoretical studies were performed with the MGB compound and complement the initial findings of ATHP major groove binding capabilities. Comparison between the relative abundance of the complexes in the mass spectra showed that ATHP : DNA is significantly favored over the ATHP : DNA : MGB state (data not shown), signifying that ATHP bound to the minor groove is preferred.

Upon confirmation of the interaction between ATHP and the major groove of FL876, the peptide concentration was increased two-fold, with respect to DNA, to assess the possibility of two peptides with one DNA (bound to the major and minor grooves). An increase in ATHP : DNA binding to 2:1 was observed at [M+5H]+5 charge state (Figure 6.3).

The experimental IMS-MS results suggest that the small difference in CCS distributions (<200 Å²) between the 1:1 and 2:1 complexes could be an indication that the peptide(s) is mostly concealed within the DNA hairpin structure. The proposed structures provide an illustration of the complex, showing that the size is mostly dictated by the DNA. An overview of the experimental and theoretical cross sections for the various complexes is provided in Appendix 3.1. Results from our IMS-MS experiments and theoretical docking studies are in good agreement with previous structural findings, in that NMR and X-ray crystallography studies have reported that the –Arg-Gly-Arg– core of the AT-hook peptide is buried into the minor groove of DNA, and that complexes are formed via
Figure 6.1. Typical mobility profiles for the [M+4H]4+ and [M+5H]5+ charge states of ATHP 1, 2, or 3 in complex with the DNA hairpin (FL876). The isotopic distribution for each complex is shown (inset) and candidate structures for each complex are proposed.
hydrogen bonds between the oxygen atoms of thymine and/or cytosine, while Van der Waals forces govern interactions with adenines. The lysines and arginines outside of this core extend tightly along the DNA forming additional points of contact as a result of electrostatic interactions with the negatively charged DNA phosphate backbone.

Comparison of the mobility distributions of the 1:1 and 2:1 peptide : DNA stoichiometries suggest that the ATHPs bind to the AT-rich regions of the DNA in a single “locked” conformation and that all three ATHPs are selectively stabilized upon binding to the pre-formed FL876 hairpin (Scheme 2). The schematic model, determined using CCSs from IMS, illustrates conformational changes of the peptide (i.e., unfolding) induced upon minor and/or major groove substrate binding. From this process we begin building a picture of the disordered-to-ordered transition of HMGA2 as it assembles with DNA, where the binding occurs indistinctly via the minor or major groove. Specifics regarding the HMGA2:DNA complex will be further investigated utilizing IMS-MS methodology in a following publication.

Scheme 2. Mechanism of ATHP attachment to DNA showing the preformed hairpin prior to peptide unfolding upon binding.
Figure 6.2. Typical mobility and isotopic distribution of DNA complexes. Single mobility peaks were observed for all complexes and charge states. Specifically, the DNA and the minor groove binding (MGB) Hoescht 33258 compound produced and [M+4H]+4 charge state. ATHP 1 binding to the DNA-Hoescht complex produced [M+4H]+4 and [M+5H]+5 charge states. Proposed structures are also shown.
Figure 6.3. Typical mobility profiles of ATHP 1, 2, or 3 in complex with the DNA hairpin (FL876). Stoichiometry of 2:1 was observed at a [M+5H]+5 charge state. The isotopic distribution for each complex is shown (inset) and candidate structures shown for the FL876 hairpin in complex with ATHP1, ATHP2 and ATHP3.

Complexes were also probed using a variety of complementary solution-phase techniques (e.g., DNA melting experiments, isothermal titration calorimetry (ITC) and fluorescence measurements) and gas-phase collision induced dissociation (CID-MS) in order to further evaluate our IMS findings. ATHP 1, 2 and 3 minor groove binding, via fluorescent titration experiments, as a function of Hoechst 33258 compound displacement was observed (Appendix 6.1). Specifically, increasing concentration of ATHP resulted
in a decreased fluorescence signal, an indication of the Hoechst stain being sequentially removed from the DNA minor groove and replaced with ATHP.

Hairpin formation was confirmed by a sharp DNA UV melting transition at 59 °C in 1xBPE (see Figure 6.4a). The ATHPs 1–3 showed a melting temperature (T$_m$) increase, indicative of binding between each peptide and hairpin. The thermal stabilization of FL876 by ATHP 1 and ATHP 2 is greater than that caused by ATHP 3.

Results from ITC analysis support observations regarding the two binding site hypothesis. Figure 6.4b shows the data retrieved from these experiments. A ‘normal’ titration curve was observed for all three peptides titrated into DNA. The enthalpy change over the titration period was measured to determine exothermic vs. endothermic binding. In fact, the DNA-binding enthalpy for ATHPs attaching in the major and minor grooves was estimated. The binding of ATHP1 to FL876 is an exothermic reaction, which yields negative enthalpies of -1.24 kcal mol$^{-1}$ and -1.30 kcal mol$^{-1}$ referring to binding sites one and two (e.g., major or minor groove), respectively. In contrast, the binding of ATHP2 to FL876 proved to be an endothermic reaction, with a positive enthalpy of 1.048 kcal mol$^{-1}$ calculated for both of the binding sites. The titration of ATHP3 into FL876 also displayed a 2:1 binding stoichiometry and the resulting exothermic reaction yielded negative enthalpies of -1.17 kcal mol$^{-1}$ and -1.71 kcal mol$^{-1}$, respectively.

Complementary gas-phase, low energy collision induced activation experiments on the [M+4H]$^{4+}$ charge state ofATHP and FL876 1:1 complexes, were used to evaluate the relative binding affinities (Figure 6.4c). As the collision energy increases, a decrease in the molecular ion signal corresponding to the complex was observed; notice that collision
Figure 6.4. a) DNA UV melting curves for uncomplexed FL876 and complexed FL876 with ATHP 1, 2 or 3 and b) ITC results determining the binding stoichiometry and affinity of ATHP 1, 2 and 3 to both the major and minor groove of FL876 DNA and c) Collision induced dissociation for FL876 complexed with ATHP1. ATHP2 or ATHP 3 (charge state and degrees of freedom were considered) with the curve derivative shown as an inset.
energy has been normalized to the charge and calculated per degree of freedom to account for mass difference and charge between the experiments, a consideration supported through previous work by Clemmers and co-worker.\textsuperscript{46} Comparison between the CID profiles of the ATHP 1-3 : FL876 complexes shows that the weakest binding profile is observed for ATHP 2 : FL876, followed by ATHP 3 : FL876 complex, and ATHP 1: FL876 complex. We interpret these CID profiles in that as the collision energy increases, multiple collisions with the collision gas (nitrogen in this case) leads to an increase in the vibrational modes of the complex eventually leading to dissociation into the main constituents (i.e., ATHP and FL876). That is, as the CID energy increases the internal energy increases until a threshold is achieved, leading to complex dissociation. The increased energy needed to begin dissociation of ATHP 1 : FL876 and ATHP 3 : FL876 distribution suggests that two distinct binding sites (e.g., minor and major groove) are occupied. The more immediate dissociation of the ATHP 2-FL876 complex suggest a predominant single binding site or two weaker attachment sites. The CID profiles combined with the fluorescent titration and ITC data indicate that ATHP 1, 2 and 3 can bind hairpin DNA and occupy both minor and major grooves.

6.5 Conclusions

The high mobility resolution and trapping capabilities of TIMS-CID-TOF MS allowed for the interrogation of ATHP 1, 2 or 3 in complex with hairpin FL876 conformational space. IMS-MS results showed that ATHP 1-3 can associate with DNA to form a complex in a “locked” mechanism enabled by interactions of the –Arg-Gly-Arg– tripeptide core and the AT-rich regions of the DNA duplex. Comparison of the ATHP 1-3 : FL876 DNA IMS profiles with candidate structures permitted the generation of a
mechanistic description of the complex formation. The binding mechanism requires the FL876 DNA to be pre-folded into a hairpin that will serve as substrate for the ATHP 1-3 attachment and transition from unstructured to structured. The experimental evidence from TIMS-MS, UV melting studies, fluorescent titration, ITC, and CID-MS experiments infer that the ATHP 1-3 peptides can indistinctively bind to the minor and major grooves of FL876 DNA.

6.6 Acknowledgements

The authors acknowledge the financial support from the National Science Foundation Division of Chemistry, under CAREER award CHE-1654274, with co-funding from the Division of Molecular and Cellular Biosciences to FFL. AG will like to acknowledge the 2017-2018 Dissertation Year Fellowship from Florida International University.

6.7 References

1. Chen, B.; Young, J.; Leng, F., DNA Bending by the Mammalian High Mobility Group Protein AT Hook 2. Biochemistry 2010, 49 (8), 1590-1595.


CHAPTER 7: PEPTIDE SEQUENCE INFLUENCE ON THE
CONFORMATIONAL DYNAMICS AND DNA BINDING OF THE
INTRINSICALLY DISORDERED AT-HOOK 3 PEPTIDE

This chapter has been submitted on February 22, 2018 for publication in Scientific Reports
7.1 Abstract

The intrinsically disordered ATHP3 was characterized at physiological conditions and in complex with DNA using single amino acid substitutions and high-resolution ion mobility spectrometry coupled to mass spectrometry (trapped IMS-MS). Results showed that ATHP3 can exist in multiple conformations at physiological conditions (at least 10 conformers were separated), where differences in the conformational space are driven by the proline cis/trans orientation and the number and location of protons. When in complex with AT rich DNA hairpins, the -RGRP- core is essential for stabilizing the ATHP3 : DNA complex at physiological conditions. In particular, the arginine in the sixth position plays the most important role during binding to AT-rich regions of hairpin DNA, in good agreement with previous NMR and X-ray data. Mobility based correlation matrices are proposed as a way to reveal differences in structural motifs across the peptide mutants based on the conformational space and relative conformer abundance.

The mammalian high mobility group protein (HMGA2, MW ~12 kDa), an intrinsically disordered protein (IDP) that aids in the regulation and expression of certain genes by influencing the remodeling of chromatin structure, is also a known oncofetal antigen, biomarker of cell transformation and metastasis, and a target in cancer therapeutics. In vivo studies have largely established the physiological role of the full length HMGA2 by monitoring the effects of overexpression and knockouts in
mice models. Specifically, HMGA2 gene amplification induces tumorigenesis, as observed in several cancers, including leukemia, tongue cancer, lung cancer, breast cancer, and pituitary adenomas. A pygmy phenotype was observed in null mice, whereas the truncated version of the protein is expressed in oversized mice. More recently, in vitro analysis implicated HMGA2 in the process of aging. HMGA2 contains three positively charged motifs – known as “AT-hooks” that bind with high affinity AT-rich DNA duplex regions and a negatively charged c-terminal tail. HMGA2 can easily form homodimers, an interaction most likely attributed to contacts forming between the acidic c-terminal tail of one protein with the basic residues of the counterpart protein, and interact with DNA and other proteins, making HMGA2 a connected protein with a large interaction network. Despite progress made using in vivo and in vitro studies, only few studies have contributed to the structural characterization (e.g., NMR and crystallography studies of ATHP2 and ATHP3 bound to duplex DNA).

Ion mobility spectrometry and mass spectrometry (IMS-MS) have shown the advantages of these techniques for structural biology, and more specifically, for the study of biomolecules and biomolecular complexes. IMS-MS has addressed specific shortcomings of current conventional methods of structural analysis (i.e. NMR and X-ray crystallography) by reducing analysis time, sample consumption, sample purity requirements and restriction on biomolecule flexibility. IMS-MS has gained
attention for the study of dynamic proteins in their native, partially folded and unfolded forms\textsuperscript{24, 30-32} and when complemented with molecular dynamic simulations has allowed the identification of specific intramolecular interactions that stabilize conformational states (i.e. proline and side-chain orientation, protonation site).\textsuperscript{33} IMS-MS can capture conformational interconversions and intermediate states for peptides, proteins, DNA and complexes.\textsuperscript{24, 25, 34} For example, in a recent study, the influence of solvent composition and time after desolvation dynamics on the ATHP3 conformational space was reported.\textsuperscript{33} At the protein level, it has been shown that phosphorylation of the c-terminal tail can play a vital role during HMGA2 protein-protein and protein-DNA interactions.\textsuperscript{35}

In the present study, the peptide sequence influence on the conformational space and DNA binding for the intrinsically disordered ATHP3 is studied using trapped IMS coupled to mass spectrometry (TIMS-CID-TOF MS) and single amino acid substitutions. A discussion on the conformational space as a function of the charge location and proline orientation and binding affinity to DNA follows.

\textbf{7.2 Results and Discussion}

Under native conditions, three charge states in positive ion mode ([M+H]\textsuperscript{+} to [M+3H]\textsuperscript{3+}) were observed for the ATHP3 and the seven amino acid substituted peptides (see corrected CCS profiles and fingerprint in Figure 7.1 a-b). A correlation matrix of the IMS fingerprints (Figure 7.1c) is used for rapid comparison and identification of the
single amino acid substitution with the smallest/largest influence in the ATHP3 mobility profile at physiological condition (first column and first row), as well as between the mutants using equation 3. For example, following the color scale, P3A and R8A are the most and R2A is the least closely related mutants to ATHP3. That is, a quick assessment of the influence of the primary sequence (e.g., single amino acid substitution) on the peptide secondary structure can be observed.

One of the advantages of using single point residue mutations in combination with IMS-MS, is the possibility to identify the conformational state of the prolines (cis/trans) as well as the charge location for the case of the basic residues by comparison of the corrected IMS profiles. While it is known that prolines in the trans position are energetically more favorable than in the cis position (95% of the time, being ~0.5 kcal mol\(^{-1}\) lower in energy with ~13 kcal mol\(^{-1}\) barrier), when substituted by alanine which is energetically more favorable in trans position (~2.5 kcal mol\(^{-1}\) lower in energy with ~20 kcal mol\(^{-1}\) barrier), allows for the determination of the ATHP3 state by looking at the conservation (proline in trans) or absence (proline in cis) of the IMS band. That is, comparison of the corrected profiles for ATHP3 and for P3A and P7A permits the assignment of the proline in ATHP3 for the different bands observed (results summarized in Table 1). Analogous, single point mutations of the basic residues allow the assignment of the charge. In the case of the ATHP3, up to seven protons can
Figure 7.1. Typical, normalized ATHP3 and corrected variant (*) mobility profiles for the $[M+H]^+$, $[M+2H]^2+$ and $[M+3H]^3+$ for intrinsic size parameters upon residue substitution to alanine (a). Peptide IMS fingerprint (b) and correlation matrix (c) are used for assessment of the primary sequence effect on the secondary structure.
be attached; however, during native conditions (pH ~ 6.7), a maximum of three protons were observed. That is, the high number of IMS bands can be attributed to different locations of the proton (i.e., protomers) in addition to the potential constraints generated from the proline configuration. Comparison of the ATHP3 profile with those of R2A, R4A, R6A and R8A permitted the assignment of the number and position of the charge in the intrinsically disordered ATHP3. Moreover, knowing the total charge and number of protons per band, the protonation probability of the N-term and the lysine at position one was calculated (difference from the charge state and number of protonated arginines). In addition to this information, the W10A substitution provides insight in the role of the tryptophan on the overall conformational space.

Inspection of Table 1 shows that ATHP 3 [M+H]+ conformers (i.e., A and B bands) differ in the orientation of P3 (trans vs cis) and the location of the charge (R8 vs N-term/K).

In the case of ATHP 3 [M+2H]2+, differences between the five bands (C-G) are also associated to the orientation of the prolines P3 and P7 and the protonation of the basic residues R2, R4, R6 and R8. In the case of ATHP [M+3H]3+, differences between the five bands (H-J) are associated to the proline P7 and the protonation of the basic residues R2,
Table 7.1. Summary of charge location and cis/trans proline configuration of the Native ATHP3 per IMS band (A-J).

<table>
<thead>
<tr>
<th>Residue</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>R2</td>
<td></td>
<td></td>
<td>N</td>
<td>N</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>N</td>
<td>+</td>
</tr>
<tr>
<td>P3</td>
<td>trans*</td>
<td>cis</td>
<td>trans</td>
<td>trans</td>
<td>trans*</td>
<td>trans</td>
<td>cis</td>
<td>Cis</td>
<td>cis</td>
<td>cis</td>
</tr>
<tr>
<td>R4</td>
<td>N</td>
<td>N</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>N</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>R6</td>
<td>N</td>
<td>N</td>
<td>+</td>
<td>+</td>
<td>N</td>
<td>N</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>P7</td>
<td>trans</td>
<td>trans</td>
<td>trans</td>
<td>trans</td>
<td>cis</td>
<td>cis</td>
<td>cis</td>
<td>Cis</td>
<td>trans*</td>
<td>trans*</td>
</tr>
<tr>
<td>R8</td>
<td>+</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>+</td>
<td>+</td>
<td>N</td>
<td>+</td>
<td>N</td>
</tr>
<tr>
<td>W10</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>N-ter/K</td>
<td>N</td>
<td>+</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>+</td>
<td>N</td>
<td>+</td>
</tr>
</tbody>
</table>

R4, R6 and R8. While in some cases the charge residue that is protonated can be uniquely assigned, for the case of the IMS band I there are at least two protonation patterns that yield the same CCS value; that is, different protonation patterns can share the same IMS band. In summary, the data presented in Table 1 shows that the intrinsically disordered ATHP3 can exist in a variety of conformations (at least 10), where the proline configuration and the charge location drive the intramolecular
interactions that stabilizes the peptide secondary structure in multiple kinetic intermediates.

Peptide-DNA complexes were formed in solution and observed in the gas-phase at \([M+4H]^{4+}\) and \([M+5H]^{5+}\) charge states under native spray conditions for all considered peptide mutants (Figure 7.2a). Beside the peptide-DNA complexes, unbound DNA and peptides were also observed at \([M+3H]^{3+}\) - \([M+4H]^{4+}\) and at \([M+H]^{+}\) - \([M+3H]^{3+}\) charge states, respectively. The relative abundance of unbound peptide and DNA was consistently higher for the \([M+2H]^{2+}\) and \([M+4H]^{4+}\) charge states, respectively. IMS profiles of the peptide-DNA complexes (Figure 7.2b and 7.2c) can be characterized by a single, broad mobility band - depicting a comparatively stable complex. The IMS peak width was consistent as a function of the charge states and the differences in CCS between the \([M+4H]^{4+}\) and \([M+5H]^{5+}\) charges suggest that no conformational rearrangements occur with the charge states (~20-100 Å²). Inspection of the peptide-DNA fingerprint plots, generated from the corrected IMS profiles, allows for a quick description of the influence of the single amino acid substitution on the structure of the peptide-DNA complex (Figure 7.2c). Further information can be extracted from the correlation matrix (Figure 7.2d). If the single amino acid substitution does not dramatically affect the corrected IMS profiles (similar fingerprint) or has a high correlation value, the most likely this amino acid has a lower role during the peptide-DNA complex formation and is not an active participant in the binding site. For
example, comparison of the ATHP3 and the R2A/R4A/R8A/W10A peptides shows higher similarities and correlation values, which suggest that these amino acids are less likely to regulate and participate in the peptide-DNA complex interaction. Moreover, P3A/R6A/P7A showed larger differences.

Figure 7.2. Typical mass spectra (a) and ‘corrected’ ion mobility profiles (b) of the native and substituted ATHP3 : DNA complexes. The IMS profile fingerprint (c) of the complexes was used to generate a correlation matrix (d). Values reported in (a) correspond to peptide : DNA : complex ratios.
and lower correlation values, which suggest that there are most likely involved in the peptide-DNA complex interaction. These observations are in good agreement with previous NMR results that suggest that the -PRGRP- core most likely interacts with the DNA.\textsuperscript{22} A crystal structure of the ATHP3 in complex with duplex B-form DNA revealed that the central core arginine residues protrude inward creating hydrogen bonds from the amino acid NH atom to the thymine oxygen atom.\textsuperscript{23}

A complementary way to evaluate the influence of each amino acid on the ATHP3 : DNA complex is through the inspection of the dissociation curve as a function of the collision energy (Figure 7.3a). The assumption is that the rupture of the intramolecular interactions that stabilizes the peptide-DNA complex using single amino acid substitutions will result in different collision energies required to disrupt the complex. Inspection of Figure 7.3a shows that the R6A peptide presents the largest change in dissociation profiles, which suggest that the arginine in the sixth position is essential for the stabilization of the ATHP3 : DNA complex in the -RGR- core, in good agreement with previous NMR observations.\textsuperscript{22,46}

The probability of peptide : DNA complex formation (or binding affinity) can also be evaluate by the differences in the relative abundance of the product (i.e., complex) with respect to the initial reactants (i.e., peptide and DNA) using equation 4 and 5 (Figure 7.3b and 7.3c). Inspection of Figure 7.3b and 3c shows that the substitutions R2A/P3A/W10A has higher binding affinity (K\textsubscript{a}) relative to the ATHP3
peptide, while the substitutions R4A/R6A/P7A has lower binding affinity compared to the ATHP3. These results suggest that R4A/R6A/P7A are directly related to the ATHP3 : DNA complex formation.

Figure 7.3. Typical CID profiles (a) showing the decrease in intensity of the [M+4H]4+ peptide : DNA complexes as a function of collision energy (eV), charge (z) and degrees of freedom (DoF). The dissociation threshold (inset) shows the absolute derivative of each CID profiles as a function of the collision energy. The binding affinity (b) and distance matrix (c) illustrates the influence of the single amino acid substitution on the peptide-DNA binding affinity.
The intrinsically disordered ATGP3 peptide in complex with DNA was studied using single amino acid substitutions and IMS-MS at physiological conditions. Seven amino acid substitutions were considered and changes in the peptide and peptide-DNA complexes conformational space as well as their relative abundance and dissociation kinetics were used to establish the amino acid positions that stabilize the ATGP3 : DNA complex formation in solution. The high resolution of the trapped IMS (TIMS) analyzer permitted the observation of multiple IMS bands for the intrinsically disordered ATGP3 peptide (at least 10 conformers). Using the corrected IMS profiles, the proline orientations (cis-trans) and charge location (basic residues and N-terminal) per IMS band were determined, providing a detailed description of the intramolecular interactions that describe the ATGP3 conformational space. These studies provided significant insight on the critical role of cis and trans proline configurations and charge locations in the populations of ATGP3 structures that are observed in physiological conditions.

The study of peptide-DNA complexes suggests that the -RGRP- core is essential for stabilizing ATGP3 : DNA complex at physiological conditions. Moreover, it was shown that weakest peptide : DNA complex is form with R6A, suggesting that this basic residue plays the most important role during binding to AT-rich regions of the DNA, in good agreement with previous NMR and X-ray data.22, 23
It should be noted that the IMS-MS workflow permits the study of multiple kinetic intermediates simultaneously, in contrast to the most abundant (or energetically favored) typically described using NMR and X-Ray measurements. This makes IMS-MS a powerful platform for the study of intrinsically disordered biomolecules and their kinetic intermediates. This work highlights the potential of IMS-MS combined with single amino acid substitutions for the evaluation of the peptide primary sequence influence on the secondary structure at physiological conditions. When combined with CID-MS, this workflow can be translated to the study of peptide complex conformations.

7.2 Experimental Methods

7.2.1 Sample preparation

Native AT-hook peptide 3 (Lys-Arg-Pro-Arg-Gly-Arg-Pro-Arg-Lys-Trp) and all amino acid substituted peptides were purchased from GenScript and used without further purification. An AT-rich DNA oligomer, denoted as FL876, sequence GGATATTGCCCGCAATATCC (C_{212}H_{270}N_{76}O_{136}P_{21}, MW 6655.1561) was purchased from Eurofins Genomics (Luxembourg City, Luxembourg) and used as received. This 22 nucleotide DNA hairpin contains a 9 base pair stem comprised of a 5 base pair AT rich region in the middle of the stem. Solvents and ammonium acetate salts utilized in this study were analytical grade or better and purchased from Fisher Scientific
(Pittsburgh, PA). A Tuning Mix calibration standard (G24221A) was obtained from Agilent Technologies (Santa Clara, CA) and used as received.

7.2.2 Peptide nomenclature

The amino acid sequence of the ATHP3 and variants are presented in Appendix 7.1, along with the nomenclature followed throughout the text. Specifically, the variant peptides are referred to by the original amino acid residue followed by their position and the replacement amino acid in a single-code nomenclature.

7.2.3 Ion Mobility Spectrometry-Mass Spectrometry

Details regarding the trapped IMS (TIMS) operation and specifics compared to traditional IMS can be found elsewhere.\textsuperscript{33, 36-39} Briefly, a custom nESI-TIMS unit was coupled to a Maxis Impact Q-TOF mass spectrometer (Bruker, Billerica, MA). The TIMS unit is run by custom software in LabView (National Instruments) synchronized with the MS platform controls. Sample aliquots (10 μL) were loaded in a pulled-tip capillary biased at 700-1200 V to the MS inlet. The nitrogen bath gas flow is defined by the pressure differential between the entrance funnel ($P_1 = 2.6$ mbar) and the exit funnel ($P_2 = 1.1$ mbar) at \textit{ca.} 294 K. A 880 kHz and 200 V\textsubscript{pp} RF potential was applied. Deflector, capillary, entrance funnel, entrance and exit analyzer voltages were 60, 50, 0, -200-0, and 60 V in positive mode (and -60, -50, 200-0, and -60 V in negative mode)
to prevent ion heating prior to IMS separation. The reduced mobility, \( K \), of an ion in a TIMS cell is described by:

\[
K = \frac{v_g}{E} = \frac{A}{(V_{\text{elution}} - V_{\text{out}})}
\]  

(1)

where \( v_g, E, V_{\text{elution}} \) and \( V_{\text{out}} \) are the gas velocity, applied electric field, elution voltage and exit analyzer voltage, respectively. After thermalization, species were eluted from the TIMS cell by decreasing the electric field in stepwise decrements (referred to as the “ramp”) and can be described by a characteristic elution voltage \( (V_{\text{elution}}) \). The mobility calibration constant \( A \) was determined using known reduced mobilities of Tuning Mix components \( (K_0\text{of }1.013, 0.835, \text{ and } 0.740 \text{ cm}^2/(\text{V.s}) \) for respective \( m/z 622, 922, \text{ and } 1222 \). The scan rate \( (S\text{r}=\Delta V_{\text{ramp}}/t_{\text{ramp}}) \) was optimized for every experiment.

The measured mobilities were converted into CCS \( (\Omega, \text{Å}^2) \) using the Mason-Schamp equation:

\[
\Omega = \frac{(18\pi)^{1/2}}{16} \frac{z}{(k_B T)^{1/2}} \left[ \frac{1}{m_i} + \frac{1}{m_b} \right]^{1/2} \frac{1}{K_0} \frac{1}{N^*}
\]  

(2)

where \( z \) is the charge of the ion, \( k_B \) is the Boltzmann constant, \( N^* \) is the number density of the bath gas and \( m_i \) and \( m_b \) refer to the masses of the ion and bath gas, respectively.

TIMS-MS spectra were analyzed using Compass Data Analysis 5.0 (Bruker Daltonik
GmbH) and TIMS Data Viewer 1.4.0.31397 (Bruker Daltonik GmbH). The IMS corrected profiles were compared using the correlation coefficient function:

\[
\text{Correl}(X,Y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2(y-\bar{y})^2}}
\]  \( (3) \)

where \( \bar{x} \) and \( \bar{y} \) are the sample means average for IMS profile 1 and 2.

7.2.4 Correction of amino acid substituted collision cross sections

For direct comparison of the mobility profiles between the ATHP 3 and single amino acid substituted peptides (e.g. Arg to Ala, Pro to Ala and Trp to Ala), the CCS profiles were adjusted based on the method previously described in references.\(^{26,40,41}\) To account for differences in CCS between N\(_2\) and He, the CCS \( \text{N}_2 = 1.0857 \times (\text{CCS}_{\text{He}}) + 81.459 \) [Å\(^2\)] conversion was used.\(^{42,43}\) This resulted in Arg to Ala, Pro to Ala and Trp to Ala substitutions to be corrected by 17.43 Å\(^2\), 2.68 Å\(^2\) and 17.43 Å\(^2\) in N\(_2\), respectively. Because the tryptophan residue CCS has not been previously reported, the Arg value was used instead considering the similarity between the geometries and steric hindrances.

Determination of binding affinities. Peptide-DNA binding affinities (\( K_a \)) were calculated, using the general equation for an association reaction, based on the peak area of the complex divided by the peak area of the unbound DNA and unbound peptide.\(^{44}\)
\[ K_a = \frac{\text{(Complex)}}{\text{(DNA)} \text{(Peptide)}} \] (4)

A distance matrix was utilized to better evaluate relative changes associated with the single amino acid substitution using the equation:

\[
D(X,Y) = \left[ \frac{K_a(\text{peptide}_X) - K_a(\text{peptide}_Y)}{K_a(\text{peptide}_{(1-8)})} \right] < 0
\] (5)

7.3 Acknowledgements

The authors acknowledge the financial support from the National Science Foundation Division of Chemistry, under CAREER award CHE-1654274, with co-funding from the Division of Molecular and Cellular Biosciences to FFL. AG will like to acknowledge the 2017-2018 Dissertation Year Fellowship from Florida International University.

7.4 References


CHAPTER 8: MICRO HETEROGENEITY OF TOPOISOMERASE IA/IB AND THEIR DNA-BOUND STATES
8.1. Introduction

DNA topoisomerases are ubiquitous enzymes, present in every organism, that manage and control DNA topology, both globally and locally.\textsuperscript{1} Topoisomerases are essential proteins for DNA strain release from supercoils during replication and transcription, thus warranting proper super helical density for essential cellular functions not disturbed.\textsuperscript{2} Recent studies have verified bacterial topoisomerase I as a viable and novel target utilized for the discovery of new antibiotics.\textsuperscript{3} A main roadblock is the lack of a three dimensional structure for the topoisomerase proteins and their DNA-bound forms. The success of traditional structural biology approaches (e.g., NMR and x-ray crystallography) have been limited to the description of subdomains or DNA-bound subdomains.\textsuperscript{4-6} As a result, there is a need for alternative approaches to describe the multistep mechanism consisting of DNA binding, cleavage and strand religation.

Native mass spectrometry (MS), and more recently in combination with ion mobility spectrometry (IMS), have provided unique insights in structural biology during the study of biomolecules and biomolecular complexes in the absence of the bulk solvent.\textsuperscript{7-8} In particular, IMS-MS has addressed specific shortcomings of current conventional methods by reducing the analysis time, sample consumption, sample purity requirements and restrictions on the biomolecule flexibility. Due to these unique advantages, IMS-MS has gained great attention for the study of protein dynamics in their native, partially folded and unfolded forms\textsuperscript{8,10,24,30-32}, and when complemented with molecular dynamic simulations, has allowed the identification of specific intramolecular interactions that stabilize conformational states.\textsuperscript{11}
IMS-MS can capture conformational interconversions and intermediate states for peptides, proteins, DNA and complexes. For example, in a recent study, the influence of solvent composition and kinetically trapped intermediates on the intrinsically disordered ATHP3 conformational space was reported. For DNA bounding proteins, IMS-MS studies have described the dimerization of the intrinsically disordered HMGA2 protein, as well as that phosphorylation of the c-terminal tail can play a vital role during HMGA2 protein-protein and protein-DNA interactions.

In this paper, we utilized a second generation trapped IMS coupled to MS for the study of topoisomerase IA and IB and their DNA bound states. The higher trapping efficiency and mass range of the new TIMS-MS instrument allowed the separation of conformational microstates for the topoisomerases and their DNA-protein complexes. In particular, first principle ion-neutral collision cross sections are reported for topoisomerase IA and IB, and their DNA bound states with a 13 base single strand DNA (ssDNA) and a 59 base stem loop DNA substrate (MBLong2). This experimental evidence shows the flexible nature of topoisomerases IA and IB.

8.2 Experimental Methods

8.2.1 Materials and reagents

A 13 base (Sub 1) and 59 base (MBLong2) DNA oligomer sequence 5’-AATGCGCTTTGGG-3’ (C_{128}H_{161}N_{49}O_{78}P_{12}, MW 4005.71) and 5’-GCCCTGAAAGATTATGGAATGCGATTAGGGT AAAGGAAGAGAGCATAATCTTTTCAGGGC-3’ (C_{582}H_{722}N_{240}O_{345}P_{58}, MW: 18396.14) was purchased from Biosearch Technologies (Petaluma, CA) and dialyzed in 100 mM ammonium acetate prior to analysis. Sub 1 is a single-stranded DNA, while MBLong2
forms a stem loop structure consisting of 16 base pair stem and 27 base loop. *E. coli* topoisomerase I (sequence: UniProt P06612) was purified and vaccinia topoisomerase IB (sequence: UniProt P68697) were purified as recombinant proteins. Water and ammonium acetate salts utilized in this study were analytical grade or better and purchased from Fisher Scientific (Pittsburgh, PA). A Tuning Mix calibration standard (G24221A) was obtained from Agilent Technologies (Santa Clara, CA) and used as received.

### 8.2.2 Trapped Ion Mobility Spectrometry – Mass Spectrometry Analysis (TIMS-MS)

Details regarding the TIMS operation and specifics compared to traditional IMS can be found elsewhere.\(^{15-19}\) Briefly, mobility separation in TIMS is based on holding the ions stationary against a moving gas using an electric field. The separation in a TIMS device can be described in the center of the mass reference frame using the same principles as in a conventional IMS drift tube.\(^{20}\) Since mobility separation is related to the number of ion-neutral collisions (or drift time in traditional drift tube cells), the mobility separation in a TIMS device depends on the bath gas drift velocity, ion confinement and ion elution parameters. The reduced mobility, \(K\), of an ion in a TIMS cell is described by:

\[
K = \frac{v_g}{E} = \frac{A}{(V_{elution} - V_{out})}
\]

where \(v_g\) and \(E\) are the velocity of the gas and the applied electric field across the TIMS analyzer region. \(V_{elution}\) is the voltage when the ions elute in the \(V_{ramp}\) sweep and \(V_{out}\) is the voltage applied at the end of the TIMS analyzer region.

A custom-built, pulled capillary nanoESI source was utilized for all the experiments. Quartz glass capillaries (O.D.: 1.0 mm and I.D.: 0.70 mm) were pulled utilizing a P-2000 micropipette laser puller (Sutter Instruments, Novato, CA) and loaded with 10 µL aliquot of the sample solution. Sample solutions consisted of 5-10 µM enzyme
in 100 mM ammonium acetate solution at native pH (pH = 6.7). For the observation of the complexes, a 1:1 ratio of 5 μM concentration of the DNA and enzyme was prepared in 100mM ammonium acetate immediately prior infusion. A typical nanoESI source voltage of +/- 600-1200 V was applied between the pulled capillary tips and the TIMS-MS instrument inlet. Ions were introduced via a stainless steel tube (1/16 x 0.020”, IDEXX Health Science, Oak Harbor, WA) held at room temperature into the TIMS cell.

Mobility calibration was performed using the Tuning Mix calibration standard (G24221A, Agilent Technologies, Santa Clara, CA) in positive ion mode (e.g., m/z = 322, K₀ = 1.376 cm² V⁻¹ s⁻¹ and m/z = 622, K₀ = 1.013 cm² V⁻¹ s⁻¹). The TIMS operation was controlled using in-house software, written in National Instruments Lab VIEW, and synchronized with the maXis Impact Q-ToF acquisition program.

Reduced mobility values (K₀) were correlated with CCS (Ω) using the equation:

\[ \Omega = \frac{\left(18\pi\right)^{1/2}}{16} \frac{z}{(k_BT)^{1/2}} \left[ \frac{1}{m_i} + \frac{1}{m_b} \right]^{1/2} \frac{1}{K_0 N^*} \]

where z is the charge of the ion, kB is the Boltzmann constant, N* is the number density and mI and mb refer to the masses of the ion and bath gas, respectively.

8.3 Results and Discussion

In order for topoisomerase enzymes to regulate DNA supercoiling they need to exist in a variety of conformations. The proposed action for type IA and IB topoisomerases are two very different modes of operation - an enzyme-bridged, strand-passage mechanism and a strand rotation model, respectively. However, structural information regarding the dynamics of topoisomerases as they process DNA supercoils is limited because of the inherent flexibility of the domains. To better understand the proposed
mechanism of DNA cleavage by topoisomerase we took advantage of the sensitivity and high mass (\(~100\ kDa\)) capabilities of the TIMS-MS technique to probe for the presence of multiple conformational isomers. We analyzed \textit{E. coli} topoisomerase IA (EcTopI) and eukaryotic \textit{vaccinia} virus topoisomerase IB (vTopIB). Both of these enzymes are of particular interest because they are structurally distinct from each other, therefore allowing us to observe two different mechanistic pathways of conformational change.

Mass spectrometry analysis of the 36 kDa \textit{vaccinia} topoisomerase IB enzyme shows the existence of two main charge state distributions (CSDs), one centered at high \textit{m/z} values (+12) and the other at lower ones (+17, +24 or +28) (Figure 8.1 and appendix 8.1). Performed at near-native pH conditions (pH=6.7), vTopIB can simultaneously assume both a low and high-charged CSD, relating to compact/native conformations or extended/denatured conformations, respectively (appendix 8.1). Mobility analysis of all observed charge states exhibited a distribution of conformations, spanning a range of 3000 to 9800 \text{Å}\(^2\), from native to partially folded, and finally to denatured states. The ‘native-like’ IMS distributions for topoIB are observed at charge states +10 to +13 (Figure 8.1). A phase transition at +13 towards conformational open states was observed, followed by more extended structures at +29 to +35. The high IMS resolving power of the TIMS analyzer allowed the observation of multiple stable IMS bands (often two broad conformers) and insight into the conformational transition pathways. Moreover, inspection of the +14 to +35 charge states exhibited a widening of IMS band distributions when compared to the +10 to +13 charge states, indicating more conformational flexibility of the extended structures (appendix 8.1).
Assessment of the x-ray structure of vTopIB, in conjunction with DNA,\textsuperscript{23} shows that a flexible linker connects the N-and C-terminal domains, allowing both ends of the

![Vaccinia Topoisomerase IB (~37 kDa)](image)

**Figure 8.1.** a) Typical mass spectra of all charge states observed under native starting solvent conditions, and (b) the mobility profiles for the native (+10 to +13) and partially folded forms (+13 to +21) of unbound vTopIB. The structure presented is a modified version pdb 3IGC,\textsuperscript{24} without the DNA substrate
enzyme to surround the DNA substrate and form contacts on opposite sides of the duplex.\textsuperscript{25} This information allows us to better understand the movements that govern the ‘native-like’ and more conformers of vTopIB. In particular, we deduce that movements of domains away from each other due to Coulombic repulsion may initially stem from this region, which in the absence of DNA, is not stabilized.

Evaluation of the 97 kDa \textit{E.coli} topoisomerase I enzyme by TIMS-MS provided information on the structural dynamics of the compact conformers. Specifically, experiments carried out at near-native conditions (pH = 6.7) probed only the ‘native-like’ states. Six charge states (e.g., +16 to +21) were populated with CCSs ranging from 5250 Å\textsuperscript{2} to 6500 Å\textsuperscript{2} (Figure 8.2a and b). A pattern of conformational change, identified by five transition states, was observed under these solvent conditions. Moreover, slight opening of the enzyme, followed by conformational change back to a more compact structure is observed, most notable for charge states +16 and +18, respectively. This peculiar trend of opening and closing, as a function of charge state, can give insight into the mechanism of conformational change and inter-molecular interactions that stabilize EcTopI. The flexible nature of EcTopI is evident by the variety of gas-phase stable conformers. The most populated ion conformer (+19) exists as a broad conformational distribution of \(~750\) Å\textsuperscript{2}. The +18 and +20 charge states are also broad, showing three and two resolved conformational isomers, respectively. Ultimately, the mobility data of the free EcTopI are in good agreement with previous studies\textsuperscript{26} which indicate that the overall arrangement of the enzyme domains can change in a dramatic way. For example, the hinge between domains II and IV may regulate the opening and closing of the N-terminal domains.\textsuperscript{4}
Figure 8.2. a) Typical mass spectra of all charge states observed under native starting solvent conditions, and (b) the mobility profiles for the native (+16 to +21) state of unbound EcTopI. The structure presented is a modified version pdb 4RUL, without the DNA substrate.

Comparison of the CSDs and mobility profiles of vTopIB and EcTopI illustrate some obvious differences, relating specifically to the disparity of each enzymes size (36 kDa vs. 97kDa, respectively). However, closer analysis reveals that charge state unfolding
of the vTopIB native conformers are more easily induced in comparison to EcTopI. We further predict that the greater conformational flexibility of free vTopIB is most likely attributed to the motion of the inter-domain hinge region,\textsuperscript{27} or the protease-sensitive bridge,\textsuperscript{28} and that greater intramolecular interactions are present in EcTopI that stabilize it in a compact form.

Gas-phase IMS-MS analysis of unbound vTopIB and EcTopI can be followed by study of the enzyme-DNA complex. The interaction of topoIB with duplex DNA occurs via a single-strand break to form a covalent bond between the catalytic tyrosine and the 3'-phosphoryl end of the broken strand. IMS studies reveal the formation of the complex at charge states +12 to +14 (Figure 8.3a and b). Interestingly, each mobility band is at least 500 Å\textsuperscript{2} wide. The presence of such broad conformational ensembles is indicative of several closely related conformational structures which have similar CCS values. We observe, however, narrowing of the mobility bands as the charge increases (e.g., +12 versus +14). Comparison of the +12, +13 and +14 charge states of both the unbound and DNA bound enzyme reveal that the complex achieves a more compact conformation. That is, the DNA stem loop induces folding of vTopIB. The formation of vTopIB into a folded C-shaped clamp, enveloping both the major and minor grooves of DNA, is strongly supported by TIMS-MS analyses.

Due to the inherent flexibility of EcTopI, structural analysis by means of traditional techniques often prove difficult. However, in 2015, a near complete x-ray structure of
Figure 8.3. a) Typical mass spectra, and (b) mobility profiles for the native states (+12 to +14) of vTopIB in complex with a 59 base stem loop DNA substrates (MBLong2), as a function of native starting solvent conditions.
EcTopI in complex with ssDNA was published, providing important insight into the multistep catalytic cycle responsible for removing negative DNA supercoils. Briefly, more details regarding the action of the N-terminal domain (ssDNA cleavage and rejoining), and the C-terminal domain (passing of the intact strand through the enzyme-bridged gate) were provided. Complementary gas-phase studies can offer another level of possible data interpretation. For example, our mass spectrometry studies revealed four charge states (+17 to +20) exhibiting a 1:2 binding stoichiometry of EcTopI enzyme to sub

![Figure 8.4. a) Typical mass spectra, and (b) mobility profiles for the native states (+17 to +20) of EcTopI in complex with two ssDNA substrates (2x sub 1), as a function of native starting solvent conditions.](image)
1 DNA, respectively (Figure 8.4a). Inspection of the IMS profiles shows that the two 13-base ssDNA substrates act to inhibit and prevent the opening of the enzyme (Figure 8.4b).

In fact, the only evidence for formation of a more open structure, as a function of charge state, is the two low abundant conformers at +20. It should be noted that 1:1 complexes were not observed. Next, we characterized the complex between the 50 base DNA stem loop (MBLong2) and EcTopI. Under native conditions, six charge states were formed (+17 to +22) at a 1:1

![Figure 8.5. a) Typical mass spectra, and (b) mobility profiles for the native states (+17 to +22) of EcTopIBin complex with a 59 base stem loop DNA substrates (MBLong2), as a function of native starting solvent conditions.](image-url)
stoichiometry (Figure 8.5a). Inspection of the corresponding conformers exhibited narrower distributions, when compared with EcTopI-2xSub1 (Figure 8.5b).

Comparison of the three topoisomerase-DNA complexes (e.g. vTopIB-MBLong2, EcTopI-2xSub1, EcTopI-MBLong2) shows that although the charge states, CCSs and overall conformations are significantly different, there are similarities. In particular, binding to DNA reduces the flexibility of the enzyme and, in the case of vTopIB-MBLong2, actually leads to slightly more compact structures.

8.4 References


APPENDICES
Appendix 2.1

Full protein sequences from which peptides were identified based on accurate m/z and mobility. Determined sequence coverage was based only on fractions analyzed.
Appendix 2.2

Peptides corresponding to targeted proteins, identified through discovery-based measurements. Observed mass, adduct and protein sequence coverage are reported.

<table>
<thead>
<tr>
<th>Protein Target</th>
<th>Peptide</th>
<th>Mass (Da)</th>
<th>Adduct</th>
<th>Protein Sequence coverage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calmodulin</td>
<td>MADQLTEEQIAEFK</td>
<td>1651.78</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Calmodulin</td>
<td>ADQLTEEQIAEFK</td>
<td>1520.74</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Calmodulin</td>
<td>ADQLTEEQIAEFKEAFSLFDK</td>
<td>2458.19</td>
<td>[M+3H]^3</td>
<td></td>
</tr>
<tr>
<td>Calmodulin</td>
<td>EAFSIFDK</td>
<td>955.47</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Calmodulin</td>
<td>EAFSIFDGDGDTITTK</td>
<td>1843.88</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Calmodulin</td>
<td>DGDGTITTK</td>
<td>906.43</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Calmodulin</td>
<td>ELGTVMR</td>
<td>804.42</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Calmodulin</td>
<td>MKDTDSEEEIR</td>
<td>1351.59</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Calmodulin</td>
<td>DTDSEEEIR</td>
<td>1092.46</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Calmodulin</td>
<td>DTDSEEEIREAFRVFDK</td>
<td>2084.97</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Calmodulin</td>
<td>EAFRVDK</td>
<td>1010.52</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Calmodulin</td>
<td>VFDKDNGYISAELR</td>
<td>1753.86</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Calmodulin</td>
<td>DNGYISAELR</td>
<td>1264.60</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Calmodulin</td>
<td>HVMTNLGEK</td>
<td>1027.51</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Calmodulin</td>
<td>LTDEEVDDEMIR</td>
<td>1348.62</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>DJ-1</td>
<td>GAEEMETVIPVDVMR</td>
<td>1674.80</td>
<td>[M+2H]^2</td>
<td>24.9</td>
</tr>
<tr>
<td>DJ-1</td>
<td>VTTHPLAK</td>
<td>865.50</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>DJ-1</td>
<td>DGLILTSR</td>
<td>873.49</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>DJ-1</td>
<td>EQENRK</td>
<td>802.39</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>DJ-1</td>
<td>GPGTSFELAIVEALNGK</td>
<td>1920.00</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Progesterone</td>
<td>GDQPAASGDSDDEPPPLPR</td>
<td>2034.88</td>
<td>[M+3H]^3</td>
<td>17.4</td>
</tr>
<tr>
<td>Progesterone</td>
<td>FYGPEGPVVFAGR</td>
<td>1515.71</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Progesterone</td>
<td>DFTPAELR</td>
<td>947.47</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Progesterone</td>
<td>ILMAINGK</td>
<td>858.50</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Progesterone</td>
<td>GLATFCLDK</td>
<td>966.48</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Progesterone</td>
<td>VFDVTK</td>
<td>707.39</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Peptide Sequence</td>
<td>Mass</td>
<td>Charge</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------------</td>
<td>----------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>Progesterone</td>
<td>YHHVHGK</td>
<td>739.38</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>ADVLSVLRQYNIQK</td>
<td>1645.92</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>ADVLSVLRQYNIQKK</td>
<td>1774.01</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>GDEVIFGEFSWPK</td>
<td>1509.71</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>TNYVVWGTGK</td>
<td>1123.57</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>RAATENIPVVVR</td>
<td>1224.69</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>IEDEECVRLDKER</td>
<td>1632.78</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>DLLGYLNGEASTSASIDR</td>
<td>1880.91</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>LMPQDWDR</td>
<td>1059.48</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>RAADEVIAEAK</td>
<td>1171.62</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>FWETLDR</td>
<td>965.46</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>AADEVLAEAK</td>
<td>1015.52</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>EGIIVQTEQIR</td>
<td>1171.62</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>TDIDDDITAIK</td>
<td>1218.60</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>SFVDAEVDVTR</td>
<td>1236.60</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>TKQPIPAAYNR</td>
<td>1257.68</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>QPIPAAYNRDYQER</td>
<td>1719.83</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>GKEETEGFK</td>
<td>1023.49</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>SVTEGASAR</td>
<td>876.43</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>DLLQDLK</td>
<td>843.47</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>FVPSDEK</td>
<td>820.40</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>QGCQRENETIQR</td>
<td>1573.76</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>WDVTVLELSYHK</td>
<td>1488.76</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>RHLDRPVFLR</td>
<td>1307.76</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>HLDRPVFLR</td>
<td>1151.66</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>FWETIDRYMVK</td>
<td>1486.73</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Parafibromin</td>
<td>FWETLDR</td>
<td>965.46</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>LSASTASELSPK</td>
<td>1189.62</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>ESPSAAGPEDKSQSK</td>
<td>1516.70</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>DKER</td>
<td>546.28</td>
<td>[M+H]^+</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>AQAEEQEEQERLQK</td>
<td>1585.77</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>AQAEEQEEQER</td>
<td>1216.53</td>
<td>[M+2H]^2</td>
<td></td>
</tr>
<tr>
<td>Protein Name</td>
<td>Mixture Peptide</td>
<td>Molecular Weight</td>
<td>Charge State</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------</td>
<td>------------------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>SAVTLPRNGR</td>
<td>1069.60</td>
<td>[M+2H]$^2$</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>SEVSETKQK</td>
<td>1034.52</td>
<td>[M+2H]$^2$</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>REER</td>
<td>588.30</td>
<td>[M+H]$^+$</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>SSATLWNSPSR</td>
<td>1204.58</td>
<td>[M+2H]$^2$</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>GPQPDR</td>
<td>668.32</td>
<td>[M+H]$^+$</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>ARLSASTASELSPK</td>
<td>1416.76</td>
<td>[M+2H]$^2$</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>ESPSAAGPEDIK</td>
<td>1086.48</td>
<td>[M+2H]$^2$</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>GPTWGR</td>
<td>672.33</td>
<td>[M+H]$^+$</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>SEVSETK</td>
<td>778.37</td>
<td>[M+H]$^+$</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>ERENEK</td>
<td>803.38</td>
<td>[M+H]$^+$</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>YEAAIQR</td>
<td>849.43</td>
<td>[M+H]$^+$</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>SREEAER</td>
<td>875.41</td>
<td>[M+H]$^+$</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>EEAEARSR</td>
<td>946.45</td>
<td>[M+H]$^+$</td>
<td></td>
</tr>
<tr>
<td>MAP-7</td>
<td>SVHRCAPAGER (CAM)</td>
<td>1239.60</td>
<td>[M+2H]$^2$</td>
<td></td>
</tr>
</tbody>
</table>
Appendix 3

Details of nESI/TIMS hardware and mobility calibration

Nano-ESI source
Custom nESI emitters were made from quartz glass capillaries (I.D. 0.7 mm, O.D. 1.0 mm) utilizing a P-2000 micropipette laser puller (Sutter Instruments, Novato, CA). The inlet capillary was a stainless steel tube (1/16 × 0.020", IDEX Health Science, Oak Harbor, WA).

TIMS cell
The TIMS cell is an electrodynamic ion funnel with three regions (see scheme below): entrance, analyzer tunnel (46 mm length), and exit. The rf voltage creates a quadrupolar field in the tunnel and dipolar field in other regions.

Mobility calibration in TIMS as described in reference 50
Multiple ion species are first trapped and thermalized at different $E$ values set up by voltage profile along the tunnel, then pushed out by “ramping” (decreasing) $E$ in steps.
with each species eluting at a characteristic voltage \((V_{\text{elution}} - V_{\text{out}})\). The total time an ion spends in the system (from gating into the TIMS cell to registration at the ToF detector) is:

\[
t_{\text{total}} = t_{\text{trap}} + \left(\frac{V_{\text{elution}}}{V_{\text{ramp}}}\right) \times t_{\text{ramp}} + t_{\text{ToF}} = t_0 + \left(\frac{V_{\text{elution}}}{V_{\text{ramp}}}\right) \times t_{\text{ramp}},
\]

where \(t_{\text{trap}}\) is the trapping time, \(t_{\text{ToF}}\) is the flight time from cell exit to detector (through MS vacuum), \(V_{\text{ramp}}\) and \(t_{\text{ramp}}\) are the voltage range and time of ramp, and \(t_0\) is the time before and after the separation step. The \(V_{\text{elution}}\) and \(t_0\) values are determined simultaneously by varying \(t_{\text{ramp}}\) from 100 to 500 ms at fixed \(V_{\text{ramp}}\) as a function of the \(m/z\) of interest.
Appendix 3.1

Collision cross sections (CCS, Å²) and resolving power (R) values for protonated histone tails measured using TIMS.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Ion</th>
<th>CCS (Å²), std. error of mean: ± 0.04%</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>K9me3</td>
<td>[M+6H]⁺</td>
<td>1303</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>[M+7H]⁻</td>
<td>1398/1406</td>
<td>203/180</td>
</tr>
<tr>
<td></td>
<td>[M+8H]⁺</td>
<td>1574/1607</td>
<td>96/102</td>
</tr>
<tr>
<td></td>
<td>[M+10H]⁺</td>
<td>1835</td>
<td>170</td>
</tr>
<tr>
<td></td>
<td>[M+11H]⁺</td>
<td>1921</td>
<td>201</td>
</tr>
<tr>
<td></td>
<td>[M+6H]⁺</td>
<td>1308</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>[M+7H]⁻</td>
<td>1405/1437</td>
<td>82/72</td>
</tr>
<tr>
<td></td>
<td>[M+8H]⁺</td>
<td>1508/1535/1574/1609</td>
<td>188/284/146/223</td>
</tr>
<tr>
<td></td>
<td>[M+9H]⁺</td>
<td>1684/1707/1751</td>
<td>167/76/233</td>
</tr>
<tr>
<td></td>
<td>[M+10H]⁺</td>
<td>1830/1839</td>
<td>218/226</td>
</tr>
<tr>
<td></td>
<td>[M+11H]⁺</td>
<td>1926/1943</td>
<td>244/183</td>
</tr>
<tr>
<td></td>
<td>[M+6H]⁺</td>
<td>1291/1357</td>
<td>128/137</td>
</tr>
<tr>
<td></td>
<td>[M+7H]⁻</td>
<td>1394/1422/1441</td>
<td>59/133/119</td>
</tr>
<tr>
<td></td>
<td>[M+8H]⁺</td>
<td>1483/1510</td>
<td>215/265</td>
</tr>
<tr>
<td>K23me3</td>
<td>[M+9H]⁺</td>
<td>1658/1702</td>
<td>240/122</td>
</tr>
<tr>
<td></td>
<td>[M+10H]⁺</td>
<td>1833</td>
<td>176</td>
</tr>
<tr>
<td></td>
<td>[M+11H]⁺</td>
<td>1948</td>
<td>229</td>
</tr>
<tr>
<td></td>
<td>[M+6H]⁺</td>
<td>1209</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>[M+7H]⁻</td>
<td>1290</td>
<td>145</td>
</tr>
<tr>
<td>K36me3</td>
<td>[M+8H]⁺</td>
<td>1395/1404/1420/1435/1447/1473</td>
<td>199/187/202/144/90/184</td>
</tr>
<tr>
<td></td>
<td>[M+9H]⁺</td>
<td>1529/1563/1572/1603</td>
<td>191/223/197/255</td>
</tr>
<tr>
<td></td>
<td>[M+10H]⁺</td>
<td>1673/1694/1715</td>
<td>222/121/159</td>
</tr>
<tr>
<td></td>
<td>[M+11H]⁺</td>
<td>1784</td>
<td>133</td>
</tr>
<tr>
<td></td>
<td>[M+6H]⁺</td>
<td>1205</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>[M+7H]⁻</td>
<td>1298/1322</td>
<td>137/85</td>
</tr>
<tr>
<td></td>
<td>[M+10H]⁺</td>
<td>1701</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>[M+11H]⁺</td>
<td>1785/1814</td>
<td>223/151</td>
</tr>
<tr>
<td></td>
<td>[M+6H]⁺</td>
<td>1207</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>[M+7H]⁻</td>
<td>1287/1310</td>
<td>239/218</td>
</tr>
<tr>
<td>K14ac</td>
<td>[M+8H]⁺</td>
<td>1396/1405/1418/1428/1435/1454/1480</td>
<td>208/156/153/111/197/117/1</td>
</tr>
<tr>
<td></td>
<td>[M+9H]⁺</td>
<td>1531/1563/1579/1611</td>
<td>247/223/198/201</td>
</tr>
<tr>
<td></td>
<td>[M+10H]⁺</td>
<td>1694</td>
<td>179</td>
</tr>
<tr>
<td></td>
<td>[M+11H]⁺</td>
<td>1786</td>
<td>192</td>
</tr>
<tr>
<td>Compound</td>
<td>M+H</td>
<td>M+2H</td>
<td>M+3H</td>
</tr>
<tr>
<td>----------</td>
<td>-----</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Species</td>
<td>Peptide</td>
<td>Mass 1</td>
<td>Mass 2</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>K9me</td>
<td>[M+6H]^{6+}</td>
<td>1212</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>[M+7H]^{7+}</td>
<td>1301/1324</td>
<td>190/221</td>
</tr>
<tr>
<td></td>
<td>[M+8H]^{8+}</td>
<td>1426/1438/1473/1499/1511</td>
<td>247/173/142/180/140</td>
</tr>
<tr>
<td></td>
<td>[M+9H]^{9+}</td>
<td>1574/1591/1628/1640</td>
<td>237/175/119/149</td>
</tr>
<tr>
<td></td>
<td>[M+10H]^{10+}</td>
<td>1704</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>[M+11H]^{11+}</td>
<td>1787</td>
<td>167</td>
</tr>
<tr>
<td></td>
<td>[M+6H]^{6+}</td>
<td>1218</td>
<td>151</td>
</tr>
<tr>
<td></td>
<td>[M+7H]^{7+}</td>
<td>1302/1309/1326/1336</td>
<td>193/73/165/69</td>
</tr>
<tr>
<td></td>
<td>[M+8H]^{8+}</td>
<td>1416/1430/1439/1453/1468/1494/1507</td>
<td>153/164/212/153/120/199/2</td>
</tr>
<tr>
<td></td>
<td>[M+9H]^{9+}</td>
<td>1561/1576/1625/1635</td>
<td>244/146/140/178</td>
</tr>
<tr>
<td></td>
<td>[M+10H]^{10+}</td>
<td>1700</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>[M+11H]^{11+}</td>
<td>1800</td>
<td>143</td>
</tr>
</tbody>
</table>
Appendix 3.2

Relative (approximate) cross sections for K9ac (dominant peaks). Lines guide the eye through trends below and above the transition region.
Appendix 3.3

TWIMS spectra for selected mixtures of me3, p, and me variants measured in N₂ using solvent (i) (solid black lines) with fits by scaled individual traces from Fig. 1 (colored lines) and their computed additions (dotted lines).

(a) K4me3/K9me3
(b) K4me3/K23me3
(c) K4me3/K36me3
Appendix 3.4

TWIMS spectra for selected variant mixtures measured in N₂ with solvent (i) using the traveling wave speeds of 650, 1000, and 1900 m/s.
Appendix 3.4 (continued)

K4me3/K23me3 8+

K4me3/K23me3 9+

K4me3/K36me3 8+
Appendix 3.4 (continued)

K27me3/K36me3 8+

T6p/S10p

K4me/K23me
Appendix 3.5

Pairwise linear correlations between transit times for ac and me3 variants at the traveling wave speeds of 650 and 1000 m/s. Variants with only three data points and p variants with partial data at $s = 1000$ m/s are not included.
Appendix 3.6

TWIMS spectra measured with different ESI solvents: (a) four me3 variants (z = 9) with solvents (i, ii, iii), (b) K23me3 (z = 8 and 9) with solvents (i, iv).
Appendix 3.7

TWIMS spectra for four me3 variants in $z = 12$ measured in N$_2$ using solvent (ii) and $s = 650$ m/s. No signal at 12+ was found for the K4me3 variant.
Appendix 3.8

TIMS spectra (on the cross section scale) measured for exemplary me3, ac, p, and me variant species using (a) $t_{\text{ramp}} = 100 - 500$ ms from MeOH/H$_2$O solutions, (b) from MeOH/H$_2$O and aqueous solutions.
Appendix 3.9

TIMS spectra for selected mixtures of p variants and traces for the two components.
Appendix 3.10

Linear correlations between transit times for me3, ac, and p variants across charge state pairs (major peaks from Fig. 1). Variants for me (with three data points) are not included.
Appendix 3.11

Linear correlations between separation parameters in TIMS for the acetylated and phosphorylated peptide variants in different charge states.

Acetylated

Phosphorylated
Appendix 3.12

Linear correlations between separation parameters in Synapt and TIMS for the acetylated and phosphorylated peptide variants.
Appendix 3.13

Linear correlations between FAIMS and TIMS (left column) or TWIMS (right column) separations for (a) acetylated and (b) phosphorylated variants, \( r^2 \) marked. To facilitate comparison, the left column in (a) repeats the data from Fig. 6 in the text.
Appendix 3.14

Linear correlations between FAIMS and TIMS (left) or TWIMS (right) separations for same acetylated (a) and phosphorylated (b) variants with different charge states selected in the two dimensions ($r^2$ marked).
Appendix 5.1

TIMS profiles as a function of the desolvation time for [M-5H]$^5$ i-motif DNA.
Appendix 5.2

TIMS profiles as a function of the desolvation time for [M-6H]^6 i-motif DNA.
Appendix 5.3

TIMS profiles as a function of the desolvation time for \([\text{M-7H}]^{\text{7}}\) i-motif DNA.
Appendix 5.4

TIMS profiles as a function of the desolvation time for \([\text{M-8H}]^8\) i-motif DNA.
Appendix 5.5

TIMS profiles as a function of the desolvation time for [M-9H]$^9$ i-motif DNA.
Appendix 6.1

Typical mass spectrum of 1:1 ATPH:DNA complex (left panel) and 2:1 ATHP:DNA complex (right panel).

- **DNA**
- **Complex ATHP : DNA (1 : 1)**
- **Complex ATHP : DNA (2 : 1)**

![Mass spectra diagram]
Appendix 6.2

Collision induced dissociation curves for the +5 charge state of FL876 complexed with ATHP 1 and 2 (charge state and degrees of freedom were considered).
Appendix 6.3

Fluorescent measurements for DNA (FL876) and Hoescht 33258 dye titrated with ATHP 1, ATHP 2 or ATHP 3.
Appendix 7.1

ATHP3 single amino acid substitutions considered and their nomenclature.

<table>
<thead>
<tr>
<th>Peptides</th>
<th>Peptide sequence and position of amino acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native</td>
<td>Lys (K) Arg (R) Pro (P) Arg (R) Gly (G) Arg (R) Pro (P) Arg (R) Lys (K)</td>
</tr>
<tr>
<td>R2A</td>
<td>Lys Ala (A) Pro Arg Gly Arg Pro Arg Lys</td>
</tr>
<tr>
<td>P3A</td>
<td>Lys Arg Ala Arg Gly Arg Pro Arg Lys</td>
</tr>
<tr>
<td>R4A</td>
<td>Lys Arg Pro Ala Gly Arg Pro Arg Lys</td>
</tr>
<tr>
<td>R6A</td>
<td>Lys Arg Pro Arg Gly Ala Pro Arg Lys</td>
</tr>
<tr>
<td>P7A</td>
<td>Lys Arg Pro Arg Gly Arg Ala Arg Lys</td>
</tr>
<tr>
<td>R8A</td>
<td>Lys Arg Pro Arg Gly Arg Pro Ala Lys</td>
</tr>
<tr>
<td>W10A</td>
<td>Lys Arg Pro Arg Gly Arg Pro Arg Lys</td>
</tr>
</tbody>
</table>
Appendix 7.2

Schematic diagram of the TIMS cell used to trap and separate ions based on differences in each species ion neutral collision cross section.
Appendix 7.3

Typical, normalized ATHP3 and variant mobility profiles for the [M+H]⁺, [M+2H]²⁺ and [M+3H]³⁺(a). Peptide IMS fingerprint (b) and correlation matrix (c) are used for assessment of the primary sequence effect on the secondary structure.
Appendix 7.4

Typical ion mobility profiles (a) of the native and substituted ATHP3 : DNA complexes. The IMS profile fingerprint (b) of the complexes was used to generate a correlation matrix (c).
Appendix 8.1

(a) Typical mass spectra of all charge states observed under native starting solvent conditions, and (b) the mobility profiles for the native (+10 to +13), partially folded (+13 to +21) and unfolded forms (+21 to +35) of unbound vTop1B. The structure presented is a modified version pdb 3IGC, without the DNA substrate.
VITA

ALYSSA GARABEDIAN

Born, Nashua, New Hampshire

2010-2013
B.A., Chemistry
Florida International University
Fort Myers, FL

2013-2016
M.S., Chemistry
Florida International University
Miami, Florida

2015-2016
Biomedical Research Initiative Summer Research Award
Florida International University
Miami, Florida

2017
Doctoral Candidate, Chemistry
Florida International University
Miami, Florida

2017-2018
Dissertation Year Fellowship (DYF)
Florida International University
Miami, Florida

PUBLICATIONS AND PRESENTATIONS


