Characterization of Highly Heterogeneous Heparin-Protein Complexes Using Novel Mass Spectrometry-Based Approaches

Yunlong Zhao

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2

Part of the Analytical Chemistry Commons

Recommended Citation

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
CHARACTERIZATION OF HIGHLY HETEROGENEOUS HEPARIN-PROTEIN COMPLEXES USING NOVEL MASS SPECTROMETRY-BASED APPROACHES

A Dissertation Presented

by

YUNLONG ZHAO

Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2017

Chemistry
CHARACTERIZATION OF HIGHLY HETEROGENEOUS HEPARIN-PROTEIN COMPLEXES USING NOVEL MASS SPECTROMETRY-BASED APPROACHES

A Dissertation Presented
by
YUNLONG ZHAO

Approved as to style and content by:

________________________________________
Igor A. Kaltashov, Chair

________________________________________
Jeanne A. Hardy, Member

________________________________________
Paul L. Dubin, Member

________________________________________
Stephen J. Eyles, Member

________________________________________
Richard W. Vachet, Department Head
Department of Chemistry
ABSTRACT

CHARACTERIZATION OF HIGHLY HETEROGENEOUS HEPARIN-PROTEIN COMPLEXES USING NOVEL MASS SPECTROMETRY-BASED APPROACHES

SEPTEMBER 2017

YUNLONG ZHAO, B.S., NANKAI UNIVERSITY
Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Igor A. Kaltashov

Heparin-like glycosaminoglycan (GAG) is a family of polysaccharide involved in variety of physiological processes. They have potentials to interact with a broad range of proteins and many of them hold crucial values in regulation of protein functions. My dissertation addresses the significance and challenges in the field of heparin-mediated studies, with a focus on the questions in biological and analytical aspects, which are largely hindered by the structural heterogeneity and function diversity of heparin. My dissertation reports the efforts I made in the past few years with respect to the development of novel analytical strategies based on a combination of mass spectrometry, ion-mobility, gas-phase chemistry and chromatography, with success in characterizing protein-GAG interacting stoichiometry and deciphering the structural code related to protein-GAG affinity.
TABLE OF CONTENTS

Page

ABSTRACT ........................................................................................................................................... iv

LIST OF TABLES................................................................................................................................... viii

LIST OF FIGURES............................................................................................................................... ix

CHAPTER

1 HEPARIN-INOVOLED RESEARCH OBJECTIVES AND BACKGROUND...................... 1

1.1 Heparin and Heparin-Related Scientific Questions ...................................................... 1
  1.1.1 Structure and Functions of Heparin-Like Glycosaminoglycan ........................................ 1
  1.1.2 Heparin-Like GAG as Therapeutic and Medication ................................................ 2
  1.1.3 Challenges of Heparin-Like GAG in Analytical Chemistry ....................................... 5

1.2 A Map of Prevailing Arsenals for Studying Heparin-Protein Interaction ......................... 6
  1.2.1 Traditional Structural Biology Strategies ................................................................. 6
  1.2.2 Physical Chemistry Strategies .................................................................................. 8
  1.2.3 Mass Spectrometry-Based Strategies ........................................................................ 9

1.3 Future Directions: State-of-Art Mass Spectrometry-Based Techniques ............................ 12
  1.3.1 Ion-Mobility Spectrometry Mass Spectrometry ........................................................ 12
  1.3.2 Novel Online LC-Native MS for Protein complexes ................................................. 13
  1.3.3 Gas-phase chemistry for manipulation of gaseous ions .......................................... 14

1.4 References ............................................................................................................................... 17

2 INTERACTIONS OF INTACT UNFRACTIONATED HEPARIN WITH ITS CLIENT PROTEINS CAN BE PROBED DIRECTLY USING NATIVE ELECTROSPRAY IONIZATION MASS SPECTROMETRY ........................................ 21

2.1 Introduction ............................................................................................................................... 21

2.2 Experimental ............................................................................................................................ 24
  2.2.1 Materials ...................................................................................................................... 24
  2.2.2 Methods ...................................................................................................................... 25

2.3 Results and Discussion ........................................................................................................... 25
  2.3.1 Characterization of protein/heparin complexes using “traditional” native ESI MS and/or ion mobility spectroscopy ...................................................................................... 25
2.3.2 Limited charge reduction of heparin polycations in the gas phase is accompanied only by minimal ion dissociation ................................................................. 26
2.3.3 Establishing the identity of AT/heparin complexes using limited charge reduction ................................................................. 29
2.3.4 Combination of limited charge reduction and ion mobility increases measurement throughput ................................................................. 31
2.3.5 Interpretation of AT/heparin mass spectrum based on the limited charge reduction data ................................................................. 33
2.4 Conclusions ........................................................................................................................................................................ 35
2.5 Acknowledgements ............................................................................................................................................................ 37
2.6 References ............................................................................................................................................................................. 44

3 ION-MOBILITY MASS SPECTROMETRY COUPLING WITH LIMITED CHARGE REDUCTION REVEALS DISTINCT MODELS OF INTERACTION BETWEEN HEPARIN AND PROTAMINE SULFATE ................................................................. 46

3.1 Introduction ............................................................................................................................................................................. 46
3.2 Experimental ........................................................................................................................................................................... 49
  3.2.1 Materials ............................................................................................................................................................................. 49
  3.2.2 Online Ion-exchange Chromatography-Native Mass Spectrometry ................................................................................ 49
  3.2.3 Ion-Mobility-Mass Spectrometry Coupling with Limited Charge Reduction ........................................................................ 50
  3.2.4 CCS Calibration ............................................................................................................................................................. 51
  3.2.5 Data Analysis ............................................................................................................................................................. 51
3.3 Results and Discussion ............................................................................................................................................................ 51
  3.3.1 Electrostatic Interaction Between Heparin and Protamine Can be Probed using WCX-MS .......................................................... 51
  3.3.2 Characterization of Complexes Formed by Protamine and dp20 using Native Mass spectrometry with Limited Charge Reduction ........................................................................................................ 52
  3.3.3 Ion-Mobility Separation and CCS measurement of Charge-Reduced Product Ions ........................................................................ 53
  3.3.4 Data from Specific m/z Regions Revealed Distinct Species with the Same Mass ........................................................................ 54
  3.3.5 Comprehensive Analysis of Stoichiometry and Conformation based on CCS-Mass Relationship ........................................... 56
3.4 Conclusions ............................................................................................................................................................................. 59
3.5 References ............................................................................................................................................................................. 68

4 MAPPING MINIMAL PROTEIN-BINDING MOTIFS USING AN ION-MOBILITY TOP-DOWN MASS SPECTROMETRY APPROACH ........................................................................................................ 70
4.1 Introduction .................................................................................................................. 70
4.2 Experimental ............................................................................................................... 72
  4.2.1 Materials and Sample Preparation ................................................................. 72
  4.2.2 Methods .............................................................................................................. 73
4.3 Results and Discussion ............................................................................................. 74
  4.3.1 General Behaviors of Unbound FGF in response to the Gas-phase Collisional
      Activation ........................................................................................................... 74
  4.3.2 Collision-induced Unfolding and Dissociation of the Complex Formed by FGF
      and Homogeneous Pentasaccharide ....................................................................... 77
  4.3.3 Probing interaction between FGF and heterogeneous decasaccharide ............ 80
4.4 Conclusions ................................................................................................................ 82
4.5 References ................................................................................................................. 93

5 DEVELOPING A NOVEL PLATFORM BASED ON ENZYMATIC DEGRADATION AND
      SIZE-EXCLUSION CHROMATOGRAPHY-MASS SPECTROMETRY FOR IDENTIFYING
      PROTEIN-BINDING MOTIFS IN HEPARIN-LIKE GLYCOSAMINOGLYCANS
      ................................................................................................................................. 95
  5.1 Introduction .............................................................................................................. 95
  5.2 Experimental .......................................................................................................... 97
  5.3 Results and Discussion ............................................................................................ 98
    5.3.1 SEC-native MS analysis of the heparinase-digested products of AT•dp20 complex 98
    5.3.2 Identification of the GAG ligand through online protein denaturation .......... 100
  5.4 Conclusions ............................................................................................................ 102
  5.5 References .............................................................................................................. 108

6 SUMMARY AND FUTURE DIRECTIONS ..................................................................... 109
  6.1 Summary ................................................................................................................. 109
  6.2 Future directions ..................................................................................................... 110

APPENDIX: CCS CALIBRATION CURVES ..................................................................... 113

BIBLIOGRAPHY ........................................................................................................... 115
LIST OF TABLES

Table 5-1. The measured mass and calculated average molecular weight at major peaks in SEC-MS of AT•dp20 degraded complex ..................107
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1-1. The wide size range and enormous diversity of sulfation patterns epitomize the extreme structural heterogeneity of heparin and related GAGs</td>
<td>15</td>
</tr>
<tr>
<td>Figure 1-2. Schematic representation of a nanogel construct utilizing heparin-based linkers to load growth factors and morphogens for delivery to tissues</td>
<td>16</td>
</tr>
<tr>
<td>Figure 2-1. ESI mass spectra of AT solutions in ammonium acetate acquired in the absence and the presence of intact heparin</td>
<td>38</td>
</tr>
<tr>
<td>Figure 2-2. Ions generated by limited charge reduction of a population of free heparin ions whose m/z values are confined to the m/z region 3190-3220</td>
<td>39</td>
</tr>
<tr>
<td>Figure 2-3. Native ESI mass spectrum of the AT/heparin mixture and three representative mass spectra of ions produced via limited charge reduction of precursor ions</td>
<td>40</td>
</tr>
<tr>
<td>Figure 2-4. Limited charge reduction mass spectra obtained using wide and narrow precursor ion selection window</td>
<td>41</td>
</tr>
<tr>
<td>Figure 2-5. Mass spectra and ion mobility mass spectra of ions produced by limited charge reduction of precursor ions at m/z 4500 isolated using a narrow and a wide selection window</td>
<td>42</td>
</tr>
<tr>
<td>Figure 2-6. Extraction of contributions of ATₙ·heparin complexes to the total ionic signal in the native ESI mass spectrum of AT/heparin mixture based on the results of limited charge reduction measurements</td>
<td>43</td>
</tr>
<tr>
<td>Figure 3-1. General structure of heparin and protamine</td>
<td>62</td>
</tr>
<tr>
<td>Figure 3-2. Online WCX-MS of dp20 alone, dp20 in the presence of protamine and protamine alone</td>
<td>63</td>
</tr>
<tr>
<td>Figure 3-3. Mass spectrum without incorporating ion isolation and mass spectrum for limited charge reduction with precursor ion selection and ion mobility mass spectrum for the limited charge reduction with precursor ions selection</td>
<td>64</td>
</tr>
<tr>
<td>Figure 3-4. Analysis of charge reduced species for the precursor selected at m/z=1600</td>
<td>65</td>
</tr>
</tbody>
</table>
Figure 3-5. Ion mobility spectra for the ion with +3 charge reduced from +4 charge precursor isolation. And comparison of CCSs of the same ion reduced from different charge states ........................................ 66

Figure 3-6. CCS vs. mass plot for all detectable ions ........................................ 67

Figure 4-1. Schematic of a gas-phase “top-down” approach to identify the protein-binding motif in unfractionated heparin ........................................ 84

Figure 4-2. Native MS of FGF alone and in the presence of synthetic dp5 .......... 85

Figure 4-3. CIU fingerprint graphs of FGF*8, FGF*7 and FGF•dp5*7 .................... 86

Figure 4-4. Fragmentation of FGF•dp5+7 induced in different unfolding states below 490 eV (70V) and above 630 eV (90V) ........................................ 87

Figure 4-5. Plots of m/z value of survived FGF•GAG complex or FGF ions versus applied activation energy ........................................ 88

Figure 4-6. Characterization of FGF/dp10 complexes using IMS-MS ................. 89

Figure 4-S1. Mass spectra of synthetic dp5 and dp10 acquired in the negative mode ........................................................................................................ 90

Figure 4-S2. IMS-MS of FGF alone and in presence of dp10 under different collisional voltage ................................................................. 91

Figure 4-S3. CIU fingerprint for FGF•dp10 at +8 and +7 charge states .......... 91

Figure 4-S4. Monitoring the mass loss of FGF•dp10+7 at different collisional voltage ........................................................................................................ 92

Figure 5-1 Workflow of approach based on enzymatic degradation to identify the minimal protein-binding motif for heparin-like GAG ........................................................................................................ 103

Figure 5-2. Schematics of SEC-native MS and SEC-denatured MS ............... 104

Figure 5-3. SEC-MS of heparinase-treated AT-dp20 complex ......................... 105

Figure 5-4. Online SEC-denatured MS tested with the AT-Arixtra complex ..... 106

Figure 5-S1. Online SEC-denatured MS of BSA ........................................ 107

Figure APPX-1. CCS calibration curves for protamine-dp20 experiment ...... 113
Figure APPX-2. CCS calibration curves for FGF•dp10 complex..............................114
CHAPTER 1
HEPARIN-INVOLEO RESEARCH OBJECTIVES AND BACKGROUND

1.1 Heparin and Heparin-Related Scientific Questions

1.1.1 Structure and Functions of Heparin-Like Glycosaminoglycan

Glycosaminoglycan is a big family of linear polysaccharide featured by repetitive units containing amino sugar and uronic acid sugar. Heparin and heparan sulfate both belongs to the same “genus” in glycosaminoglycan (GAG) family, which are ubiquitously found in human body. I borrow the term from taxonomy because both heparin and heparan sulfate have high similarity in terms of the structure and physical properties. The repetitive disaccharide unit of heparin or heparan sulfate consists of a glucosamine and a glucuronic acid or iduronic acid (which differ from each other with respect to the chirality of C5) (Figure 1-1). Sequence polydispersity arises from the biosynthesis of heparin-like GAGs, in which enzymatically controlled modifications of precursor linear polysaccharides lead to tremendous variability in the levels and patterns of O- and N-sulfation and deacetylation. Interestingly, although we believe these two species of GAG can do exactly the same thing in the human body because they can interact with broad range of proteins through the sulfate groups and the basic groups in proteins, they actually have very different physiology-relevant emphases. Heparan sulfate is synthesized in many cells and has diverse functions in physiology ranging from tissue development to immune regulation\textsuperscript{1,2} and is associated with many diseases such as cancer\textsuperscript{3} and neurodegenerative diseases\textsuperscript{4,5}. For example, matured fibroblast cells present
heparan sulfate in the form of proteoglycans on the extracellular matrix and it can regulate the function of fibroblast growth factor family, facilitating the affinity to fibroblast growth factor receptors\(^6\). In contrast, heparin is only synthesized in the mast cells, so it dominantly serves in blood as a coagulant regulator\(^8\). Therefore, heparan sulfate has larger variety of the chain lengths (for anchoring with matrix proteins); and heparin contains higher degrees of sulfation (for effective action on coagulant factors). It needs to be clarified here that we often use the term “heparin” to represent all heparin-like GAGs or use the heparin molecules of higher modification variety (which is cheaper as well) as a model molecule for developing analytical methods that can be applied for other heparin-like GAGs or other heterogeneous polyanions.

1.1.2 Heparin-Like GAG as Therapeutic and Medication

1.1.2.1 Heparin as Anticoagulant Reagents

Since the mystery among heparin, antithrombin and factor Xa has gradually been unveiled, heparin became one of the most effective anticoagulant reagents used for treatment of thrombosis or supplied in the circumstance when blood clotting needs to be prevented, such as surgical operations or intravenous therapies. Although heparin as an conventional active pharmaceutical ingredient (API) has been used for decades, the downstream demand of it is still growing.\(^12\) Effective heparin-based therapeutics includes full-length unfractionated heparin (UFH), the enzymatic degraded low-molecular-weight heparin (LMWH) and the synthetic pentasaccharides designed (commercially available as Arixtra\(^\text{TM}\)) with different
efficacy and safety based on clinical data\textsuperscript{13-17}. The coagulant cascade is initiated by activation of coagulant factors via proenzyme cleavage and terminated by interaction with antithrombin (AT), an inhibitor to many coagulant factors (such as thrombin or factor Xa). The presence of heparin will dramatically facilitate the inhibitory function of AT. The mechanism of the heparin-mediated factor Xa or thrombin inhibition is still embroiled in a controversy between structural biologists and biophysicists. From the prospective of structural biology, heparin binding allosterically induces the conformational change of AT, making its inhibitory loop accessible to factor Xa/thrombin, followed by a typical serpin-protease mechanism. This point is supported by X-ray studies and will be illustrated in more detail in section 2.1. And biophysicists highlight the role of heparin as a bridge to either stabilize the transient complex formed by AT and factor Xa/thrombin or improve the chance of collision between the two proteins by reducing the degree of diffusional freedom.

Regardless of which mechanism is accepted, it has been established that there is positive relevance between heparin-binding affinity and the anticoagulant effects. And the binding affinity of heparin to proteins under physiological conditions (i.e. neutral pH and modest ionic strength) is affected by many factors including the length of heparinoids and the degree of sulfation. Although it is still a question that whether all heparin-binding proteins require a heparin motif consisting in a specific sulfation pattern (assuming some of interactions may just rely on a non-specific electrostatic interaction), for AT a particular sulfation pattern within the binding region is definitely required for retrieving the maximum binding
affinity. Relevant studies have shown that a special 3-O-sulfation at one glucosamine unit are critical to modulating the binding affinity\textsuperscript{18,19} but its presence is rare in naturally-occurring heparin samples, which dominantly consist of glucosamine with 6-O-sulfation, 2-N-sulfation and iduronic acid with 2-O-sulfation. That explains why only 30\% of heparin can bind to AT and no AT-binding affinity has been observed in heparan sulfate, the latter containing even fewer sulfation sites. The synthetic strategy to produce the heparin oligosaccharides with unique structure and sequence (e.g.

1.1.2.2 Heparin-Based Therapeutics Deliver System

Although the function and therapeutic value of heparin mainly resides in the anticoagulant effect through interaction with AT, many other proteins with clinical values have the potential to bind to heparin. Taking advantage of the functional diversity of heparan sulfate and synthetic chemistry, we can always design and build heparin-based materials to trap therapeutic proteins or nucleic acids. Heparin-based delivery systems feature high capacity of client protein and excellent degradability. One important area where a heparin-based delivery system holds enormous promise is in regenerative medicine, where their potential for controlled growth factor delivery is being actively explored.\textsuperscript{23-25} One of the possible designs of a nanogel is to utilize heparin chains electrostatically attached to polycationic segments of elastin-like polypeptide for loading of growth factors and morphogens with subsequent delivery to the damaged tissue (such as the poly-
lysine/heparin/payload protein complex highlighted in Figure 1-2). In addition, understanding the ability of heparin-like GAGs to sequester and release growth factors will translate to advances in the design of tissue engineering scaffolds. On the other hand, developing novel heparin-based nanomaterials with advanced properties (such as temperature-sensitive control) expands the applications of heparin in the field of therapeutics.26,27

1.1.3 Challenges of Heparin-Like GAG in Analytical Chemistry

One of the most highly desired objectives in the field of heparin-mediated protein regulation is to understand and define the structure/function relationship and finally decipher the sulfation code within the protein-heparin interactome. However, the structural diversity exhibited in heparin-like GAG makes this objective difficult to achieve even compared to the complex paradigm of protein-protein interactions. The structures of proteins are usually tightly controlled at the genetic level, and enzymatic modifications (such as phosphorylation, which plays roles of on/off switches in signaling) are typically limited to one (or a handful) per polypeptide chain. Contrary to this, heparin-like GAGs exhibit a dramatically higher level of structural diversity, and the segments conferring affinity to specific targets frequently are not well defined. Nevertheless, extensive efforts have been made searching for specific and precisely defined GAG sequences capable of recognizing protein targets with high affinity.28 This early success in the discovery of Arixtra™ inspired numerous subsequent searches for well-defined recognition sites in other GAG-interacting proteins. Many of the more recent studies have led to a more
nuanced view of GAG structural motifs and their affinity for specific proteins (e.g., a set of sulfate groups at strategic locations as opposed to a well-defined chemical structure\textsuperscript{29}.

Besides the uncertainty in the mechanism of protein modulation, the heterogeneity in samples of heparin-like GAG also brings about challenge in analytical strategy itself. Countless modifications make the simple task such as mass measurement of protein-GAG complexes impossible. Measuring the mass would help determine the binding stoichiometry, evaluate the binding affinity and characterize the structure. The general ideas to overcome this challenge are i) to use a synthetic GAG with pure sequences for numerous activity tests and the mechanistic studies or ii) to seek better separation strategies or powerful analytical tools, which are reviewed in the following sections.

1.2 A Map of Prevailing Arsenals for Studying Heparin-Protein Interaction

In this section, I enumerate several popular tools, which have been used in heparin-related studies for decades.

1.2.1 Traditional Structural Biology Strategies

The mechanism of a modulator (protein or other ligand) affecting protein function is mirrored by its impact on protein structure. Traditional tools such as X-ray crystallography and NMR has been successful in answering this type of questions by defining localization of ligand-binding sites and enabling structural comparisons. High resolution structures are achieved by crystalizing the ligand-bond protein to fix a snapshot of the all-atom coordination and the electron density
map (including the intensity and phasing) generated based on the X-ray diffraction data is used to build up the structural model. This approach has been performed in order to completely understand how heparin-like GAG regulates the function of proteins. And in all of these cases, a synthetic oligosaccharide was utilized. The heparin-binding site of AT and the allostERIC activation model in the inhibition of Factor Xa was also acquired using this approach, where a fully sulfated, pure pentasaccharide was used. From the crystal structures we clearly see that AT activation upon pentasaccharide binding is the crucial step for Xa inhibition. This mechanism agrees with the earlier discovery that the ternary complex of heparin-AT-Xa does not require a direct interaction between heparin and the protease Xa, while the ternary complex formed by AT-heparin-thrombin does. The structures of the ternary complex were solved by two groups in the same year; they both used mimetic heparin polysaccharide containing two negative charged domains, which enabled them to bind AT and thrombin individually. As another giant family of heparin-binding proteins, the fibroblast growth factors (FGF) initiate signaling related to cell proliferation and differentiation through the binding to the fibroblast growth factor receptor (FGFR), which can be facilitated by the mediation of heparan sulfate. A series of FGF•FGFR structures in the presence or absence of heparin oligomer were resolved for different protein subspecies. Some dynamic regions (e.g. the GAG units unbound or nonspecifically bound to the protein) were likely missing from the electron density map. Nevertheless, crystallography is still the most reliable approach to understanding the role of certain modification pattern in protein binding (by tracking the hydrogen bonds and salt bridges). Compared to
crystallography, NMR can generate structural models embracing more kinetic/dynamic information. In solution, protein structure in the presence of GAG can also be resolved using NMR, which provides some critical insights regarding heparin-mediated FGF-FGFR interaction and thrombosis.\textsuperscript{36-38} It should be noted, that structural modeling can be interpreted in an unbiased fashion only if the relationship between protein function and the structural variety of heparin-like GAG has been elucidated through other empirical and experimental approaches. It is critical to choose an appropriate binding element to generate a snapshot that may also represent the scenarios for other related binders.

\subsection*{1.2.2 Physical Chemistry Strategies}

The hypothesis that the functional effects of heparin arise from specific interaction between positive residues of proteins and sulfate groups of heparin-like GAGs can be demonstrated using model saccharides and solving the structure. If true, the redundant sulfate-containing disaccharide may not participate in the protein regulation based in the “key-to-lock” model. Physical chemistry fills the gap caused by a universal electrostatic interaction with the polyanion-particle interaction model emphasizing the roles of charge density in binding affinity and many heparin-mediated effects, such as protein aggregation\textsuperscript{39,40}. This approach provides better insights regarding the universal basis of heparin/protein interaction in solution. Many tools for biophysical chemistry studies have also been used including capillary electrophoresis (for measurement of mobility), light scattering techniques (for measurement of size distribution) and isothermal calorimetric
titration (for measurement of binding affinity). These techniques are usually tolerant to the structural variety to some extent and can deal with heterogeneous sample directly without extra separation. One fruit we have harvested from this area is the knowledge that binding affinity (perhaps contributed by non-specific electrostatic interaction) is affected by environmental factors including pH and ionic strength, which is valuable information for us in choosing a condition in order to strengthen the global binding affinity and/or minimize non-specific binding.

### 1.2.3 Mass Spectrometry-Based Strategies

Mass spectrometry holds several advantages when being applied to the heparin-related analysis: i) moieties rich in sulfate provide numerous negative charges (in the electrospray source) and enhances the detection of the ions (in the negative ion mode); ii) all successfully ionized heparin molecules can be displayed simultaneously in a mass spectrum and those peaks representing different masses can be distinguished as long as the resolving power and dynamic range of the mass analyzer can tolerate and accurately reflect the heterogeneity in the sample. Mass spectrometry is a perfect analytical tool for many applications including GAG sequence mapping, protein epitope mapping and GAG-bound protein structure characterization.

### 1.2.3.1 Mass Spectrometry-Based “Proteomics”

The robustness of mass spectrometry is tightly associated with the most straightforward tasks: to characterize the identity of GAG molecules in terms of the mass, structure and sometimes the abundance in the background of many other
species. The applications for this technology may range from completely profiling heparin-like GAG in naturally-occurring samples to identifying the heparin segment interacting with partner proteins. The area of glycosaminoglycanomics aiming at analyzing the GAG sample extracted from different tissues, cell types and individuals borrows the following means from regime of proteomics.

(1) Similar to the “bottom-up” strategy in proteomics, unfractionated heparin-like GAGs need to be digested to small segments in order to be analyzed effectively\(^1\). The segments can be further restrained if the purpose is to screen the segments that can tightly bind to a protein. The digested mixture as a library often passes through a protein-immobilized column.

(2) The high-throughput analysis relies on a LC separation prior to the MS analysis. Similar to the platform of proteomics, the LC separation can be combined with MS in an online fashion\(^2\). The types of separation for GAG can be RPC, HILIC and SEC.

(3) Deconvolution of the spectra is based on database search for a structural assignment at MS level. Recently, the Zaia group utilized an optimized online SEC-MS approach to analyze a GAG sample and successfully identified species with the chain length up to 30 saccharide units\(^3\).

(4) Identification of GAG segments can be done at the MS/MS level to fulfill advanced ion monitoring and structural analysis. For example, Linhardt and his collaborators in China developed a LC-MRM method to quantitatively analyze disaccharide from cell culture. Like proteomics employing MS/MS for peptide sequencing, cross-ring fragmentation of GAG can allow us to localize the
modifications. The Amster group introduced electron-detached dissociation (EDD) technique into this field, which can generate a huge number of fragment ions consisting in the cross-ring fragmentations and does not induce sulfate shedding. However, EDD has to be carried using an incredibly long acquisition time and hence it is impossible to run with online LC separation. The Sharp group recently developed a CID-based online LC-MS/MS platform where the labile sulfate groups for CID are derived to isotopic-labeled acetylated group during the sample preparation step.

1.2.3.2 Native Mass Spectrometry for Protein-GAG Complex

In native mass spectrometry, the solution conformation and non-covalent structure of protein complexes can be maintained to a large extent. One of the merits of using native mass spectrometry for characterization of protein-GAG complex in the gas-phase is that the electrostatic interactions between protein and GAG are strengthened and hence the complexes can withstand harsh desolvation conditions. An intrinsic drawback of native mass spectrometry is the modest mass accuracy in the high m/z regions, so it usually hinders native mass spectrometry being primarily chosen for identification of the ions with significant mass variation. However modern mass spectrometer (e.g. FT-ICR) dramatically improves the resolving power and mass accuracy for a wide range of heterogeneous analytes. For example, AT was incubated with a combinatorial hexasaccharide library and was analyzed in a manner of intact complexes under the native condition. As a result, peaks represent AT interacting with different dp6 members displayed in the
spectrum and have adequate resolution to determine the number of sulfate groups and ammonium adduct. A semi-quantitation was achieved based on the peak areas\textsuperscript{46}.

Another potential value of native mass spectrometry is that the ions with near-native structure can be further manipulated in the gas phase, which reflects meaningful information regarding the structure and stability. In another work related to AT interacting with pentasaccharides, the collisional cross section (i.e. the average projection area of a tumbling ion) of native ion of complex was estimated using ion-mobility spectrometry as a supplemental measurement of native mass spectrometry\textsuperscript{47}.

1.3 Future Directions: State-of-Art Mass Spectrometry-Based Techniques

Although current tools described in previous sections hold the keys to vast array of heparin-related questions, the limitations of those techniques present analytical challenges caused by heterogeneous structure and diverse function of heparin-like GAG. To push the boundaries toward the better elucidation of structure/diversity relationship, I will explore some emerging mass spectrometry-based techniques in this dissertation. The mainly relevant techniques are discussed as following, which aim at increasing a dimension of separation an may be promising to solve heterogeneity-related problems.

1.3.1 Ion-Mobility Spectrometry Mass Spectrometry

In traveling wave ion-mobility spectrometry, the gas-phase mobility of an ion is measured based on its arrival time when passing through a plate-stacking cell in
the buffering gas. Mobility of an ion can be converted to CCS, which is a reflection of the particle size and structural features. It is still controversial whether the structural model predicted based on measured CCS values and MD simulations can be trusted. Nevertheless, ion-mobility spectrometry is absolutely a powerful separation tool even for the molecules with small structural differences. Interestingly, ion-mobility measurements are enabled for both precursor ions and product ions, in a MS/MS experiment. Li’s group has reported work involving IMS to locate a single epimeric amino acid within a peptide by comparing arrival times for fragment ions 48.

For heparin and heparin-protein complexes, structural variety may lead to a MS peak convolution. The ion-mobility separation will potentially increase the resolution by increasing an extra dimension of separation. This feature will be demonstrated and reiterated in following chapters (Chapter 2 through 4).

1.3.2 Novel Online LC-Native MS for Protein complexes

Native MS can be combined with size exclusion chromatography (SEC) and ion exchange chromatography (IXC) in an online fashion. Our group is one of the pioneers to explore the feasibility of SEC-MS 49 and IXC-MS 50 and its potential applications. This technique has now drawn quite a lot of attention from many groups in industry. It can be used to characterize highly heterogeneous sample such as those proteins containing combinations of post-translational modifications.

It has been reviewed previously that SEC-MS is a perfect tool for high-throughput profiling low-molecular-weight heparin, because the species with
different chain length are retained on the column for different time and can be identified individually. This feature should work for protein-heparin complex as well under the native condition in order to reduce the peak convolution. Online SEC-native MS will be an important tool for achieving our goal in the fourth chapter.

### 1.3.3 Gas-phase chemistry for manipulation of gaseous ions

The last approach to enhance the analytical dimensions is the use of quadrupole. In another words, gas-phase chemistry including collision-induced activation and electron-transfer-induced limited charge reduction can be applied on a mass-selected population. And the latter has been verified to be promising in determining the mass for heterogeneous protein systems. The most intriguing aspect of this is that the gas-phase chemistry can be combined with other two separation-tools mentioned above and give rise to a three-dimensional analytical platform. The combination of charge-reduced or activated ions and ion-mobility separation will be reiterated in Chapter 2,3,4.
Figure 1-1. The wide size range (1-50 kDa) and enormous diversity of sulfation patterns epitomize the extreme structural heterogeneity of heparin and related GAGs.
Figure 1-2. Schematic representation of a nanogel construct utilizing heparin-based linkers to load growth factors and morphogens (labeled as HBPs, heparinbinding proteins) for delivery to tissues.
1.4 References


CHAPTER 2

INTERACTIONS OF INTACT UNFRACTIONATED HEPARIN WITH ITS CLIENT PROTEINS CAN BE PROBED DIRECTLY USING NATIVE ELECTROSPRAY IONIZATION MASS SPECTROMETRY*

2.1 Introduction

Glycosaminoglycans (GAGs) are ubiquitous linear polysaccharides that carry out diverse functions in the human body,\(^1\) with heparin being perhaps one of the best known GAGs due to its widespread use in medicine as an anticoagulant.\(^2\) In addition to its critical involvement in the blood coagulation cascade, heparin (as well as its next-of-kin heparan sulfate\(^3\)) possesses a broad range of other clinically relevant properties.\(^4,5\) Despite the widely acknowledged therapeutic potential of heparin and heparin-like GAGs in clinical fields ranging from oncology to neurodegenerative diseases to tissue engineering,\(^6\text{-}^{10}\) their use remains disappointingly limited outside of the anticoagulant field.\(^11\) One of the major reasons for that is the extreme structural heterogeneity of heparin, which makes it very difficult to decipher the molecular mechanisms of its interaction with nearly half-a-thousand proteins comprising GAG interactome.\(^12\) Even the mundane task of determining the composition and binding stoichiometry of heparin/protein complexes remains presently unattainable (unless intact heparin is substituted with relatively homogeneous short fragments\(^13\text{-}^{16}\) or synthetic mimetics\(^17\)). Sequence polydispersity of heparin and related GAGs arises from their biosynthesis, in which

\* This chapter is an authorized reprint of the article published by Zhao, et al. in Analytical Chemistry 88 1711-1718 (2016).
enzymatically-controlled modifications of precursor linear polysaccharides lead to truly intimidating variability in the levels and patterns of O- and N-sulfation and deacetylation. These countless modification patterns (which appear to be stochastic, but are in fact exquisitely regulated and biofunctional), as well as the chain length variation, present a truly Herculean task vis-a-vis analytical characterization of heparin and its interaction with client proteins.

Among the many experimental techniques used in biophysics and structural biology to probe biopolymer interactions, native electrospray ionization mass spectrometry (ESI MS)\textsuperscript{18-20} stands out as a versatile tool that is not limited by the physical size of the non-covalent assemblies and is capable of observing multiple species in solution as long as the distinction can be made on the basis of mass. However, in a case of highly heterogeneous samples native ESI MS (as well as ESI MS measurements carried out under denaturing conditions) fail to provide meaningful information due to extensive overlap of ionic peaks corresponding to species with different masses but similar or indeed identical $m/z$ values; in extreme cases these overlaps may give rise to a continuum distribution of ionic signal without any discernable features.\textsuperscript{21} Historically, complexity reduction of ESI mass spectra of heterogeneous samples was achieved by introducing a separation (chromatographic) step into the workflow,\textsuperscript{22} which requires the use of non-denaturing chromatography (such as size exclusion), if the separation step is to be combined with native ESI MS.\textsuperscript{23} Instead of (or sometimes in addition to) introducing a solution-phase separation step, ion mobility can provide an orthogonal separation dimension in the gas phase to aid MS analysis of complex samples.\textsuperscript{24,25}
Interpretation of highly congested mass spectra can also be aided by using ultra-high resolution mass analyzers capable of resolving even very closely spaced ionic peaks,\textsuperscript{26} although this approach is typically used for relatively low-molecular weight analytes, while its applications in the analysis of heterogeneous biopolymers remain limited. Finally, spectral congestion can be frequently brought to an acceptable level by reducing the charge states of biopolymer ions via e.g. proton transfer reactions in the gas phase.\textsuperscript{27}

Recently, we introduced an approach that combines complexity reduction (mass-selection of a narrow distribution of ionic species from a heterogeneous mixture) and either electron transfer to or electron capture by the selected ions without subsequent dissociation to induce partial reduction of the ionic charge.\textsuperscript{28} The resulting spectra are easy to interpret, leading to correct mass assignment for biopolymer assemblies as large as 0.5 MDa.\textsuperscript{29} Using the electron-based charge reduction schemes as opposed to proton-transfer reactions allows this approach to be applied to macromolecular ions in which the majority of cationizing agents are not protons (\textit{e.g.,} synthetic polymers and nanoparticles). Since the gas-phase GAG polycation formation usually proceeds via alkali metal\textsuperscript{30,31} or ammonium adduct formation,\textsuperscript{15} electron-based charge reduction should also be well-suited for heparin and its complexes with the client proteins. In the present work we use this approach to obtain meaningful information on complexes formed by intact unfractonated heparin with a paradigmatic heparin-binding protein, antithrombin-III (AT). Complexes of multiple stoichiometries (up to three AT molecules per single heparin chain) are detected alongside free (unbound heparin). We also demonstrate that a
Combination of this technique with ion mobility separation in the gas phase allows the total analysis time to be decreased dramatically. The ability to detect distinct heparin/protein complexes demonstrated in this work expands the reach of native ESI MS into the field where analytical options have been very limited so far. This development will undoubtedly have a significant impact not only by catalyzing the studies of the mechanisms governing the interactions of heparin-like GAGs with their client proteins, but also by assisting the design of new and effective drug delivery systems utilizing heparin as a scaffold in areas ranging from oncology to tissue regeneration.32-34

2.2 Experimental

2.2.1 Materials

Porcine heparin (from mast cells) was purchased from Sigma-Aldrich Chemical Company (St. Louis, MO) and used without any purification, fractionation or enzymatic treatment. High heparin-affinity fraction of antithrombin-III (AT) was a generous gift from Professor Paul Dubin (UMass-Amherst). All solvents, buffers and salts were of analytical grade or higher. The stock heparin solution was prepared by directly dissolving weighed powder in water and the stock protein solution was cleaned via buffer exchanging into 30 mM solution of ammonium acetate using ultrafiltration. The heparin/AT mixture was prepared immediately before the MS analysis without extended incubation. The final concentrations of AT and heparin were 0.025 mg/mL and 0.03 mg/mL, respectively.
2.2.2 Methods

All MS and ion mobility measurements were carried out with a Synapt G2-Si HDMS (Waters Corp., Milford, MA) mass spectrometer equipped with a nanospray ionization source. To minimize collisional activation of ions transferrin the ESI interface region, the following parameters were kept at minimal values: sampling cone voltage, trap CE, trap DC bias and transfer CE. Ion selection prior to limited charge reduction in the trap was carried out by setting quadruple parameters for different width of ion transmission window: LM resolution, 2 – 4; HM resolution, 15; and offset voltage, 0. To trigger electron transfer reactions, the trap wave height was set at 0.5 V. ETD recharging current was also optimized (typically 8 – 20 A) for sufficient intensity of charge-reduced species. Post-acquisition processing of MS and IMS data was carried out using MassLynx 4.1 and DriftScope 2.6, respectively (Waters Corp., Milford, MA). Curve fitting was done using OriginPro (Origin Labs, Northampton, MA).

2.3 Results and Discussion

2.3.1 Characterization of protein/heparin complexes using “traditional” native ESI MS and/or ion mobility spectroscopy

Native ESI MS provides an elegant way to monitor interactions of short and relatively homogeneous heparinoids with proteins; however, application of this technique to larger systems (such as intact heparin) is not straightforward. Figure 2-1 shows native ESI mass spectra of AT in heparin-free solution and AT/heparin mixture. The mass spectrum of heparin-free AT contains contributions from two glycoforms, which are commonly referred to as α- and β-AT (molecular weights 57.9
and 55.7 kDa, respectively), with the latter accounting for over 90% of the total ionic signal. The mass spectrum of the AT/heparin mixture displays abundant ionic signal, while at the same time showing that the signal of free (unbound) AT is nearly eliminated, suggesting that strong binding takes place between the protein molecules and the GAG chains. However, the ionic signal above \( m/z \) 4,000 is a completely unresolved continuum distribution that spans several thousand \( m/z \) units. Even though the distribution is not monotonic, and some features can be discerned, it is impossible to obtain any meaningful information related to the ionic masses. Likewise, IMS analysis of the AT/heparin mixture reveals the presence of several semi-distinct (partially overlapping) ionic populations (see Figure 2-S1 in), but the additional separation afforded by the mobility measurements in the gas phase clearly is not sufficient for obtaining interpretable IM-MS patterns from which the masses of putative AT/heparin complexes could be deduced.

### 2.3.2 Limited charge reduction of heparin polycations in the gas phase is accompanied only by minimal ion dissociation

We approached the task of measuring the masses of AT/heparin complexes initially by isolating relatively narrow ionic populations in the gas phase followed by their charge reduction using electron transfer reactions, a technique initially tested on extensively glycosylated proteins, and later applied to large protein complexes and aggregates. Electron-driven charge reduction of the polycationic species in the gas phase is ideally suited for heparin, as ESI-generated heparin polycations are largely alkali metal- and ammonium adducts, rather than protonated species, which is likely to make the use of proton transfer reactions for
the charge reduction purposes ineffective. However, electron transfer to a heparin (or heparin-containing) polycation generates a free radical, which may induce ion fragmentation in the gas phase. If the charge reduction is not accompanied by collisional activation of ions, such fragmentation events are unlikely to lead to dissociation (physical separation) of the fragments in the gas phase when applied to globular proteins and protein complexes, where multiple hydrogen bonds reinforce ternary and quaternary structures. In fact, ion dissociation was never observed in our previous work when electron-based (both transfer and capture) limited charge reduction was applied to proteins, their complexes and aggregates under conditions that minimize collisional activation. However, GAG chains are linear and unlikely to collapse to globular structures upon transfer from solution to the gas phase in the ESI interface. Therefore, generating a free radical in the process of electron-driven charge reduction may conceivably result in dissociation (fragmentation followed by physical separation of the fragments), unless the affected chemical bond is part of the saccharide ring (in which case a cross-ring cleavage would be required for dissociation).

To evaluate the likelihood of such unintended dissociation of heparin and heparin-containing polycations in the gas phase when collisional activation is minimized, we applied the limited charge reduction technique to the ions representing free (unbound) heparin in the mass spectrum of the AT/heparin mixture. Figure 2-2 shows the results of limited charge reduction of ions confined to a narrow ionic population (m/z 3190-3230) that was selected from the region of the AT/heparin mass spectrum populated by free heparin ions. The vast majority of
product ions appear in the $m/z$ region above that of the precursor ions; only a small fraction of ions representing less than 1% of the total ionic signal appear below the low-mass cutoff for the precursor ion selection window (see the colored region in the inset in Figure 2-2) and, therefore, are fragment ions. In addition to these low-$m/z$ fragments, some of the product ions populating the region of the spectrum above the mass selection window can also be fragment ions (with low charge states). However, analysis of their $m/z$ values suggests that the vast majority of them are simply products of ionic charge reduction, and if any dissociation does accompany this process, it results only in a very small mass change.

This analysis was carried out by assigning a random integer charge state $Z$ to the precursor ion population within the selection window and calculating the masses of charge-reduced species $Z-1$, $Z-2$, etc. The results of these calculations are shown in Figure 2-2 with colored arrows, each color corresponding to a specific mass and charge state of the precursor ions (as indicated in Figure 2-2). Changing the position of the precursor ion selection window will, of course, allow a different set of free heparin molecules to be detected, and complete “cataloging” of the free heparin species would require the precursor ion selection window to be moved across the entire $m/z$ region of interest (a point that will be addressed later). In the mass spectrum of the products of limited charge reduction shown in Figure 2-2 there are four different groups of heparin polycations with $m/z$ ratios falling within the precursor ion selection window, ranging from 16 to 26 kDa. The mass spectrum does not contain any “orphan” ions in the $m/z$ region above that of the precursor ion selection window that cannot be assigned as products of charge reduction without
dissociation. Similar conclusions can be drawn from the analysis of charge-reduced heparin ions whose precursors were selected from different $m/z$ windows suggesting that no significant dissociation of heparin polycations occurs as a result of electron-driven limited charge reduction under conditions that minimize collisional activation. Even lower dissociation yields are expected for heparin-protein assemblies, where strong electrostatic interactions are expected to provide additional structural reinforcement in the gas phase.

2.3.3 Establishing the identity of AT/heparin complexes using limited charge reduction

Even though straightforward use of native ESI MS or IM-MS of the AT/heparin mixture does not allow meaningful information to be extracted from the unresolved mass spectra, some spectral features could nonetheless be seen in the high $m/z$ region that hint at the presence of protein/GAG complexes with different stoichiometries, each giving rise to a continuum ionic signal. For example, ions in the $m/z$ region 4500-5500 (immediately above that of AT ions) may represent 1:1 AT/heparin complexes, while the ability of a single heparin chain to accommodate more than one protein may give rise to ions with higher $m/z$ values. In order to confirm that and identify the regions corresponding to different protein loading, precursor ions were selected within broad (100 $m/z$ units) windows and subjected to limited charge reduction. Three representative mass spectra of ions produced by limited charge reduction of precursor ions representing the three distinct regions in the native ESI mass spectrum of the AT/heparin mixture are shown in Figure 2-3. The $m/z$ values of ion peaks from the well-defined charge ladder representing the
precursor ions at highest m/z (selection window 6850-7050) can be readily used to obtain the masses of ions populating this region of the mass spectrum (189, 196, 203, 210 kDa, corresponding to AT₃-heparin complex). The analysis of the charge ladder representing the precursor ions at lowest m/z values (4750-4850) is also relatively straightforward, yielding four ionic species in the 62–77 kDa mass range (corresponding, as expected, to AT·heparin complex, *vide supra*; see the panel labeled “m/z = 4800” in the Supporting Information for more detail). However, the mass spectrum of the products of charge reduction of precursor ions confined to the isolation window m/z 5950-6050 (purple trace in Figure 2-3) is very convoluted and cannot be interpreted unequivocally without making any assumptions.

This overwhelming spectral complexity is likely to be due to the presence of multiple AT/heparin complexes whose m/z values fall within a relatively wide precursor ion selection window. In order to circumvent this problem and to simplify the data analysis process, another mass spectrum was collected after narrowing the width of the precursor ion selection window ten-fold (m/z 5995-6005). The resulting mass spectrum (the cyan trace in Figure 2-4) is indeed greatly simplified compared to the original spectrum and contains contributions from only three distinct ionic species corresponding to AT₂·heparin complexes with different heparin chain lengths (14.6, 20.6 and 26.6 kDa, respectively), while the majority of other charge-reduced species that were prominent in the original spectrum (purple trace in Figure 2-4) are filtered out. For example, a peak at m/z 6435 is clearly absent from the charge ladder generated upon narrowing the precursor ion selection window, suggesting that its own precursor lies outside of the narrow
precursor ion selection window \((m/z 5995-6005)\); a straightforward analysis yields an ionic mass of 180 kDa, corresponding to a 3:1 AT/heparin complex (see black arrows in Figure 2-4). A complete set of the limited charge reduction data (obtained by moving the precursor ion selection window over the entire \(m/z\) region of interest) is presented in the Supplementary Material section.

2.3.4 Combination of limited charge reduction and ion mobility increases measurement throughput

Although narrowing down the precursor ion selection window as a means of limiting the number of different ionic species leads to high-quality MS data, where mass assignment can be readily (and unequivocally) made, it has one significant drawback. Indeed, it results in a dramatic increase of the overall data acquisition time, as a significant number of experiments are required to cover the entire \(m/z\) region of interest to reveal all contributors to the unresolved ionic signal in the native ESI mass spectrum \((e.g., \text{it would require 300 measurements to cover the entire } m/z \text{ range } 4000-7000 \text{ in native ESI mass spectrum of the AT/heparin mixture if the width of the precursor ion selection window were kept at } 10 \text{ } u, \text{ as was done to acquire data shown in Figure 2-4). To avoid this clearly unacceptable data acquisition time scale, one may use wide precursor ion selection windows for } m/z \text{ regions where relatively modest spectral complexity allows such data to be readily interpreted } (e.g., m/z > 6500 \text{ in Figure 2-3}), \text{ and resort to narrow selection windows only when necessary (vide supra).}

Another possibility to increase the throughput without sacrificing the quality lies with incorporating ion mobility in the experimental work flow. Even when
relatively narrow precursor ion isolation windows are used, adding ion mobility
dimension provides an additional advantage of higher resolution, which allows
better distinction to be made among the closely spaced charge-reduced species
representing different precursor ions (Figure 2-5A). This advantage, however, is
not critical, as the ion peaks populating the m/z region above 5500 u in Figure 2-5A
can be readily distinguished from another even without taking the mobility data
into consideration, although the 2-D data presentation (using both m/z and the
mobility dimensions) does allow the charge state assignment of the product ions to
be verified by localizing on a single diagonal a group of ions having the same mass
but different number of charges on a single diagonal. More importantly, inclusion of
the ion mobility step in the measurement scheme allows meaningful information to
be extracted even when a wide m/z window is used for selecting the precursor ions
(Figure 2-5B). Even though no distinction can be made in this case among the
charge-reduced species along either m/z or the drift time axes alone, the distinct
ionic signals are clearly discernable in the 2-D diagram, revealing the presence of
1:1 AT/heparin complexes with relatively short heparin chains (ranging from 4.5
kDa to at least 13.5 kDa). At the same time, these “broad-band” experiments provide
substantial improvements vis-a-vis measurements’ throughput, as the data
acquisition time for a single measurement is independent of the window’s width
(both mass spectra shown in Figure 2-5 were acquired within 20 min, but a wider
precursor ions selection window allows fewer experiments to be carried out to
cover the entire m/z region of interest).
2.3.5 Interpretation of AT/heparin mass spectrum based on the limited charge reduction data

Application of the limited charge reduction to ions selected from different regions of the native ESI mass spectrum of AT/heparin mixture allows various spectral features to be readily assigned as contributions of complexes with different stoichiometry, as well as free heparin. The mass and charge state ranges obtained for these species in the course of limited charge reduction measurements allow the deconvolution of the unresolved mass spectrum of the AT/heparin mixture to be carried out (Figure 2-6). The most important assumption made here was that the distribution of charges for ions representing an AT$_x$·heparin complex at each specific stoichiometry fits normal (Gaussian) distribution, which had been shown before to work well with homogeneous protein ions generated by native ESI.\textsuperscript{39-41} However, unlike homogeneous populations of protein ions, the individual charge states within each AT$_x$·heparin complex are not expected to be resolved, prompting us to model the entire population of such ions as a Gaussian curve in the inverse ($z/m$) space. The initial positioning of each curve was based on the mass spectral data provided by limited charge reduction measurements, as shown at the top of Figure 2-6. This was followed by supervised minimization aiming at achieving the best fit of the sum of all curves with the experimental data, allowing the contribution of each AT$_x$·heparin complex (as well as unbound heparin) to the total ion signal in the mass spectrum of the AT/heparin mixture to be determined.

Although it may be tempting to relate the relative abundance of each type of complex to its fractional concentration in solution, it is likely that at least some distortion exists due to the differences in ionization efficiencies, which are known to
be influenced by a variety of factors (the major one being relative hydrophobicity of the species in solution\textsuperscript{42,43}). Nevertheless, semi-quantitative conclusions can be made on the influence of various intrinsic and extrinsic characteristics on the efficiency of formation of protein/GAG complex formation with a specific stoichiometry by monitoring the intensity change of the corresponding ionic signal in response to variation of either environmental factors (\textit{e.g.}, solution ionic strength, temperature and pH) or the binding partners’ characteristics (\textit{e.g.}, protein modifications or mutations, alteration of GAG length of extent of sulfation, \textit{etc.}).

Another important information that can be uniquely extracted from the limited charge reduction measurements presented in this work is the range of GAG chain lengths that can accommodate a certain number of proteins (such protein loading information is particularly valuable in the design of protein drug delivery systems utilizing GAGs as scaffolds). For example, the 1:1 AT/heparin complexes (signal component colored in red in Figure 2-6) incorporate relatively short heparin chains (up to 25 kDa), while complexes with higher stoichiometry use notably longer heparin chains (up to 36 kDa and 49 kDa for 2:1 and 3:1 complexes, respectively). The implicit assumption made here is that a single AT molecule cannot accommodate more than a single heparin chain, which allowed us to determine the stoichiometry of AT/heparin complexes based on their masses. While there is unequivocal proof that AT has only a single heparin-binding site,\textsuperscript{44,45} other heparin clients may exhibit more promiscuous behavior and accommodate a larger number of GAG chains, making it difficult to deduce the stoichiometry of the protein/heparin complex from its mass alone. This problem may be circumvented by introducing a
separation step (either size exclusion or ion exchange LC) in the experimental routine presented in this report (this work is currently underway in our laboratory). Finally, the ability to obtain information on charge state distributions of ionic populations representing various protein/GAG complexes allows the conformational stability of the polypeptide components of such complexes to be evaluated. Not surprisingly, the charge state distributions of all AT/heparin complexes observed in this work are narrow, and the ionic charge density is low, suggesting that a compact conformation is maintained in solution. However, this may not be universally true, and the shapes of charge state distributions of ions representing polypeptide/GAG complexes may provide important clues vis-à-vis the nature of the interaction (e.g., binding to a well-defined “epitope” on the protein surface vs. polyanion/polycation interaction that maximizes the electrostatic attraction by keeping both chains unfolded, which is likely to be the case for heparin interaction with protamines or protamine-inspired polycations). Finally, the ability to determine the average charge of ions representing a protein/GAG complex raises a very intriguing question whether this information can be used to estimate physical dimensions of these complexes in solution, an approach that had been used in the past to estimate the size of globular proteins and protein complexes, and was later extended to unfolded proteins, including intrinsically disordered species.

2.4 Conclusions

Mass measurements of large highly heterogeneous macromolecular ions and their non-covalent complexes traditionally presented a challenge for native ESI MS;
even though this technique had been successfully used in the recent past to measure masses of very large protein assemblies, the best results are usually obtained when working with highly homogeneous targets. In contrast, the progress had been relatively modest when the analytical targets are (bio)polymers whose synthesis is controlled enzymatically or semi-synthetically, rather than genetically. In this regard, heparin and heparin-like GAGs constitute a particularly “difficult” class of compounds that exhibit truly intimidating degree of heterogeneity: it challenges the very notion of molecular mass, which loses its clear and intuitive meaning when applied to highly heterogeneous species, such as heparin. As a result, analytical characterization of heparin and heparin-like GAGs (as well as their complexes with client proteins) were traditionally focused on extracting characteristics averaged across the entire ensemble; an alternative approach focuses on short homogeneous mimetics as heparin surrogates, which willingly sacrifices the unparalleled diversity of intact native heparin. In contrast, a combination of native ESI MS and limited charge reduction in the gas phase allows meaningful information to be obtained on individual protein/GAG complexes, including protein loading and the range of the GAG chain lengths. While the exhaustive characterization of all ionic species contributing to the continuum signal in native ESI could be time consuming, introduction of ion mobility measurements into the experimental scheme allows the measurements be carried out in a broad-band mode, leading to a significant increase in throughput (as had been observed in other instances when utilization of wide precursor ion selection windows increased the measurement throughput). The analytical capabilities of this method of monitoring GAG/protein interactions will be
expanded further by incorporation of an on-line separation step prior to ESI MS analysis: recently we demonstrated that limited charge reduction can be used in the on-line format with ion exchange and size exclusion chromatography (manuscript in preparation). Utilization of an extra separation step that does not compromise the integrity of non-covalent associations will allow complex GAG/protein systems (where two or more different proteins are tethered to a single chain) to be interrogated in a meaningful fashion. This will undoubtedly open up a host of exciting and previously unavailable opportunities to study interactions between heparin-like GAGs and a variety of their client proteins, a boon to both fundamental studies of GAGs, and the design of drug delivery systems relying on GAGs as versatile scaffolds.

2.5 Acknowledgements

The authors are grateful to Prof. Paul Dubin (UMass-Amherst) for inspiring discussions and to Prof. Cedric E. Bobst (UMass-Amherst) for technical help and advice.
Figure 2-1. ESI mass spectra of 4 µM AT solutions in 30 mM ammonium acetate (pH 6.0) acquired in the absence (grey trace) and the presence of 0.03 mg/mL intact heparin (black trace). Charge states are labeled for ions representing the β-glycoform of AT (ionic species of the α-form are labeled with circles).
Figure 2-2. Ions generated by limited charge reduction of a population of free heparin ions whose m/z values are confined to the m/z region 3190-3220. Colored arrows indicate positions of charge reduced species corresponding to precursor ions with different masses and charge states whose m/z values fall within the selection window. Asterisk indicates an ionic peak that could not be assigned as a product of charge reduction alone (and, therefore, must represent fragment ions).
Figure 2-3. Native ESI mass spectrum of the AT/heparin mixture (black trace) and three representative mass spectra of ions produced via limited charge reduction of precursor ions whose m/z values fall within 4450-4550 (red), 5950-6050 (purple) and 6850-7050 (blue) m/z regions.
Figure 2-4. Limited charge reduction mass spectra obtained using wide (purple) and narrow (cyan) precursor ion selection window. The ionic species labeled with A, B and C represent resolved charge-reduced species whose precursor ions’ m/z values fall within both the broad and the narrow selection windows (calculated masses are 138.1, 132.1 and 126.1 kDa, respectively). The black arrows indicate positions of several charge-reduced species and their common precursor ion that had been filtered out by switching for the broad to the narrow selection window.
Figure 2-5. Mass spectra (cyan traces) and ion mobility mass spectra (2-D diagrams) of ions produced by limited charge reduction of precursor ions at m/z 4500 isolated using a narrow selection window (window width 10 $u$, panel A) and a wide selection window (window width 100 $u$, panel B). The resolved charge-reduced species are circled and their masses/charge states labeled.
Figure 2-6. Extraction of contributions of AT₃-heparin complexes to the total ionic signal in the native ESI mass spectrum of AT/heparin mixture based on the results of limited charge reduction measurements. The number of AT molecules in each ionic species is indicated on each curve; the black line is the summation of all individual contributions (best fit to the experimental data shown in gray). The colored bars at the top of the graph show m/z ranges for each species provided by limited charge reduction measurements.
2.6 References

CHAPTER 3
ION-MOBILITY MASS SPECTROMETRY COUPLING WITH LIMITED CHARGE REDUCTION REVEALS DISTINCT MODELS OF INTERACTION BETWEEN HEPARIN AND PROTAMINE SULFATE†

3.1 Introduction

Heparin as one of the best known glycosaminoglycans (GAG) is the most negatively charged natural polymer to our knowledge with mass ranging from 3kDa to 23kDa (Figure 3-1A). It can interact with a broad range of proteins including many coagulant factors and other clinically relevant binding partners. Heparin (usually in the form of a heparinase digestion mixture with low molecular weight) are widely used in medicine as an anticoagulant. Protamine is a family of small arginine-rich proteins (50-100 amino acids) synthesized in eukaryotic sperm cells for packaging chromosomal DNA. The highly dense positive charge also allows protamine to potentially interplay with heparin, thus one derived drug, protamine sulfate, also simplified as protamine in this paper, has been widely used as a natural antidote of heparin, for instance, during the closing of a surgery in order to restore the level of platelet aggregation that has been originally suppressed by low molecular weight heparin. The interaction between heparin and protamine has also been utilized for making heparin sensors, protein therapeutic delivery systems and tissue engineer scaffolds. The sequence of protamine is genetically encoded, varying in different species (Figure 3-1B). It may result in alternative secondary

† This entire chapter will be submitted as a manuscript for publication after further edition.
structures and affinity with heparin. A complete analysis of above information has a practical value in the sense that quality control and comparability study can be achieved for the protamine produced from different sources and batches. Some techniques for particle size analysis (e.g. dynamic light scattering and analytical ultracentrifugation) have been proved to be a valuable tool to probe the behavior of protamine and heparin interaction\textsuperscript{10}. However, these methods were limited to estimate the size of particles in nanoscale (with diameter 30~60nm) and it lacks accuracy for approaching stoichiometry when the complex is initiated in the early stage. And routine structural characterization tools (e.g. X-ray crystallography and nuclear magnetic resonance), although can create models for protamine itself and protamine/DNA complex\textsuperscript{11,12}, is very challenge to solve a meaningful model of protamine/heparin complex. One reason is that these tools have low tolerance of heterogeneity\textsuperscript{13} in a sample like protamine/heparin mixture, where extensively variable sulfation patterns of heparin and diverse sequence of protamine may give rise to many species with different stoichiometry and structure.

Traveling wave ion-mobility mass spectrometry (TWIM-MS) has been widely used in mass analysis and structural characterization of macromolecules. Measurement of arrival time provides information correlated to the physical size of the ionized molecules. Intriguingly, it can be directly performed to resolve the components in a sample containing multiple species, such as a polymer with broad mass distribution, or a mixture of protein oligomers differing at the number of subunits and topology\textsuperscript{14,15}. However, TWIM-MS-based approach has been limited to the well-resolved species whose mass and charge state can be determined
accurately for the purpose of collision cross section (CCS) calculation. And therefore a direct CCS determination is impossible when analytes are poorly resolved under native condition such as glycosylated proteins and most likely in the case of protamine and heparin complexes.

Previously, we reported that electron transfer-induced limited charge reduction (LCR) of a selected ionic population was an unparalleled solution in order to decrease the bias of charge state assignment caused by peak convolution in a mass spectrum acquired from heterogeneous samples such as glycoprotein\textsuperscript{16}, protein complex\textsuperscript{17} and aggregation\textsuperscript{18}. We recently reported another case with this technique to characterize stoichiometry of the complexes formed by antithrombin and unfractionated heparin under the native condition. Without doing limited charge reduction, the mass spectrum was fully unresolved and only displayed several bad-shaped ionic clusters with nearly a continuous m/z, because of insufficient desolvation and extensive degree of ammonium ion adduct\textsuperscript{19}. Nevertheless, limited charge reduction allowed us to resolve all species existing in the same isolation window and reconstruct the distribution of the complexes through placing isolation window across the entire m/z range. In the same work, we also demonstrated that use of a post-LCR ion-mobility separation dramatically enhanced the resolution of charge-reduced species. The different charge-reduced populations were readily separated in the drift time domain even though the mass spectrum itself is very convoluted when the isolation window is broad.

In the current paper we investigated a mixture of protamines and heparin oligosaccharides under the native condition using TWIM coupling with limited
charge reduction. Our data provides novel insights of stoichiometry and conformation for the protamine/oligosaccharide complexes during the early stage of the particle formation.

3.2 Experimental

3.2.1 Materials

Protamine sulfate was provided by United States Pharmacopeia. Heparin oligosaccharide isolated through size exclusion chromatography was purchased from Iduron (UK). We used the fraction predominantly consisting of the oligosaccharide with twenty saccharide units, thus referred as dp20. Protamine and dp20 powder were dissolved into distilled water (Millipore, MA) initially to make stock solutions then diluted as 10mg/ml with 150mM ammonium acetate at pH=6.9. We mixed protamine and dp20 at 1:1 ratio (g/g) and directly used them for chromatography analysis or native mass spectrometry without extra time for incubation.

3.2.2 Online Ion-exchange Chromatography-Native Mass Spectrometry

Ion-exchange chromatography was carried on using Agilent 1200 HPLC system equipped with a weak cation exchange column ProPac-100 (Thermo Fischer). The column was equilibrated with 10mM ammonium bicarbonate at pH=8.5 prior to sample injection. Gradient elution was applied by increasing the concentration of ammonium bicarbonate up to 500mM at the same pH. No extra salt was added into the mobile phase. A 7T Bruker SolariX FT-ICR mass spectrometer
was used for online mass spectrometry analysis. To get better ionization performance, the inlet flow for electrospray source is reduced to 20µl/min via a splitter after the fluid pass through the column. Data acquisition was triggered with Hystar software.

### 3.2.3 Ion-Mobility-Mass Spectrometry Coupling with Limited Charge Reduction

The protamine dp20 mixture was introduced to the gas phase using Synapt G2S equipped with nanospray source. In order to create a gentle source condition with well ionization efficiency, capillary voltage was 1.5 kV, source temperature remained at 80 °C, cone voltage was 10 V. Limited charge reduction was carried on basically following a modified procedure for regular electron transfer dissociation. Briefly, a narrow ionic population within specific m/z window was isolated using quadrupole first, LM resolution was set to 3.9 and HM resolution was set to 12. Then electron transfer reagent 1,3-dicyanobenzene was introduced into the trap to interact with analyte ions and induce their limited charge reduction. The vacuum and traveling wave height in the trap cell were optimized so that a remarkable charge reduced populations can be observed. For ion-mobility separation of charge reduced ions, IMS wave velocity and height were set to 300 m/s and 25 V, respectively; vacuum reached 3.5 mbar in helium cell and 2.5 mbar in IMS cell. Trap DC bias was ramped to 35 V in order to achieve sufficient ion transmission.
3.2.4 CCS Calibration

A couple of standard proteins with small CCS and mass value including native cytochrome C, BLG and denatured ubiquitin was run at the same condition as for the protamine dp20 mixture. Calibration curve was created based on IM-MS following the regular protocol described in reference (15).

3.2.5 Data Analysis

Online IXC-MS data were processed using DataAnalysis software (Bruker Daltonics) and LCR-IM-MS data were analyzed using Masslynx and Driftscope software (Waters). CCS plot fitting was done using Origin Pro (Northampton, MA).

3.3 Results and Discussion

3.3.1 Electrostatic Interaction Between Heparin and Protamine Can be Probed using WCX-MS

Online WCX-MS enables a real-time monitoring of eluents as ionic strength increases, although the use of ammonium bicarbonate may not have the same performance of separation as sodium chloride does. We observed that dp20 eluted immediately before the gradient elution step started, indicating that dp20 was negatively charged at pH 8.5 and was not retained on the column (Figure 3-2, the top panel). In contrast, protamine itself only eluted at high ionic strength. Extracted ion chromatogram rendered two fractions in the protamine sample based on their retention times (Figure 3-2, the bottom panel, left column). The late eluted fraction represented the protamine species with higher mass and more positively charged residues. However, the retention of protamine on the column was fully eliminated in
the presence of dp20 (Figure 3-2, the middle panel). We found protamine species with higher mass were eluted together with dp20. The result clearly suggested that a charge neutralization and an interaction occurred between polyanions and polycations, although we did not observe ionic signals corresponding to the mass of protamine dp20 complexes (>9000Da) under the applied harsh desolvating conditions chosen for the online experiment. Interestingly, the signals of several protamine species with lower masses (like the one shown in the yellow trace in the bottom panel of Figure 3-2) were invisible. That is perhaps because the ionic signals of less abundant species were suppressed during ionization if they are not separated by ion-exchange chromatography.

3.3.2 Characterization of Complexes Formed by Protamine and dp20 using Native Mass spectrometry with Limited Charge Reduction

To examine whether the observed charge neutralization can be attributed to a direct interaction between protamine and dp20, we analyzed the mixture under native conditions using mass spectrometry. The mass spectrum of the mixture displayed extensive ionic signals but only a small portion of them were partially resolved (below 1500 m/z) and they corresponded to the unbound dp20. All ionic signals from m/z=1500 to m/z=3000 were completely unresolved and suggested an ionic cluster with signals at every single m/z value (as shown in the top panel of Figure 3-3). To measure the mass of the cluster, we isolated precursor ions at m/z=2300 and induced limited charge reduction. A series of well-resolved product ions appeared in the m/z region from 2500 to 5000. Interpretation of those charge-reduced ions was done by following the same approach described in previous
work. The intensity of charge-reduced ions was sufficient for us to assign charge states based on the spacing between two correlated peaks and to calculate their mass then. As a result, three ionic species with different mass were characterized in the isolated ionic population (as indicated in the middle panel of Figure 3-3). It was straightforward to determine the 6,900Da and 9,200Da species to be two complexes of protamine and dp20 at 1:1 ratio based on the individual mass range for dp20 (3,360-5,760Da) and for protamine (3,000-4,500Da), respectively. The third mass at 11,500Da must be a ternary complex however we cannot assure the ratio of protamine and dp20. Obviously, either the complex with a single protamine chain and two dp20 chains or the one with a dp20 chain and two protamine chains can both show masses of 11,500Da. This is a common issue that we cannot just base on the mass value itself to assign a complex stoichiometry when the theoretical mass ranges of two heterogeneous binders are largely overlapping.

3.3.3 Ion-Mobility Separation and CCS measurement of Charge-Reduced Product Ions

First, we investigated a precursor ion with relative low m/z (1600) using limited charge reduction. As shown in Figure 3-4, we observed five distinct ionic populations in the ion mobility mass spectrum. The ionic signals within the m/z region from 1490 to 1650, originated from the isolated precursor, were divided into two populations in the drift time domain, centering at 4.3ms and 5.18ms respectively. They were subjected to limited charge reduction and then generated other three populations, being further separated according their drift times and m/z values. By a manually deconvoluting those peaks, we can promptly find two sets of
charge ladders corresponding to the ions with molecular weight 4800Da and 6400Da respectively (labeled in different colors and summarized in Figure 3-4). Clearly, the involvement of ion-mobility increased the resolving power for the different ionic species by providing an extra dimension of separation. It was also noted that the peaks of both precursor ions and charge reduced product ions in ion-mobility mode remained broaden and displayed more signals below the desired m/z values (set in the quadrupole) compared to those in the spectrum acquired when ion-mobility separation was off. These tailing signals were probably induced by the shedding of sulfate group under an enhanced DC bias voltage applied in the trap cell, which was required for a sufficient ion transmission through the ion-mobility cell against the buffering nitrogen gas.

Next we calculated the collision cross sections for each observed ionic population (listed in the table under Figure 3-4). To be noted, a precursor ion and its multiple charge-reduced states had very intimate CCS values (standard deviation < 0.2) suggesting that the physical size of a polymer ion remained the same during the charge-stripped process in the gas phase. Therefore, the similarity of CCS between different charge states can be a criterion for checking whether correct ionic species were picked up as one charge ladder representing the same precursor ion.

3.3.4 Data from Specific m/z Regions Revealed Distinct Species with the Same Mass

To identify the complexes in the mixture, the precursor ions at other m/z value across the entire region where ionic signals displayed were isolated and subjected to limited charge reduction. We compared the IM-MS at precursor at
m/z=2400 with the mass spectrum without ion mobility at the m/z=2300 (the bottom two panels of Figure 3-3). This small discrepancy on precursor m/z value selection (i.e. 2300 vs. 2400) was kept only to provide better alignment of peaks due to the consideration of the tailing in the spectrum acquired in the ion-mobility mode. From the ion-mobility mass spectrum, some additional information can be extracted. First, a new charge ladder representing a 13,800Da precursor ion with +6 charge state was exclusively identified in the ion-mobility mode. We can clearly noticed that its reduced charge states (+4<sup>±6</sup>) (representing the +4 product ion reduced from the +6 precursor ion) was thoroughly overlapped in the m/z domain with the product ions reduced from another precursor charge state (+2<sup>±3</sup>) of 6,900Da, however these two species can be separated in the drift time domain due to the remarkable differences in charge. Similarly, IM-MS enabled the distinction between two ionic populations, (+3<sup>±6</sup>) of 13,800Da and (+2<sup>±4</sup>) of 9,200Da.

Although significantly driven by charge difference between individual populations, the IM-based separation can also distinguish two populations with differences in physical size (CCS values) when they had the same mass and charge. Indeed, for the 11,500Da species, two populations of charge series presented with different drift time values, corresponding to two species with the same mass but different CCS values. We also observed two populations of distinct CCS for multiple charge states of 13,800Da ions. We hypothesized that the second population was not converted from the other in the gas phase. To verify this hypothesis, we investigated the same charge state +3<sup>±4</sup> generated from four different precursors. As illustrated in Figure 3-5A, it was that the population with larger CCS (referred to
as population B) did not appear until the mass reached a value between 8,800Da and 9,600Da and then the relative abundance of B continued to grow as the mass increases. This observation strongly suggested that the population B, representing a conformation or stoichiometry differing from A, can only formed over a specific mass range, reflecting the presence of the species in the solution. We also observed that the population A had more tailing signals than the population B, suggesting that the population A was composed of higher degree of heparinoid component, which offers more opportunity of mass decrease related to sulfate group shedding. For example, the population A could represent the complex with one protamine and two chains of dp20, and the population B could represent the complex with two protamine and one dp20 because the theoretical minimal mass of the latter stoichiometry is about 9360Da (3000+3000+3360) and just falls in the proposed mass range determined upon above. In a conclusion, ion-mobility-enabled analysis on limited charge reduction data provides additional hints on the potential stoichiometry and conformation of the complex.

3.3.5 Comprehensive Analysis of Stoichiometry and Conformation based on CCS-Mass Relationship

This approach only focused on the well-resolved charge-reduced populations to extract the CCS, mass and charge information of species that coexist within the precursor isolation windows and overlooked the ions without good separation, including the precursor ions themselves. For the precursors at m/z=2400 (Figure 3-3, the bottom panel) as an example, the ionic signals falls within 2200-2400 are very convoluted thus we were unable to precisely extract the drift time value for
each populations. One remaining questions is whether the CCS is unique for a specific ion and not related to the charge state of precursor. To elucidate this point, we compared the CCS of the same ion generated from three different charge reduction processes in the gas phases, presented in Figure 3-5B. For a specific molecule of 12,000Da, the +3 charge ion (m/z=4000) can be reduced from three precursor ions with different charge states +6, +5, +4 (see the ionic populations of at m/z=2000, 2400, 3000 respectively) and all of these +3 products produced consistent CCS values (about 11.16nm²) no matter what precursor it was born from), indicating that they represented the same specie in terms of mass and stoichiometry. The consistency of CCS also happened on +4 charge ions except for the +4 precursor itself, of which the CCS is not available due to the convoluted signals fallen within the selection window. These observations strongly suggests an universal conclusion that if a molecule is ionized to different positive charge states, the CCS of the ion at charge state \( n \) is identical to the CCS of the ion at \( (n+i) \) charge state when it is reduced to charge state \( n \) in the gas phase. In another word, the CCS of one charge state is relatively independent to the pathway in which the ion is generated. Besides in the case of complexes formed by protein and linear polymer, this conclusion has also been validated for other types of macromolecule such as native single proteins including albumin and \( \beta \)-lactoglobulin (data not shown) and denatured ubiquitin\(^{20} \).

Because the mass and CCS of a complex were independent to the displayed charge states, we can plot the CCS values of all available ions versus their masses to elucidate the relation between stoichiometry distribution and the mass of the
complex, as shown in Figure 3-6, where the two ions with the same mass but significant difference on CCS values (such as population A and B in Figure 3-5A) are in different colors. All of those data points were fitted by three individual trends, in which the CCS value increases monotonically as mass increases. The first curve on the bottom covers a relatively low and narrow mass range (5,000Da-7,000Da). It may represent unbound dp20 and dp20 non-specifically binding with contaminant molecules (such as PEG). Because dp20 is a chain with fixed number of saccharides, the CCS increase can only be attributed to number of sulfate group (or ammonium adduct) increasing. The enhanced repulsion between negatively charged sulfate groups makes the chain less compact. The other two curves represent protamine and dp20 complexes. Interestingly, based on the wide mass range of the middle curve (6,000 – 24,000 Da), it may be a combination of multiple stoichiometries. The same types interpretation of IM-MS data can be seen in previous IM-MS study on the mechanically interlocked polymers mixture, where the complexes contain a polymer chain with variable mass and a number of macrocycles.21 Therefore, based on the experimental mass range, we hypothesize that the middle trend is the protamine-heparin 1:1, 1:2 and 1:3 complexes and the top trend is the protamine-heparin 2:1 and 3:1 complexes. The rationale of this assignment can be verified through their potential structure features, which is discussed as follow.

There are several studies that suggested that certain types of protamine has an irregular helical structure with all charged arginine side chains toward outside22. An X-ray crystallography12 and NMR11 study also revealed that the side chains on the protamine helix can interact with the grooves of double strand nucleic acids.
Another recent molecular dynamic study illustrated the interface between protamine and heparin\textsuperscript{23}. All these data imply that a protamine molecule is very compact compared to a dp20 molecule with the same mass. Therefore two different models regarding how protamine and dp20 can form complexes are proposed upon the two trends (in orange and cyan in Figure 3-6) respectively. One model contains one or two protamine as the building center and all dp20 chains wrapping around the protamine core. The second model contains multiple protamine molecules bridged by one or two heparin chains. Obviously the model II has a more extended conformation (as shown in the cartoon of Figure 3-6). Both model I and model II may represent the early stage of the growth of protamine/heparin-involved particles at various ratios of mixture. And they can only coexist when the concentration of protamine and dp20 is low. If we keep increasing the concentration of either on or both, it will induce the formation of larger complex, giving rise to an insoluble aggregation eventually. In model I, the complex can grow large under the shielding from excess amount of dp20. Model II is favored when protamine is in excess. However the extended structure of model II is quite unstable and it does not allow the complex to expand too much, otherwise, the hydrophobic surface of protamine may easily facilitate precipitation immediately\textsuperscript{24}.

3.4 Conclusions

Native ESI-IM-MS provides an elegant means to probe the topology of small protein units aggregates, such as the discovery of the various models of β-amyloid related complex formation\textsuperscript{25}, however it is not straightforward to apply this
approach to analysis of a heterogeneous systems. We divided the task of dissecting entire ionic cluster into multiple small portions, only focusing on the ions falling within a narrow isolation window. Through identifying and recording mass, charge and CCS (and the relative abundance for a validation purpose) for all populations in the small portion of cluster using limited charge reduction, we can finally reconstruct a complete picture depicting the distribution of each type of complexes from the giant unresolved cluster. This limited charge reduction is the only “magnifying scope” that allows us to look at a fully unresolved spectrum in a meaningful way.

Electron transfer induced charge reduction occurs very efficiently in proteins and other ions with higher positive charges without creating significant fragmentation. Unlike proteins, heparin-involved protamine cannot attract ample electrons due to Coloumb’s force and it would be difficult to generate adequate charge-reduced states that will be utilized for accurate mass and charge assignment. And the measurement of the CCS for charge-reduced populations may fill this gap by providing an extra criterion to examine the initial charge state assignment. Another unique property of protamine dp20 complex is that its lower charge state has an enlarged CCS in most cases, differing from what we previously learned from studies on proteins, which follows an opposite law that the CCS gets smaller when charge is reduced\textsuperscript{20,26}. One possible reason is that the electrons initially transferred to the complex will release part of the GAG chain wrapped on the protamine and the increased repulsion between highly dense negative charges of GAG makes the chain
more extended, if we assume the conformation of GAG chain makes the major contribution to the CCS of the entire complex.

Although we tried to correlate the CCS values to some structure models, we are aware of that we would never build a robust structure model without other in silico modeling approaches such as molecular dynamics simulation as well. We can regardless treat the CCS-mass profile as a fingerprint, which records unique information in terms of the distribution of different types of complexes in the solution. By monitoring the CCS-mass fingerprint in response to the change of experimental parameters (e.g., the ratio of mixture and the length of saccharide) and the conditions of the solution (e.g., ionic strength and pH), we can evaluate the interactions between protamine and heparin, which is critical to both basic studies on mechanism of protamine/heparin aggregation and the practical applications such as design of protamine-heparin based delivery system and method development for quality control or comparability study in the filed of pharmaceutics. In addition, the approach we described in this chapter can be applied to other heterogeneous system for structural characterization.
Figure 3-1. General structure of heparin and protamine. (A) Template of low molecular weight heparin. The red color indicates where the modifications (acetylation and sulfation) perhaps occur. (B) The primary structure of protamine sulfate. It is a mixture of subtypes P1 to P4, encoded in different genes.
Figure 3-2. Online WCX-MS of dp20 alone (top), dp20 in the presence of protamine (middle) and protamine alone (bottom). UV traces and base peak chromatogram are shown in the left column. The mass spectra of analytes were shown in the right column are the average of scans within time windows indicated by the arrows in the left column. The asterisk sign indicates contaminant (PEG) peaks.
Figure 3-3. Top: mass spectrum without incorporating ion isolation; Middle: mass spectrum for limited charge reduction with precursor ion selection at \( m/z = 2300 \); Bottom: ion mobility mass spectrum for the limited charge reduction with precursor ions selection at \( m/z = 2400 \). The number in the basket indicates how many species were observed.
Figure 3-4. Analysis of charge reduced species for the precursor selected at m/z=1600.
Figure 3-5. (A) Ion mobility spectra (drift time v.s. intensity) for the ion with +3 charge reduced from +4 charge precursor isolation. All peaks are fitted into Gaussian curves. (B) Comparison of CCSs of the same ion reduced from different charge states.
Figure 3-6. CCS v.s. mass plot for all detectable ions (including precursor and charge reduced ions). All points are fitted in three individual curves with given equations. Proposed stoichiometry and conformation of the protamine dp20 complex are illustrated in cartoon. Red color indicates the +4 ion precursor that is not distinguished from other ions at m/z=3000.
3.5 References


CHAPTER 4

MAPPING MINIMAL PROTEIN-BINDING MOTIFS USING AN ION-MOBILITY TOP-DOWN MASS SPECTROMETRY APPROACH‡

4.1 Introduction

Heparin is a highly sulfated glycosaminoglycan (GAG) and is the most densely negative charged natural-occurring polymer. It is known for its roles in regulation of coagulant cascades. Heparin is a broad-spectrum anticoagulant reagent that has been used in many circumstances ranging from treatment of thrombosis diseases to avoidance of blood clots during surgical operation and intravenous therapy¹⁻³. In addition, heparin-like GAG can bind to a large numbers of proteins through a general polyanion-polycation interaction, exemplified by γ-interferon⁴ and the family of growth factors (GF). It is well-established that growth factors can trigger numerous signals correlated to cell proliferation and differentiation when binding to growth factor receptors (GFR)⁵ and the GF-GFR dependent signaling can be facilitated by the involvement of heparan sulfates, a heparin-like GAG that anchors on the surface of cellular matrix. Although the majority of heparin-related application still revolve around its anticoagulant effects, the use of heparin to build nanomaterials (see section 1.1.1.2) has become into an emerging technique in medical fields such as tissue engineer and therapeutic protein delivery⁶.

A prerequisite to understanding the mechanism of heparin/protein interaction is to learn the structural detail in the “binding motif” where specific

‡ This entire chapter with be submitted as a manuscript for publication after further edition.
modification patterns (including O-sulfation, N-sulfation and N-acetylation) within a limited number of saccharide units are required for interaction with client proteins. The bottom-up approach to identification of the binding motif has been successfully explored in past decades: a library of oligosaccharide with combinatory segments was applied on a client protein-conjugated mobile phase for separation based on the affinity. The pool containing the strongest affinity species was analyzed by mass spectrometry and tandem mass spectrometry. However, comprehensive analysis of binding motif is extremely time-consuming as the heterogeneity in the samples caused by the variety of chain lengths and modifications increases. For these reasons, our knowledge of binding motif of heparin is only available in a few of cases.

Herein we describe an alternative top-down approach that can be applied on the complex formed by protein and unfractionated glycosaminoglycan and enable a rapid estimation on number of saccharide and sulfate groups within the required minimal length of saccharide units without requiring complicated sample preparation steps. Although the intact heparin is heterogeneous, the saccharide units and sulfate groups that actually contribute to the binding are always limited to a small number of combinations. In addition, electrostatic interactions are strengthened when samples are introduced into the gas-phase, where the shielding effect of solvent molecules has been minimized. We hypothesize that heparin chain segments that are not bonded to the protein surface will fragment very readily upon collisional activation (via both glycosidic bond cleavage and sulfate “shedding”), while the units involved in interaction with the protein will remain bound to it (although they may lose “external” sulfate groups not stabilized by salt bridges with
the protein basic groups) (Figure 4-1). However, as a potential pitfall, unexpected protein unfolding during collisional activation may damage the protein conformation residing in the binding region and lead to a GAG releasing from protein surface before being cleaved. To the date, the relation between unfolding and dissociation of protein-polyanion complex in response to collisional activation has not been explored. In this work, we will investigate the feasibility of this approach on a well-studied system: acidic fibroblast growth factor (FGF) and oligosaccharide heparin with fixed number of saccharide units. We are also reporting the relevant behaviors of gaseous ion of protein-heparin complex under collision activation.

4.2 Experimental

4.2.1 Materials and Sample Preparation

Recombinant human FGF-1 protein (FGF) was purchased from GenScript. Synthetic heparin mimetic pentasaccharide Arixtra™ (dp5) was a gift from Prof. Robert Linhardt at Rensselaer Polytechnic Institute. Heterogeneous heparin enzymatic degraded products with five disaccharide units (dp10) was purchased from Iduron (UK). Prior to mass spectrometry analysis, FGF was incubated with dp5 or dp10 in buffer containing 100mM ammonium acetate at pH 7. All proteins for CCS calibration were purchased from Aldrich Sigma (St. Louis, MO).
4.2.2 Methods

The mixture of FGF and dp5 or dp10 was directly analyzed using Waters Synapt G2Si mass spectrometer equipped with nanospray ion source (Milford, MA). Capillary voltage was set to 1.2V. Sampling cone voltage was 20V. Source temperature was set to 30 °C. In the ion-mobility mode, traveling-wave of the IMS cell was tuned up in order to get good drift time peak shape and position. Trap DC voltage was kept at 35V.

Gas-phase mass shedding and protein unfolding were monitored through MS/MS experiment induced by collision with the argon gas. Ionic population representing FGF alone or FGF-GAG complex was isolated using quadrupole. Collision voltage was applied prior to ion-mobility separation and it was gained stepwise with 10V interval until it reached 120V. This voltage frame was chosen based on the property of ion and instrumentation condition. If applied collision voltage goes beyond certain experimental frame, the ions with overwhelming kinetic energy will be immediately scattered and unable to arrive to the TOF analyzer to generate any meaningful MS/MS data. Ion-mobility mass spectra under different energy were acquired for 2 to 5 minute for each to make sure sufficient signal-to-noise ratio can be achieved for analysis. Acquired data were interpreted in two different ways: 1) collision-induced unfolding (CIU) fingerprint graph was used to describe the unfolding profile of the protein in the gas phase against different activation energy in the presence or absence of GAG ligands; 2) collision-induced dissociation (CID) was used for monitoring the mass loss of the protein-GAG complex responding to the bond cleavage (including S-O and glycosidic bonds in
GAG and peptide bonds in protein). CIU/CID followed my IMS/MS measurement has been applied for characterization of a protein binding to ligands of structural variety such as lipid.

Collision cross section (CCS) of an ion was acquired based on the drift time value at the apex of a peak in the ion-mobility spectrum. To convert drift time to CCS value, we took advantage of a calibration curve generated by acquiring the ion mobility mass spectra of a series of native protein with established CCS values including Cytochrome C, β-Lactoglobulin, Bovine serum albumin and Alcohol dehydrogenase under equal instrument condition. Charge state and mass of ions were considered to eliminate the effect of electric field on drift times before plotting them against the CCS values. The detailed steps for CCS calibration was described in the standard protocol for protein complex CCS calibration.

4.3 Results and Discussion

4.3.1 General Behaviors of Unbound FGF in response to the Gas-phase Collisional Activation

The observed spectrum of FGF alone displayed three charge states corresponding to a mass at 15,900 Da (Figure 4-2A, the top panel). The CCS value upon the major charge state (+8) is 18.34nm², close to the number extracted from the crystal structure, revealing that the protein has a similar conformation presented in the native state. We investigated the unfolding profile of FGF protein by stepwise increasing the collision voltage in the trap, where selected FGF ions can be activated and all metastable populations can be captured by IM-MS. As the
collision energy was increased, two other **major** populations of FGF at +8 charge state were generated, displaying enlarged CCS values compared to the native state. These two populations represent two intermediate conformations during the gas-phase unfolding. The first intermediate has modest size expansion (14% CCS enhancement) and it implies that the protein backbone is rearranged to an extensive conformation in certain region(s); thus it is referred as a "partially activated" state. The second intermediate has 30% expansion compared to the native CCS and it may represents a state where the protein globally loses its structure features, thus referred as a “fully unfolded” state. The transition between different states can be monitored using the CIU fingerprint (**Figure 4-3A**). For instance, the +8 charge state of FGF can present its partially activated state at an applied energy as low as 32eV. In this chapter, the applied energy is defined to be the product of applied trap collision voltage and the charge state of the selected ion. And the fully unfolded state appeared when the applied energy was above 100eV. It became an exclusive population when the applied energy reached 220eV and the populations of first two states were completely converted to the fully unfolded state.

The measured CCS values of a FGF in other charge states are reasonable as well. FGF$^{+7}$ had smaller CCS value (17.5nm$^2$) than FGF$^{+8}$ because of the reduced Coulomb repulsion amongst positive charges. Relevant molecular dynamic simulation studies have suggested that during collision-induced unfolding the conformational change of a gaseous protein ion is associated with the migration of charged atoms such as protons and salt adducts$^{12,13}$. As a reasonable hypothesis, the conformation of different intermediate states and the energy threshold to induce
the transition between different states are likely to be affected by the actual surface charge. Indeed, given our results shown in the CIU fingerprint (Figure 4-3, A vs. B), both of the two intermediate states of FGF still existed in the lower charge state FGF$^+7$; and they had relatively smaller CCS values compared to the higher charge state FGF$^+8$. We also observed that FGF$^+7$ was more resistant to collision-induced activation than FGF$^+8$. For FGF$^+7$, the energies triggering the first transition (from native state to the partially activated state) and the second (from the partially activated state to the fully unfolded state) were definitely improved compared to those for FGF$^+8$, because we can clearly see that the partially activated state of FGF$^+7$ did not appear until 200eV (vs. 32eV for FGF$^+8$) and it remained as the dominant population across the entire energy range in our experimental frame (up to 770eV). In addition, the population of the fully unfolded state remained in modest abundance (<10% of the total ionic intensity). We anticipated that the energy that was capable to induce majority of fully unfolded FGF$^+7$ is out of the experimental frame.

Collisional activation also caused ubiquitous protein fragmentation for both +7 and +8 charge states in company with protein unfolding. It should be noted that even in the case of +8 ion, where protein unfolding is easier and protein fragmentation is more dramatic, a population of parent ions survives and can be distinguished from the ionic signals of fragmentation in the ion-mobility mass spectra, as shown in Figure 4-S2A.
4.3.2 Collision-induced Unfolding and Dissociation of the Complex Formed by FGF and Homogeneous Pentasaccharide

The rationale of the top-down approach for identifying FGF-binding heparin segments was initially tested using a synthetic pentasaccharide (dp5), which contains a homogeneous structure and modification pattern. Although this molecule was originally designed as a mimetic of a heparin segment for antithrombin modulation, it can bind to FGF as well because it consists in critical structural features related to the FGF-binding affinity\(^{14}\) (highlighted in Figure 4-2B). The interaction between dp5 and FGF can be directly probed using native mass spectrometry (Figure 4-2A). The spectrum of FGF in the presence of dp5 shows that FGF and dp5 can only form 1:1 complex (FGF•dp5) and there is no evidence suggesting other stoichiometries are present. The CCS for the complex based on the +7 charge state is 17.80 nm\(^2\) and this value is very close to that of unbound FGF at 17.50 nm\(^2\), suggesting that this minute CCS increment (+1.7%) is contributed by the size of dp5 itself when it presents on the protein surface and does not induce major conformational change of the protein.

Similar to unbound FGF, when (FGF•dp5)\(^+7\) was treated with collisional activation both protein unfolding and fragmentation were observed. However, we also observed that an intense peak in the product ion spectra gradually shifted to the lower m/z as the collisional energy increases, starting from the m/z value corresponding to precursor (FGF•dp5)\(^+7\), and finally stopped by the m/z value corresponding to unbound FGF\(^+7\). This peak represents the FGF•GAG complex whose saccharide components are stripped from the complex successively and the protein component still remains intact during this process. To examine how binding
to dp5 or GAG fragments affects the protein unfolding pathway, we focused on a population of FGF•GAG (without taking any unbound FGF population) in each spectrum for extraction of CIU fingerprint. It clearly shows that the energies for both transitions are lower compared to those of unbound FGF+7 (Figure 4-3, B vs. C). We can conclude that interaction with dp5 does not change the number of intermediate states or significantly affect their conformations but it inhibits protein unfolding by stabilizing individual states.

The correlation between the protein conformation and the mass loss were studied based on average mass spectra scanned within ion-mobility arrival time of each state individually. For the native state (Figure 4-4A, grey trace), the spectrum under mild activation (350eV) displays a group of equally spaced peaks (Δm=22Da) corresponding to the mass of FGF•dp5 plus various number of sodium adducts. The sodium ions are extensively attached onto dp5 molecule, observed in the spectrum of dp5 alone (Figure 4-S1A). When the applied energy was increased to 490eV, two more groups of peaks with 79Da less than the original group appear, indicating the loss of sulfate groups. For the native states, we did not observe other reasons that caused mass decreases rather than sulfate shedding. The mass spectra of the complex in the partially activated state (Figure 4-4A, blue trace) are identical to those in the native state against the applied energy under 560eV. When the applied energy went above 560eV, the native population was completely converted to the partially activated state; meanwhile some ionic signals appeared corresponding to the gradual mass decrease in a larger interval than the one caused by sulfate shedding. When the applied voltage is above 700eV, two distinct populations of
fragments became apparent, one representing FGF with pentasaccharide residues containing two sulfate groups (m/z ~ 2430) and the other with trisaccharide residues (m/z ~ 2390). It suggests that the unbound saccharide units can be removed from the FGF surface in the partially activated state through glycosidic bond cleavage (Figure 4-4B, blue trace). In the fully unfolded state (Figure 4-4B, red trace), GAG fragmentation occurred dramatically and an unbound FGF peak (m/z = 2283) became dominant. These results demonstrate that as long as the complex stays native or partially activated, the heparin-binding domain of FGF is intact and the specific interactions including H-bonds formed between protein residues and the sulfate or carboxyl groups on GAG are normally maintained. In the fully unfolded conformation, essential protein structural features for accommodation of GAG ligand are lost and therefore the remaining saccharide units can be rejected immediately. These results also suggest that “non-essential segment” in the native state of the gaseous ion may be collapsed onto the protein surface via long-distance electrostatic forces. These non-specific interactions play important roles in stabilize the native conformation. In the partially activated state, the structural features related to the non-specific interaction are disrupted and therefore the non-essential segment happens to detach from the protein surface after being cleavage.

The m/z values of the dominant peaks are plotted against applied energy for all three states (Figure 4-5). As a negative control, unbound FGF*8 does not show any collision-dependent mass decrease (Figure 4-5). The value of (FGF•dp5)*7 in the partially activated state (blue open dots) keeps decreasing and then remains
unchanged at two values, corresponding to FGF plus dp5 with two sulfate groups (5,2,0) and FGF plus dp3 (two uronic acids and one glucosamine) with two sulfate groups (3,2,0), respectively. This latter represents the last GAG components that can reside in FGF that is not fully unfolded and it is assumed to be the minimal segment to achieve specific binding. Indeed, this number is very close to the predicted binding segment (3,3,0) based on previously confirmed trisaccharide motif that is directly involved in the interaction with FGF (Figure 4-1A).

4.3.3 Probing interaction between FGF and heterogeneous decasaccharide

We hypothesize that this established approach for identifying the binding motif of heparin can generate a constant result no matter the heparinoid chain length and heterogeneity. FGF was incubated with decasacharide (dp10) containing diverse sulfation pattern, which was isolated from the native heparin degradation products using size-exclusion chromatography, and was analyzed using native mass spectrometry. The stoichiometry of the complex formed by FGF and dp10 cannot be determined upon the acquired mass spectrum alone because the peaks are rather broad and partially convoluted due to the intrinsic mass heterogeneity of dp10 (Figure 4-6A and Figure 4-S1B). Ion-mobility can considerably enhance the confidence of charge state assignment for the reasons that (1) adjacent ions with different charge can be fully resolved based on their drift time and (2) the CCSs for all tentative charge states of the same specie are supposed to be constant. In order to separate ions that have different charges but still overlap, for instance the population at m/z 2700, a mild collisional voltage (50V) was applied to all ions and
it was expected to make the overlapped ions distinguishable through differential protein activation because the ion with higher charge actually bear higher collisional energy. As a result, one series of ions, which are supposed to have higher charge, readily shift into the region of the field with higher drift times. They represent a ternary complex with two FGF molecules bridged by one dp10 chain (FGF₂•dp10), which is consistent with previous observations\textsuperscript{15,16}, while the remaining series of ions that are relatively inert to activation represent the 1:1 complex (FGF•dp10)(\textbf{Figure 4-6B}). The coexistence of two stoichiometries are verified by a post-IM CID for the ionic populations at m/z=2700 (\textbf{Figure 4-6C}).

The 1:1 complex FGF•dp10 at +7 and +8 were individually isolated and were followed with stepwise increased collisional voltage. The complex shows the same behaviors as FGF•dp5 (\textbf{Figure 4-S2} and \textbf{Figure 4-S3}). Briefly, GAG fragmentation occurred with protein unfolding. In the presence of dp10, protein in the lower charge state had the smallest transition energy between different folding states and the mass loss of the complex in the partially activated state was much more gentle compared to that in the fully unfolded state. Based on the mass loss profile of (FGF•dp10)+7 in the partially activated state, the energy threshold for removing non-specifically attached saccharide units is about 95V (665eV), similar to the value in FGF•dp5 complex. However, the mass of GAG remaining in the protein complex falls in the range corresponding to the tetrasaccharide (\textbf{Figure 4-S4}). We also observed a peak at m/z 2664, representing the GAG-free FGF with +6 charge, when collisional voltage was above 90V (630eV) and this charge-reduced product ion only had one major population with a CCS value as small as the native protein. Therefore,
it is very intriguing that the protein does not need to be fully unfolded in order to release the remaining GAG-FGF interactions but it can be refolded to an even smaller size only if one charge on the protein is stripped off by the leaving GAG. However, this pathway is only found in (FGF•dp10)* and the reason is still not clear.

We also investigated the FGF₂•dp10 ions under collisional voltage. Contrary to our expectation, prior to the energy reaching the threshold to trigger CID of bound dp10, FGF₂•dp10 follows a typical pathway of collision-induced protein-protein complex dissociation with asymmetric charge partitioning₁⁷: one protein monomer is rapidly unfolded and released from remaining binary complex. Therefore, this gas-phase approach may not be used for estimation of the binding motif in a heparinoid-bridged ternary complex as long as the proteins directly contact to each other.

4.4 Conclusions

This approach is a potential tool for estimating the number of sulfate and saccharide units within the binding motif of GAG. As suggested from our FGF•dp5 data, FGF binding will afford the heparinoid segment protection from the collision-induced saccharide degradation locally, while the segments not involved in the interaction will be removed. However, the performance of this approach really depends on the energy threshold for different effects induced by collision. In general, shorter heparinoid chains required lower energy to achieve the last step, where all unnecessary components of GAG are removed. A lower charge state
should be chosen for CID to avoid malice protein unfolding; otherwise no meaningful fragmentation data can be generated.

The protein unfolding intermediates can be easily monitored using IMS. However this is not a strict approach that can reveal details of the protein structure. Detailed simulations must be carried on in future, so we will know the location where the unfolding happens. Nevertheless, the approach described in this chapter allows us to monitor the status of the ions manipulated in the gas phase.
Figure 4-1. Schematic of a gas-phase “top-down” approach to identify the protein-binding of motif in unfractionated heparin
Figure 4-2. (A) Native MS of FGF alone and in the presence of synthetic dp5. (2) The GAG components involved in the FGF-binding (highlighted in purple) are predicted based on the known binding motif identified in standard dp4 (within the brackets)
Figure 4-3. CIU fingerprint graphs of (A) FGF$^{+8}$, (B) FGF$^{+7}$ and (C) FGF•dp5$^{+7}$.
Figure 4-4. Fragmentation of FGF•dp5+7 induced in different unfolding states (A) below 490 eV (70V) and (B) above 630 eV (90V).
Figure 4-5. Plots of m/z value of survived FGF•GAG complex or FGF ions versus applied activation energy.
Figure 4-6. Characterization of FGF/dp10 complexes using IMS-MS: (A) mass spectrum of FGF/dp10 mixture; (B) distinguish of overlapping populations using pre-IM activation followed IMS-MS; (C) verification of the identities of two populations with the same m/z using post-IM CID.
Figure 4-S1. Mass spectra of (A) synthetic dp5 and (B) dp10 acquired in the negative mode.
Figure 4-S2. IMS-MS of FGF alone and in presence of dp10 under different collisional voltage. Critical changes regarding the mass and conformation are labeled on the graphs.

Figure 4-S3. CIU fingerprint for FGF•dp10 at +8 and +7 charge states
Figure 4-S4. Monitoring the mass loss of FGF•dp10⁺⁷ at different collisional voltage. The last survived complex is probed and compared with the one deduced from FGF•dp5⁺⁷.
4.5 References


13. Popa, V., Trecroce, D. A., McAllister, R. G. & Konermann, L. Collision-Induced Dissociation of Electrosprayed Protein Complexes: An All-Atom Molecular


CHAPTER 5

DEVELOPING A NOVEL PLATFORM BASED ON ENZYMATIC DEGRADATION AND SIZE-EXCLUSION CHROMATOGRAPHY-MASS SPECTROMETRY FOR IDENTIFYING PROTEIN-BINDING MOTIFS IN HEPARIN-LIKE GLYCOSAMINOGLYCANS

5.1 Introduction

Heparin-like glycosaminoglycan play crucial roles in regulation of diverse functions. However deciphering the mechanisms of heparin-mediated regulation is challenging due to the tremendous structural variety of heparin chains, defined by modifications including O-sulfation, N-sulfation and N-deacetylation within a disaccharide repetitive unit. The modification is controlled enzymatically by mechanisms that are still unclear, which leads to combinatorial modification patterns presented in naturally-occurring heparin-like GAG samples. In the last chapter, we described a mass spectrometry-based approach to identify the minimal binding motif of heparin (defined by the number of sulfate group and saccharide unit within the interface), which utilizes a “top-down” idea to chop off the unessential components (including GAG chains and the sulfate groups that are not involved in the interaction) from a protein-heparin complex. Mass measurement of the surviving complex indicates the minimal binder. However, as we have shown in the last chapter, the gas-phase collision-induced fragmentation of GAG in a

§ The data reported in this chapter will be in a manuscript for publication.
meaningful way is limited to a simple system consisting of PTM-free protein and the heparin chain shorter than 10 saccharide units.

In this chapter, we introduce an alternative procedure in which the degradation of GAG component is induced in a manner of enzymatic digestion. Heparinase-based heparin degradation used to have a routine application to create low-molecular-weight heparin\textsuperscript{1-3}. We expected that heparinase would degrade every disaccharide unit along the GAG chain gradually except in the region where the client protein is interacting. In order to achieve this goal in an unbiased fashion, the family II of heparinase is chosen from all candidates for the following advantages: (1) it has the least preference to the cleavage site and can digest along the polysaccharide chain in both sulfate-rich domain and neutral domain; (2) only two saccharide units one each side of the scissile bond (P1 P2 | P1’ P2’) of GAG chain are required for initiating the enzymatic recognition\textsuperscript{4,5} so that it can access the saccharide adjacent to the region protected by client protein and “clean out” the unbound components thoroughly. In pursuit of high performance of analysis, we developed an online size-exclusion chromatography (SEC) native mass spectrometry (MS) platform, which embraces the benefits from the both techniques (Figure 5-1A). It allows the surviving protein complex to be isolated from the degraded GAG products and to subject to a real-time ion detection by MS. Native MS has been successfully used for sorting the binders from a combinatorial oligosaccharide complex upon the affinity to a protein\textsuperscript{6}. Based on our previous experience, the performance of native MS including resolution and dynamic range for detection in a heterogeneous sample can be significantly improved vis-à-vis the
incorporation of online chromatography\textsuperscript{7}. And the elution profile for individual species can be extracted and it allows us to monitor the dynamic impact on each other\textsuperscript{8}. In addition, although the digestion is carried on in the solution phase, we can still take advantage of the gas-phase chemistry to get rid of the unessential sulfate in the protein-interacting domain, which cannot be removed in the step of heparinase treatment.

In this work, we test this approach for the binding motif for antithrombin, which has been well established (Figure 5-1B) in previous studies\textsuperscript{9,10}, in the complex with heparin oligosaccharide with 20 units.

### 5.2 Experimental

Antithrombin-III (AT) dominantly consisting of the $\alpha$-form (with four glycosylation chains) was purchased from Haematologic Technologies (Essex Junction, VT). Heparin oligosaccharide dp20 was purchased from Iduron (UK). In order to form a strong complex, AT and dp20 were incubated in 1:5 molar ratio at room temperature for 30 min in the solution containing 30mM ammonium acetate at pH 7. Heparinase II was added into the mixture to a final concentration 8 u/ml and incubated for another 30 min at 35 degree. The sample was immediately transferred in to a sample vial and ready for SEC-MS analysis.

The platform of SEC-native MS is shown in Figure 5-2A. Chromatography is carried on in Agilent 1100 equipped with autosampling system. The flow rate for SEC (TSK G3000) separation is set to 0.3 ml/min. A fluid splitter was employed to reduce the flow prior to injection to the source of Bruker SolariX FT-ICR mass
spectrometer (Billerica, MA). In source CID voltage was set to 60V. Sample injection and data acquisition was triggered from computer using Hystar software. GAG-ligand can be dissociated from the protein complex by mixing with methanol solution containing 12% acidic acid. The solution was pumped into a fluid mixture by another HPLC at the same flow rate (0.3 ml/min), where the SEC fluid was mixed with the denaturing solution allowing the protein complex to be denatured and release the GAG ligand. The released GAG was then subjected to the mass spectrometer for an accurate mass measurement (Figure 5-2B). Online LC-MS data was processed using DataAnalysis software.

5.3 Results and Discussion

5.3.1 SEC-native MS analysis of the heparinase-digested products of AT•dp20 complex

The digestion mixture passed through SEC column with two major retention times displayed on the UV trace (Figure 5-3A, grey trace). The early peak (eluted at 19-22min) represents the protein-GAG protein mixture and the late peak (eluted at 28-30min) represents the pieces of disaccharide degraded from the intact GAG chain. There are some early-eluted species with very low UV absorbance (25-28min) corresponding to the GAG components with longer polysaccharide chain produced by incomplete digestion. The Base Peak Chromatogram (BPC) reflects the time when ionic signal was actually detected by MS and it shows an obvious delay (~2 min) compared to the UV profile. Upon the BPC profile, the average mass spectra within three continuous time windows in the early major peak are shown in Figure 5-3B. We clearly saw that during such a short period less than 1.5 min, three
species were eluted one after another and successfully detected by MS. The first spectrum represents AT•dp20, and perhaps other unresolved species. The second spectrum displays many peaks corresponding to the complexes formed by AT and GAG with different mass ranging from dp6 to dp10. They were produced from the partial digestion of GAG in the presence of AT. The third spectrum is only comprised of the GAG-free AT. We did not observe any AT-involved complex that contains less than 6 saccharide units during any period of elution. The order of the peak elution also agrees with chain length of the GAG binding to the protein.

Based on the spectrum with +16 ions of AT in complex with short chain oligosaccharide (i.e. AT•dp6+16, AT•dp8+16, AT•dp10+16), only one sharp and abundant peak was observed for each complex (see Figure 5-3C for a zoom-in view), corresponding to (6,6,1), (8,9,1) and (10,12,1) respectively. (In the standard nomenclature system of GAG, (X,Y,Z) represents the structure in terms of the number of saccharide (X), the number of sulfate groups (Y) and the number of acetylation groups (Z).) These numbers indicate the specific modification pattern that plays the most important roles in AT-binding. Mass accuracy of our measurement is shown in Table 5-1. The binders with other patterns either leave from the protein in the early stage or have the unessential sulfate groups to be shed in the gas-phase during in-source CID. The elution profiles extracted for the three peaks indicate that AT•(6,6,1) is the most abundant complex that survived from heparinase treatment and in-source purging (Figure 5-3D). It is well known that AT can specifically interact with a piece of GAG containing a pentamer (5,6,1). Although additional sulfate groups and saccharide units may also enhance the binding affinity,
the six sulfates presented in Figure 5-1B are most important and represent the minimal requirement to trigger strong and specific interaction, which has also been demonstrated in a crystal structure. Our results are highly consistent with what have been identified previously in terms of the numeric modification pattern. In our harsh condition, all unessential sulfate groups are shed, even some of those that may contribute to the specific or non-specific interactions. As a conclusion, the binder for a protein can be determined in a one-step measurement with high mass accuracy; while the traditional ways to identify the binder of strong affinity is very time-consuming.

5.3.2 Identification of the GAG ligand through online protein denaturation

Measuring the mass of intact complex using native MS is promising however the lack of the isotopic peaks for a highly charged ion of protein-GAG complex (e.g. AT•(6,6,1)+16 in Figure 3C) probably leads to inaccurate mass measurements. For this reason, we designed an experimental setup (illustrated in Figure 5-2B) that is able to dissociate the remaining minimal binder from the protein complex via an online denaturing. It allows the binder itself to be measured using FT-MS, which gives more accurate results compared to the measurement of non-covalent complex using native MS. First, we test whether a protein can undergo unfolding by mixing with the denaturing solution after the native protein is from SEC column. The mass spectra for the SEC peaks of BSA are shown in Figure 5-S1, where the low m/z of charge envelops of both monomeric and dimeric BSA apparently indicates that the protein has an unfolded conformation during electrospray. Then we tested whether
the unfolded AT can release a small GAG ligand, Arixtra\textsuperscript{TM} (5,8,0). It is a synthetic mimic of the pentameric binder with the same modification pattern, except for an additional sulfate group replacing the acetylation group in the naturally-occurring binder (5,6,1) or (5,7,1). The UV trace of the chromatogram is identical to the one without mixing with denaturant because the denaturation step occurs after UV detection. Only one peak is visible in the chromatogram, characterized by the AT•Arixtra complex according to its retention time (Figure 5-4A, grey trace). The identity of this peak is also verified based on the native mass spectra (Figure 5-4B, green trace), in which no free AT is observed. In contrast, no AT•Arixtra is survived in denatured MS (Figure 5-4B, blue trace), suggesting that the ligand is dissociated from the complex. During the sample preparation, Arixtra was added in excess to the sample so there is another peak eluted at \(~25\) min corresponding to the excess Arixtra, which cannot be seen in UV trace because Arixtra does not have UV absorption. The mass spectrum of Arixtra was acquired in the negative mode in order to reach the best sensitivity, in which the isotopic peaks of Arixtra can be clearly observed (Figure 5-4B, orange trace). Extracted ion chromatogram for Arixtra display two peaks (Figure 5-4A, orange trace), one (at \(~25\) min) representing the excess fraction not involved in the interaction with AT and another (at \(~20\) min) representing the fraction in AT•Arixtra complex as an attached ligand. Therefore, we can probe the identity of GAG ligands within the same time window when the protein complex is eluted using this online denaturing approach.
5.4 Conclusions

The existence of unbound AT suggests that there is a competition between AT and heparinase in terms of the binding affinity to heparin chains. Generally, for an enzyme, a higher substrate turnover rate will make the affinity for the substrate weaker ($K_M$), therefore heparinase digestion should be carried out under conditions where the enzyme has highest turnover rate, in order to avoid stripping the GAG chain off from the surface of AT. We described in the previous chapter, non-specific electrostatic interaction is remarkable in the gas-phase and likely to force the entire heparin chain to collapse on the surface of the protein, which prevent the saccharide with weak interaction from being removed. In the solution phase, this accessibility issue has been circumstance. Our data strongly suggest that the regions of dp20 that does not contribute to the strong interaction can be accessed by heparinase for cleavage.
Figure 5-1 (A) Workflow of approach based on enzymatic degradation to identify the minimal protein-binding motif for heparin-like GAG. (B) Previously identified minimal binding segment for antithrombin-III; the critical
Figure 5-2. Schematics of (A) SEC-native MS and (B) SEC-denatured MS.
Figure 5-3. SEC-MS of heparinase-treated AT-dp20 complex. (A) SEC profile of heparinase-treated AT-dp20 complex. (B) The average mass spectra for different retention times at T1: 20.4-20.8 min; T2: 20.8-21.2 min; T3: 21.2-21.6 min. (C) Zoom-in view of the mass spectrum for T2. (D) Elution profile for four species detected by SEC-MS.
Figure 5-4. Online SEC-denatured MS tested with the AT-Arixtra complex: (A) UV trace (grey) and extracted ion profile of AT (blue) and Arixtra (orange) extracted based on the peaks in denatured MS. (B) Average mass spectrum under native condition (green) and denatured condition (blue) within the chromatogram peak at T1 and under denatured condition with the chromatogram peak at T2 (orange).
Figure 5-S1. Online SEC-denatured MS of BSA: (A) UV and BPC traces of SEC chromatogram of BSA, dimer and monomer peaks are indicated with arrows; (B) average mass spectra of BSA dimer (top) and monomer (bottom) under denatured condition (blue) and monomer under native condition (green).

Table 5-1. The measured mass and calculated average molecular weight at major peaks in SEC-MS of AT•dp20 degraded complex

<table>
<thead>
<tr>
<th>Unit: Da</th>
<th>(6,6,1)</th>
<th>(8,9,1)</th>
<th>(10,12,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>measured</td>
<td>1534.4</td>
<td>2111.4</td>
<td>2689.3</td>
</tr>
<tr>
<td>calculated</td>
<td>1534.2</td>
<td>2111.7</td>
<td>2689.1</td>
</tr>
</tbody>
</table>
5.5 References


6.1 Summary

Heparinoid molecules bind to heparin-binding proteins through both the long-distance electrostatic attraction between opposite charges and the short-distance interactions including hydrogen bonding and salt-bridge bond. Native MS is an excellent tool to probe these non-covalent interactions. We developed multiple strategies particularly for this system based on the special properties of heparin-protein complexes. First, protein and heparin bind more tightly when they are introduced to the gas phase, so we can take advantage of the strong electrostatic interaction to carry out lossless ion manipulations in the gas phase. For examples, we selectively chopped off the unbound GAG component using collision-induced activation for identifying the number of saccharides and sulfate groups that reside on the binding interface (Chapter 4); and we also performed limited charge reduction to reconstruct the mass and charge values for protein-heparin complex ions whose binding stoichiometry does not change during reaction with the electron donors (Chapter 2 and 3). Second, the structure of gaseous ions introduced by native electrospray is largely similar to that in the solution and allow us to study the conformation of protein complex in the gas-phase using IM-MS. With CCS measurement, for examples, we were able to capture a near native structure of FGF-oligosaccharide complex and other two intermediate states created by collision-induced unfolding (Chapter 4); and we also observed two distinct models for
protamine-oligosaccharide complexes (Chapter 3) based on the charge-reduced populations. Third, the presence of GAG chains brings about high degree of heterogeneity in the complexes and it triggers the development of “top-down” strategies to fetch the unique elements contributed to the binding affinity. Technically, to reduce the heterogeneity, we built multiple two-dimensional platforms by incorporating a separation tool with MS detection such as quadruple isolation (Chapter 2 and 3), online chromatography (Chapter 5) and ion-mobility separation (Chapter 2,3,4).

It should be noted that all strategies we describe in previous chapters are not limited to the system of heparin-protein complexes; it has potential to be applied to other analytes with polydispersed mass distribution. Online SEC-MS and IXC-MS are powerful tools for quantitatively analysis of non-covalent protein-protein complexes with multiple stoichiometries such as the complexes formed by monoclonal antibody and antigens or Fc receptors. Limited charge reduction can be used for interpreting unresolved spectra (MS or IM-MS) of samples containing proteins with heterogeneous modifications (e.g. antibody drug conjugation, PEGylated proteins). Our results in Chapter 2 through Chapter 5 provide valuable data and examples showing how those strategies can work as novel solutions to the heterogeneity-related problems.

6.2 Future directions

Beyond my dissertation studies described in previous chapters, there are three major directions we can explore in the future. Because CCS is an average
projection area of a molecule at all orientations, it can reflect anisotropy to some extent. As a result, the measured CCS values for a protein (e.g. FGF, AT and protamine) in the presence of GAG chain are correlated to the structure of the gaseous ions. However the specific structure of a protein complex is still not clear. In future, we will perform molecular dynamic (MD) programs such as GROMAC, VMD to simulate the structures during protein aggregation or protein unfolding under collisional activation over a period of time. We will also extract CCS values from several snapshots obtained from MD and find the one that is closest to the experimental CCS. This approach can eventually help us to understand what domain of FGF is unfolded in the partially activated state and to confirm that the binding interface is not damaged in this state.

In Chapter 4 and 5, we only focus on identifying the binding motif on heparinoid chain. We are also interested in mapping the heparin-binding epitope on the protein using MS-based footprinting such as hydrogen/deuterium exchange (HDX), covalent labeling or cross-linking. HDX-MS has been a well-established technique to probe the protein-ligand interface assuming the ligand can change the solvent-exposed surface of the protein. We hypothesize that the region of protein that binds to heparin is less solvent-exposed in the presence of heparin and shows a lower deuterium uptake. This approach has never been reported by others so we can test it with well-studied heparin-binding proteins such as FGF and AT.

At last, as shown in Chapter 5, the degradation of heparin using heparinase allows us to generate the minimal segment of GAG that is bound to AT. We also demonstrate that analyzing the saccharide units dissociated from the complex
through protein denaturation can achieve better mass accuracy compared to analyzing the mass of AT•GAG complex using native MS; the latter requires higher mass range and more delicate desolvation condition. In future, we will develop an approach to achieve this goal with higher controllability. First, we will immobilize the protein on beads and pack them into a preparation column. Then we will load a mixture of heparinoid molecules and wash with the equilibration buffer. Next, we will wash the column with the solution containing proper concentration of salt to get rid of the nonspecifically bound species. Heparinase digestion step will be carried out by directly injecting the enzyme into the column and leave at 33°C for adequate time. Then we will wash out the degraded unbound products with the washing buffer and elute the tightly bound species by increasing the concentration of salt in the mobile phase. The eluted fraction can be analyzed using FT-MS after desalting and lyophilizing the sample. We can test this approach with AT because immobilized AT is a matured technique for screening heparin fraction with high anticoagulant effect. The in-column digestion is also a routine approach that has been performed in the process of GST-fusion protein purification. Therefore we assume that this approach we propose here is feasible and may give us some results with higher accuracy and less bias.
APPENDIX

CCS CALIBRATION CURVES

CCS calibration was carried out following the standard protocol that can be found in *Nature Protocols* 3, 1139 - 1152 (2008). The curves are plotted using reported CCS of standard proteins (y-axis) vs. the corrected drift time measured in the experimental condition (t”, x-axis).

Standard proteins were selected based on the required mass range of the sample to be analyzed. For the protamine-do20 experiment (see Chapter3), the standard proteins are Cytochrome C and monomeric β-Lactoglobulin. For the FGF•dp5/dp10 experiments the standard proteins are Cytochrome C, β-Lactoglobulin, BSA and ADH. All standard proteins were analysis under native conditions.

![CCS calibration curve](image)

**Figure APPX-1. CCS calibration curves for protamine-dp20 experiment**
Figure APPX-2. CCS calibration curves for FGF•dp10 complex

\[ y = 6.7904x - 1.0426 \]

\[ R^2 = 0.99917 \]
BIBLIOGRAPHY


Greinacher, A., Alban, S., Omer-Adam, M. A., Weitschies, W. & Warkentin, T. E. Heparin-induced thrombocytopenia: a stoichiometry-based model to explain the differing immunogenicities of unfractionated heparin, low-molecular-


Rhomberg, A. J., Ernst, S., Sasisekharan, R. & Biemann, K. Mass spectrometric and capillary electrophoretic investigation of the enzymatic degradation of

Richard, B., Swanson, R. & Olson, S. T. The signature 3-O-sulfo group of the anticoagulant heparin sequence is critical for heparin binding to antithrombin but is not required for allosteric activation. The Journal of biological chemistry 284, 27054-27064 (2009).


