Washington University in St. Louis

Washington University Open Scholarship

Mass Spectrometry-Based Strategies for Biomolecular Structure Analysis: Steroid Metabolites Structures and Protein-Protein Interactions

Yuetian Yan
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Part of the Chemistry Commons

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Chemistry

Dissertation Examination Committee:
Michael L. Gross, Chair
Timothy E. Holy, Co-Chair
Gaya K. Amarasinghe
Kevin Moeller
John-Stephen Taylor

Mass Spectrometry-Based Strategies for Biomolecular Structure Analysis: Steroid Metabolites
Structures and Protein-Protein Interactions
by
Yuetian Yan

A dissertation presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2015
St. Louis, Missouri

Table of Contents

List of Figures

Chapter 5

Chapter 6

Chapter 7

List of Tables

List of Schemes

Acknowledgments

First and foremost, I'd like to thank my PhD advisors, Professors Michael L. Gross and Timothy E. Holy, for their support and guidance during my five years of thesis research. Particularly, having spent most of my time in the mass spectrometry lab led by Dr. Gross, I truly appreciate the experience of getting trained in an outstanding research environment with many erudite scientists and hard-working colleagues. It is a culture worth emulating.

I thank my collaborators, Dr. Xiaoyan Fu from Department of Anatomy and Neurobiology, Washington University; Dr. Andy Wowor and Prof. James Cole from Department of Chemistry, and Prof. Debra Kendall from Department of Pharmaceutical Sciences, University of Connecticut; Dr. Adrienne Tymiak, Dr. Guodong Chen, Dr. Richard Huang, and Dr. Hui Wei from PCO BDAS, Bristol-Myers Squibb, for their collaborations on interesting research projects.

I thank the entire staff scientists in our facility, especially Don L. Rempel and Henry W. Rohrs for their assistance in data modeling and instrument troubleshooting. All of them provided technical instructions to me during my PhD training.

I thank all my colleagues, both current and past (from 2009 to 2015), in our Mass Spectrometry Resource for their support and friendship.

I thank my Dissertation Advisory and Examination Committee members, Prof. Kevin Moeller, Prof. John-Stephen Taylor, and Prof Gaya Amarasinghe for their guidance in my thesis research.

Finally, I thank my parents, Wenxin Yan and Meihong Shen, for her support to my chosen research career.

<div align="right">Yuetian Yan</div>

Washington University in St. Louis

May 2015

Abstract of the Dissertation

Mass Spectrometry-Based Strategies for Biomolecular Structure Analysis: Steroid Metabolites
Structures and Protein-Protein Interactions
by
Yuetian Yan
Doctor of Philosophy in Chemistry
Washington University in St. Louis, 2015
Professor Michael L. Gross, Chair
Professor Timothy E. Holy, Co-Chair

 Mass spectrometry is an important method for studying the structure of both small molecules and large biomolecules (e.g., proteins). The majority of the applications prior to 1970 were focused on small molecules, owing to the limited ionization methods which posed difficulties in producing gas-phase ions for large biomolecules then. Beginning in the 1980's, with the introduction of new ionization methods (ESI and MALDI), the applications have gradually switched to biological science measuring large bioorganic molecules. Today, with the developing interest in metabolomics and proteomics, and ongoing improvement in MS-based techniques, mass spectrometry is extensively applied in the study of both small and large molecules.

The research presented in this thesis falls into two main parts, which focus on the application of MS in (1) structural analysis of steroid metabolites and (2) characterization of protein-protein interactions. In the first part, combinations of different MS methods are adopted and used to solve the structures of unknown steroid metabolites, which are the pheromones responsible for mouse communication in mouse urine. This part includes three chapters, the first two of which discuss the method development of using MS to study the structure of steroid metabolites; and the third chapter presents the application of the MS methods in solving a newly discovered steroid pheromone, which is determined as a sex-specific hormone. In the second part, two MS-

based strategies, namely, hydrogen-deuterium exchange (HDX) and fast photochemical oxidation of proteins (FPOP), are applied in two studies of protein-protein interactions, including: (1) dimerization of SecA, which is a motor protein in bacteria translocation pathway; and (2) interface mapping of EGFR binding to Adnectin1. In the first chapter in Part 2, we used HDX MS to characterize the dimer interface of SecA, and, meanwhile, detected a conformational change from open to closed forms at the pre-protein binding domain upon dimerization. This conformational change provided leads for the active form of SecA. In the second chapter in Part 2, we applied FPOP, which is modified to suit therapeutic protein formulation conditions, to map the epitope of Adnectin1-EGFR interaction at amino acid residue level. The epitope identified agrees with that from both HDX study and crystallography results, presenting more evidence of the capability of FPOP in epitope mapping.

These five studies on characterization of steroid metabolites and protein-protein interactions show the successful application of mass spectrometry in the structural study of both small molecules and large proteins. Furthermore, there's a great potential for study of more complex systems.

Chapter 1

Introduction: Mass Spectrometry Assists Structural Analysis of Steroid Metabolites Important in Mouse Communication

1.1 Abstract

Among mice, social behavior, such as sex, social status, and identity is organized via pheromones and other odor cues in chemical signaling. The molecular identity of these signals, however, remains largely unknown. Recently, a group of steroid conjugates in mouse urine was found to serve as important ligands for the pheromone-sensing neurons. Although two of those steroid metabolites were structurally analyzed with NMR and mass spectrometry, many more remains structurally unresolved. These steroid metabolites, which occur in trace amounts in mouse urine and elsewhere and, therefore, are difficult for NMR study, are challenging to characterize. In the first part of this thesis, we explored the application of a combination of MS tools for determining structure of steroid metabolites. In the area of steroid conjugates analysis, mass spectrometry dates back to the 1950s, first using electron ionization (EI). With further development of mass spectrometry, other suitable methods for structure characterization became available for the study of steroid conjugates, including different types of tandem mass spectrometry. In this chapter, we give an introduction to the techniques and instrumentation in mass spectrometry that can be applied in structural study of steroid metabolites; we also review the history of steroid conjugates study by mass spectrometry.

1.2 Social Communication of the Mouse

Animals are social. They use their sense of smell to identify and learn the status of other animals. Natural cues such as urine-containing metabolites inform about reproductive status, social dominance, and levels of stress[1]. By investigating various bodily fluids, such as urine, an animal can discriminate between individuals and "communicate".

For many mammals, olfactory sensations derive from at least two anatomically and molecularly distinct neural pathways, the main and accessory olfactory systems. The main olfactory system is composed of main olfactory epithelium (MOE) that contains olfactory sensory neurons (OSNs). Chemical cues are detected in the MOE after they bind to G protein-coupled receptors on the cilia of OSNs. For a long time, the main olfactory system was believed to utilize only volatile chemical cues and trigger innate physiological or behavior responses[2-4]. Recently, evidence appeared showing that murine OSNs can also respond to small, nonvolatile major histocompatibility complex (MHC) peptide ligands that lead to changes in behavior[5]. Similarly, the accessory olfactory system is composed of the vomeronasal organ (VNO) epithelium, which contains vomeronasal sensory neurons (VSNs). Odorants that are dissolved in nasal mucous gain access to the VNO via an active pumping mechanism[6] and then bind to the G protein-coupled receptors located on VSNs. Although it was presumed that the VNO preferentially responds to large-molecular-weight, nonvolatile molecules, it has also been shown to respond to volatile chemosensory stimuli in vitro[7, 8].

Unlike the main olfactory system, the accessory olfactory system plays a more narrowly specialized role for social communication. Previous studies showed that VNO lesions lead to deficits in reproductive and territorial behaviors in mice[9-13]. More than two decades ago, Wysocki and his colleagues studied the importance of VNO by using a dental burr to enlarge the

incisive foramen and then remove the entire VNO (VNX)[14]. Many following-up studies[9] have been published from the same group, reporting VNX impaired abilities including ultrasonic vocalizations in males, hormone surges in response to cues from females, male-mouse courtship and sexual behavior, territorial marking and inter-male aggression, and maternal behaviors in females. More recently, several other groups genetically mutated mice to form a homozygous deficiency in Trp2, a cation channel expressed in the VNO. These Trp2 mutants reveal a striking reduction in the electrophysiological response to pheromones in the VNO, and, therefore afford the opportunity to examine the role of VNO in the generation of innate social behaviors in mice. The studies[10, 11, 13] showed that Trp2 mutant males and nursing females fail to display aggression on intruder males, Trp3 mutant female showed reduction in lactating behavior, and trp2 mutant male also lost sex discrimination and initiated sexual behaviors toward both males and females. All these findings suggest that VNO-mediated pheromone input regulates important sexual and social behaviors in mice.

1.3 Vomeronasal Agonists Carry Pheromone Information

Pheromones are chemical cues that are released by animals to regulate their social interactions within the same species. Although there are many known ligands found for main olfactory sensory neurons, only a small number of agonists carrying pheromone information for VSNs have been identified thus far, in contrast to the many chemosensory receptors found in the rodent nose. Behaviorally, the most well characterized source of VSN ligand is urine, because mice employ urine investigation as a tool to discriminate between individuals and urine also comes available in relatively large quantities. Among the known chemicals carrying pheromonal information, several volatile molecules[8, 15] were identified to be VSN agonists. In addition, nonvolatile major histocompatibility complex (MHC) class I peptides[16] and major urinary

proteins (MUPs)[17] have been reported to activate selectively VSNs. Altogether, these identified vomeronasal agonists are far less numerous than the amount of chemosensory receptors found in the rodent nose (more than 1300[18]). Also, these molecules at saturating concentrations were reported together to stimulate fewer than 15% of the neurons, whereas hundred-fold diluted urine stimulated 30-40% of the neural population[7]. Both facts indicate that there are more compounds that are responsible for most of the activity in mouse urine unidentified.

In an effort to search for the unidentified source responsible for most of the activity in dilute urine, Nodari et al.[19] adopted a de novo screen of active constituents of mice urine by presenting chromatographic fractions of urine to VSNs and recording their neuronal firing responses. As a result, a large family of compounds that collectively accounts for much of the activity in female mouse urine was identified, most of which contain a sulfate group and two are structurally characterized as sulfated steroids[20] by using nuclear magnetic resonance (NMR) and mass spectrometry (MS). Indeed, several lines of evidence suggest that urine contains other active, but unknown, sulfated steroids. In fact, the urine from female mice of the BALB/c strain, contains sulfated compounds estimated to produce 80% of the total activation of these sensory neurons[19].

Unfortunately, like many metabolites, many of the unknown cues in natural stimuli are present in very low abundance, restricting the possibility of using NMR analysis for structure identification. Therefore, more sensitive methods are needed, and the first part of this thesis describes the exploration of various combinations of MS methods to determine steroid-metabolite structure, and demonstrates its successful application to solve one unknown and putative female sex pheromone.

1.4 Mass Spectrometry

Mass spectrometry (MS) is a microanalytical technique that can be used to determine the elemental and isotopic signature of an analyte and provide some aspects of its molecular structure. The simplest mass spectrometer is composed of an ion source, a mass analyzer, and a detector. In a typical MS procedure, a sample, which might be solid, liquid, or gas, is ionized to form ions in gas phase before they are separated according to their mass-to-charge ratios (m/z). The majority of applications prior to 1970 were analysis of small molecules, because only analytes having significant vapor pressure could produce gas-phase ions by the only commonly available ionization techniques at that time, electron ionization (EI) and chemical ionization (CI). After 1970, the capabilities of MS were expanded by the development of desorption/ionization (D/I) techniques, which enables generating gas-phase ions directly from a condensed phase. However, it was not until the late 1980's, when John Fenn invented electrospray ionization (ESI)[21], Fanz Hillenkamp and Michael Karas developed matrix-assisted laser desorption ionization (MALDI)[22], that MS become vastly applied in biological science for measuring bioorganic molecules, including peptides, proteins, carbohydrates, and DNA or RNA fragments.

Today, EI and CI continue to play very important roles when combined with gas chromatography (GC/MS), whereas ESI and MALDI are mainly responsible for the dominant role of MS in the biological sciences. GC/MS now is a mature technique in analysis of small molecules, supported by several large online databases. ESI and MALDI are even more routinely used in both small and large molecule analysis owing to easy sample handling and the requirements for only small sample quantity.

Besides the availability of various ionization techniques, the success of MS in science today is also attributable to combining different mass analyzers to afford hybrid mass

spectrometers that provide high-mass resolving power measurements. In conducting this thesis research, three hybrid mass spectrometers were mainly used; they all involve ESI (including nanospray) combined with linear ion trap/Orbitrap, GC-quadrupole/time-of-flight (TOF), and MALDI-TOF/TOF. These are described in the following sections.

1.4.1 Quadrupole/Time-of-Flight Mass Spectrometers

1.4.1.1 Quadrupole Mass Analyzer

As indicated by its name, a quadrupole mass analyzer (sometimes called quadrupole mass filter or QMF) is composed of four parallel rods that ideally form a hyperbolic cross section. Each opposing rod pair is connected together and held at the same polarity. A radio frequency (RF) and direct current (DC) voltage are applied between one pair of rods and the other. Extracted ions from an ion source travel down the quadrupole between the rods along the longitudinal axis toward the detector. Only ions of a certain *m/z* will reach the detector for a given U (magnitude of the DC potential) and V (magnitude of the RF signal), given a constant AC frequency, and other ions will collide with the rods owing to unstable trajectories. This is mathematically modeled with the use of the Mathieu differential equation[23]. And this property of the QMF enables a quadrupole either to function as a mass selector in which only ions with specific m/z can be transmitted, or to scan a mass spectrum by increasing or decreasing U and V at a fixed ratio from a pre-established minimum to a maximum value.

Selective transmission of ions over a narrow *m/z* by the quadrupole is most often used in tandem MS (MS/MS) analysis mode to select a precursor ion for study. One example instrument designed for such analysis is a triple quadrupole mass spectrometer, which is composed of a linear set of three quadrupoles in series. The first (Q_1) and third (Q_3) quadrupoles act as mass filters, and the second (Q_2) quadrupole is operated in the RF-only mode (transmits ions of all *m/z*
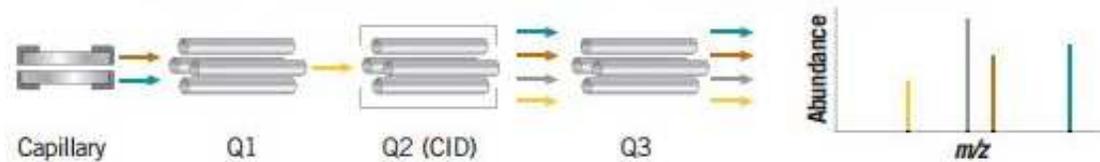
Figure 1- 1. A schematic of a triple quadrupole mass spectrometer (QQQ-MS). Q1 and Q3 represent two mass filters for precursor and fragment ion selection while Q2 (collision cell) creates fragment ions via collision-induced dissociation (CID). Taken from http://www.particlesciences.com/news/technical-briefs/2009/mass-spectrometry-bioanalysis.html.

values) to serve as a collision cell[24, 25]. Q_1 is used for selection of the precursor ions and transmitting them to Q_2, where collisions of the precursor ions with an appropriate concentration of collision gas (Ar, He, or N_2) occurs, and product ions are formed. These product ions are passed to Q_3 where they may be fully scanned to give a mass spectrum or transmitted over a narrow range to achieve "selected reaction monitoring" whereby duty cycle and sensitivity are increased (Figure 1-1).

In modern mass spectrometry, quadrupoles are very often used in hybrid mass spectrometers. The tandem quadrupole-TOF (QTOF) mass spectrometer uses a time-of-flight device as the second mass analyzer and has the advantage of improved mass accuracy to the nearest 0.1 millimass unit vs. using a quadrupole for the second mass analyzer.

1.4.1.2 Time-of-flight Mass Spectrometry

Time-of-flight (TOF) mass spectrometry is a method that separate and detect ions of different *m/z* in a field-free drift path after the application of a constant acceleration voltage (U) for all ions. After the initial acceleration with a high electrical potential, we can equate two types of energies as:

$E = qU = \frac{1}{2} mv^2 = E_k$ Eq. 1

The velocity (*v*) of the ion in the flight tube, therefore, can be obtained.

$$v = \sqrt{\frac{2qU}{m}} \qquad\qquad \text{Eq. 2}$$

Because the field-free drift tube is at a fixed length (*l*), the time that an ion takes to reach the detector depends on its velocity, and, therefore, its mass-to-charge ratio. By measuring this time and using the known experimental parameters, the *m/z* of the ion can be determined from the following equation deduced from Eq. 2:

$$\frac{m}{z} = \frac{2U}{v^2} = \frac{2U}{(l/t)^2} \qquad\qquad \text{Eq. 3}$$

If the TOF is coupled with a reflectron, the kinetic energy dispersion in the direction of ion flight can be corrected, and as a result the mass resolving power is significantly improved. A modern Q-TOF instrument can have mass resolving power of 60,000 over a wide mass range.

1.4.2 Linear Ion Trap/Orbitrap Mass Spectrometers

1.4.2.1 Linear Ion Trap

The linear ion trap, or linear quadrupole ion trap (LIT or LTQ, Figure 1-2a), uses a set of quadrupole rods to confine ions radially (by employing an RF potential) and a static electrical potential applied to end electrodes to confine the ions axially[26]. A common instrument from Thermo Fisher, the LTQ,
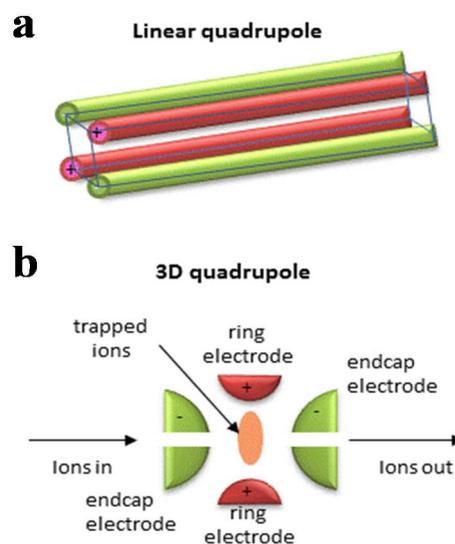
Figure 1- 2. a. The linear ion trap. b. The chamber of the 3D quadruple trap. Taken from http://employees.csbsju.edu/hjakubowski/classes/ch331/protstructure/olcompseqconform.html.

makes use of a transmission quadrupole that can be used as a selective mass filter, but unlike a transmission quadrupole, it can also function as an actual trap by creating a longitudinal potential well for the ions[27]. The LTQ has the advantages of larger ion storage capacity than the 3D quadrupole trap (Figure 1-2b). These properties allow LTQ to exceed the triple quadrupole in MS/MS analysis to achieve a better signal-to-noise ratio by accumulating ions in MS^2 before acquiring a mass spectrum of the product ions. Furthermore, this trapping property of the LTQ makes MS^3 or further possible without further complex construction, but this is also true of 3D traps. This advantage in tandem mass spectrometry is especially advantageous in the structural analysis of small molecules, which will be the focus of the first part of my thesis.

1.4.2.2 Orbitrap Mass Spectrometry

The orbitrap, which is a modification based on the Kingdon trap[28], is a type of Fourier transform mass spectrometer. The orbitrap is composed of two coaxial axisymmetric electrodes, the inner spindle-shaped electrode and an outer barrel-shaped surface (Figure 1-3). A constant electric potential is imposed between the two electrodes, and, therefore, ions can be trapped in the space between them by balancing the electrostatic attraction to the inner electrode by the centrifugal force of the spinning ions. Furthermore, the opposing surfaces of the axisymmetric coaxial electrodes are nonparallel, leading to varying electric field between the two
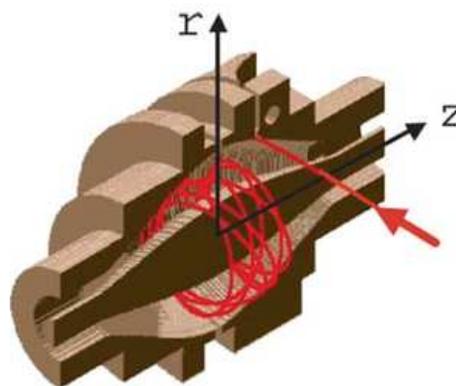
Figure 1- 3. Cutaway view of the Orbitrap mass analyzer. Ions are injected into the Orbitrap at the point indicated by the red arrow. Taken from Perry et al. 2008. Mass Spectrometry Reviews. Page 664.

10

surfaces along the z-axis, reaching a minimum at the center of the orbitrap. The electric-field vectors are nonparallel, with decreasing projection of the electric field onto the z-axis toward the center of the trap, which is the force driving the ions toward the center of the trap along the z-axis. Therefore, besides cycling around the inner electrode, the ions also have a natural tendency to oscillate axially (along the z-axis) in the orbitrap. Owing to properties of quadro-logarithmic potential[29], this axial motion is harmonic, with its angular frequency dependent on its *m/z*:

$$\omega_z = \sqrt{k\frac{q}{m}} \qquad \text{Eq. 4}$$

where *k* is the force constant/ field curvature. This oscillation feature allows the orbitrap to function as a mass spectrometer, and the ion signal is detected as a time-domain signal by using image current detection as sensed by the halves of the barrel electrode; the image current is then Fourier transformed to frequency-based signal. Image current detection is also at the heart of FT ICR mass spectrometers, and was first utilized with that instrumentation over 40 years ago.

The orbitrap analyzer can be interfaced to a linear ion trap, which together provide very high resolving power (100,000 at *m/z* 400) and efficient MS/MS capability. Therefore, much of the research reported here in this thesis (both the first and second parts) was conducted with LTQ-Orbitrap interfaced to capillary HPLC.

1.5 Gas-Phase Fragmentation

Mass Spectrometry studies its subjects in their ionic forms in the gas phase. In the ion source, an atom or molecule is converted into an ion in the gas phase by adding or removing charged particles, such as electrons or protons. The neutral substance is either admitted to the gas phase prior to or accompanying ionization. In Electron ionization (EI) and chemical ionization (CI), the excess energy introduced upon ionization almost always leads to fragmentation of the molecule.

This fragmentation pattern can be viewed as a "chemical fingerprint" and is useful in the study of molecular structure. In other soft ionization processes, such as electrospray ionization (ESI), very often only molecular ions are formed. In such situations, the molecular ions can be purposefully fragmented by collision-induced-dissociation with collision gas in a collision cell, which is built before the mass analyzer, to help understand the structure. Such gas-phase fragmentation remains very important and widely used in modern mass spectrometry for study of both small molecules and large biomolecules.

1.5.1 Electron Ionization

Electron ionization (EI) is a widely used ionization method in mass spectrometry that uses energetic electrons to interact with gas-phase molecules to produce ions. In an EI ion source, electrons are produced by electrically heating a thin metal filament to a temperature at which thermionic emission occurs. The electrons are then accelerated to 70 eV, concentrated into a beam, and directed toward the sample gas which is introduced in the ion source perpendicular to the beam. The electric field around the sample neutral molecules undergoes large fluctuation because of the close passage of highly energetic electrons, and, therefore, causes the sample molecules to ionize and fragment. The following gas phase reaction describes the ionization process.

$$M + e^- \rightarrow M^{+\bullet} + 2e^- \qquad \text{Eq. 5}$$
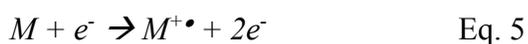
The ionization process is often followed by predictable fragmentation. The fragmentation pattern of the molecular ion can be used to deduce the structural features of the analyte. Knowledge in organic chemistry, such as electron-triggered cleavages and rearrangements and chemical bond stability, are used to suggest or to rationalize the fragmentation pathways. Many analytes have a unique "chemical fingerprint", therefore EI has been widely used in structure studies of natural

products and, more recently, in metabolism. Furthermore, there are commercially available libraries containing hundreds of thousands of reference EI mass spectra which can be used to facilitate identification of unknown compounds.

Today, EI is mostly used in GC-MS instruments because it is most easily adapted to gas-phase sample ionization. In the study of small molecule structure characterization, we used one such instrument to show the application of EI in the structural study of unknown steroid metabolites and its value in structure elucidation.

1.5.2 Tandem Mass Spectrometry

Desorption/ionization (D/I) has greatly advanced mass spectrometry in several application areas, the most important of which is the identification of large bioorganic molecules. Now electrospray ionization (ESI) and matrix-assisted laser desorption/ionization coupled mass spectrometers are the two most popular instruments in a mass spectrometry lab. These ionization methods produce protonated molecules ($[M + H]^+$) in the positive-ion mode and deprotonated molecules ($[M - H]^-$) in the negative-ion mode of the instrument, or in some cases cationized/decationized molecules. Most of the ions formed by such soft ionization techniques are even-electron species and, therefore, do not undergo fragmentation when formed by soft ionization. Thus, although they are useful in providing the molecular mass of the analyte, they are also disappointing in not providing much, if any, structural information. This problem is solved by combining tandem mass spectrometry with the soft ionization technique, like ESI or MALDI.

There are many methods used to fragment ions, and they result in different types of fragmentation. In the following, we will introduce the two tandem mass spectrometry methods

that were used in the research presented in this thesis; they are low-energy collision-induced dissociation and high-energy charge-remote fragmentation.

1.5.2.1 Collision-induced Dissociation

Collision-induced dissociation (CID) is most often what is being used in a tandem mass spectrometry experiment. It is achieved by introducing neutral target molecules or atoms into the collision cell and inducing collisions with the accelerated analyte ions in the gas phase. As a result, energy is added to the ions as internal energy that leads to subsequent bond breakage and fragmentation of the ions.

CID of Peptides and Proteins

The ability to generate fragment ions in tandem mass spectrometry (or MS/MS) is very important for the identification of an analyte with high confidence. MS/MS is now widely used in protein mass spectrometry, and is almost indispensable for identification of
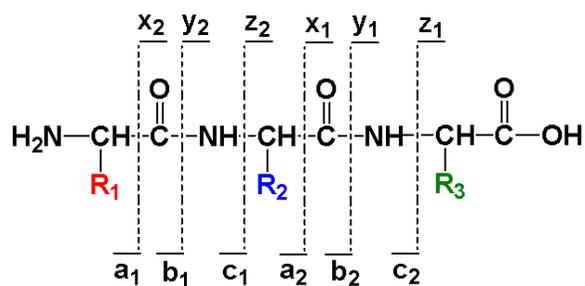
Figure 1- 4. The fragmentation sites of peptide during gas-phase fragmentation.

peptides and proteins. Peptide ions that are fragmented with CID produce a series of product ions, namely b ions if the charge is retained on the N-terminus and y ions if the charge is retained on the C-terminus (Figure 1-4). Other fragment ions, indicated by a, c, x, or z in Figure 1-4, can also be formed in gas-phase with other fragmentation methods, but these methods were not utilized in this thesis. This application of CID MS/MS in peptide identification is important in the second part of this thesis.

CID of Small Molecules

14

The application of CID for small molecules, especially in structure identification, however, remains limited. This is partly because structural chemists turn to NMR in structure studies in cases where sample quantity is not limited. More importantly, structural elucidation with CID is complicated and sometimes incomplete. Although some basic theories and rules of fragmentation have been proposed, the deduction of fragmentation remains difficult when facing complicated molecules including steroids. Despite such difficulties in data analysis, CID mass spectrometry is an important tool in studying complex biological samples including biological fluids. A subject for our study in this thesis, mouse urine, is a challenging sample from which fractionating and obtaining sufficient analyte for NMR analysis is most times difficult. In the first part of this thesis, we focused on MS/MS and MS/MS/MS of steroid sulfates to understand their general fragmentation patterns, and discovered several fragmentation rules that we applied to the characterization of one unknown steroid metabolite.

1.5.2.2 Charge-remote Fragmentation

Charge-remote fragmentation (CRF) is a class of gas-phase decompositions that occurs remote from the charge site[30-32]. The two important criteria that enable charge-remote fragmentation are (1) a charge fixed at one end of the molecule and (2) high-energy collisional activation (> 10,000 eV in the laboratory frame). The ionization method is not a prerequisite for CRFs, although the mostly frequently used ionization method for generating precursor ions was fast atom bombardment (FAB). Provided sufficient energy is deposited on a molecule, other complementary ionization methods can also be utilized to ionize samples prior to CRF.

Although it is generally agreed that CRF proceeds similarly to gas-phase thermal reactions and the charge is not involved in the fragmentation, the detailed mechanisms and energetic are still debated. For fatty acids, charge-remote cleavage of C-C bonds can be explained by a 1,4-H_2

elimination mechanism (Scheme 1-1[33]). On the other hand, a radical mechanism initialized by a homolytic C-C bond cleavage may also be operative (Scheme 1-2[34, 35]). Other mechanisms[36, 37] were also proposed later. However, none of them can explain all the data on CRF. Charge-remote fragmentation has been proven useful in the structure characterization of long-chain or poly-ring molecules[38-42]. Early CRF experiments were performed mostly on sector instruments that are capable of providing high collision energies. However, there has been a marked decline in the application of high-energy CID with the demise of magnet sector instruments and the current dominance of ion traps, QTOFs, and hybrid ion-trap FT spectrometers that employ low-energy CID. Recently, interest resumed in CRF with the introduction of TOF/TOF instruments that provide high collision energies[43-46]. The introduction of a first TOF and an accelerating voltage before the collision cell in the design of the TOF/TOF instruments enabled precursor ions to gain a high kinetic energy, which is usually between 1 to 20 kV (e.g., 20 kV accelerating voltage equals to 20 keV kinetic energy for a singly charged species). As a result, the precursor ions have high kinetic energy and high-energy collision is allowed with neutral gases when entering the collision cell. As part of the research in the first part of the thesis, we showed the application of MALDI TOF/TOF instrumentation to induce CRF and
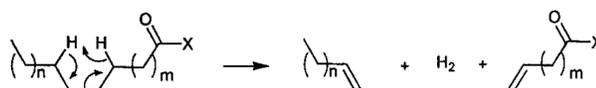
Scheme 1- 1. 1,4-H_2 elimination mechanism proposed by Gross and co-workers. X may correspond to O^- or OLi_2^+.
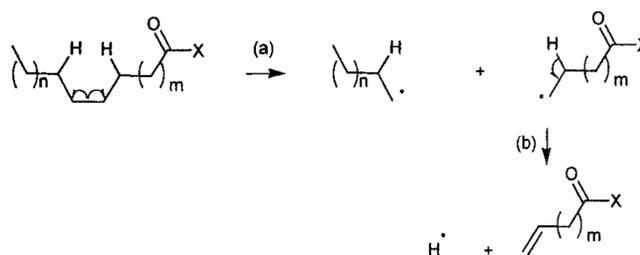
Scheme 1- 2. Homolytic bond-fragmentation mechanism. X may correspond to O^- or OLi_2^+.

enable structural characterization of steroid metabolites.

1.6 MS-Based Structural Analysis of Steroid Conjugates

Biological MS dates back to the 1950s, when steroid was one of the first class of biomolecules analyzed by MS[47]. In the early studies of steroids, EI was the dominant ionization method and later was coupled with GC. GC/MS offers very good chromatographic separation of analytes even those with minor structural differences (e.g., isomers). Meanwhile, EI produces informative and often characteristic fragmentation, therefore the structural information is forthcoming. However, for the majority of the steroid analysis, derivatization is required to form volatile compounds to be analyzed by GC[48]. Because the sample injection volume (1-5 μL typically) is small, intense analyte cleanup and concentration prior to analysis are often required to provide sufficient intensity of analyte signals for detection. EI spectra are often complicated, and made even more so with derivatization that leads to more fragments of the derivative itself. Despite the difficulties, GC/MS is still routinely used in structure analysis, partly due to the availability of a vast EI online data base for structural identification of analytes.

The introduction of the desorption-ionization techniques of fast atom bombardment[49, 50] and ESI[51, 52] bypassed the problems of analyte volatility and thermal stability, and allowed direct coupling with liquid chromatography (LC) for analyte separation. The development of tandem mass spectrometry (MS/MS)[53, 54] about the same time (1980a) afforded a new approach for structural characterizations of steroid conjugates with mass spectrometry. This approach used FAB ionization on magnetic-sector instruments and high collision energies to generate charge-remote fragmentation. The approach was demonstrated effective in resolving structures of bile salt conjugates and other steroids[40-42, 55]. Today, we are using direct-inlet ESI rather than FAB, and sector instruments have been replaced with ion traps, QTOFs, and hybrid ion-trap FT

spectrometers that employ low-energy CID. Over the years, low-energy CID has seen some application in structure characterization of steroid sulfates[56], but the data interpretation often stops at accurate mass and functional group fragmentation to deduce known steroid metabolites in a sample mixture. Recently, high-energy CID[43, 44] has been resumed with TOF/TOF instruments, and applied in some areas for structural characterization[45, 46, 57, 58].

1.7 Purpose of this Study

In the first part of the thesis research, we utilized all the different MS approaches mentioned above for the structure characterization of steroid metabolites in Chapter 2 and Chapter 3. We then applied this combination of MS methods in Chapter 4 in the study of mouse urine to solve the structure of an unknown steroid sex pheromone that is important in mouse communication.

1.8 References

1.   Ferrero, D. M., and Liberles, S. D. (2010) The secret codes of mammalian scents, *Wiley interdisciplinary reviews. Systems biology and medicine 2*, 23-33.

2.   Kobayakawa, K., Kobayakawa, R., Matsumoto, H., Oka, Y., Imai, T., Ikawa, M., Okabe, M., Ikeda, T., Itohara, S., Kikusui, T., Mori, K., and Sakano, H. (2007) Innate versus learned odour processing in the mouse olfactory bulb, *Nature 450,* 503-508.

3.   Restrepo, D., Arellano, J., Oliva, A. M., Schaefer, M. L., and Lin, W. (2004) Emerging views on the distinct but related roles of the main and accessory olfactory systems in responsiveness to chemosensory signals in mice, *Hormones and behavior 46*, 247-256.

4.   Shepherd, G. M. (2006) Behaviour: smells, brains and hormones, *Nature 439*, 149-151.

5.   Spehr, M., Kelliher, K. R., Li, X. H., Boehm, T., Leinders-Zufall, T., and Zufall, F. (2006) Essential role of the main olfactory system in social recognition of major

histocompatibility complex peptide ligands, *The Journal of neuroscience : the official journal of the Society for Neuroscience 26*, 1961-1970.

6.   Meredith, M., Marques, D. M., O'Connell, R. O., and Stern, F. L. (1980) Vomeronasal pump: significance for male hamster sexual behavior, *Science 207*, 1224-1226.

7.   Holy, T. E., Dulac, C., and Meister, M. (2000) Responses of vomeronasal neurons to natural stimuli, *Science 289*, 1569-1572.

8.   Leinders-Zufall, T., Lane, A. P., Puche, A. C., Ma, W., Novotny, M. V., Shipley, M. T., and Zufall, F. (2000) Ultrasensitive pheromone detection by mammalian vomeronasal neurons, *Nature 405*, 792-796.

9.   Wysocki, C. J., and Lepri, J. J. (1991) Consequences of removing the vomeronasal organ, *The Journal of steroid biochemistry and molecular biology 39*, 661-669.

10.   Leypold, B. G., Yu, C. R., Leinders-Zufall, T., Kim, M. M., Zufall, F., and Axel, R. (2002) Altered sexual and social behaviors in trp2 mutant mice, *Proceedings of the National Academy of Sciences of the United States of America 99*, 6376-6381.

11.   Stowers, L., Holy, T. E., Meister, M., Dulac, C., and Koentges, G. (2002) Loss of sex discrimination and male-male aggression in mice deficient for TRP2, *Science 295*, 1493-1500.

12.   Halpern, M., and Martinez-Marcos, A. (2003) Structure and function of the vomeronasal system: an update, *Progress in neurobiology 70*, 245-318.

13.   Kimchi, T., Xu, J., and Dulac, C. (2007) A functional circuit underlying male sexual behaviour in the female mouse brain, *Nature 448*, 1009-1014.

14.   Wysocki, C. J., Nyby, J., Whitney, G., Beauchamp, G. K., and Katz, Y. (1982) The vomeronasal organ: primary role in mouse chemosensory gender recognition, *Physiology & behavior 29*, 315-327.

15.   Novotny, M. V. (2003) Pheromones, binding proteins and receptor responses in rodents, *Biochemical Society transactions 31*, 117-122.

16.   Leinders-Zufall, T., Brennan, P., Widmayer, P., S, P. C., Maul-Pavicic, A., Jager, M., Li, X. H., Breer, H., Zufall, F., and Boehm, T. (2004) MHC class I peptides as chemosensory signals in the vomeronasal organ, *Science 306*, 1033-1037.

17.   Chamero, P., Marton, T. F., Logan, D. W., Flanagan, K., Cruz, J. R., Saghatelian, A., Cravatt, B. F., and Stowers, L. (2007) Identification of protein pheromones that promote aggressive behaviour, *Nature 450*, 899-902.

18.   Mombaerts, P. (2004) Genes and ligands for odorant, vomeronasal and taste receptors, *Nature reviews. Neuroscience 5*, 263-278.

19.   Nodari, F., Hsu, F. F., Fu, X., Holekamp, T. F., Kao, L. F., Turk, J., and Holy, T. E. (2008) Sulfated steroids as natural ligands of mouse pheromone-sensing neurons, *The Journal of neuroscience : the official journal of the Society for Neuroscience 28*, 6407-6418.

20.   Hsu, F. F., Nodari, F., Kao, L. F., Fu, X., Holekamp, T. F., Turk, J., and Holy, T. E. (2008) Structural characterization of sulfated steroids that activate mouse pheromone-sensing neurons, *Biochemistry 47*, 14009-14019.

21.   Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules, *Science 246*, 64-71.

22.   Strupat, K., Karas, M., and Hillenkamp, F. (1991) 2,5-Dihydroxybenzoic acid: a new matrix for laser desorption—ionization mass spectrometry, *International Journal of Mass Spectrometry and Ion Processes 111*, 89-102.

23.   Gerald, T. (2012) Ordinary Differential Equations and Dynamical Systems, *Providence: American Mathematical Society*.

24.   Ortega-Gadea, S., Bernabe-Zafon, V., Simo-Alfonso, E. F., Ochs, C., and Ramis-Ramos, G. (2006) Characterization of industrial alkylpolyphosphonates by infusion electrospray ionization-ion trap mass spectrometry with identification of the impurities by tandem capillary zone electrophoresis, *Journal of mass spectrometry : JMS 41*, 23-33.

25.   Carrol, D. I., Dzidic, I., Horning, E. C., and Stillwell, R. N. (1981) Atmospheric pressure Ionization mass Spectrometry, *Applied Spectroscopy Reviews 17*, 377-406.

26.   Douglas, D. J., Frank, A. J., and Mao, D. (2005) Linear ion traps in mass spectrometry, *Mass spectrometry reviews 24*, 1-29.

27.   Tabert, A. M., Goodwin, M. P., and Cooks, R. G. (2006) Co-occurrence of boundary and resonance ejection in a multiplexed rectilinear ion trap mass spectrometer, *Journal of the American Society for Mass Spectrometry 17*, 56-59.

28.   Kingdon, K. H. (1923) A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures, *Physical Reviews 21*, 408.

29.   Makarov, A. (2000) Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis, *Analytical chemistry 72*, 1156-1162.

30.   Adams, J. (1990) Charge-remote fragmentations: analytical applications and fundamental studies, *Mass spectrometry reviews 9*, 141-186.

31.   Gross, M. L. (1992) Charge-remote fragmentations: method, mechanism and applications, *Int J Mass Spectrom Ion Processes 118/119*, 137-165.

32.   Adams, J., and Songer, M. J. (1993) Charge-remote fragmentations for structural determination of lipids, *Trends Anal Chem 12*, 28-36.

33.   Jensen, N. J., Tomer, K. B., and Gross, M. L. (1985) Gas-phase ion decomposition occurring remote to a charge site, *J Am Chem Soc 107*, 1863-1868.

34.   Wysocki, V. H., Ross, M. M., Horning, S. R., and Cooks, R. G. (1988) Remote-site (charge-remote) fragmentation, *Rapid communications in mass spectrometry : RCM 2*, 214-217.

35.   Wysocki, V. H., and Ross, M. M. (1991) Charge-remote fragmentation of gas-phase ions: mechanistic and energetic considerations in the dissociation of long-chain functionalized alkanes and alkenes, *Int J Mass Spectrom Ion Processes 104*, 179-211.

36.   McAdoo, D. J. (1988) Ion-neutral complexes in unimolecular decompositions, *Mass spectrometry reviews 7*, 363-393.

37.   Contado, M. J., Adams, J., Jensen, N. J., and Gross, M. L. (1991) A charge-remote allylic cleavage reaction: Mechanistic possibilities, *Journal of the American Society for Mass Spectrometry 2*, 180-183.

38.   Gross, M. L., Chess, E. K., Lyon, P. A., Crow, F. W., Evans, S., and Tudge, H. (1982) Triple analyzer mass spectrometry for high resolution MS/MS studies, *Int J Mass Spectrom Ion and Ion Physcis 42*, 243-254.

39.   Tomer, K. B., Crow, F. W., and Gross, M. L. (1983) Location of double-bond position in unsaturated fatty acids by negative ion MS/MS, *J. Am. Chem. Soc. 105*, 5487-5488.

40.   Tomer, K. B., Jensen, N. J., Gross, M. L., and Whitney, J. (1986) Fast atom bombardment combined with tandem mass spectrometry for determination of bile salts and their conjugates, *Biomedical & environmental mass spectrometry 13*, 265-272.

41.   Cheng, M. T., Barbalas, M. P., Pegues, R. F., and McLafferty, F. W. (1983) Tandem mass spectrometry: structural and stereochemical information from steroids, *J. Am. Chem. Soc. 105*, 1510-1513.

42.   Kingston, E. E., Beynon, J. H., Newton, R. P., and Liehr, J. G. (1985) The differentiation of isomeric biological compounds using collision-induced dissociation of ions generated by fast atom bombardment, *Biomedical mass spectrometry 12*, 525-534.

43.   Cotter, R. J. (2013) High energy collisions on tandem time-of-flight mass spectrometers, *Journal of the American Society for Mass Spectrometry 24*, 657-674.

44.   Cornish, T. J., and Cotter, R. J. (1993) A curved-field reflectron for improved energy focusing of product ions in time-of-flight mass spectrometry, *Rapid communications in mass spectrometry : RCM 7*, 1037-1040.

45.   Trimpin, S., Clemmer, D. E., and McEwen, C. N. (2007) Charge-remote fragmentation of lithiated fatty acids on a TOF-TOF instrument using matrix-ionization, *Journal of the American Society for Mass Spectrometry 18*, 1967-1972.

46.   Pittenauer, E., and Allmaier, G. (2009) High-energy collision induced dissociation of biomolecules: MALDI-TOF/RTOF mass spectrometry in comparison to tandem sector mass spectrometry, *Combinatorial chemistry & high throughput screening 12*, 137-155.

47.   Sjovall, J. (2004) Fifty years with bile acids and steroids in health and disease, *Lipids 39*, 703-722.

48.   Horning, E. C. (1968) Gas phase analytical methods for the study of steroid hormones and their metabolite. In Gas Phase Chromatography of Steroids, edited by Eik-Nes, K. B., Horning, E. C., *Berlin, Germany: Springer Verlag*.

49.   Barber, M., Bordoli, R. S., Sedgwick, R. D., and Tyler, A. N. (1981) Fast atom bombardment of solids (FAB): A new ion source for mass spectrometry, *J Chem Soc Chem Commun 7*, 325-327.

50.   Barber, M., Bordoli, R. S., Elliott, G. J., Sedgwick, R. D., and Tyler, A. N. (1982) Fast atom bombardment mass spectrometry, *Anal. Chem. 54*, 645A-657A.

51.   Yamashita, M., and Fenn, J. B. (1984) Electrospray ion source. Another variation on the free-jet theme, *J. Phys. Chem. 88*, 4451-4459.

52.   Yamashita, M., and Fenn, J. B. (1984) Negative ion production with the electrospray ion source, *J. Phys. Chem. 88*, 4671-4675.

53.   Cooks, R. G., and Glish, G. L. (1981) Mass-spectrometry mass-spectrometry, *Chem. Eng. News 59*, 40-52.

54.   Levsen, K., and Schwarz, H. (1983) Gas-phase chemistry of collisionally activated ions, *Mass spectrometry reviews 2*, 77-148.

55.   Tomer, K. B., and Gross, M. L. (1988) Fast atom bombardment and tandem mass spectrometry for structure determination: remote site fragmentation of steroid conjugates and bile salts, *Biomedical & environmental mass spectrometry 15*, 89-98.

56.   Strahm, E., Saudan, C., Sottas, P. E., Mangin, P., and Saugy, M. (2007) Direct detection and quantification of 19-norandrosterone sulfate in human urine by liquid chromatography-linear ion trap mass spectrometry, *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences 852*, 491-496.

57.   Sun, G., Yang, K., Zhao, Z., Guan, S., Han, X., and Gross, R. W. (2007) Shotgun metabolomics approach for the analysis of negatively charged water-soluble cellular metabolites from mouse heart tissue, *Analytical chemistry 79*, 6629-6640.

58.   Pittenauer, E., and Allmaier, G. (2009) The renaissance of high-energy CID for structural elucidation of complex lipids: MALDI-TOF/RTOF-MS of alkali cationized triacylglycerols, *Journal of the American Society for Mass Spectrometry 20*, 1037-1047.

Chapter 2*

Mass Spectrometry Combinations for Structural

Characterization of Sulfated-Steroid Metabolites

2.1 Abstract

Steroid conjugates, which often occur as metabolites, are challenging to characterize. One application is female-mouse urine, where steroid conjugates serve as important ligands for the pheromone-sensing neurons. Although the two with the highest abundance in mouse urine were previously characterized with mass spectrometry (MS) and NMR to be sulfated steroids, many more exist but remain structurally unresolved. Given that their physical and chemical properties are similar, they are likely to have a sulfated steroid ring structure. Because these compounds occur in trace amounts in mouse urine and elsewhere, their characterization by NMR will be difficult. Thus, MS methods become the primary approach for determining structure. In this chapter, we show that a combination of MS tools is effective for determining the structures of sulfated steroids. Using 4-pregnene analogs, we explored high-resolving power MS (HR-MS) to determine chemical formulae; HD exchange MS (HDX-MS) to determine number of active, exchangeable hydrogens (e.g., OH groups); methoxyamine hydrochloride (MOX) derivatization MS, or reactive desorption electrospray ionization with hydroxylamine to determine the number of carbonyl groups; and tandem MS (MS^n), high-resolution tandem MS (HRMS/MS), and GC-MS to obtain structural details of the steroid ring. From the fragmentation studies, we deduced three major fragmentation rules for this class of sulfated steroids. We also show that a combined MS approach is effective for determining structure of steroid metabolites, with important implications for targeted metabolomics in general and for the study of mouse social communication in particular.

2.2 Introduction

Determining the structure of steroid conjugates is a major challenge in metabolomics. A particular need is to understand animal communication. Mammals use their sense of smell to identify and learn the status of other animals. Natural cues such as urine contain metabolites informative about reproductive status, social dominance, and levels of stress[1]. In most cases, however, the identity of the signaling compounds remains unknown. A recent study[2] used the responses of olfactory neurons in the mouse vomeronasal organ, specializing in the detection of "social odors" or "pheromones" produced by other animals, to assay the activity of purified urine fractions. The structures of two active compounds were determined[3], by using a combination of mass spectrometry (MS) and nuclear magnetic resonance (NMR), were shown to be sulfated steroids of the glucocorticoid family. Because steroids are among the main regulators of mammalian physiology[4], their metabolites are attractive candidates for conveying many aspects of an animal's status. Indeed, several lines of evidence suggest that urine contains other active, but unknown, sulfated steroids. For the urine from female mice of the BALB/c strain, sulfated compounds account for an estimated 80% of the total activation of these sensory neurons[2].

Unfortunately, like many metabolites, many of the unknown cues in natural stimuli are present in much lower abundance than the two compounds identified by Hsu et al.[3], and consequently present a greater challenge for structural analysis. MS is particularly apt for obtaining structural information from small sample quantities, but we felt that a single MS method would not be adequate. Thus, the purpose of the research reported in this chapter is to explore the effectives of various combinations of MS methods to determine steroid-sulfate structure.

For many years, steroidal hormones have been analyzed by gas chromatography-mass spectrometry (GC-MS) since Sweeley and Horning's[5] first success in 1960. More recently,

steroids were also analyzed by using HPLC coupled with tandem mass spectrometry (LC-MS/MS)[6-12]. GC has the advantages over LC of higher chromatographic resolution, affording specificity for minor structural differences, as well as a vast EI online database for structural identification of analytes. GC-MS, however, requires some knowledge of the analytes' structure, because GC's reliance on volatility often necessitates functional-group-specific derivatization. In contrast, LC-MS/MS requires less initial knowledge of the structure of the sample, as derivatization is not a requirement for ionization. This eases the handling requirement for the sample and reduces possible losses of material during sample preparation.

In this chapter, we describe a combination of methods, taking advantage of both GC-MS and LC-MS/MS, which constitute a MS-based "tool box" for studying structures of sulfated steroids. Although each tool is not new, we wished to explore whether a combination would offer an effective approach to steroid identification. To demonstrate, we focused on a set of six commercial or custom-synthesized 4-pregnene steroids, a key class of compounds because they are central in steroid biosynthetic pathways in the mouse[13]. The 4-pregnenes sulfates were chosen because they are structurally similar to previously identified sulfated steroids found in mouse urine and active for mouse communication. We also anticipate that more unknown sulfated steroids belonging to this category exist in mouse urine. The combination we selected include different MS-based approaches including 1) high mass-resolving-power MS (HR-MS) to determine chemical formula; 2) HD exchange MS (HDX-MS) to determine number of active hydrogens (e.g., –OH groups); 3) methoxyamine hydrochloride (MOX) derivatization MS, or online derivatization with hydroxylamine by reactive desorption electrospray ionization, to determine the number of carbonyl groups; and 4) tandem MS (MS^n), high-resolution tandem MS (HRMS/MS), and 5) GC-MS to probe details of the steroid-ring structure. From the

fragmentation studies, we can draw some generalized fragmentation rules for sulfated-steroid CID MS/MS. This combination of MS-based methods provides an opportunity to "footprint" steroid sulfates metabolites, providing a foundation for sequel studies of mouse social communication.

2.3 Materials and Methods

2.3.1 Materials

Sulfated steroids, including 4-pregnen-11β, 21-diol-3, 20-dione 21-sulfate (termed SS425, standing for sulfated steroid 425, the number of which is the molecular mass of [M - H]⁻), 4-pregnen-11β, 17, 21-triol-3, 20-dione 21-sulfate (SS441) and 4- pregnen-17, 21-diol-3, 11, 20-trione 21-sulfate (SS439), and steroids, including 5β- pregnan-11β, 21-diol-3, 20-dione, 5α-pregnan-3α, 11β, 21-triol-20-oneand 4- pregnen-11β, 20β, 21-triol-3-one were purchased from Steraloids Inc (Newport, RI), and dissolved in methanol to 5 mM as stock solution. Strata-X 33u polymeric reversed phase 60 mg/3 m for solid phase extraction (SPE) was purchased from Phenomenex (Torrance, CA). All other chemicals and reagents were purchased from Sigma Aldrich (St. Louis, MO).

2.3.2 Sulfation of Steroids

Sulfation methods to synthesize sulfated steroids from steroids were modified from that of Kornel et al.[14] to suit for micro-synthesis. Acidified pyridine was prepared by adding 0.3 mL chlorosulfonic acid dropwise to 5 mL dry pyridine while cooling on an ice bath. To sulfate steroids 5β- pregnan-11β, 21-diol-3, 20-dione, 5α- pregnan-3α, 11β, 21-triol-20-one and 4-pregnen-11β, 20β, 21-triol-3-one, 5 µL of 5 mM stock steroid solution was dried by using a SpeedVac at room temperature, and the residue redissolved in 80 µL pyridine. Acidified pyridine

(30 μL) was then added, and the mixture was shaken for 5 min, as optimized, following which saturated $NaHCO_3$ solution was used to neutralize the reaction solution, which was then dried in a SpeedVac at room temperature. The residue was redissolved in 3 mL H_2O and desalted with SPE columns, and the final product was eluted with 1 mL methanol. Sulfation of 5β- pregnan-11β, 21-diol-3, 20-dione and 4- pregnen-11β, 20β, 21-triol-3-one occurred at C21, as indicated by LC-MS/MS results (data not shown), leading to single products named SS427 (standing for sulfated steroid 425, the number of which is the molecular mass of $[M - H]^-$) and iSS427 (standing for isomer SS427), respectively; sulfation of 5α- pregnan-3α, 11β, 21-triol-20-one happened at either C3 or C21, referred to as 3-SS429 and 21-SS429 respectively, which were separated with a custom-packed C18 column before infusion into ESI source for MS/MS analysis.

2.3.3 Mass Spectrometry

2.3.3.1 Electrospray Ionization Mass Spectrometry (ESI-MS):

Sample stock solutions were diluted to 1 μM and infused to an ESI source at 3 μL/min. Both low-mass-resolving power and high-mass-resolving power ESI-MS were conducted in the negative-ion mode on a LTQ Orbitrap Velos and a LTQ Orbitrap from Thermo Scientific (West Palm Beach, FL), equipped with the Xcalibur operating system. Spray voltage in the negative ion mode was set to 3.5 kV; capillary voltage and temperature was at -32 V and 275 ºC; and the tube lens was -222 V. Tandem MS (MS^n) experiments were carried out after low energy CID by using a relative collision energy ranging from 30 to 40% of the maximum and high energy HCD with 150 eV collision energy.

2.3.3.2 Reactive Desorption Electrospray Ionization (Reactive DESI):

All DESI experiments were carried out by using an electrosonic spray ionization (ESSI) source to generate charged microdroplets, and a silica capillary to introduce the sample[15, 16]. The ESSI sprayer was aimed to the outlet of the sample introduction capillary at an angle of 30°-45°, and the horizontal distance between the two was ~0.5 mm. The solvent for ESSI consisted of methanol/water in DESI and methanol/water/5% hydroxylamine/0.05% acetic acid in reactive DESI, and was injected at 10 µL/min, comparing to 4 µL/min of the sample flow rate. The control DESI experiments were done in the negative-ion mode, whereas the reactive DESI experiments were done in the positive-ion mode. The spray voltage was set at 5 kV for positive ions and -5 kV for negative ions; the nebulizing gas (N_2) pressure for ESSI was 80 psi. MS detection was on a Thermo Finnigan LCQ DecaXP plus. Capillary temperature and voltage was set at 250 ºC and 21 V, and the tube lens was 10 V for positive ions; they were set to 250 ºC, -47 V, and -50 V, respectively, for negative ions.

2.3.3.3 Gas Chromatography-Mass Spectrometry (GC-MS):

The sulfate group of sulfated steroids was removed by solvolysis as described in[17], under conditions for micro-synthesis. Acidified ethyl acetate was prepared by shaking 1 M sulfuric acid in water and ethyl acetate (V:V= 1:2) vigorously for 2-3 min and then discarding the water phase. The dried sample was dissolved in acidified ethyl acetate to a concentration of 1 µM to give a total of 5 mL and incubated for ~ 6 h at 38 ºC, followed by neutralization with ammonium hydroxide. The reaction time was optimized for both SS425 and SS441 to 6 h. The desulfated sample was then dried, and redissolved in 15 µL pyridine, to which 15 µL of *N,O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA) with 1% TMS was added and incubated at 70 ºC for 1 h for SS425 and 2 h for SS441. Derivatized samples were then loaded into an Agilent

(Santa Clara, CA) 7693 autosampler, and 1 µL injections were analyzed on an Agilent 7890A gas chromatograph system interfaced to an Agilent 7200 accurate-mass Q-TOF mass spectrometer. The samples were injected with a 10:1 split. The GC column was a DB-5MS (12 m, 0.25 mm i.d., 0.33 um film coating, P.J. Cobert St. Louis, MO). A linear temperature gradient was used: the initial temperature of 80° was held for 2 min and increased to 280 at 15°/min. The temperature was held at 280° for 5 min. The samples were ionized by electron ionization (EI), and the source temperature, electron energy, and emission current were set at 280 °C, 70 eV and 300 uA, respectively. The injector and transfer line temperatures were also 280°C.

2.3.4 Hydrogen Deuterium Exchange (HDX) and Methoxyamine Hydrochloride (MOX) Derivatization

HDX of sulfated steroids was initiated by dissolving a dried sample in D_2O, to give a sample concentration of ~ 2 µM, at neutral pH. The same volume of acetonitrile was added prior to conducting ESI MS and MS/MS. MOX derivatization involved 18 h incubation of dried sample (20 µL × 1 uM) in 100 µL MOX reagent (2% in pyridine) at 60 ºC. The product solution was then dried under nitrogen and re-dissolved in 20 µL MeOH and infused to the ESI source for analysis.

2.4 Results and Discussion

2.4.1 Electrospray Ionization (ESI-MS), HDX, and MOX Derivatization

Given that these sulfated steroids contain a negative charge on the sulfate group, all ESI mass spectra and product-ion spectra (MS^n) spectra were obtained in the negative-ion mode to give an [M - H]$^-$ molecular ion. High-resolving-power mass measurement gave mass accuracy within 5

ppm for both ^{32}S and ^{34}S isotopic peaks, based on the molecular ion formulae (Table 2-1), as expected.

Upon HDX, the molecular ion profile cleanly shifted 1 Da for SS425, indicating the existence of one exchangeable proton in the compound. For SS441, the shift is 2 Da, indicating two exchangeable protons. For an unknown, its accurate mass would reveal the number of oxygens; that number, minus those in the sulfate group, indicates the maximum number of exchangeable sites, provided there are no nitrogens in the molecule. Under neutral conditions, hydrogens alpha to a carbonyl group exchange very slowly and will not be "counted" here.

MOX derivatization converts the carbonyl group to an oxime (C=NOCH3), leading to an increase of 29 Da in mass for every carbonyl of the compound. Thus, HDX and MOX derivatization afford a count of the number of ketone or aldehyde groups (Table 2-1 indicates two carbonyl groups in both SS425 and SS441).

Table 2- 1. Outcomes of the application of HRMS, HDX, MOX derivatization, reactive DESI, and HCD MS/MS of SS425 and SS441

Sulfated steroids	Structure	HR/MS		HDX MS (# of –OH)	MOX deriv. (# of C=O)	Reactive DESI (# of C=O)	HCD MS/MS (-SO$_4$$^-$?)
		^{32}S (error)	^{34}S (error)				
SS425		425.1641 (3 ppm)	427.1670 (3 ppm)	426 (1)	483 (2)	491/492/493 (2)	97 (yes)
SS441		441.1584 (1 ppm)	443.1549 (3 ppm)	443 (2)	499 (2)	507/508/509 (2)	97 (yes)
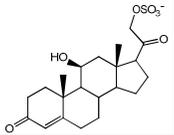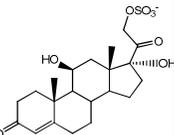
2.4.2 Reactive Desorption Electrospray Ionization (Reactive DESI)

As an alternative to MOX derivatization in counting carbonyl groups in sulfated steroids, we followed Huang et al.[18] in performing reactive DESI on steroids in the positive-ion mode, utilizing hydroxylamine and ambient-pressure ionization to derivatize steroidal carbonyl groups. Hydroxylamine was used instead of methoxyamine because it is more volatile than the latter and is compatible with on-line derivatization and detection. Unlike traditional DESI, we introduced the sulfated steroid samples in a capillary with a constant flow instead of depositing them on a plate. In this process, protonated hydroxylamine ions are formed during electrospray and carried in microdroplets toward the substrate where they react. The proposed product ions include a positively charged tetrahedral intermediate (hydroxylamine adduct, with a mass increase of 34 Da) and an oxime formed by loss of water from the tetrahedral intermediate (Scheme 2 in literature[18]). Because the reaction with hydroxylamine ions give a more readily protonated group in lieu of the origianal carbonyl group, the derivatization adds a positive charge to the molecule, counterbalanced in part by the negative charge of the highly acidic sulfate group. Provided there is more than one carbonyl group in the compound, we expect the products to be zwitterions
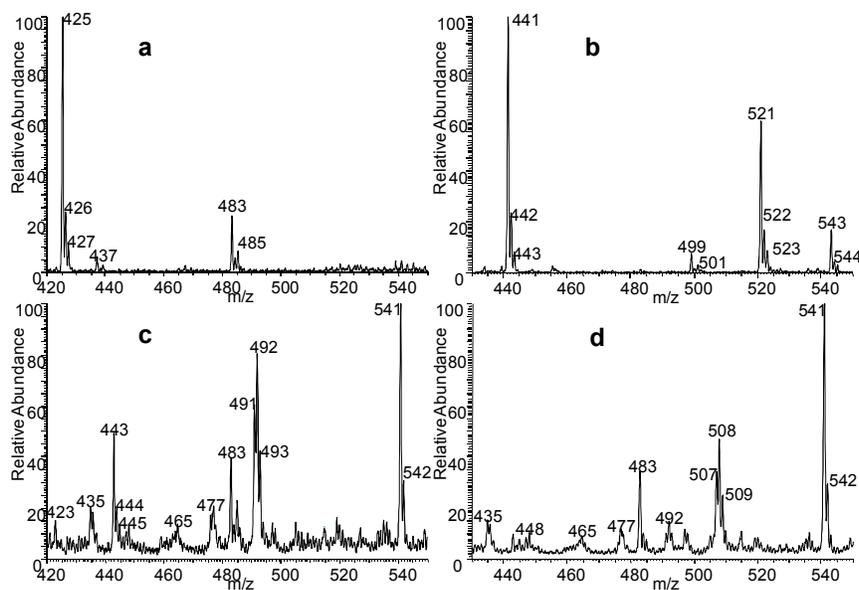
Figure 2- 1. DESI mass spectra in the negative-ion mode for (a) SS425, (b) SS441. Reactive DESI mass spectra in the positive-ion mode of (c) SS425, (d) SS441.

34

both in solution and in a mass spectrometer ion source operating in the positive-ion mode.

As shown in Table 1 and Figure 2-1, product ions from reactive DESI on SS425 and SS441 are of m/z 493 and 509, respectively, both with mass increases of 68 Da, indicating the presence of two carbonyl groups in each compound. Unlike Huang et al.'s study[18], the oxime product wasn't observed in either reaction, and the tetrahedral intermediate product was accompanied by an oxidation reaction that resulted in products with 1 or 2 Da decrease in mass, such as those at m/z 491/492 and m/z 507/508 in reactive DESI of SS425 and SS441, respectively. These differences likely result from the different instrumentation and DESI setup (we used ESSI instead of DESI). Although the mechanisms for the formation of the oxidized species of the tetrahedral intermediates is not a focus of this study, a possible explanation is that the ESI capillary (in this case, DESI capillary) is an electrolytic half-cell where redox reactions occur under the conditions of DESI[19].

2.4.3 HCD MS/MS of ESI Produced Steroid Sulfates

Verifying the presence of a sulfate group to account for at least one sulfur atom in the compounds is straightforward. With HCD MS^2, the product-ion spectra of SS425 and SS441 are dominated by ions of m/z 97, HOSO$_3^-$, confirming that one sulfur atom is part of a sulfate group. One example is the HCD product-ion spectrum (MS^2) of SS425 (Figure 2-2). Aside from the ion of m/z 97, the ion of m/z 80 is SO$_3^-$, offering more
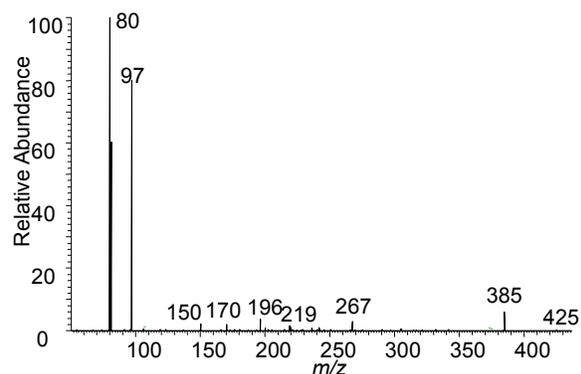
Figure 2- 2. HCD fragmentation of the [M − H]$^-$ of m/z 425 of SS425.

confirmation for the presence of a sulfate group. The formation of HSO_4^- and SO_3^- is expected and was reported previously in product-ion spectra of many steroidal sulfates[20-23]. The proton-transfer mechanism as proposed in[24], occurs with the elimination of a ketene and a ketone. However, other vinyl or phenyl sulfates, such as estron-3-yl 3-sulfate, do not undergo bisulfate elimination to form HSO_4^- in mass spectrometric fragmentation[22, 23, 25], because that reaction would form a prohibitively high-energy benzyne. We term the formation of the product ion $HOSO_3^-$ of m/z 97 upon HCD MS/MS as the "97" rule, which indicates the presence of a sulfate as a signature product ion. Although HCD gives highly obvious evidence for a sulfate substituent via the abundant product ions of m/z 80 and 97, the other product ions, requiring higher dynamic range, provide incomplete structural details of the sulfated steroid. Therefore, we utilized an alternative tandem MS method with CID fragmentation in an ion trap, which prevented observation of the m/z 80 and 97 ions and provided higher dynamic range for product ions arising from ring cleavages.

2.4.4 Multi-stage CID (MSⁿ) of ESI-Produced Steroid Sulfates

Today's MS/MS technology allows the low-energy, negative-ion CID spectra of metabolites such as SS425 and SS441
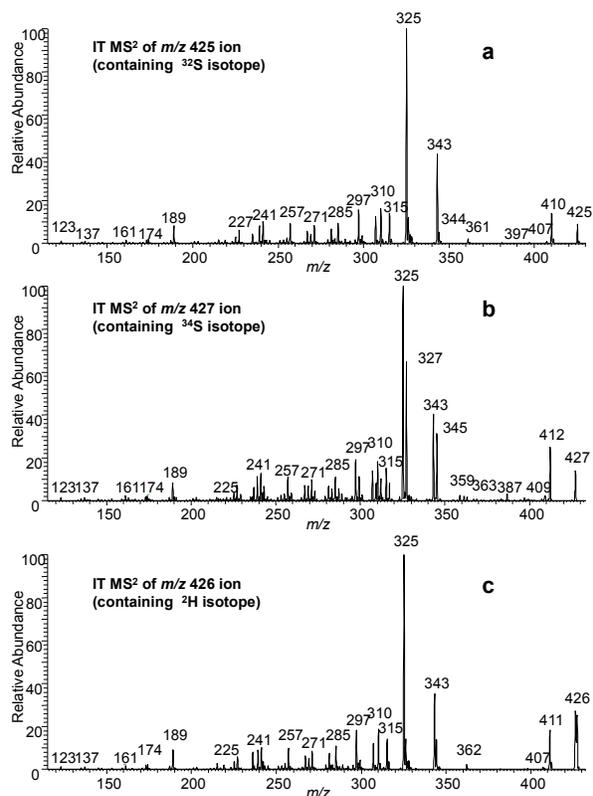
Figure 2- 3. Product-ion spectra of the [M - H]⁻ ions of SS425 of (a) m/z 425 (^{32}S isotopolog) (b) m/z 427 (the ^{34}S isotopolog), (c) the ion of m/z 426 (^{2}H isotopolog) after HDX.

36

(presented in Figures 2-3a and 2-4a) to be obtained with accurate *m/z* measurement (Table 2-2). We were able to obtain product-ion (MS^2) spectra (Figure 2-3b and 2-4b) from precursor ions of *m/z* 427 for SS425 and 443 for SS441, which are a mixture of isotopologs containing approximately 53% ^{34}S. These spectra also confirm the presence/absence of sulfate group. For example, the single ion of *m/z* 410 in Figure 2-3a must contain the sulfate group, consistent with the calculated elemental composition and the shift of 2 Da to m/z 412 in Figure 2-3b. More significantly, the product ions of *m/z* 325 and 343 must be

Figure 2- 4. Product-ion spectra of the [M - H]$^-$ ions of SS441 of (a) *m/z* 441 (^{32}S isotopolog) (b) *m/z* 443 (^{34}S isotopolog), and (c) *m/z* 443 after H/DX (^{2}H isotopolog).

formed by losses including the sulfate groups as the *m/z* 325 ion now appears as both *m/z* 325 (losses of $H_2^{34}SO_3 + H_2O$) and 327 ions (losses of $H_2^{32}SO_3 + H_2O$); the latter retaining two ^{13}C atoms. This is consistent with accurate mass measurements for ions of *m/z* 343 and 325 (Table 2-2).

To assign with more confidence the compositions of the product ions, MS^2 was also conducted on samples where the -OH group has become deuterated (Figure 2-3c and 2-4c). By comparing product ions in Figure 2-3a and 2-3c, or Figure 2-4a and 2-4c, any mass shift indicates, in the absence of HD scrambling, the number of –OH groups retained in the related product ion.

Noteworthy is the lack of any significant shifts of the m/z 325 and 343 ions (Figure 2-3c), indicating that the loss of H_2SO_3 and H_2O now occur as those of $HDSO_3$ and H_2O. The results suggest that the loss of H_2SO_3 involves the H transfer from –OH on C11 to the sulfate group while transferring the charge from the sulfate group to the oxygen on C11.

Table 2- 2. Calculated fragment-ion elemental compositions based on accurate mass detection of CID fragmentations with $HRMS^2$ for SS425 and SS441.

a. SS425			b. SS441		
m/z	Calc. mass	Comp.	*m/z*	Calc. mass	Comp.
410.1374	410.1394	$C_{20}H_{26}O_7S$	426.1354	426.1354	$C_{20}H_{26}O_8S$
343.1890	349.1904	$C_{21}H_{27}O_4$	423.1479	423.1483	$C_{21}H_{27}O_7S$
325.1785	325.1798	$C_{21}H_{25}O_3$	411.1479	411.1483	$C_{20}H_{27}O_7S$
315.1943	315.1955	$C_{20}H_{27}O_3$	343.1909	343.1915	$C_{21}H_{27}O_4$
310.1552	310.1563	$C_{20}H_{22}O_3$	331.1909	331.1915	$C_{20}H_{27}O_4$
297.1839	297.1849	$C_{20}H_{25}O_2$	315.1596	315.1602	$C_{19}H_{23}O_4$
285.1475	285.1485	$C_{18}H_{21}O_3$	313.1803	313.1809	$C_{20}H_{25}O_3$
271.1684	271.1693	$C_{18}H_{23}O_2$	297.1487	297.1496	$C_{19}H_{21}O_3$
257.1529	257.1536	$C_{17}H_{21}O_2$	282.1256	282.1261	$C_{18}H_{18}O_3$
241.1582	241.1587	$C_{17}H_{21}O$	189.0921	189.0921	$C_{12}H_{13}O_2$
189.0911	189.0910	$C_{12}H_{13}O_2$	138.9710	138.9709	$C_2H_3O_5S$
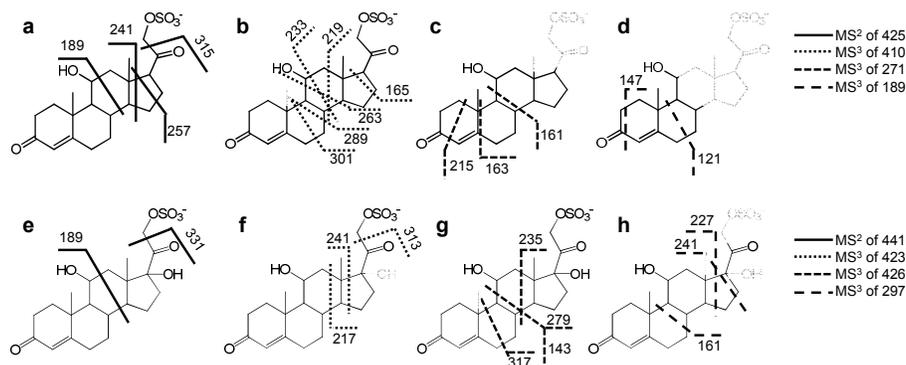
Figure 2- 5. Sites of fragmentation for the (a) molecular ion and (b-d) various product ions of SS425. Sites of fragmentation for the (e) molecular ion and (f-h) various product ions of SS441. Those parts of the structures drawn in fainter lines represent the regions lost as neutrals in the MS/MS-induced fragmentation.

This combination of isotopic labeling and accurate *m/z* measurement in the product-ion spectra permits a fragmentation pathway scheme to be proposed. Those schemes (Scheme 2-1a and b), and reiterated in a simpler form in Figures 2-5a and 2-5e) are informative of the steroid-ring structure (the carbons are numerically labeled according to convention on structure of SS425 in Scheme 1a for convenience). Additional support and complementary information to characterize the steroid ring were found in a MS^3 experiment and are

Scheme 2-1. Proposed fragmentation pathways for (a) SS425 and (b) SS441 upon CID MS/MS. Ion structures are consistent with fragmentation and accurate masses of product ions, and they are proposed to facilitate the interpretation of the mass spectra and are not based on standard ion-chemistry criteria.
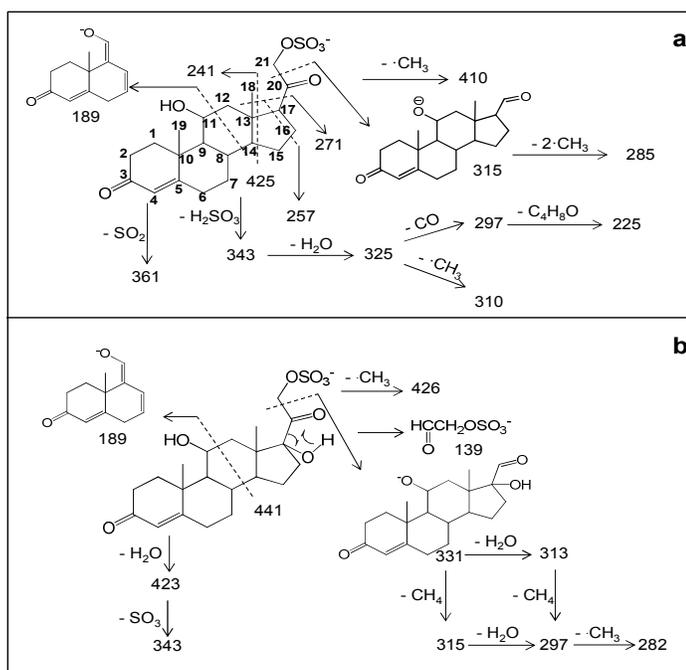
reiterated in Figures 2-5b, c, d, f, g, and h. The fragmentation pathways proposed here for either MS^2 or MS^3 are not validated by molecular orbital theory or isotope labeling. Instead, the ion structures are proposed according to their accurate masses to be consistent with the fragmentation. The purpose is to show correlations in the fragmentation patterns of compounds from the same class of sulfated steroids and to provide insights on how CID fragmentation can be applied to track functional groups and their changes on the ring. One outcome is that the ring topology of steroid sulfates of this type can be characterized by the rich fragmentation seen in

both MS^2 and MS^3 experiments, as illustrated in Figure 2-5. Collisional activation, however, does not always produce all possible fragmentations of the ring, in which case, the specific functional group location cannot be made. For example, the location of C17-OH in SS441 cannot be established because there's no cleavage for C16-C17 bond; in which case if SS441 were an unknown compound, the –OH group can only be vaguely assigned to C16 or C17.

2.4.5 Functional Group Location based on CID MS/MS

One goal of interpreting tandem MS and accurate m/z data is to locate a functional group on the steroid ring. The fragment ions of m/z 189 found for both SS425 and SS441 are depicted to contain steroid rings A and B, as well as C11. The formation of ions of m/z 189 is not simply due to the two homolytic cleavages of C11-C12 and C8-C14 bonds (scheme 2-1), but also must involve a proton transfer from C11-OH to the weak SO_3^- base, possibly followed by loss of the H_2 and $C_9H_{14}O_5S$. Although the reasoning to assign structure is similar to that used by Hsu et al.[3], and also supported by chemical composition from accurate mass and the lack of a sulfate group by ^{34}S-MS^2, we want to emphasize that we are not conducting a mechanism study but rather to use pathways to study substitutions on C11. To determine whether there is a correlation between the presence of a hydroxyl group on C11 and the occurrence of this product ion of m/z 189 in MS^2, four more sulfated steroids, namely, iSS427, SS427, 21-SS429 and SS439, were prepared, and their product-ion spectra were studied to search for similar cleavages. As expected, three of the four compounds, all of which possess a hydroxyl group at C11, gave product ions of m/z 189, 191 and 193 (Figure 2-6 a, b, and c), respectively, as a result of the cleavage of the C11-C12 and C8-C14 bonds. By contrast, the product-ion spectrum of SS439 contains no such product ion, theoretically of m/z 187, owing to this cleavage, because C11 is now a carbonyl group (Figure 2-6d). Thus, we summarize this correlation between the presence of C11-OH and

the product ion of *m/z* 189/191/193 as the "C11-OH" rule, indicating that only when there's a hydroxyl group on C11 will this fragmentation happen in sulfated steroids. On the other hand, when interpreting product-ion spectra of unknown sulfated steroids, MS^3 should be conducted to fragment the *m/z* 189/191/193 ion, if present, to confirm its structure before invoking this rule and making conclusions. Although there are many other carbon sites on the steroid ring to accommodate a hydroxyl group and likely other fragmentation rules for steroids with hydroxyl group at those sites, C11, based on biological relevance, is an important position

Figure 2- 6. CID fragmentation in the negative-ion mode of the $[M - H]^-$ of (a) iSS427, (b) SS427, (c) 21-SS429, and (d) SS439.

on the steroid ring structure because it is often hydroxylated by 11β-hydroxylase during steroid hormone synthesis in mouse, and sometimes followed by further oxidation[26]. These differences are readily detected by MS^2.

Another important functional group, located at C20, is similarly deciphered by tandem MS. A derivative of many cholesterol steroids formed in steroid hormone synthesis involves reduction of cholesterol side-chain length and the hydroxylation of C21[13] by two important steroid enzymes, cholesterol 20–22 desmolase and 21-hydroxylase. C20 retains a -OH or =O after C20-C22 cleavage with 20-22 desmolase action[27], and hydroxylated C21 could be then sulfated by a

41

sulfatase to form sulfated steroids. As a result, sulfated steroids occurring biologically often differ at C20 as either a C=O or a C-OH. Again, this difference can be revealed by a product ion formed by the cleavage of bond C20-C21 and loss of a neutral molecule CH_2SO_4 to give a product $[M - H^+ - 110]^-$, which only occurs when C20 is C=O. For example, as in Figure 2-3a, the loss of 110 from the precursor ion of m/z 425 leads to a product ion of m/z 315 (relative abundance of more than 20%). Accurate mass of the m/z 315 ion from $HRMS^2$ (Table 2-2) and $^{34}S\text{-}MS^2$ (Figure 2-3b) indicates this ion forms by the loss of CH_2SO_4, allowing us to write an ion structure (Scheme 2-1a). The same fragmentation occurs for SS441. More support for this fragmentation chemistry comes from the product-ion spectra of iSS427, SS427, 21-SS429 and SS349, for which the loss of 110 occurs for the latter three (Figure 2-6 b, c, d), all of which have an C20 oxidized to C=O. In contrast, when C20 retains a hydroxyl group, this cleavage is inhibited. As shown in Figure 2-6a, no obvious product ion at m/z 317 is present. Thus, we summarize this as the "-110" rule, which addresses the correlation between C20 =O and fragmentation C20-C21 upon CID.

2.4.6 Gas Chromatography Mass Spectrometry (GC-MS)

GC combined with EI-MS is a powerful approach for the identification of unknown compounds, steroids being one important class of them. GC is particularly advantageous because it provides effective separation, which is needed for the analysis of complex biological mixtures in metabolomics. Further, EI mass spectra has always been informative in structure studies, but it is even more so with today's high-resolving power MS instruments that give accurate m/z for all fragments at speeds sufficient for GC separation. Exhaustive studies were reported by others to characterize the EI mass spectra of steroids and delineate fragmentation[28]. Furthermore, online databases of EI mass spectra have expanded tremendously during the past three decades. For

example, Wiley Registry 9th edition contains approximately 662,000 EI mass spectra of 592,000 compounds; and Wiley also published a database "Mass Spectra of Physiologically Active Substances: Including drugs, steroid hormones, and endocrine disruptors" and contains 4182 mass spectra and chemical structures of steroids. Thus, structural studies of unknown steroids can again take advantage of GC-MS, and, thus, we included it in our MS approach for unknown sulfated steroids.

Figure 2- 7. Mass spectra in the GC-MS mode for derivatized (a) SS425 and (b) SS441.

To utilize GC-MS, we had to remove chemically the sulfate group from the two standards, SS425 and SS441, and we then derivatized the products for GC/MS analysis. Solvolytic removal of the sulfate group was successful for nanomoles of samples, and the product yields were greater than 50% for both SS425 and SS441 (as determined by LC/MS). Removing the sulfate group on sulfated steroids is important for GC/MS analysis because the resulting samples are considerably more volatile. It is noteworthy, however, that solvolysis may convolute quantitative analysis for complex biological fluids for which a requirement may be to compare free versus

43

sulfated steroids. For that purpose, the combination of LC/MS and GC/MS for may be more suitable for quantitative analysis. After BSTFA/TMS derivatization of the desulfated steroids, we analyzed 0.1 nmol of sample. This low requirement of sample is compatible with studying low-abundance, urinary sulfated steroids, an application

Scheme 2- 2. EI fragmentation pathways of derivatized SS441. Ion structures are proposed only to be consistent with accurate masses and fragmentation outcomes.
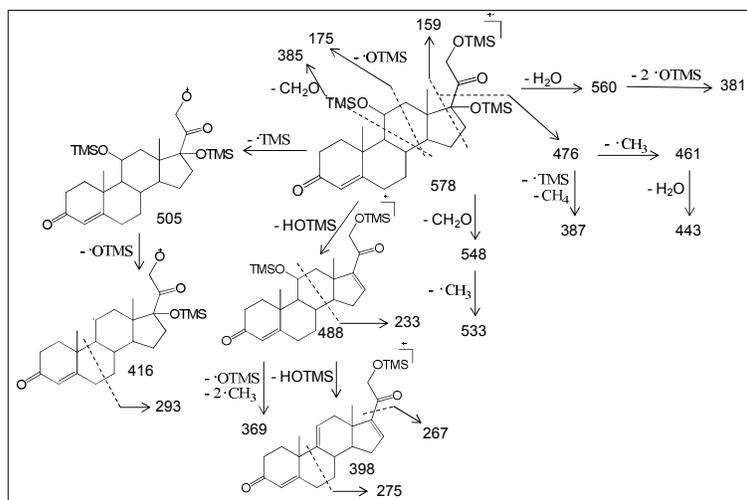
of this methodology. Accurate-mass mass spectra for SS425 and SS441 (Fig 2-7a and b) were searched in NIST MS database. A match of corticosterone bis(trimethylsilyl) ether was found for SS425, with a high probability score of 93%. However, no match was found for SS441. Although this structure may appear in other databases, we turned to the EI fragmentation pathway of SS441, based on accurate masses of fragments, to extract signature fragmentations that assist in identification. As before, the ion structures in this scheme are not established but only proposed to be consistent with accurate masses and fragmentation outcomes, and to organize our conclusions on how EI fragmentation might report on the structural details of a steroid ring. Scheme 2-2 shows that under EI fragmentation, cleavages in the ring, mostly near the functional groups, reveal more detailed information about the ring structure. Therefore, the rich fragmentation from EI-MS complements that from ESI-MS[n], and, taken together, allows one or several structure candidates to be proposed for an unknown.

2.5 Conclusions

A difficulty in assigning the structures of steroid sulfate metabolites as constituents in natural stimuli is their low abundance, precluding the application of most structural tools. To deal with this problem, we implemented a combined approach including HRMS to give the chemical formula, and HDX, MOX derivatization and reactive DESI to count various functional groups. The combination also includes MS^n and GC-MS to resolve

Figure 2- 8. Three fragmentation rules of sulfated steroids under low energy CID.

the arrangement of the functional groups and ring structure. Three major fragmentation rules were deduced in this study (Figure 2-8), the "97" rule, the "C11-OH" rule and the "-110" rule, collectively addressing the three metabolic important carbon sites for this family of compounds from the mouse. Proper chemical manipulation and derivatization permit the analysis of sulfated steroids by GC-MS to take advantage of its highly effective separation and the availability of online database searching to confirm structural assignments, taking advantage of the complementary and rich EI fragmentation. This MS-based combined approach consumes approximately 0.1 nmol of sample for ESI-MS with direct infusion when analyzed in the negative-ion mode and 1 nmol was required for simple chemical modification and derivatization for GC-MS. An outcome is that the ring topology of steroid sulfates of this type can often be well characterized, except when an important fragmentation is missing and only vague assignments can be made. In addition, although the fragmentation "rules" will likely change for other classes of steroids, this approach should be general for rare steroid metabolites, affording

one or several putative structure that can be synthesized and its MS properties compared with those of the unknown.

Another important application may be premarin, a pharmaceutical from urine of female pregnant horses, which is comprised of a large number of steroid sulfates, many still resisting identification.

2.6 Acknowledgements

2.7 References

1.   Ferrero, D. M., and Liberles, S. D. (2010) The secret codes of mammalian scents, *Wiley Interdiscip Rev Syst Biol Med 2*, 23-33.

2.   Nodari, F., Hsu, F. F., Fu, X. Y., Holekamp, T. F., Kao, L. F., Turk, J., and Holy, T. E. (2008) Sulfated steroids as natural ligands of mouse pheromone-sensing neurons, *J Neurosci 28*, 6407-6418.

3.   Hsu, F. F., Nodari, F., Kao, L. F., Fu, X. Y., Holekamp, T. F., Turk, J., and Holy, T. E. (2008) Structural Characterization of Sulfated Steroids That Activate Mouse Pheromone-Sensing Neurons, *Biochemistry-Us 47*, 14009-14019.

4.   Falkenstein, E., Tillmann, H. C., Christ, M., Feuring, M., and Wehling, M. (2000) Multiple actions of steroid hormones - A focus on rapid, nongenomic effects, *Pharmacological Reviews 52*, 513-555.

5.   Sweeley, C. C., and Horning, E. C. (1960) Microanalytical separation of steroids by gas chromatography, *Nature 187*, 144-145.

6.   Stanczyk, F. Z., and Clarke, N. J. (2010) Advantages and challenges of mass spectrometry assays for steroid hormones, *J Steroid Biochem Mol Biol 121*, 491-495.

7.   Kang, S., Park, S., Kim, M. J., Oh, S. M., Chung, K. H., and Lee, S. (2013) A sensitive and selective LC-MS/MS analysis coupled with an online sample enrichment technique for H295R steroidogenesis assay and its application in the investigation of the effect of sildenafil on steroidogenesis, *Analytical and bioanalytical chemistry 405*, 9489-9496.

8.   Fayad, P. B., Prevost, M., and Sauve, S. (2013) On-line solid-phase extraction coupled to liquid chromatography tandem mass spectrometry optimized for the analysis of steroid hormones in urban wastewaters, *Talanta 115*, 349-360.

9.   Keevil, B. G. (2013) Novel liquid chromatography tandem mass spectrometry (LC-MS/MS) methods for measuring steroids, *Best practice & research. Clinical endocrinology & metabolism 27*, 663-674.

10.   Galuska, C. E., Hartmann, M. F., Sanchez-Guijo, A., Bakhaus, K., Geyer, J., Schuler, G., Zimmer, K. P., and Wudy, S. A. (2013) Profiling intact steroid sulfates and unconjugated steroids in biological fluids by liquid chromatography-tandem mass spectrometry (LC-MS-MS), *The Analyst 138*, 3792-3801.

11.   Couchman, L., Vincent, R. P., Ghataore, L., Moniz, C. F., and Taylor, N. F. (2011) Challenges and benefits of endogenous steroid analysis by LC-MS/MS, *Bioanalysis 3*, 2549-2572.

12.   Soldin, S. J., and Soldin, O. P. (2009) Steroid hormone analysis by tandem mass spectrometry, *Clinical chemistry 55*, 1061-1066.

13.   Payne, A. H., and Hales, D. B. (2004) Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones, *Endocr Rev 25*, 947-970.

14.   Kornel, L., Kleber, J. W., and Connie, J. W. (1964) Studies on steroid conjugates: II. Chemical synthesis and characterization of sodium cortisol-21-sulfate and sodium tetrahydrocortisol-3,21-disulfate, *Steroids 4*, 67-75.

15.   Takats, Z., Wiseman, J. M., Gologan, B., and Cooks, R. G. (2004) Electrosonic spray ionization. A gentle technique for generating folded proteins and protein complexes in the gas phase and for studying ion-molecule reactions at atmospheric pressure, *Anal Chem 76*, 4050-4058.

16.   Miao, Z., and Chen, H. (2009) Direct analysis of liquid samples by desorption electrospray ionization-mass spectrometry (DESI-MS), *Journal of the American Society for Mass Spectrometry 20*, 10-19.

17.   Burstein, S., and Lieberman, S. (1958) Hydrolysis of ketosteroid hydrogen sulfates by solvolysis procedures, *J Biol Chem 233*, 331-335.

18.   Huang, G., Chen, H., Zhang, X., Cooks, R. G., and Ouyang, Z. (2007) Rapid screening of anabolic steroids in urine by reactive desorption electrospray ionization, *Anal Chem 79*, 8327-8332.

19.   Van Berkel, G. J. (1997) The electrolytic nature of electrospray, *Electrospray Ionization Mass Spectrometry, 2nd Edition; Cole, R.B., Ed; Wiley: New York, NY, USA*, 65-105.

20.   Chatman, K., Hollenbeck, T., Hagey, L., Vallee, M., Purdy, R., Weiss, F., and Siuzdak, G. (1999) Nanoelectrospray mass spectrometry and precursor ion monitoring for quantitative steroid analysis and attomole sensitivity, *Analytical Chemistry 71*, 2358-2363.

21.   Gaskell, S. J. (1988) Evaluation of the Quantitative-Analysis of a Steroid Sulfate Using Fast Atom Bombardment and Tandem Mass-Spectrometry, *Biomedical and Environmental Mass Spectrometry 15*, 99-104.

22.   Weidolf, L. O., Lee, E. D., and Henion, J. D. (1988) Determination of boldenone sulfoconjugate and related steroid sulfates in equine urine by high-performance liquid chromatography/tandem mass spectrometry, *Biomed Environ Mass Spectrom 15*, 283-289.

23.   Zhang, H., and Henion, J. (1999) Quantitative and qualitative determination of estrogen sulfates in human urine by liquid chromatography/tandem mass spectrometry using 96-well technology, *Anal Chem 71*, 3955-3964.

24.   Attygalle, A. B., Garcia-Rubio, S., Ta, J., and Meinwald, J. (2001) Collisionally-induced dissociation mass spectra of organic sulfate anions, *Journal of the Chemical Society-Perkin Transactions 2*, 498-506.

25.   Rudewicz, P., and Straub, K. M. (1986) Rapid Structure Elucidation of Catecholamine Conjugates with Tandem Mass-Spectrometry, *Analytical Chemistry 58*, 2928-2934.

26.   Domalik, L. J., Chaplin, D. D., Kirkman, M. S., Wu, R. C., Liu, W. W., Howard, T. A., Seldin, M. F., and Parker, K. L. (1991) Different isozymes of mouse 11 beta-hydroxylase produce mineralocorticoids and glucocorticoids, *Molecular endocrinology 5*, 1853-1861.

27.   Burstein, S., and Gut, M. (1976) Intermediates in the conversion of cholesterol to pregnenolone: kinetics and mechanism, *Steroids 28*, 115-131.

28.   Gaskell, S. J. (1983) Analysis of steroids by mass spectrometry, *Methods Biochem Anal 29*, 385-434.

Chapter 3*

High-Energy Collision-Induced Dissociation by MALDI TOF/TOF Causes Charge-Remote Fragmentation of Steroid Sulfates

3.1 Abstract

A method for structural elucidation of biomolecules dating to the 1980s utilized high-energy collisions (~10 keV, laboratory frame) that induced charge-remote fragmentations (CRF), a class of fragmentations particularly informative for lipids, steroids, surfactants, and peptides. Unfortunately, the capability for high-energy activation has largely disappeared with the demise of magnetic sector instruments. With the latest designs of tandem time-of-flight mass spectrometers (TOF/TOF), however, this capability is now being restored to coincide with the renewed interest in metabolites and lipids including steroid-sulfates and other steroid metabolites. For these metabolites, structure determinations are required at concentration levels below that appropriate for NMR. To meet this need, we explored CRF with TOF/TOF mass spectrometry for two groups of steroid sulfates, 3-sulfates and 21-sulfates. We demonstrated that the current generation of MALDI TOF/TOF instruments can generate charge-remote-fragmentations for these materials. The resulting collision-induced dissociation (CID) spectra are useful for positional isomer differentiation and very often allow the complete structure determination of the steroid. We also propose a new nomenclature that directly indicates the cleavage sites on the steroid ring with carbon numbers.

3.2 Introduction

Steroids are among the main regulators in mammalian physiology[1], and as a result, their metabolites are attractive candidates for conveying many aspects of an animal's status. Steroid sulfates were first recognized as naturally occurring metabolites in 1938 when Schachter and Marrian[2] isolated estrone sulfate from the urine of pregnant mares. Since then, many more steroid sulfates have been isolated from different biological sources and identified. Until the 1960s, they were considered as biologically inactive end-products of metabolism appropriate for elimination[3]. Many efforts have been made since then to search for the physiological role for steroid sulfates, and these efforts revealed the significance of sulfated steroids for structural, signaling and regulatory functions[4-8].That they are biologically active is made clear by the commonly used drug Premarin, which is a complex mixture of steroid sulfates. Recently, with the emergence of lipidomics and metabolomics, the interest in steroid conjugate analysis is again growing. An example is the work of one of the authors (T. Holy), who found steroid sulfates of the glucocorticoid family as "social odors" or "pheromones" in mouse urine[9, 10]. Given the low levels of these materials in mammalian fluids (at sub-micro molar concentration in mouse urine[9]), we need improved methods for structure determinations of steroid sulfates to deal with materials at concentrations below the levels at which NMR can be used.

In early studies of analysis of steroids and their conjugates, GC/MS with electron ionization (EI) was the dominant approach[11-13]. Although EI affords rich fragmentation, its utility is limited because most steroids require derivatization to form volatile, thermally stable analytes[14-17], sometimes compromising sensitivity. In the 1980s, the introduction of fast atom bombardment (FAB)[18] and tandem mass spectrometry (MS/MS)[19, 20], afforded a new approach for structural characterizations of steroid conjugates with mass spectrometry, but the approach required

magnetic-sector instruments and high collision energies to generate charge-remote fragmentation. The concept of charge-remote (originally called remote-site fragmentation) was first described in 1983 as observed on a unique three-sector mass spectrometer[21, 22]. This instrumentation was later used to demonstrate the charge-remote fragmentations of a series of bile acids and bile salt conjugates[23] and other steroids[24, 25]. Tomer and Gross[26] then reported in 1988 that both charge-remote and charge-proximate fragmentations occur for bile salts, steroid sulfates, and glucuronides, showing that nearly complete structure determination of steroid rings could be achieved. Unfortunately, there has been a marked decline in the application of high-energy CID with the demise of sector instruments and the current dominance of ion traps, QTOFs, and hybrid ion-trap FT spectrometers that employ low-energy CID. Although the latter has been applied over years in structural characterization of organic compounds, including steroid sulfates[10, 27], the structural elucidation is complicated and incomplete because the major fragment ions usually result from losses of small molecules, and MS^3 is often required to get a more complete set of fragment ions. Furthermore, low-energy CID spectra tend to show more rearrangement reactions and are not as reproducible as high-energy spectra. Because low-energy CID strongly depends on instrument characteristics and settings, the prospects of building an MS^2 database for structure searches has become more complex[28].

The introduction of TOF/TOF instruments, specifically designed for high-energy CID[29, 30], has triggered a "renaissance" of interest in high-energy fragmentation processes. MALDI coupled to an appropriate TOF/TOF-instrument can accommodate the need for sufficient mass resolving power, precursor-ion selection, and high collision energies for the structural analysis of biomolecules[31-34]. The first evidence comes from Trimpin et al.[31] who showed that metal-cationized fatty acids desorbed from a solvent-free preparation undergo CRF in a TOF/TOF.

Additional evidence is from Allmeier and Pittenauer[33, 34] who induced CRF of trialkylglycerols (TAGs). Other applications of MALDI TOF/TOF were also reported for carbohydrates and peptides[35, 36]. Whereas fatty acids have been investigated by MALDI TOF/TOF and their CID spectra are well understood, steroids and their conjugates have been far less investigated[37]. In this chapter, we demonstrate, for the first time to our knowledge, the application of MALDI TOF/TOF instrumentation to induce CRF and enable structural characterization of sulfated steroids. The results are comparable to those obtained on the older tandem magnetic sector instruments and strongly suggest that TOF/TOF instruments offer a new opportunity for lipid and steroid structure-determination studies and for lipidomics in general.

3.3 Experimental Procedure

3.3.1 Materials

All steroids and sulfated steroids, cholesteryl sulfate, 3β-hydroxy-5-pregnen-20-one-3-sulfate (termed SS395, for sulfated steroid 395, the number of which is the molecular mass of [M - H]⁻), etiocholan-3α-ol-11,17-dione sulfate (SS383), 4-pregnen-11β,21-diol-3,20-dione-21-sulfate (SS425), 4- pregnen-11β, 17,21-triol-3,20-dione-21-sulfate (SS441), 4-pregnen-17,21-diol-3,11,20-trione-21-sulfate (SS439), 5β-pregnan-11β,21-diol-3,20-dione, and 4-pregnen-11β, 20β,21-triol-3-one were purchased from Steraloids Inc (Newport, RI), and dissolved in methanol to give a 5 mM stock solution. SS427 and iSS427 were prepared by sulfating 5β-pregnan-11β, 21-diol-3,20-dione and 4-pregnen-11β,20β,21-triol-3-one as described previously[38]. The structures of all steroid sulfates used are listed in Figure 3-1. All other chemicals and reagents were purchased from Sigma Aldrich (St. Louis, MO).
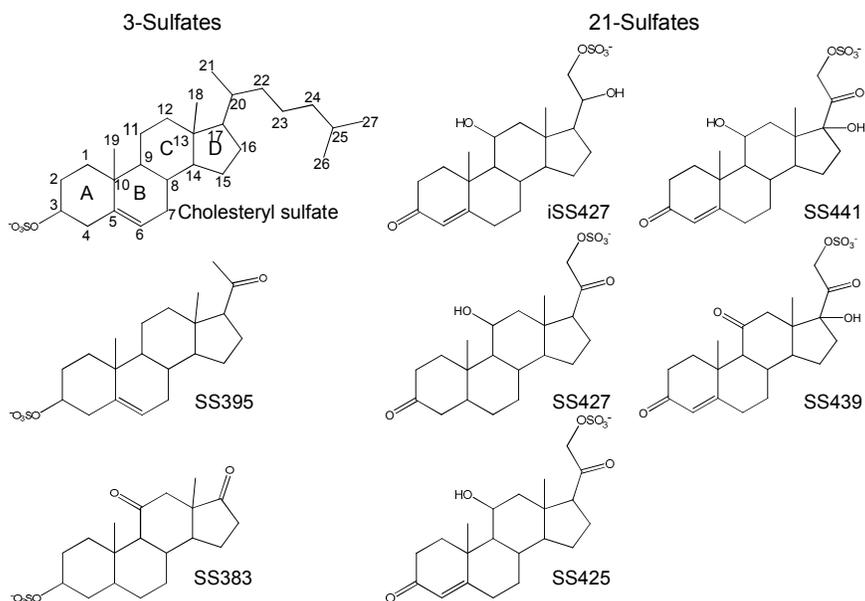
Figure 3- 1. Structures of steroid sulfates investigated.

3.3.2 Mass Spectrometry

A MALDI-TOF/TOF mass spectrometer (SpiralTOF™, JEOL Ltd., Tokyo, Japan) was used for all measurements. This MALDI-TOF/TOF mass spectrometer has unique multi-turn spiral ion optics with a 17 m long flight path to provide ultrahigh mass resolving power for MALDI-TOF measurements[39-41]. An ion gate positioned at the 15 m point in the spiral ion flight path provided mono-isotopic precursor-ion isolation for tandem mass spectrometry upon high-energy CID[42, 43]. Selected monoisotopic precursor ions undergo 20 kV collisions with the helium collision gas at a pressure of 1.7×10^{-4} Pa. Ions exiting the collision chamber underwent 9 kV post-acceleration and focusing to ensure efficient detection of low m/z fragments.

A 349 nm Nd-YLF laser operated at a laser repetition rate of 1 kHz was used to desorb and ionize the samples. The extraction delay was optimized to 50 nsec. α-Cyano-4-hydroxycinnamic acid (CHCA) matrix was dissolved in 1:1 water/acetonitrile containing 0.1% trifluoroacetic acid at a concentration of 10 mg/mL (Trifluoroacetic acid, normally removed for the negative-ion

mode detection, caused the signal intensity to decrease by two fold compared to its absence . It was utilized here to show its compatibility with negative-ion mode detection and to keep open the possibility for adoption when analyzing a biological sample mixture in both the positive- and negative-ion modes.). The steroid sulfate and matrix solutions were mixed together 1:1 by volume. A 1.0 µL aliquot of this mixture was deposited onto the MALDI target plate for analysis with the MALDI-TOF/TOF mass spectrometer.

3.4 Results and Discussion

We used MALDI TOF/TOF to investigate the fragmentation of two groups of steroid sulfates, namely 3-sulfates and 21-sulfates. These two groups were chosen because hydroxyl sites on neutral steroids suitable for sulfonation by steroid sulfotransferase are commonly present at carbons 3 and 21 (or 17) for C-21 steroids (or C-19 steroids). Furthermore, they have the sulfate group fixed at either end of the steroid[44], allowing us to demonstrate the charge-remote nature of the fragmentation. The fixed charge at one end of the precursor ion after ionization should allow CRF to be induced remotely as the predominant reaction channel under high-energy CID for both 3-sulfates and 21-sulfates. With the fixed sulfate group, these steroid sulfates can be readily desorbed and ionized by MALDI as [M – H]⁻ species (Figure 3-2a shows
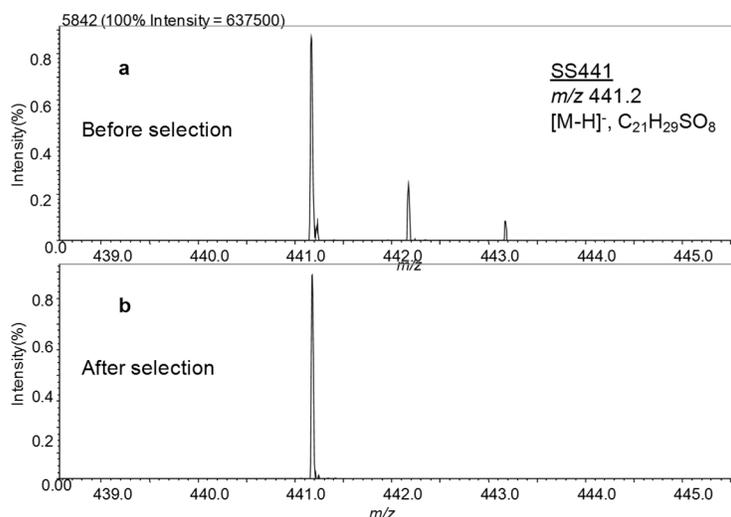
Figure 3- 2. Monoisotopic precursor-ion selection of SS441 for TOF/TOF Measurements in Spiral TOF.

one example of the negative-ion mass spectrum of SS441). With the SpiralTOF, accurate mass measurements are achieved in MS1 scans, and monoisotopic precursor-ion selection is made possible for fragmentation (Fig 3-2b), allowing straightforward interpretation of the product-ion spectra (TOF/TOF). The major fragmentations for members of each subclass are similar and often involve through-ring cleavages, as is described below.

Table 3- 1. Major fragment ions from high-energy CID of [M - H]⁻ ions of 3-sulfates
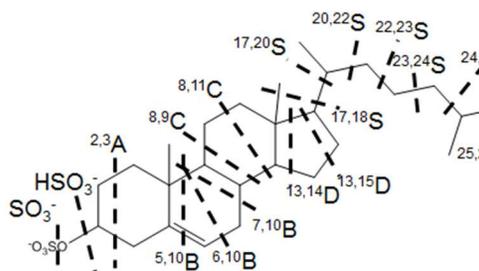
Steroid sulfate	SO_3^- HSO_3^-	$^{2,3}A$	$^{5,10}B$	$^{6,10}B$	$^{7,10}B$	$^{8,9}C$	$^{8,11}C$	$^{13,14}D$	$^{13,15}D$	$^{17,18}S$
Cholesteryl sulfate	80,97	123	N/A	N/A	217	243	257	311	325	337
SS395	80,97	123	N/A	N/A	217	243	257	311	N/A	337
SS383	80,97	123	191	204	N/A	245	N/A	327	341	N/A

3.4.1 Fragmentation Nomenclature

Given that this fragmentation by through-ring cleavages may be general, we provide a new nomenclature designating the steroid ring cleavages. Shown for cholesteryl sulfate (Table 3-1) is the nomenclature, where each cleavage is named for the ring in which it occurs and the positions of two carbon atoms constituting the two bonds that must be broken to cause a through-ring cleavage (the designated carbon atoms are to the left of the cleavage, and ring A is on the left, ring D on the right, as per convention). For example, a cleavage that breaks ring B at C9-C10 and C7-C8 is designated as $^{7,10}B$. In addition to major ring cleavages, high-mass ions are formed by cleavages of the side chain attached to ring D, remote from the charge site (Figure 3-3).

Cleavages occurring at the side chain are named by using the two carbon positions at the C-C bond being cleaved. For example, the cleavage at bond C20-C22 is named as ^{20,22}S ("S" refers to the side-chain). Note that the cleavages are denoted with a simple dashed line, because in charge remote fragmentations the charge is fixed and the product ion always contains the charge.

3.4.2 3-Sulfates

We investigated three steroids substituted with sulfate at the 3-position (A ring) by MALDI

Figure 3- 3. High-energy CID fragmentation in the negative-ion mode of 3-sulfates: a. cholesteryl sulfate, b. SS395, and c. SS383. The highly broadened [M – H]⁻ signals indicate this precursor is off scale presumably because the product-ion yield is low and dispersed over many fragments.

TOF/TOF and obtained highly informative MS/MS data, with rich fragmentation occurring through the steroid ring (Table 3-1). As indicated in the product-ion spectra (Table 3-1 and Figure 3-3), 3-sulfates undergo predominantly charge-remote fragmentation upon high-energy collisions, providing information about the nature of the substituent and its location, especially if the substituent is an alkyl chain. Taking cholesteryl sulfate as an example, we find rich fragmentation of the ring structure and the side chain at C17 (Figure 3-3a) to give a detailed

59

"foot-print" of the ring junctions and the side-chain branching of this compound. For example, the mass difference for the B-ring cleavage, ^{7,10}B, and the C ring cleavage, ^{8,9}C, is 26 Da, which corresponds to C_2H_2 and shows that the ring junction is unsubstituted. Comparing the three 3-sulfates, we find that processes ^{2,3}A, ^{8,9}C, and ^{13,14}D occur for the three compounds, and the mass differences between ^{2,3}A and ^{8,9}C are 120 Da (C_9H_{12}) for cholesteryl sulfate and SS395 whereas for SS383, the difference is 122 Da (C_9H_{14}), consistent with the absence of a double bond in the latter. Similarly, the differences between ^{8,9}C and ^{13,14}D are consistent with the substitution patterns (Table 3-1 and Figure 3-3). For cholesteryl sulfate and SS383, the differences between ^{13,14}D and ^{13,15}D are 14 Da, as expected whereas for SS395, the ^{13,15}D fragment is missing, and replaced by an ion m/z 323 formed by loss of C_4H_8O by an unknown mechanism. On the contrary, the appearance of product ions from cleavages across the B ring are significantly different for SS383 and the other two, mainly because the C5-C6 bond is a single bond for SS383 and a double bond for the other two (Figure 3-3 a, b). A fragment ion is formed by the nearby cleavage, ^{7,10}B, of both cholesteryl sulfate and SS395, although this fragment is missing in the product-ion spectrum of SS383 (Figure 3-3c). Besides B-ring cleavages, cleavages of the C ring show the expected differences among the three compounds. For example, ^{8,11}C is absent in SS383 fragmentation, consistent with the presence of the C11 keto group that prevents this cleavage (Table 3-1), and the differences between ^{8,9}C and ^{8,11}C for cholesteryl sulfate and SS395 is 14 Da, consistent with a CH_2 group at position 11. This detailed analysis of ring cleavages was not presented in the early papers on CRF of steroids, and the prospects for a nearly complete structural analysis were not understood. Importantly, branching sites on the side chain are easily determined by examining the pattern for C_nH_{2n+2} eliminations, because cleavages

happening at either side of a branch point are more facile than those occurring via double cleavages at a branch point.

Table 3- 2. Major fragment ions from high-energy CID of [M - H]⁻ ions of 21-sulfates
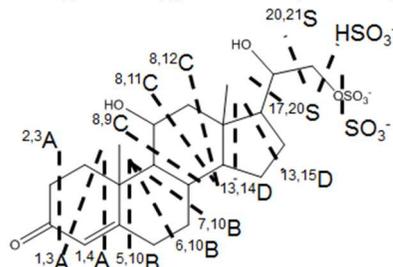
Steroid sulfate	$^{2,3}A$	$^{1,3}A$	$^{1,4}A$	$^{5,10}B$	$^{6,10}B$	$^{7,10}B$	$^{8,9}C$	$^{8,11}C$	$^{8,12}C$	$^{13,14}D$	$^{13,15}D$	$^{17,20}S$	$^{20,21}S$
iSS427	383	371	N/A	317	303	291	263	235	221	180	167	139	110
SS427	383	371	357	315	301	288	261	233	219	178	165	N/A	110
SS425	381	369	N/A	315	301	289	261	233	219	178	165	N/A	110
SS441	381	367	N/A	331	317	305	N/A	N/A	235	N/A	N/A	139	110
SS439	395	383	N/A	329	N/A	303	N/A	N/A	235	N/A	181	139	110

Note: * All 21-sulfates produce fragment ions at m/z 80 and 97, representing SO_3^- and HSO_4^- respectively.

    ** $^{2,3}A$ and $^{1,3}A$ product ions from SS441 associated with a consecutive lost of H_2O, as revealed in *m/z*.

In addition to the common fragment ions due to ring or side-chain cleavage, some unexpected fragments are also formed; these are consistent with our early findings on CRF of steroid sulfates [26]. For example, besides the losses of neutral molecules in which the C17 substituent is cleaved with H-transfer, cleavages also occur to eliminate directly a CH_3 radical, presumably forming a distonic ion, rather than H transfer and elimination of neutral molecules (this cleavage is termed $^{17,18}S$, as shown in Table 3-1). The mechanism for this cleavage was previously proposed to occur through a six-member transition state (Scheme 1 in ref 26). Although obtaining detailed mechanistic data is not the purpose of this paper and must await future work, we show the

fragmentation for clarity. Another example is the fragmentation of the three 3-sulfates to undergo

a CH₄ loss upon CID. This process most likely arises from the loss of an angular methyl group

(C18 or C19) with H rearrangement and indicates the presence of methyl substituents.

3.4.3 21-Sulfates

We investigated five different compounds with different substitutions on the ring or the side

chain, as listed in Table 3-2. Similar to 3-sulfates, 21-sulfates give rich charge-remote

fragmentation as cross-ring cleavages and of the side chain upon high-energy CID (Figure 3-4).

Ions of *m/z* 80 and 97 are also abundant in product-ion spectra of 21-sulfates, representing

SO_3^- and HSO_4^- from releases of small, charged sulfur-containing groups. Although we adopted
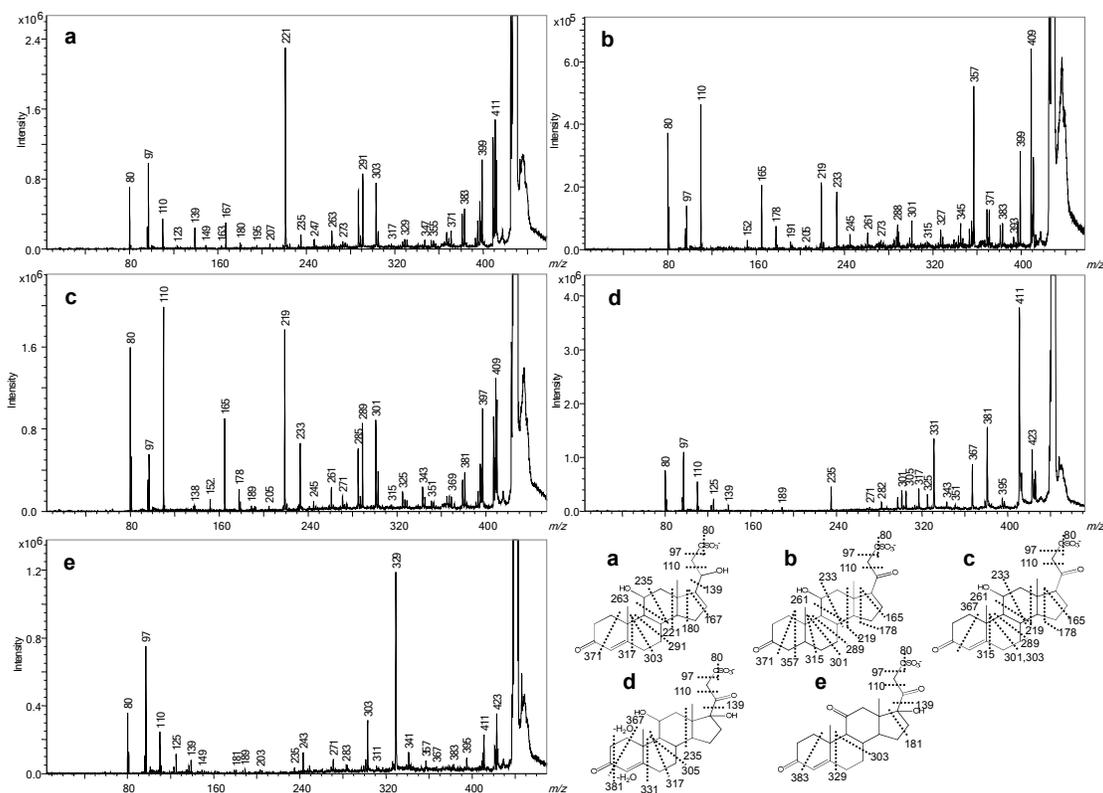
Figure 3- 4. High-energy CID fragmentation in the negative-ion mode of 21-sulfates:
a. iSS427, b. SS427, c. SS425, d. SS441, and e. SS439. The highly broadened [M −
H]⁻ signals indicate this precursor is off scale presumably because the product-ion
yield is low and dispersed over many fragments.

the same nomenclature for all the cleavages of 21-sulfates as for the 3-sulfates, the designations for 21-sulfates are of portions of the steroid that are lost as neutrals whereas for the 3-sulfates, they are part of the fragment ion. The scheme in Table 2 shows an example of the cleavages of SS427.

The CID spectra of [M - H]$^-$ ions can be used to delineate clearly the locations and type of different substitutions. For example, the positional isomers, iSS427 and SS427 differ in the position of a double bond; iSS427 has the double bond between C4 and C5, whereas SS427 has the unsaturation located at C20 as a keto group. As a result, their product-ion spectra (Figure 3-4 a, b) show that ions formed by ^{1,4}A cleavage are among the most abundant fragments formed from SS427 but undetectable for iSS427. The same types of fragmentations involving rings B, C, and D occur for both compounds, but they show a consistent mass shift of 2 Da. In addition, the C20 keto groups in SS427 and in SS425 (Figure 3-4 b, c) prevent the ^{17,20}S cleavage, but replacement of the keto group with C20-OH in iSS427 promotes this fragmentation and gives the ion of m/z 139 (Figure 3-4a). An m/z 139 ion is observed, however, although less abundant, in the product-ion spectra of SS441 and SS439 (Figure 3-4 d, e), indicating that the effect of a C20 keto group on ^{17,20}S is eased when substituted by a hydroxyl group at C17 (the only difference between SS441 and SS425). Although SS441 differs from SS425 only by the C17 hydroxyl group, it undergoes fewer cleavages of the C and D rings, suggesting that a C17 hydroxyl group suppresses cleavages of the C and D ring. Likewise, SS439 also undergoes nearly no cleavages across the C and D ring under CID.

Expulsions of neutral molecules, CH$_4$ or CO, are dominant reactions for most 21-sulfates being studied. The loss of CH$_4$ is similar to that in 3-sulfates, which likely involves an angular methyl group. Loss of CO is also a charge-remote cleavage on the A ring of 21-sulfates, which occurs

similarly at the D ring of SS383, a 3-sulfate steroid. The other two 3-sulfates undergo no such fragmentation owing to the lack of keto groups on the ring system. Besides losses of CH_4 and CO, SS441 also gives abundant ions by the loss of H_2O, which are followed by other consecutive fragmentations such as [2,3]A and [1,3]A, as indicated in Table 3-2.

We wish to note that the cleavages are not always as simple as indicated by the dashed lines. Sometimes the cleavage occurs as designated (e.g., [1,3]A, [1,4]A, [7,10]B, [8,11]C, [8,12]C, and [13,15]D) whereas other cleavages occur as designated but are accompanied by additional H transfer(s). Consider, for example, [2,3]A. A simple through-ring cleavage would give loss of C_2H_2O, but the actual fragmentation is to lose C_2H_4O. The mechanistic reasons for these differences are interesting and remain to be elucidated. Nevertheless, structural assignments can still be made in the absence of a complete understanding of the ion chemistry.

3.4.4 Approximate Detection Limits

In addition to studying the effectiveness of MS/MS fragmentation for structure elucidation, we evaluated the detection limit of the 3- and 21-sulfates on this MALDI TOF/TOF instrument, requiring sufficient fragmentation information for structure characterization. We took one example from each group of steroids, cholesteryl sulfate and SS425, and evaluated the signal-to-noise ratio (S/N) of the base peaks in their product-ion spectra when different amounts of sample were applied to the MALDI plate. As shown in Table 3-3, S/N of the base peak at *m/z* 311 in the product-ion mass spectrum of cholesteryl sulfate decreased from 200 to 80 when the applied sample amount is reduced from 25 pmol to 4 pmol. Further reduction of the sample applied to 1 pmol caused the S/N for detecting the ion of *m/z* 331 to decrease to 20. Although the intensity for the base peak detection is still acceptable with 1 pmol of sample application, some of lower-abundance product ions disappeared below the S/N at this condition (e.g., the ion of *m/z* 217),

making a structure assignment incomplete. Similarly, for SS425, S/N of the base peak at *m/z* 219 decreased from 500 to 60 in the product-ion spectrum when the sample applied was decreased from 25 to 1 pmol; further, the lower abundance fragment ions began to disappear with 1 pmol of sample application. Although the S/N tracks only approximately the applied sample quantity, we can conclude that the approximate detection limit sufficient for structure assignment of steroid sulfates is low pmol levels, with some variations for different samples. This low sample-size requirement demonstrates the applicability of MALDI TOF/TOF for analyzing biological samples where only trace amounts can be isolated (e.g., SS425 was shown to exist in mouse urine at micro-molar concentrations in a previous study[9]).

Table 3- 3. Signal-to-noise ratios of base peaks in the product-ion spectra of cholesteryl sulfate and SS425 as a function of the amount of sample applied.

Steroid sulfate	Cholestery sulfate				SS425			
Base peak (m/z)	311				219			
Applied amount (pmol)	25	10	4	1	25	10	4	1
S/N	200	100	80	20	500	400	200	60

3.5 Conclusion

The effectiveness of charge-remote-fragmentation for primary structure identification of lipids and steroids has been known for some time. The approach can determine locations of double bonds, identify various functional groups, locate branching and ring junctions directly from product-ion spectra. Instrumentation and ionization methods traditionally used for CRF, however, have become rare, diminishing the importance of CRF in structural studies. It continues to be recognized and invoked, however, as one class of fragmentation, a "thermal-like"

process that occurs with little or no involvement of a charge or radical site. Low-energy MS/MS fragmentation of steroid sulfates[45] produce fewer cross-ring cleavages and is less informative than the spectra from high-energy activation, which induces extensive fragmentation across the steroid ring at the ring junctures.

The ability to induce CRF coupled with the advantage of monoisotopic precursor-ion selection afforded by new TOF/TOF instruments, permits relatively straight-forward interpretation of product-ion spectra because all product ions are monoisotopic. These characteristic ions formed under high-energy CID are nearly identical to those observed many years ago on tandem sector instruments when the sample was introduced by FAB ionization. Cleavages through the steroid rings occur whether the steroid is a 3- or a 21-sulfate, consistent with the classification that the processes are "charge-remote" as would be expected because the intervening rigid steroid ring makes it impossible for the charge site to interact with the reaction site. Although the earlier measurements on steroids proved that the fragmentation is indeed charge-remote and showed the nature of the remote side chain, the work in this chapter also shows that there is more information in these spectra (e.g., on ring junctions and substituent locations).

In spite of the commonality of ring cleavages, 3-sulfates can be easily differentiated from 21-sulfates by the low m/z product ions. For example, 3-sulfates give a product ion of m/z 123 by ^{2,3}A cleavage, whereas 21-sulfates give product ions of m/z 110 arising from the side-chain cleavage, ^{20,21}S. 3-Sulfates containing alkyl side chains on the D ring can also be differentiated from 21-sulfates, because the former is rich in fragments in the high mass range, owing to cleavages of the side chain, to give product ions that reflect its structure and locate the branch points. Locating functional groups is also straightforward by considering the mass differences between nearby cleavages. An important caveat is that certain substitutions on the ring (e.g., a

C11 keto group) can prevent nearby fragmentations. In this case, the absence of cleavages is also informative and should be taken into account when elucidating unknown steroid structures.

We project that an important application of this approach will be in imaging steroids, lipids, and related materials.

3.6 Acknowledgements

3.7 References

1.	Falkenstein, E., Tillmann, H. C., Christ, M., Feuring, M., and Wehling, M. (2000) Multiple actions of steroid hormones--a focus on rapid, nongenomic effects, *Pharmacological reviews 52*, 513-556.

2.	Schachter, B., and Marrian, G. F. (1938) The isolation of estrone sulfate from the urine of pregnant mares, *J. Biol. Chem. 126*, 663-669.

3.	Wang, D. Y., and Bulbrook, R. D. (1968) Steroid sulphates, *Advances in Reproductive Physiology 3*, 113-146.

4.	Strott, C. A. (2002) Sulfonation and molecular action, *Endocrine reviews 23*, 703-732.

5.	Noordam, C., Dhir, V., McNelis, J. C., Schlereth, F., Hanley, N. A., Krone, N., Smeitink, J. A., Smeets, R., Sweep, F. C., Claahsen-van der Grinten, H. L., and Arlt, W. (2009) Inactivating PAPSS2 mutations in a patient with premature pubarche, *The New England journal of medicine 360*, 2310-2318.

6.	Gibbs, T. T., Russek, S. J., and Farb, D. H. (2006) Sulfated steroids as endogenous neuromodulators, *Pharmacology, biochemistry, and behavior 84*, 555-567.

7.	Strott, C. A., and Higashi, Y. (2003) Cholesterol sulfate in human physiology: what's it all about?, *Journal of lipid research 44*, 1268-1278.

8.	Hill, M., Parizek, A., Kancheva, R., and Jirasek, J. E. (2011) Reduced progesterone metabolites in human late pregnancy, *Physiological research / Academia Scientiarum Bohemoslovaca 60*, 225-241.

9.	Nodari, F., Hsu, F. F., Fu, X. Y., Holekamp, T. F., Kao, L. F., Turk, J., and Holy, T. E. (2008) Sulfated steroids as natural ligands of mouse pheromone-sensing neurons, *J Neurosci 28*, 6407-6418.

10. Hsu, F. F., Nodari, F., Kao, L. F., Fu, X. Y., Holekamp, T. F., Turk, J., and Holy, T. E. (2008) Structural Characterization of Sulfated Steroids That Activate Mouse Pheromone-Sensing Neurons, *Biochemistry-Us 47*, 14009-14019.

11. Ryhage, R., and Stenhagen, E. (1960) Mass spectrometry in lipid research, *Journal of lipid research 1*, 361-390.

12. Budzikiewicz, H., Djerassi, C., and Williams, D. H. (1964) Structure Elucidation of Natural Products by Mass Spectrometry. Vol. II, *steroids, Terpenoids, Sugars and Miscellaneous classes*.

13. Sjovall, J. (2004) Fifty years with bile acids and steroids in health and disease, *Lipids 39*, 703-722.

14. Wudy, S. A., Hartmann, M., and Homoki, J. (2002) Determination of 11-deoxycortisol (Reichstein's compound S) in human plasma by clinical isotope dilution mass spectrometry using benchtop gas chromatography-mass selective detection, *Steroids 67*, 851-857.

15. Reddy, S., Iden, C. R., and Brownawell, B. J. (2005) Analysis of steroid conjugates in sewage influent and effluent by liquid chromatography-tandem mass spectrometry, *Analytical chemistry 77*, 7032-7038.

16. Bean, K. A., and Henion, J. D. (1997) Direct determination of anabolic steroid conjugates in human urine by combined high-performance liquid chromatography and tandem mass spectrometry, *Journal of chromatography. B, Biomedical sciences and applications 690*, 65-75.

17.   Bowers, L. D., and Sanaullah. (1996) Direct measurement of steroid sulfate and glucuronide conjugates with high-performance liquid chromatography-mass spectrometry, *Journal of chromatography. B, Biomedical applications 687*, 61-68.

18.   Barber, M., Bordoli, R. S., Sedgwick, R. D., and Tyler, A. N. (1981) Fast Atom Bombardment of Solids (Fab) - a New Ion-Source for Mass-Spectrometry, *J Chem Soc Chem Comm 7*, 325-327.

19.   Cooks, R. G., and Glish, G. L. (1981) Mass-Spectrometry Mass-Spectrometry, *Chem Eng News 59*, 40-52.

20.   Levsen, K., and Schwarz, H. (1983) Gas-Phase Chemistry of Collisionally Activated Ions, *Mass Spectrom Rev 2*, 77-148.

21.   Gross, M. L., Chess, E. K., Lyon, P. A., Crow, F. W., Evans, S., and Tudge, H. (1982) Triple Analyzer Mass-Spectrometry for High-Resolution Ms/Ms Studies, *Int J Mass Spectrom 42*, 243-254.

22.   Tomer, K. B., Crow, F. W., and Gross, M. L. (1983) Location of Double-Bond Position in Unsaturated Fatty-Acids by Negative-Ion Ms/Ms, *J Am Chem Soc 105*, 5487-5488.

23.   Tomer, K. B., Jensen, N. J., Gross, M. L., and Whitney, J. (1986) Fast atom bombardment combined with tandem mass spectrometry for determination of bile salts and their conjugates, *Biomedical & environmental mass spectrometry 13*, 265-272.

24.   Cheng, M. T., Barbalas, M. P., Pegues, R. F., and Mclafferty, F. W. (1983) Tandem Mass-Spectrometry - Structural and Stereochemical Information from Steroids, *J Am Chem Soc 105*, 1510-1513.

25.   Kingston, E. E., Beynon, J. H., Newton, R. P., and Liehr, J. G. (1985) The differentiation of isomeric biological compounds using collision-induced dissociation of ions generated by fast atom bombardment, *Biomedical mass spectrometry 12*, 525-534.

26.   Tomer, K. B., and Gross, M. L. (1988) Fast atom bombardment and tandem mass spectrometry for structure determination: remote site fragmentation of steroid conjugates and bile salts, *Biomedical & environmental mass spectrometry 15*, 89-98.

27.   Strahm, E., Saudan, C., Sottas, P. E., Mangin, P., and Saugy, M. (2007) Direct detection and quantification of 19-norandrosterone sulfate in human urine by liquid chromatography-linear ion trap mass spectrometry, *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences 852*, 491-496.

28.   Werner, E., Heilier, J. F., Ducruix, C., Ezan, E., Junot, C., and Tabet, J. C. (2008) Mass spectrometry for the identification of the discriminating signals from metabolomics: current status and future trends, *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences 871*, 143-163.

29.   Cotter, R. J. (2013) High energy collisions on tandem time-of-flight mass spectrometers, *Journal of the American Society for Mass Spectrometry 24*, 657-674.

30.   Cornish, T. J., and Cotter, R. J. (1993) A curved-field reflectron for improved energy focusing of product ions in time-of-flight mass spectrometry, *Rapid communications in mass spectrometry : RCM 7*, 1037-1040.

31.   Trimpin, S., Clemmer, D. E., and McEwen, C. N. (2007) Charge-remote fragmentation of lithiated fatty acids on a TOF-TOF instrument using matrix-ionization, *Journal of the American Society for Mass Spectrometry 18*, 1967-1972.

32.   Sun, G., Yang, K., Zhao, Z., Guan, S., Han, X., and Gross, R. W. (2007) Shotgun metabolomics approach for the analysis of negatively charged water-soluble cellular metabolites from mouse heart tissue, *Analytical chemistry 79*, 6629-6640.

33.   Pittenauer, E., and Allmaier, G. (2009) The renaissance of high-energy CID for structural elucidation of complex lipids: MALDI-TOF/RTOF-MS of alkali cationized triacylglycerols, *Journal of the American Society for Mass Spectrometry 20*, 1037-1047.

34.   Pittenauer, E., and Allmaier, G. (2009) High-energy collision induced dissociation of biomolecules: MALDI-TOF/RTOF mass spectrometry in comparison to tandem sector mass spectrometry, *Combinatorial chemistry & high throughput screening 12*, 137-155.

35.   Narouz, M. R., Soliman, S. E., Fridgen, T. D., Nashed, M. A., and Banoub, J. H. (2014) High-energy collision-induced dissociation tandem mass spectrometry of regioisomeric lactose palmitic acid monoesters using matrix-assisted laser desorption/ionization, *Rapid communications in mass spectrometry : RCM 28*, 169-177.

36.   Baum, F., Fedorova, M., Ebner, J., Hoffmann, R., and Pischetsrieder, M. (2013) Analysis of the endogenous Peptide profile of milk: identification of 248 mainly casein-derived peptides, *Journal of proteome research 12*, 5447-5462.

37.   Hailat, I., and Helleur, R. J. (2014) Direct analysis of sterols by derivatization matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and tandem mass spectrometry, *Rapid communications in mass spectrometry : RCM 28*, 149-158.

38.   Yan, Y., Rempel, D. L., Holy, T. E., and Gross, M. L. (2014) Mass Spectrometry Combinations for Structural Characterization of Sulfated-Steroid Metabolites, *Journal of the American Society for Mass Spectrometry 25*, 869-879.

39.   Satoh, T., Tsuno, H., Iwanaga, M., and Kammei, Y. (2005) The design and characteristic features of a new time-of-flight mass spectrometer with a spiral ion trajectory, *Journal of The American Society for Mass Spectrometry 16*, 1969-1975.

40.   Satoh, T., Tsuno, H., Iwanaga, M., and Kammei, Y. (2006) A new Spiral Time-of-Flight Mass Spectrometer for High Mass Analysis, *J. Mass Spectrom. Soc. Jpn. 541*, 11-17.

41.   Satoh, T., Sato, T., and Tamura, J. (2007) Development of a high-Performance MALDI-TOF mass spectrometer utilizing a spiral ion trajectory, *Journal of The American Society for Mass Spectrometry 18*, 1318-1323.

42.   Satoh, T., Sato, T., Kubo, A., and Tamura, J. (2011) Tandem Time-of-Flight Mass Spectrometer with High Precursor Ion Selectivity Employing Spiral Ion Trajectory and Improved Offset Parabolic Reflectron, *Journal of The American Society for Mass Spectrometry 22*, 797-803.

43.   Kubo, A., Satoh, T., Itoh, Y., Hashimoto, M., Tamura, J., and Cody, R. (2013) Structural Analysis of Triacylglycerols by Using a MALDI-TOF/TOF System with Monoisotopic Precursor Selection, *Journal of The American Society for Mass Spectrometry 24*, 684-689.

44.   Strott, C. A. (1996) Steroid sulfotransferases, *Endocrine reviews 17*, 670-697.

45.   Yan, Y., Rempel, D. L., Holy, T. E., and Gross, M. L. (2014) Mass Spectrometry Combinations for Structural Characterization of Sulfated-Steroid Metabolites, *Journal of the American Society for Mass Spectrometry*.

Chapter 4

Structural Characterization of A Novel Steroid Metabolite:

A Molecular Code for Identity in the Vomeronasal System

4.1 Abstract

In social interactions, individuals are recognized by visual, auditory, or olfactory cues, but the fundamental basis of recognition is poorly understood. Olfactory recognition exploits emitted chemical cues, but identifying the key signals among thousands of compounds remains a major challenge. To address this need, Holy et al. developed a new technique, Component-Activity Matching, to select candidate ligands that "explain" patterns of bioactivity across diverse complex mixtures. Using this technique, 23 components from mouse urine were found as candidates to explain firing rates in seven of eight functional classes of vomeronasal sensory neurons. And one putative female sex pheromone, was identified to stimulate a class of neurons selective for females. This compound accounted for much of the neuronal activity of some female strains, and greatly enhanced investigatory behavior of males. Here, we report the structural characterization of this unknown compound. Mass spectrometric approaches, including high-resolution mass spectrometry, HDX and other chemical derivatization followed by ESI mass spectrometry, and tandem mass spectrometry are used to identify this compound, which we named as cortigynic acid. The structural identity of the compound is later confirmed with custom synthesis based on the proposed structure.

4.2 Introduction

Mammals explore the chemical world with several olfactory modalities[1]. The accessory olfactory system (AOS) emphasizes the detection of social cues, sometimes called pheromones, which regulate behavior among members of the same species[2, 3]. Behavioral data have shown that pheromones are present in several scent sources: secretory glands (lacrimal, salivary, and preputial), urine, and feces[2, 4]. Mouse urine excites widespread activity among vomeronasal sensory neurons (VSNs)[5-8] and is the best behaviorally characterized source of chemical cues for mammalian social communication[2]. Although recent progress has been made toward identifying the molecular nature of pheromone cues by purifying individual ligands from mouse urine[9-11], the identities of olfactory ligands inside urine cues are largely unknown.

Recently, Holy lab developed a novel approach, Component-Activity Matching (CAM), for systematically and exhaustively defining a "short list" of candidate vomeronasal ligands that encode identity. This approach exploits the striking differences among natural stimuli across different sexes and strains of mice[8] in a forward screen for compounds that may drive activity in vomeronasal neurons, without the need for laborious sample purification. The hypothesis is that these different neuronal activity patterns could be caused by different ligands present in one or more of urine samples. Thus, the molecular correlates of identity might therefore be discoverable by seeking compounds whose abundance could "explain" these patterns of physiological responses. Using this approach, a small set of constituents whose concentrations match a pattern of neuronal responsiveness across samples were identified. Later, they focused on a class of neurons that responded to urine from females of all strains, which did not respond to any male strains. One best CAM candidate that explains this pattern of neuronal activity was purified. It is proven to account for one-fourth of neuronal activity of female C57BL/6J mouse urine.

In this chapter, we report the structural characterization of this best CAM candidate, which we call M377, with mass spectrometry. Different mass spectrometry methods are used, including high-resolution MS, tandem MS and HDX MS. Based on the proposed structure, a standard compound was custom-synthesized and used to confirm the identity of the unknown compound. Preliminary behavior studies indicate that this M377 (cortigynic acid) is a sex pheromone that is unique in female mouse urine. The identification of its structure introduces a new type of molecular cues for identity in a mammalian olfactory system.

4.3 Experimental

4.3.1 Materials

1, 4-pregnadien-11β, 17, 20-triol-3-one 21-carboxylic acid were purchased from Steraloids inc (Newport, RI), All other chemicals and reagents were purchased from Sigma Aldrich (St. Louis, MO).

4.3.2 Hydrogen/Deuterium Exchange (H/DX)

H/DX of M377 was initiated by dissolving a dried sample in D_2O at neutral pH, to give a sample concentration that, based on ion count of similar structures, was ~ 2 μM. The same volume of acetonitrile was added prior to conducting ESI MS.

4.3.3 Periodic Acid Reaction

The periodic acid reaction with M377, as well as a comparison standard 1, 4-pregnadien-11β, 17, 20-triol-3-one 21-carboxylic acid (Steraloids, Newport, RI), was initiated by mixing 20 μL of sample (~1 μM in MeOH) with 80 uL of 0.5% periodic acid. The reaction mixtures were then shaken for 20 minutes before nanoLC/MS analysis.

4.3.4 Mass Spectrometry

Stock solution of M378 and synthetic compound was diluted to ~1 µM and infused to an ESI source at 3 µL/min. ESI-MS were conducted in both the negative-ion mode on a LTQ Orbitrap Velos and the positive-ion mode on a LTQ Orbitrap from Thermo Scientific (West Palm Beach, FL), equipped with the Xcalibur operating system. In the negative ion mode, spray voltage was set at 3.5 kV; capillary voltage and temperature was at -32 V and 275 ºC; and Tube lens was -222 V. In the positive ion mode, spray voltage was set at 4.5 kV; capillary voltage and temperature was at 49 V and 250 ºC; and Tube lens was 100V. Tandem MS (MSn) experiments were carried out after low energy CID by using a relative collision energy ranging from 30 to 40% of the maximum.

4.4 Results and Discussion

4.4.1 Chemical Formula Deduction of M377 with High Resolution MS

High-resolution mass spectra (HR-MS) and H/DX mass spectra were obtained to determine elemental composition and the presence of particular chemical groups. HR-MS in the negative-ion mode gave an
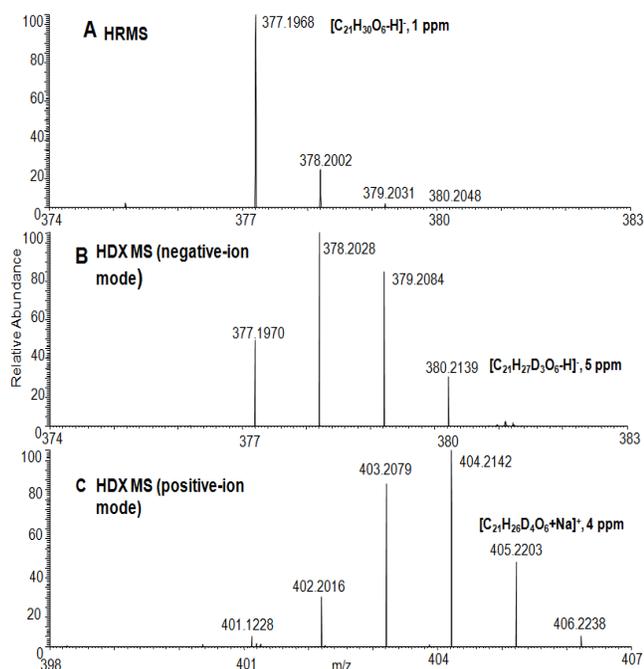
Figure 4- 1. HRMS mass spectra for M377 (A) in the negative ion mode (B) in the negative ion mode after HDX (C) in the positive ion mode after HDX.

78

accurate mass of 377.1968 for molecular ion [M-H]⁻ (Figure 4-1A), consistent within 1ppm with the formula $C_{21}H_{29}O_6^-$. This chemical formula is also supported by the isotopic peak distribution, with the ^{13}C peak being ~20% of the ^{12}C peak. H/DX MS was performed in both negative and positive ion modes, showing that the maximum mass shift after hydrogen/deuterium exchange is 3Da in negative ion mode and 4 Da in positive ion mode. This indicates that there are four exchangeable sites (hydroxyl groups) in the molecule, of which one gets ionized in the negative ion mode for detection, suggesting the presence of a carboxylic group (Figure 4-1B and C).

4.4.2 Structural Characterization of M377 with Tandem MS

For further structural elucidation of M377, low-energy negative-ion collision-induced-dissociation (CID) spectra were acquired at both low and high resolving power (Figure 4-2A and B), allowing elemental composition to be calculated for product ions (Table 4-1). For example, one of the product ions at m/z 189.0922 is calculated to be ($C_{12}H_{13}O_2^-$), with mass accuracy within 1ppm. Interestingly, this product ion was also observed in a previous study [11], so here we performed MS³ on this product ion. The result (Figure 4-3A) showed an identical MS³ spectrum. Thus, the ion of m/z 189 ($C_{12}H_{13}O_2^-$) has the structure depicted
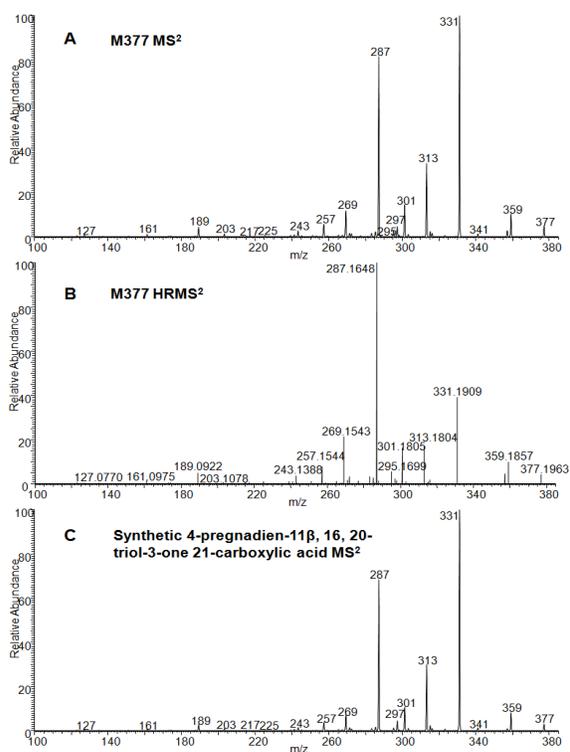
Figure 4- 2. CID fragmentation in the negative ion mode of (A) M377, (B) M377 with accurate mass detection, and (C) 4-pregnadien-11β, 16, 20-triol-3-one 21-carboxylic acid.

79

in Scheme 4-1. Another example is one of product ions at m/z 331.1909 ($C_{20}H_{27}O_4^-$) (Figure 4-3B), formed by the loss of CH_2O_2 by cleavage of the C20-C21 bond. This fragment ion supports our hypothesis of a -COOH group at this position. Similarly, an ion of m/z 301.1805 (Figure 4-3C) is attributed to the loss of $C_2H_4O_3$, from the cleavage of the C17-C20 bond, which could be explained by a consecutive C20-CH(OH)- group next to the C21–COOH group. This interpretation is also supported by the MS^3 data of ions at m/z 331, which led to fragment ions at m/z 301 (Figure 4-3B). Also, given the known structure of the ion of m/z 189 ($C_{12}H_{13}O_2^-$), the fragment ion
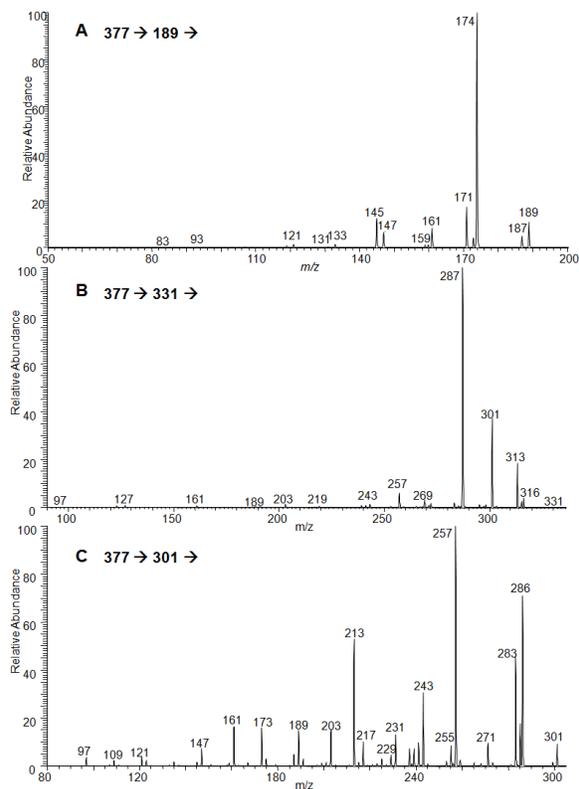
Figure 4- 3. MS^3 in the negative ion mode of (A) product ion at m/z 189 from M377 MS^2 (B) product ion at m/z 331 from M377 MS^2 (C) product ion at m/z 301 from M377 MS^2.

of m/z 203 ($C_{13}H_{15}O_2^-$) can be assigned as the cleavage in C ring which differs by one carbon site from that producing m/z 189. Similarly, fragment ions of m/z 243 ($C_{16}H_{19}O_2^-$) and m/z 257 ($C_{17}H_{21}O_2^-$) were formed by cleavages of the D ring. When compared to ions of m/z 189, this indicated that there were no –OH group substitutions on C12, C13, C14 or C15 (Scheme 4-1).

Thus, the location of each functional group can be assigned as indicated in Scheme 4-1, with one exception: the –OH group located either on C16 or C17. Based on the difference (C_2H_4O) of product ions of m/z 257 and 301, the cleavage sites that led to them can only be separated by two carbons, C16 and C17, one of which bears the last unassigned –OH group. This interpretation is

also supported by MS^3 analysis of an ion of *m/z* 301 (Figure 4-3C), which showed an abundant

Table 4- 1. High-resolution Mass Measurements of Major Fragment ions.

measured mass (Da)	calculated mass (Da)	elemental composition of ion	Elemental composition of lost molecule
377.1963	377.1970	$C_{21}H_{29}O_6^-$	N/A
359.1857	358.1864	$C_{21}H_{27}O_5^-$	H_2O
331.1909	331.1915	$C_{20}H_{27}O_4^-$	CH_2O_2
313.1804	313.1809	$C_{20}H_{25}O_3^-$	CH_4O_3
301.1805	301.1809	$C_{19}H_{25}O_3^-$	$C_2H_4O_3$
295.1699	295.1704	$C_{20}H_{23}O_2^-$	CH_6O_4
287.1648	287.1653	$C_{18}H_{23}O_3^-$	$C_3H_6O_3$
269.1543	269.1547	$C_{18}H_{21}O_2^-$	$C_3H_8O_4$
257.1544	257.1547	$C_{17}H_{21}O_2^-$	$C_4H_8O_4$
243.1388	243.1391	$C_{16}H_{19}O_2^-$	$C_5H_{10}O_4$
203.1078	203.1078	$C_{13}H_{15}O_2^-$	$C_8H_{14}O_4$
189.0922	189.0921	$C_{12}H_{13}O_2^-$	$C_9H_{16}O_4$
161.0975	161.0972	$C_{11}H_{13}O^-$	$C_{10}H_{16}O_5$

fragment of *m/z* 257. However, due to the lack of additional fragmentation between the C16 and

C17, MS^2 gave little indication of which of these locations should be preferred. Therefore, we

carried out a chemical degradation with periodic acid, which reacts with 1,2-diols by cleaving the

C-C bond and forming an aldehyde. Consequently, periodic acid would react with M377 if the –

OH was on C17 (forming a 1,2-diol at C17 and C20), but not if the –OH was on C16. Both

M377 and a standard compound 1, 4-pregnadien-11β, 17, 20-triol-3-one 21-carboxylic acid

(P0750-000, Figure 4-4A) were reacted with periodic acid. P0750-000 served as a positive

control, and the chemical derivatization products were detected by positive-ion mode MS. As expected, P0750-000 reacted with periodic acid and generated a reaction product of m/z 301.1798 $(C_{19}H_{24}O_3)^+$ (Figure 4-4A). In contrast, M377 did not react with periodic acid, with only the unreacted $[M+H]^+$ of m/z 379.2115 present (Figure 4-4B). Therefore, we confirmed that, unlike P075-000, M377 has a 1,3–diol instead of a 1,2–diol. Since C20 has a –OH group (as described above), the unassigned –OH group must be positioned at C16 instead of C17. Therefore, we proposed the structure of M377 to be 4-pregnadien-11, 16, 20-triol-3-one 21-carboxylic acid, also known as 16-hydroxycorticosterone 20-hydroxy-21-acid.

To test the validity of this structural interpretation, we obtained a custom synthesis of this proposed compound, as a epimeric mixture at carbon 20. Product-ion mass spectra of the synthetic and endogenous compounds are nearly indistinguishable (Figure 4-2A and C), and MS^3 on major MS^2 fragment ions from both compounds are also consistent (data not shown), indicating identical structures of the two. Further, the two stereoisomeric components of the synthetic epimeric mixture were quantified by HPLC as having a concentration ratio of 10:1; of which the less abundant component matched the endogenous compound in elution time and

Scheme 4- 1. Fragmentation processes proposed for the $[M-H]^-$ ion of M377.
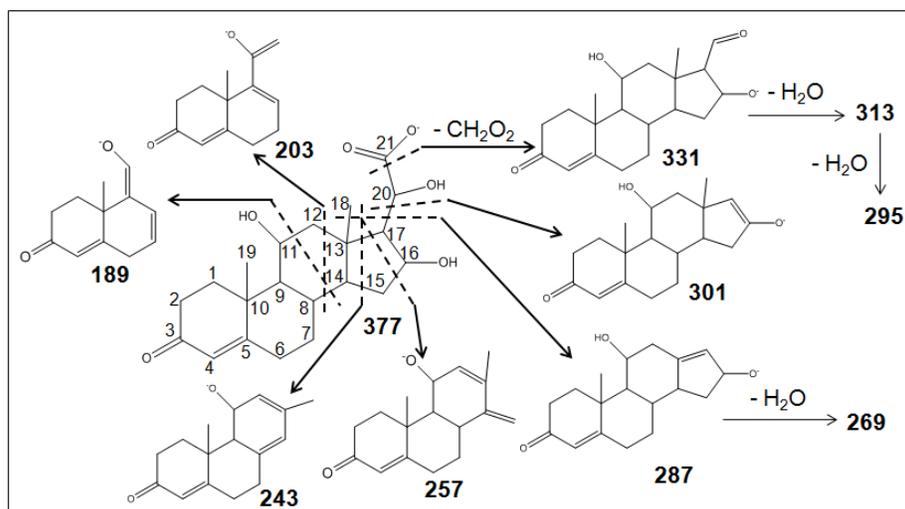
showed the same neuronal activity as endogenous M377. The structure of M377 has several distinctive features. To our knowledge, biological activity of a steroid carboxylic acid has not been previously reported. As a 21-carbon steroid with a hydroxyl group at the 11th position, this structure likely derives from metabolism of a glucocorticoid[(12)]. However, the hydroxyl group at

Figure 4- 4. HRMS mass spectra for products of periodic reaction with (a) 1, 4-pregnadien-11β, 17, 20-triol-3-one 21-carboxylic acid, (b) M377.

position 16 is a motif characteristic of a female sex steroid, estriol, which in humans is only produced in significant amounts during pregnancy[(13)]. We therefore named this new VSN ligand "cortigynic acid."

4.5 Conclusion

The structural elucidation of M377 as a vomeronasal code for identity helped support the newly developed strategy, CAM, as a validated fast unbiased screen for VNO ligands. Without requiring any purification, CAM identified a single compound, out of 1,634 possibilities, to explain responses of a neuronal type selective for urine from BALB/c females. Structural analysis of this compound, M377, to be a novel ligand, cortigynic acid. One of the most intriguing aspects of this structure is the presence of a hydroxyl group on C16, a motif known

from the estrogen compound estriol. Unlike the C17 position, a hydroxyl group at C16 is less common among urinary metabolites. One possible source for the enzymatic activity that adds this hydroxyl to C16 of a steroid is a P450 16 hydroxyl enzyme[14]. In fact, this 16α-hydroxylase activity in mouse liver microsomes was found in many female inbred mice, including the BALB/c and C57BL/6J mouse strains used in our current study.

Preliminary behavioral study showed cortigynic acid to be a potential first cue for encoding "sex" from urinary cues for the mouse vomeronasal system. Cortigynic acid was present at high concentration only in intact adult female mice urine, and was relatively constant across strains. In behavioral tests, cortigynic acid substantially and significantly increased male investigatory behavior. When added to male urine, this one ligand increased the duration of investigation bouts by more than two-fold. Based on results with a close structural analog, we speculate that the specific interest in this compound integrates multiple structural features, including the 16-hydroxyl moiety specific to female mice. This is the first known female-selective ligand in mouse urine. These results take a major step towards a comprehensive understanding of the molecular code for identity in the vomeronasal system.

4.5 Acknowledgements

4.6 References

1. Ma, M. (2007) Encoding olfactory signals via multiple chemosensory systems, *Critical reviews in biochemistry and molecular biology 42*, 463-480.

2. Halpern, M., and Martinez-Marcos, A. (2003) Structure and function of the vomeronasal system: an update, *Progress in neurobiology 70*, 245-318.

3. Dulac, C., and Torello, A. T. (2003) Molecular detection of pheromone signals in mammals: from genes to behaviour, *Nature reviews. Neuroscience 4*, 551-562.

4. Kimoto, H., Haga, S., Sato, K., and Touhara, K. (2005) Sex-specific peptides from exocrine glands stimulate mouse vomeronasal sensory neurons, *Nature 437*, 898-901.

5. Holy, T. E., Dulac, C., and Meister, M. (2000) Responses of vomeronasal neurons to natural stimuli, *Science 289*, 1569-1572.

6. He, J., Ma, L., Kim, S., Nakai, J., and Yu, C. R. (2008) Encoding gender and individual information in the mouse vomeronasal organ, *Science 320*, 535-538.

7. Stowers, L., Holy, T. E., Meister, M., Dulac, C., and Koentges, G. (2002) Loss of sex discrimination and male-male aggression in mice deficient for TRP2, *Science 295*, 1493-1500.

8. Tolokh, II, Fu, X., and Holy, T. E. (2013) Reliable sex and strain discrimination in the mouse vomeronasal organ and accessory olfactory bulb, *The Journal of neuroscience : the official journal of the Society for Neuroscience 33*, 13903-13913.

9. Nodari, F., Hsu, F. F., Fu, X., Holekamp, T. F., Kao, L. F., Turk, J., and Holy, T. E. (2008) Sulfated steroids as natural ligands of mouse pheromone-sensing neurons, *The Journal of neuroscience : the official journal of the Society for Neuroscience 28*, 6407-6418.

10.   Chamero, P., Marton, T. F., Logan, D. W., Flanagan, K., Cruz, J. R., Saghatelian, A., Cravatt, B. F., and Stowers, L. (2007) Identification of protein pheromones that promote aggressive behaviour, *Nature 450*, 899-902.

11.   Hsu, F. F., Nodari, F., Kao, L. F., Fu, X., Holekamp, T. F., Turk, J., and Holy, T. E. (2008) Structural characterization of sulfated steroids that activate mouse pheromone-sensing neurons, *Biochemistry 47*, 14009-14019.

12.   Wang, T., Shah, Y. M., Matsubara, T., Zhen, Y., Tanabe, T., Nagano, T., Fotso, S., Krausz, K. W., Zabriskie, T. M., Idle, J. R., and Gonzalez, F. J. (2010) Control of steroid 21-oic acid synthesis by peroxisome proliferator-activated receptor alpha and role of the hypothalamic-pituitary-adrenal axis, *The Journal of biological chemistry 285*, 7670-7685.

13.   Raju, U., Bradlow, H. L., and Levitz, M. (1990) Estriol-3-Sulfate in Human Breast Cyst Fluid - Concentrations, Possible Origin, and Physiological Implications, *Annals of the New York Academy of Sciences 586*, 83-87.

14.   Noshiro, M., Lakso, M., Kawajiri, K., and Negishi, M. (1988) Rip locus: regulation of female-specific isozyme (I-P-450(16 alpha) of testosterone 16 alpha-hydroxylase in mouse liver, chromosome localization, and cloning of P-450 cDNA, *Biochemistry 27*, 6434-6443.

Chapter 5

Introduction: Protein-Protein Interactions, Hydrogen/Deuterium Exchange, Fast Photochemical Oxidation of Proteins for Mass Spectrometry-Based Protein Biophysics

5.1 Abstract

Knowledge about the structural and biophysical properties of proteins when alone or in complexes with other ligands in solution is important in understanding many biological processes. Such knowledge is also needed for therapeutic drug development. Recent evolution and improvement of MS-based approaches are making it powerful for the study of protein chemistry. Two most popular MS-based protein footprinting strategies are hydrogen deuterium exchange and fast photochemical oxidation of proteins, which enable the study the biophysical properties of proteins in solution as well as their interactions with binding partners. The principles of these two techniques are introduced in this chapter as a prelude to the Chapter 6 and 7, where we report their application in some protein biophysics studies.

5.2 Protein-Protein Interaction

Protein-protein interactions (PPIs) play a central role in regulating many biological processes. The interactions are fundamentally characterized as stable or transient, both of which can be weak or strong. The most common stable protein interactions are between the subunits of a protein complex, which in many cases can now be isolated and purified. The transient interactions often require changes in conditions, such as phosphorylation or external factors that affect conformations, to trigger the interaction. When in contact with their binding partners, transiently interacting proteins control many cellular processes[1], including protein modification, transport, folding, signaling, and cell cycling.

Proteins bind to each other via a combination of chemical forces, including van der Waals forces, hydrophobic interactions, and salt bridges. The binding interfaces could occur over large surfaces spanning hundreds of residues or be represented by a small peptide formed by proteolysis. The binding affinity is influenced by, but not limited to, the types of interaction and the size of the binding domain. Other factors, such as ionic strength, pH, and presence of other ligands, may also affect the binding affinity.

Malfunction of the protein-protein interactions can result in disease[2, 3]. Because PPI plays a significant role in life cycles, considerable efforts have been made to discover what features of a protein are "druggable"; and modulation of PPI has emerged as a promising strategy to treat human diseases and is receiving growing attention by the scientific community. Many of the drugs on the market were developed by using these principles, and they treat a multitude of diseases.

As the foundation of drug design, PPI and its understanding are key to drug development. In recently years, the technologies to characterize protein-protein interactions have rapidly matured,

one of which being mass spectrometry-based approaches[4]. These methods not only facilitate drug development, but also help to improve our knowledge in biological networks.

5.3 MS-Based Protein-Protein Interaction Studies

A common strategy that is adopted in mass spectrometry-based methods for investigating the conformational properties of proteins in solution is to introduce covalent modifications into proteins and use ESI or MALDI MS, often coupled with separation, to read out the results[5]. More specifically, this strategy involves measuring the labeling reaction rate at local regions of a protein to ascertain information about the local solvent accessibility and conformation. This approach can probe protein conformational changes associated with protein-protein or protein-ligand interactions and protein dynamics. In the last decade, a variety of different covalent labeling techniques have been

Figure 5- 1. Three types of hydrogens in proteins. (1) black: amide hydrogens; (2)white: hydrogens attached to sidechain functional groups; (3) hydrogens bonded to carbons.

developed and applied in this area of study[6-8]. The work described here in the second part of the thesis focuses on two covalent labeling methods combining with MS in studying protein biophysics, namely, hydrogen deuterium exchange (HDX) and fast photochemical oxidation of proteins (FPOP).

5.4 Hydrogen Deuterium Exchange

Hydrogen deuterium exchange measures the integration of deuterium into a protein(s) by relying on the natural phenomenon of hydrogen exchange. It was traditionally used in conjunction with

NMR analysis[9], and its use with MS was implemented shortly after the development of ESI[10, 11]. Important developments in HDX MS have followed through the years[7, 12-19], now it is well-established and more useful than expected in protein biophysics studies, offering unparalleled limits of detection, low sample consumption, and the promise of single amino-acid resolution. HDX monitors the exchange rate of amide hydrogens on the protein backbone. There are three types of hydrogens[20] in the protein, namely the amide hydrogen, the hydrogens attached to the functional groups of side chain, and the hydrogens covalently bonded to carbon (Figure 5-1[21]). The exchange rates of these three types of hydrogens are significantly different. Hydrogens attached to the functional groups of side chains exchange fastest (half lives of ms), and are easily back-exchanged to hydrogen in HDX by returning the protein to a protic solvent. On the contrary, hydrogens bonded to carbon atoms are essentially not exchangeable with deuterium in D_2O. The exchange rate of amide hydrogens is in the middle (half lives of seconds to hours); therefore, they are measurable on a lab time-scale. The characteristics of HDX allow us to measure readily the extent of deuterium uptake of amide hydrogens on the backbone and monitor protein dynamics.

5.4.1 Effect of pH and Temperature on HDX rates

The HDX reaction can be catalyzed by either acid or base. Thus, the rate of HDX is sensitive to pH -- one unit change in pH leads to ten-fold change in the exchange rate[22]. Although the intrinsic exchange rates of backbone hydrogens are different and dependent on solvent accessibility and hydrogen bonding, the overall exchange rate reaches the slowest value near pH 2.5 (Figure 5-2A). Furthermore, like most other reactions, temperature also affects the rate of HDX[23, 24], which decreases ten-fold by lowering temperature from room temperature (25 ℃) to 0 ℃ at pH 2.5. Therefore, HDX can be "quenched" by adding acidic solution to the exchanging

medium to bring down its pH to 2.5 while cooling in ice water. This slows considerably back exchange and allows capture of the HDX information while conducting digestion and MS analysis. The use of quench conditions is important because analysis of the deuteration is almost always performed in protic media. After the quench, the overall half life of HDX is longer but still much shorter than other covalent labeling strategies. Therefore, MS analysis following HDX is often kept short, within approximately10 min, to minimize the amount of back-exchange and yet obtain sufficient deuterium uptake to assure reliable measurement of it.
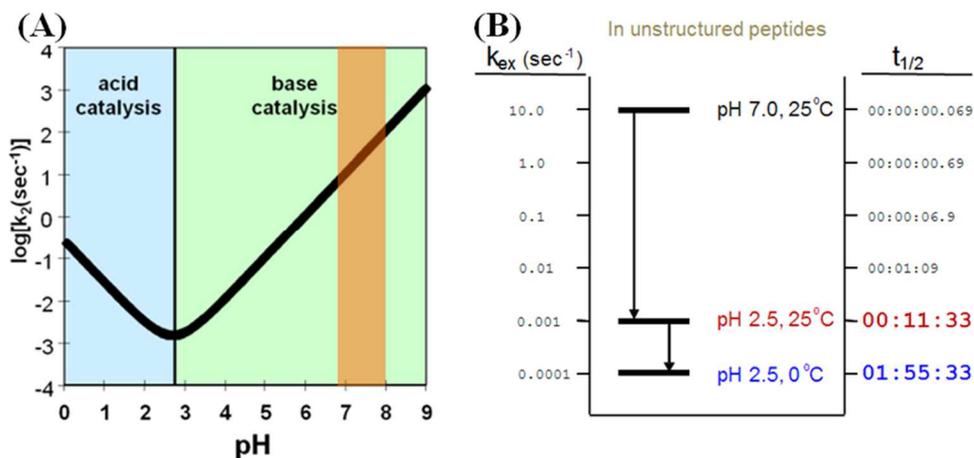
Figure 5- 2. The rates of H/DX depend on pH (A) and temperature (B). Taken from HXMS website: http://www.hxms.neu.edu/research/tutorial_theory.htm

5.4.2 Exchange Kinetics of HDX (EX1 vs. EX2)

Key to the application of HDX in structural studies is that conformational properties of the protein modulate the N-H to N-D conversion rate. Two important factors limiting HDX of amide sites are hydrogen bonding and solvent protection. In protected regions, the overall rate constant k_{HDX} is much smaller than k_{ch}, which is the rate constant for exchange in unstructured peptides. Nevertheless, those buried amide hydrogens can still undergo HDX with measurable rates by exchanging during short-lived, transitional "open" conformations, which are created by

conformational fluctuations of the protein. Such structural fluctuations can either be determined globally or localized for a protein segment by conducting proteolysis prior to analysis. The model for most hydrogen exchange of native proteins in D_2O is:

$$N - H_{cl} \underset{k_{cl}}{\overset{k_{ch}}{\rightleftharpoons}} N\text{-}H_{op} \xrightarrow{} N - D \qquad\qquad \text{Eq. 5-1}$$

where k_{op} is the rate constant of protein unfolding/opening motion and k_{cl} is the rate constant of protein folding/closing motion. Eq. 5-1 is typically discussed with two limiting regimes referred to as EX1 and EX2[25, 26].

EX1 is characterized by $k_{ch} \gg k_{cl}$, so that k_{HDX} is given by the unfolding/opening rate constant: $k_{HDX} = k_{op}$. EX1 kinetics can be seen as a cooperative unfolding event involving several residues in a protein region, all of which exchange before refolding/closing. In such situations, the protein or protein regions give multiple groups of deuterium uptake distribution after HDX, making data interpretation difficult.

EX2 is characterized by $k_{ch} \ll k_{cl}$, leading to $k_{HDX} = (k_{op}/k_{cl})k_{ch}$. Under this condition, the probability of HDX occurring during a single opening event is small, many opening/closing cycles may occur before an isotope exchange happens on a given amide. Proteins undergoing this behavior show a single major group of deuterium uptake distribution after HDX in the mass spectra and that distribution broadens and moves to higher m/z with time of exchange. Under physiological conditions, most proteins, including those discussed in this thesis, undergo EX2 exchange kinetics and only a few undergo EX1 kinetics naturally. Intermediate cases that exhibit aspects of both regimes are also known[26, 27].

5.4.3 HDX MS-Based Protein Structure and Dynamics Study

Continuous labeling is the strategy most often applied in studying protein structure, dynamics, and interaction. In continuous labeling, a native protein is exposed to D_2O, and the deuterium uptake level is monitored as a function of exchange time (often ranging from seconds to hours). As the exchange time increases, the extent of deuterium labeling gradually increases in the protein. Protein regions that spend more time in open conformations undergo faster exchange than those more tightly folded. Based on Eq. 5-1, HDX primarily monitors protein structural dynamics, and dynamics is closely connected to protein structure (folded regions are generally less dynamic than disordered regions).

Because deuterium is one unit in mass higher than hydrogen, the amount of D incorporation can be calculated by measuring the mass increase of a protein/peptide after HDX. Therefore, the relative D uptake level can be calculated by Eq. 5-2:

$$D\% = \{[m(P) - m(N)]/[m(F) - m(N)]\} \times 100\% \qquad \text{Eq. 5-2}$$

where m(P), m(N) and m(F) are the centroid values of partially deuterated protein/peptide, nondeuterated protein/peptide, and fully deuterated protein/peptide, respectively. By plotting D% against T (time), a HDX kinetic curve for protein/peptide is generated, and can be used for comparison among different states.

HDX kinetics can be measured either at the global level or, more commonly, at the peptide level. Peptide-level analysis is especially important because protein conformational changes occur at small regions, such as at protein-protein interaction site, where one may seek to locate the specific interface with higher spatial resolution than observing changes at the global level. HDX coupled with enzymatic digestion (using acidic proteases) prior to mass spectrometry analysis affords such peptide-level resolution. Subsequently, similar analysis with HDX kinetic curves at
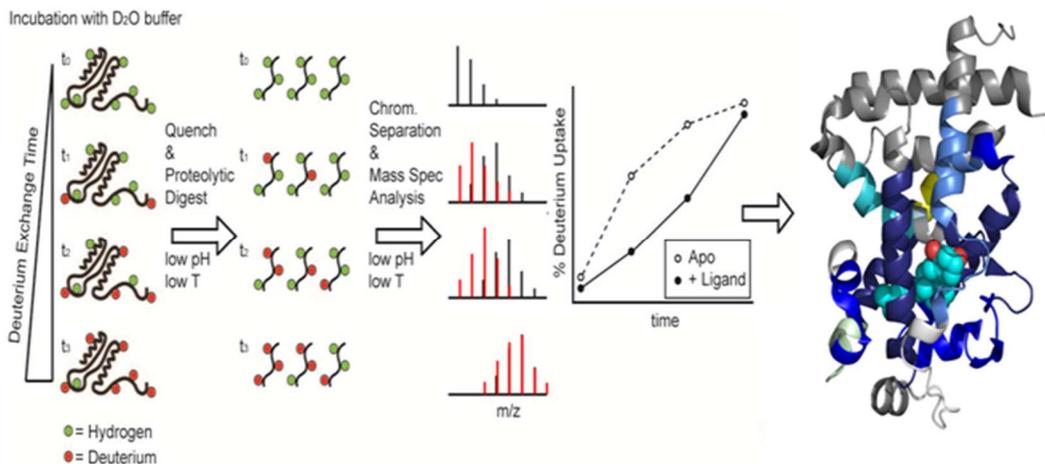
Figure 5- 3. H/DX workflow that provides "peptide-level" information. Taken from Chalmers et al. 2011. Expert Review Proteomics.

peptide-level can be used to compare among different states (i.e., apo vs. holo) to locate the regions that undergo conformational changes (i.e., protection) in different states (Figure 5-3). Residue-level HDX analysis is sometimes performed with electron-transfer dissociation to refine the spatial resolution to the amide level. However, this application is difficult and limited by instrumentation and fragmentation efficiently. It is not considered further in this thesis.

Today, HDX MS is widely used in protein biophysics and in drug development, such as epitope mapping, protein-stability studies, and protein-ligand binding studies (for reviews, see[28-30]). For a protein-protein interaction study, usually two treatments of the protein, bound and unbound, are prepared and subjected to continuous labeling. The kinetic curves of most abundant peptides from both treatments are plotted together and compared, as depicted in Figure 5-3. Regions that undergo no conformational change upon binding will show no difference in extent of deuterium incorporation. Regions that show significant protection against D labeling are likely to be located at or near the binding interface.

Furthermore, this traditional continuous labeling technique, HDX MS-based ligand binding and protein stability experiments can be used in other applications. For example, Gross et al.[31-34]

developed a method for monitoring protein-ligand interactions by MS, titration and hydrogen deuterium exchange (PLIMSTEX). PLIMSTEX curves show the amount of exchanged hydrogens against ligand concentration at a fixed exchange time. From the shape of the curve, binding affinity and stoichiometry can be uncovered through least-squares fitting analysis of the PLIMSTEX titration curve. Fitzgerald and coworkers[35-42] developed a method in studying protein stability, namely, stability of unpurified proteins from rates of HD exchange (SUPREX). SUPREX uses denaturant to affect protein thermodynamic stability, and its curves display the number of exchanged hydrogens as a function of denaturant concentration for a fixed exchange time. SUPREX profiles have a sigmoidal appearance resembling unfolding curves as measured by optical methods. These two methods are not applied in studies in this thesis, and therefore are not further discussed.

5.5 Fast Photochemical Oxidation of Proteins

The concept of applying oxidative labeling coupled with MS footprinting in protein biophysics study is very similar to that of continuous labeling HDX MS. However, more elaborate means than just sample dilution with a solvent are required to initiate the labeling reaction. Oxidative labeling utilizes hydroxyl radicals, which can be
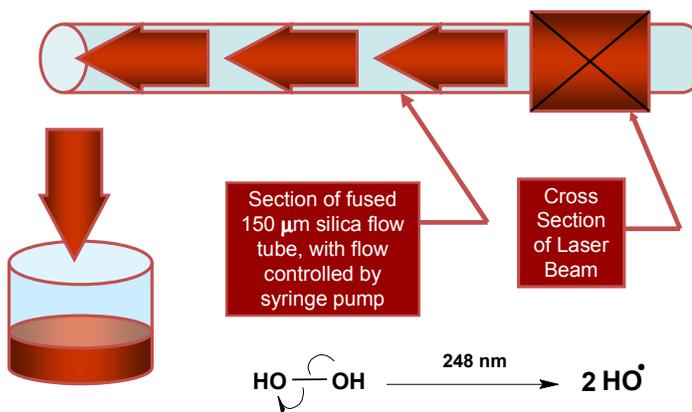
Figure 5- 4. Flow system for Fast Photochemical Oxidation of Proteins (FPOP). The frequency of the pulsed laser beam and the sample flow rate are adjustable to control the single labeling of the sample.

generated by different means, to react with certain amino acids in the protein[43-46]. Fast oxidative labeling (millisec time scale) was first established by Chance and coworkers[44] by radiolytic modification. Photochemical labeling initiated by laser irradiation of H_2O_2 was later developed by different groups[47-49]. Of those methods, only fast photochemical oxidation of proteins (FPOP), developed by Hambly and Gross[49], uses a flow system and controls the radical life to ~ 1 μs by introducing a radical scavenger to prevent excessive labeling (Figure 5-4).

FPOP labeling monitors protein conformation and solvent accessibility by exposing the protein to hydroxyl radicals for a short (~ μs) time and measuring the modification extent. In an FPOP experiment, the sample is mixed with 15 mM of H_2O_2 and a chosen radical scavenger (conventionally 20 mM of Gln), and then flowed through a silica tube (ID: 150 μM) which has a clear window facing laser irradiation (Figure 5-4). The laser is a pulsed 248 nm KrF laser and its frequency is adjusted along with the sample flow rate in order to avoid double shots of the same sample fraction. Upon irradiation, H_2O_2 is photolyzed to form hydroxyl radicals, which then reacts with the scavenger and the protein. The presence of the scavenger controls the radical life time to ~ μs, and, therefore, the protein sample is only shortly exposed to the hydroxyl radicals. Because of this short exposure and fast reaction, only one single state, which is often the native state, of the protein is "captured" and labeled. As a result of the sampling of a single conformation of the protein, the labeling products (+0, +16, +32 etc. Da states) present in a Poisson distribution[50]. Other developments with FPOP involves using different radicals (e.g., $SO_4^-\bullet$ and I•) to react with different amino acids; such developments expanded the original design of FPOP and demonstrates its versatility. In this thesis, however, we only discuss FPOP labeling in protein-protein interaction characterization with the original •OH design.

The resulting modification of an amino acid reacting with hydroxyl radicals is generally thought to be independent of the technique used for radical production[51-53]. Intrinsic reactivity of the functional groups in proteins with hydroxyl radicals have been extensively studied over the years[43-46]. Previous studies showed that 14 of the 20 common amino acids in the protein react with hydroxyl radicals[43] (these commonly modified residues and their corresponding reaction rate constants are listed in Table 5-1. The oxidation products are numerous, but the most common modification is +16 (+oxygen or substitution of an H by an OH). One example of a hydroxyl radical reaction with an amino acid is shown in Figure 5-5. Besides labeling the side chain of the protein amino-acid residues, hydroxyl radicals can also nonspecifically cleave protein backbones. However, such cleavages occur 10-1000 times slower than side-chain modifications[43, 44, 46], and they can be further minimized by modifying labeling conditions.

Table 5- 1. Rate Constants for Reaction of 14 Amino Acids with Hydroxyl Radical at near Neutral pH. Take from Chance et al. 2005 Anal. Chem.

substrate	rate ($M^{-1}s^{-1}$)	pH
Cys	3.5×10^{10}	7.0
Trp	1.3×10^{10}	6.5-8.5
Tyr	1.3×10^{10}	7.0
Met	8.5×10^{9}	6-7
Phe	6.9×10^{9}	7-8
His	4.8×10^{9}	7.5
Arg	3.5×10^{9}	6.5-7.5
Ile	1.8×10^{9}	6.6
Leu	1.7×10^{9}	~ 6
Val	8.5×10^{8}	6.9
Gln	5.4×10^{8}	6.0
Lys	3.5×10^{8}	6.6
Ser	3.2×10^{8}	~ 6
Glu	2.3×10^{8}	6.5
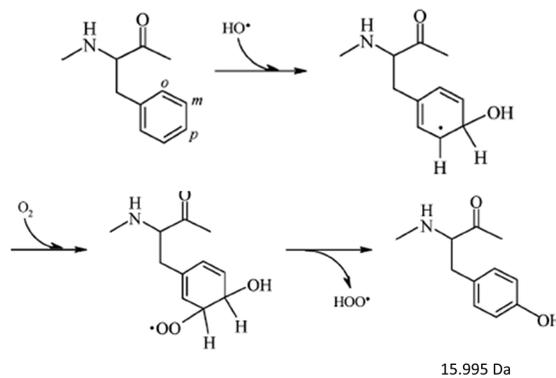
Figure 5- 5. Reaction of hydroxyl radical with phenylalanine. Taken from Chance et al 2007 Chem. Rev.

FPOP experiments rely on peptide mapping with high-resolution LC/MS and MS/MS to identify the site of modification in a protein/peptide. Unlike HDX, hydroxyl radical labeling is complicated and heterogeneous at different residues and/or sites of a protein; therefore, performing MS/MS on the digested modified peptides is important for identifying the modification. Because •OH labeling is irreversible, complex sample handling, such as long digestions and LC/MS analysis, can be performed. Nevertheless, data interpretation remains difficult. There is now some commercial software available for identifying modification type and calculating the modification extent; however, they are far from perfect and a lot of manual examination is still required after software data processing.

5.6 The Complementary Roles of HDX and FPOP

Both HDX and FPOP footprinting have the advantage of probing protein conformations in the solution-phase with no known protein size limit. These experiments can provide a comprehensive picture of biological change related to conformational changes as affected by mutation[54, 55], aggregation[56, 57], and ligand binding[58-60].

Both HDX and FPOP labeling methods have their relative strengths and weaknesses. For example, HDX labeling is reversible and, therefore, requires rapid, low-temperature digestion, separation and MS analysis, which sometimes limits the number of protein identifications especially with complex samples having undergone post-translational modification. On the other hand, FPOP labeling does not require specialized LC-MS-based strategies to preserve the label because it is irreversible and, therefore, is more flexible to accommodate complex sample treatment and LC-conditions tuning to achieve as complete as possible peptide identification and good sequence coverage. However, FPOP labeling has the difficulty of being heterogeneous in both labeling subjects and products, generating products with varied molecular weights and

retention times in the LC-MS readout. Thus, FPOP imposes considerable more difficulty on data analysis. HDX labeling is homogeneous, and its data analysis is much easier.

HDX and FPOP labeling provide complementary information to each other on protein conformations. For example, FPOP labeling is very fast whereas HDX labeling is relatively slow. In footprinting, short labeling time can be advantageous in probing low-amplitude motions that occur on fast timescales, but at the same time, short labeling limits the range of solution-phase conformations that can be probed, which becomes less of a problem for longer labeling times. Besides labeling speed, HDX and FPOP labeling also provide orthogonal information regarding the labeling subjects. HDX labeling occurs on the protein backbone, while FPOP targets the protein side-chains. Therefore, they may identify conformational changes caused by different interactions.

In the second part of this thesis, we present the application of both HDX and FPOP labeling to study protein structure and dynamics. In chapter 6, we studied the dimerization of SecA protein with HDX MS. The controversial interface was solved along with new findings of accompanying conformational changes upon dimerization, which provided leads for the active form of SecA. In chapter 7, we focus on mapping the epitope of Adnectin1-EGFR interaction at amino acid residue level by FPOP. The epitope identified agrees with that from both HDX study and crystallography results, presenting more evidence of the capability of FPOP in epitope mapping.

5.7 References

1. Alberghina, L., Mavelli, G., Drovandi, G., Palumbo, P., Pessina, S., Tripodi, F., Coccetti, P., and Vanoni, M. (2012) Cell growth and cell cycle in Saccharomyces cerevisiae: basic regulatory design and protein-protein interaction network, *Biotechnol Adv 30*, 52-72.

2. Metz, J. T., and Hajduk, P. J. (2010) Rational approaches to targeted polypharmacology: creating and navigating protein-ligand interaction networks, *Curr Opin Chem Biol 14*, 498-504.

3. Charbonnier, S., Gallego, O., and Gavin, A. C. (2008) The social network of a cell: recent advances in interactome mapping, *Biotechnol Annu Rev 14*, 1-28.

4. Pacholarz, K. J., Garlish, R. A., Taylor, R. J., and Barran, P. E. (2012) Mass spectrometry based tools to investigate protein-ligand interactions for drug discovery, *Chem Soc Rev 41*, 4335-4355.

5. Fitzgerald, M. C., and West, G. M. (2009) Painting proteins with covalent labels: what's in the picture?, *J Am Soc Mass Spectrom 20*, 1193-1206.

6. Xu, G., and Chance, M. R. (2007) Hydroxyl radical-mediated modification of proteins as probes for structural proteomics, *Chem Rev 107*, 3514-3543.

7. Engen, J. R. (2009) Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS, *Analytical chemistry 81*, 7870-7875.

8. Gau, B., Garai, K., Frieden, C., and Gross, M. L. (2011) Mass spectrometry-based protein footprinting characterizes the structures of oligomeric apolipoprotein E2, E3, and E4, *Biochemistry 50*, 8117-8126.

9. Dyson, H. J., and Wright, P. E. (2004) Unfolded proteins and protein folding studied by NMR, *Chem Rev 104*, 3607-3622.

10.   Chowdhury, S. K., Katta, V., and Chait, B. T. (1990) Probing conformational changes in proteins by mass spectrometry, *J Am Chem Soc 112*, 9012-9013.

11.   Katta, V., and Chait, B. T. (1991) Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry, *Rapid Commun Mass Spectrom 5*, 214-217.

12.   Smith, D. L., Deng, Y., and Zhang, Z. (1997) Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry, *J Mass Spectrom 32*, 135-146.

13.   Engen, J. R., and Smith, D. L. (2001) Investigating protein structure and dynamics by hydrogen exchange MS, *Analytical chemistry 73*, 256A-265A.

14.   Kaltashov, I. A., and Eyles, S. J. (2002) Studies of biomolecular conformations and conformational dynamics by mass spectrometry, *Mass Spectrom Rev 21*, 37-71.

15.   Kaltashov, I. A., and Eyles, S. J. (2002) Crossing the phase boundary to study protein dynamics and function: combination of amide hydrogen exchange in solution and ion fragmentation in the gas phase, *J Mass Spectrom 37*, 557-565.

16.   Hoofnagle, A. N., Resing, K. A., and Ahn, N. G. (2003) Protein analysis by hydrogen exchange mass spectrometry, *Annu Rev Biophys Biomol Struct 32*, 1-25.

17.   Eyles, S. J., and Kaltashov, I. A. (2004) Methods to study protein dynamics and folding by mass spectrometry, *Methods 34*, 88-99.

18.   Garcia, R. A., Pantazatos, D., and Villarreal, F. J. (2004) Hydrogen/deuterium exchange mass spectrometry for investigating protein-ligand interactions, *Assay Drug Dev Technol 2*, 81-91.

19.   Wales, T. E., and Engen, J. R. (2006) Hydrogen exchange mass spectrometry for the analysis of protein dynamics, *Mass Spectrom Rev 25*, 158-170.

20.   Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange, *Proteins 17*, 75-86.

21.   Morgan, C. R., and Engen, J. R. (2009) Investigating solution-phase protein structure and dynamics by hydrogen exchange mass spectrometry, *Curr Protoc Protein Sci Chapter 17*, Unit 17 16 11-17.

22.   Coales, S. J., E, S. Y., Lee, J. E., Ma, A., Morrow, J. A., and Hamuro, Y. (2010) Expansion of time window for mass spectrometric measurement of amide hydrogen/deuterium exchange reactions, *Rapid Commun Mass Spectrom 24*, 3585-3592.

23.   Hvidt, A., and Linderstrom-Lang, K. (1954) Exchange of hydrogen atoms in insulin with deuterium atoms in aqueous solutions, *Biochim Biophys Acta 14*, 574-575.

24.   Englander, S. W., and Kallenbach, N. R. (1983) Hydrogen exchange and structural dynamics of proteins and nucleic acids, *Q Rev Biophys 16*, 521-655.

25.   Krishna, M. M., Hoang, L., Lin, Y., and Englander, S. W. (2004) Hydrogen exchange methods to study protein folding, *Methods 34*, 51-64.

26.   Konermann, L., Tong, X., and Pan, Y. (2008) Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches, *J Mass Spectrom 43*, 1021-1036.

27.   Xiao, H., Hoerner, J. K., Eyles, S. J., Dobo, A., Voigtman, E., Mel'cuk, A. I., and Kaltashov, I. A. (2005) Mapping protein energy landscapes with amide hydrogen exchange and mass spectrometry: I. A generalized model for a two-state protein and comparison with experiment, *Protein Sci 14*, 543-557.

28.   Engen, J. R., Wales, T. E., and Shi, X. (2011) Hydrogen Exchange Mass Spectrometry for Conformational Analysis of Proteins, *Ency. Anal. Chem.* .

29.   Wei, H., Mo, J., Tao, L., Russell, R. J., Tymiak, A. A., Chen, G., Iacob, R. E., and Engen, J. R. (2014) Hydrogen/deuterium exchange mass spectrometry for probing higher order structure of protein therapeutics: methodology and applications, *Drug Discov Today 19*, 95-102.

30.   Huang, R. Y., and Chen, G. (2014) Higher order structure characterization of protein therapeutics by hydrogen/deuterium exchange mass spectrometry, *Anal Bioanal Chem*.

31.   Zhu, M. M., Rempel, D. L., Du, Z., and Gross, M. L. (2003) Quantification of protein-ligand interactions by mass spectrometry, titration, and H/D exchange: PLIMSTEX, *J Am Chem Soc 125*, 5252-5253.

32.   Sperry, J. B., Shi, X., Rempel, D. L., Nishimura, Y., Akashi, S., and Gross, M. L. (2008) A mass spectrometric approach to the study of DNA-binding proteins: interaction of human TRF2 with telomeric DNA, *Biochemistry 47*, 1797-1807.

33.   Zhu, M. M., Rempel, D. L., and Gross, M. L. (2004) Modeling data from titration, amide H/D exchange, and mass spectrometry to obtain protein-ligand binding constants, *J Am Soc Mass Spectrom 15*, 388-397.

34.   Mei, Z. M., Chitta, R., and Gross, M. L. (2005) PLIMSTEX: a novel mass spectrometric method for the quantification of protein–ligand interactions in solution, *International journal of Mass Spectrometry 20*, 213-220.

35.   Dai, S. Y., and Fitzgerald, M. C. (2006) Accuracy of SUPREX (stability of unpurified proteins from rates of H/D exchange) and MALDI mass spectrometry-derived protein

unfolding free energies determined under non-EX2 exchange conditions, *J Am Soc Mass Spectrom 17*, 1535-1542.

36.   Dai, S. Y., Gardner, M. W., and Fitzgerald, M. C. (2005) Protocol for the thermodynamic analysis of some proteins using an H/D exchange- and mass spectrometry-based technique, *Analytical chemistry 77*, 693-697.

37.   Powell, K. D., and Fitzgerald, M. C. (2001) Measurements of protein stability by H/D exchange and matrix-assisted laser desorption/ionization mass spectrometry using picomoles of material, *Analytical chemistry 73*, 3300-3304.

38.   Powell, K. D., Ghaemmaghami, S., Wang, M. Z., Ma, L., Oas, T. G., and Fitzgerald, M. C. (2002) A general mass spectrometry-based assay for the quantitation of protein-ligand binding interactions in solution, *J Am Chem Soc 124*, 10256-10257.

39.   Powell, K. D., Wales, T. E., and Fitzgerald, M. C. (2002) Thermodynamic stability measurements on multimeric proteins using a new H/D exchange- and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry-based method, *Protein Sci 11*, 841-851.

40.   Roulhac, P. L., Powell, K. D., Dhungana, S., Weaver, K. D., Mietzner, T. A., Crumbliss, A. L., and Fitzgerald, M. C. (2004) SUPREX (Stability of Unpurified Proteins from Rates of H/D Exchange) analysis of the thermodynamics of synergistic anion binding by ferric-binding protein (FbpA), a bacterial transferrin, *Biochemistry 43*, 15767-15774.

41.   Tang, L., Hopper, E. D., Tong, Y., Sadowsky, J. D., Peterson, K. J., Gellman, S. H., and Fitzgerald, M. C. (2007) H/D exchange- and mass spectrometry-based strategy for the thermodynamic analysis of protein-ligand binding, *Analytical chemistry 79*, 5869-5877.

42. Wang, M. Z., Shetty, J. T., Howard, B. A., Campa, M. J., Patz, E. F., Jr., and Fitzgerald, M. C. (2004) Thermodynamic analysis of cyclosporin a binding to cyclophilin a in a lung tumor tissue lysate, *Analytical chemistry 76*, 4343-4348.

43. Xu, G., and Chance, M. R. (2005) Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting, *Analytical chemistry 77*, 4549-4555.

44. Maleknia, S. D., Brenowitz, M., and Chance, M. R. (1999) Millisecond radiolytic modification of peptides by synchrotron X-rays identified by mass spectrometry, *Analytical chemistry 71*, 3965-3973.

45. Stadtman, E. R., and Levine, R. L. (2003) Free radical-mediated oxidation of free amino acids and amino acid residues in proteins, *Amino Acids 25*, 207-218.

46. Garrison, W. M. (1987) Reaction Mechanisms in the Radiolysis of Peptides, Polypeptides, and Proteins, *Chem Rev 87*, 381-398.

47. Sharp, J. S., Becker, J. M., and Hettich, R. L. (2004) Analysis of protein solvent accessible surfaces by photochemical oxidation and mass spectrometry, *Analytical chemistry 76*, 672-683.

48. Aye, T. T., Low, T. Y., and Sze, S. K. (2005) Nanosecond laser-induced photochemical oxidation method for protein surface mapping with mass spectrometry, *Analytical chemistry 77*, 5814-5822.

49. Hambly, D. M., and Gross, M. L. (2005) Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale, *J Am Soc Mass Spectrom 16*, 2057-2063.

50.   Gau, B. C., Chen, H., Zhang, Y., and Gross, M. L. (2010) Sulfate radical anion as a new reagent for fast photochemical oxidation of proteins, *Analytical chemistry 82*, 7821-7827.

51.   Takamoto, K., and Chance, M. R. (2006) Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes, *Annu Rev Biophys Biomol Struct 35*, 251-276.

52.   Xu, G., and Chance, M. R. (2004) Radiolytic modification of acidic amino acid residues in peptides: probes for examining protein-protein interactions, *Analytical chemistry 76*, 1213-1221.

53.   Xu, G., Takamoto, K., and Chance, M. R. (2003) Radiolytic modification of basic amino acid residues in peptides: probes for examining protein-protein interactions, *Analytical chemistry 75*, 6995-7007.

54.   Yang, J., Garrod, S. M., Deal, M. S., Anand, G. S., Woods, V. L., Jr., and Taylor, S. (2005) Allosteric network of cAMP-dependent protein kinase revealed by mutation of Tyr204 in the P+1 loop, *J Mol Biol 346*, 191-201.

55.   Yamamoto, T., Izumi, S., Ohmae, E., and Gekko, K. (2004) Mass spectrometry of hydrogen/deuterium exchange of Escherichia coli dihydrofolate reductase: effects of loop mutations, *J Biochem 135*, 487-494.

56.   Kheterpal, I., and Wetzel, R. (2006) Hydrogen/deuterium exchange mass spectrometry--a window into amyloid structure, *Acc Chem Res 39*, 584-593.

57.   Yao, Z. P., Tito, P., and Robinson, C. V. (2005) Site-specific hydrogen exchange of proteins: insights into the structures of amyloidogenic intermediates, *Methods Enzymol 402*, 389-402.

58.   Burke, J. E., Karbarz, M. J., Deems, R. A., Li, S., Woods, V. L., Jr., and Dennis, E. A. (2008) Interaction of group IA phospholipase A2 with metal ions and phospholipid vesicles probed with deuterium exchange mass spectrometry, *Biochemistry 47*, 6451-6459.

59.   Hamuro, Y., Coales, S. J., Morrow, J. A., Molnar, K. S., Tuske, S. J., Southern, M. R., and Griffin, P. R. (2006) Hydrogen/deuterium-exchange (H/D-Ex) of PPARgamma LBD in the presence of various modulators, *Protein Sci 15*, 1883-1892.

60.   Chalmers, M. J., Busby, S. A., Pascal, B. D., He, Y., Hendrickson, C. L., Marshall, A. G., and Griffin, P. R. (2006) Probing protein ligand interactions by automated hydrogen/deuterium exchange mass spectrometry, *Analytical chemistry 78*, 1005-1014.

Chapter 6*

Hydrogen Deuterium Exchange Mass Spectrometry for the Analysis of SecA Dimerization in Solution

6.1 Abstract

The Sec pathway mediates protein translocation across the inner membrane of bacteria. SecA is a motor protein that drives preprotein translocation through the SecYEG channel. SecA reversibly dimerizes under physiological conditions but different dimer interfaces have been observed in SecA crystal structures. Here, we have used biophysical approaches to address the nature of the SecA dimer that exists in solution. We have taken advantage of the extreme salt sensitivity of SecA dimerization to compare the rates of hydrogen-deuterium exchange of the monomer and dimer and compared the result to the study of effects of single alanine substitutions on dimerization affinity performed by our collaborators in U Connecticut. Our results support the antiparallel dimer arrangement observed in the crystal structure of *Bacillus subtilis* SecA. Additional residues lying within the preprotein binding domain and the C-terminus are also protected from exchange upon dimerization, indicating linkage to a conformational transition of the preprotein binding domain from an open to a closed state. In agreement with this interpretation, normal mode analysis demonstrates that the SecA dimer interface influences the global dynamics of SecA such that dimerization stabilizes the closed conformation.

6.2 Introduction

In bacteria, the majority of secretory preproteins are translocated through a general secretion (Sec) pathway that contains a Sec translocase complex comprising the integral membrane channel, SecYEG, and the cytosolic ATPase motor protein, SecA[1, 2]. SecA binds preproteins, associates with the SecYEG channel, and harnesses energy from ATP hydrolysis to drive conformational changes that lead to preprotein translocation[3-6]. SecA is a large, 102 kDa multifunctional protein that is comprised of several domains: nucleotide binding domains I and II (NBD I and II), a preprotein binding domain (PBD), and a C-domain that is comprised of an α-helical scaffold domain (HSD), an α-helical wing domain (HWD) and a carboxyl-terminal linker (CTL) (Figure 6-1)[7]. The PBD and HWD contribute to the formation of a binding groove for the signal peptide region of the preprotein[8-12].

Figure 6- 1. Structural domains of SecA. The structure of *E. coli* SecA (PDB 2VDA[8]) with domains indicated by color: NBD I (blue), PBD (purple), NBD II (green), HSD (red) and HWD (cyan). In this monomer structure the PBD adopts an open conformation.

SecA exists in a monomer-dimer equilibrium that is sensitive to salt concentration and temperature[13-16]. The cellular concentration of SecA is 5-8 μM[17, 18], and its dissociation constant is 0.28 μM in 200 mM KCl[16]. Thus, in the absence of ligands, SecA likely exists as a dimer in the cytoplasm[13, 16]. It has been suggested that SecA functions as a dimer during preprotein translocation because a cross-linked SecA dimer[19, 20] and a genetically produced SecA dimer[21] are active for translocation. Nevertheless, the oligomeric state of SecA during preprotein translocation remains controversial. Some studies indicate that

dissociation of SecA is favored in the presence of phospholipids[18, 22] or synthetic signal peptides[23], and upon SecYEG binding[24]. Other reports support an active monomeric form of SecA, indicating that the monomeric SecA mutant is functional[25] whereas a disulfide cross-linked dimer is nonfunctional and its reduced monomeric counterpart is functional[26].

The relative positioning of the two protomers in the SecA dimer is also unclear. Although the structure of the SecA protomer obtained from different bacterial species is highly conserved in crystal structures, multiple dimeric interfaces have been identified[27-31] among the five crystal structures of the SecA dimer reported to date (Figure 6-2). Both parallel[29] and antiparallel[27, 28, 30, 31] dimer orientations are observed and these structures contain different dimerization interfaces. For example, the dimer interface of *Escherichia coli* SecA (PDB 2FSF[31]) lies on the opposite side relative to the dimer interface in one of the *Bacillus subtilis* SecA structures (PDB 1M6N[27]). It is difficult to distinguish crystal packing contacts from biologically relevant protein-protein interfaces[7, 32, 33], and it remains unclear which of these crystal structures, if any,
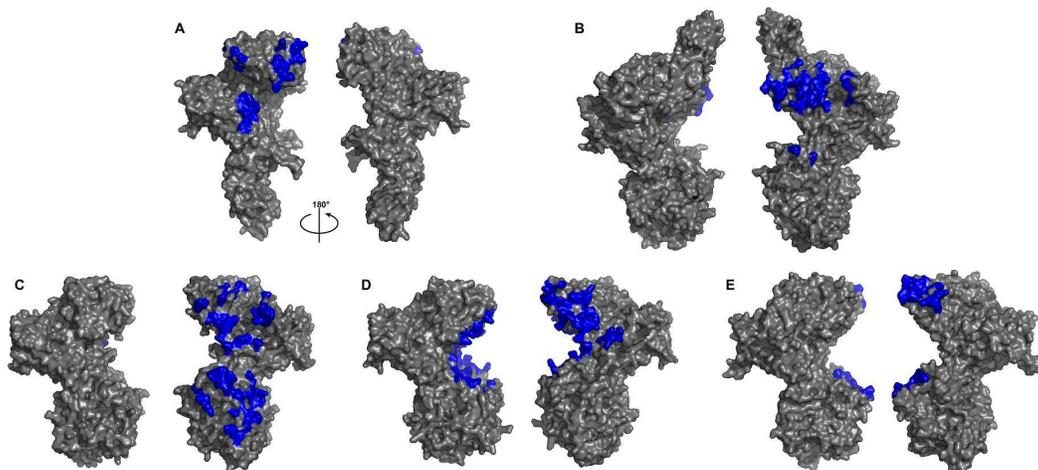
Figure 6- 2. Alternative dimer interfaces in SecA. The dimer interfaces in different structures of SecA are indicated in blue with the protomers in two orientations. A) *E. coli* SecA (PDB 2FSF)[31], B) *T. thermophilus* (PDB 2IPC)[29], C) *B. subtilis* (PDB 1M6N)[27], D) *B. subtilis* (PDB 2IBM)[30], and E) *M. tuberculosis* (PDB 1NL3)[28]. The cyan circles in B and C indicate the location of the N-terminal region.

corresponds to the physiologically relevant SecA dimer. Single particle cryo-electron microscopy measurements support an antiparallel dimer structure[14] and fluorescence resonance energy transfer distance measurements show the greatest agreement with the 1M6N antiparallel dimer[34, 35]. A cross-linking study led to the proposal of a novel interface that involves residues from the NBD I, PBD, and HSD[36].

Interestingly, the PBD is found to adopt multiple conformations. A closed form that brings the PBD and HWD in close proximity is found in several SecA dimer crystal structures[27-29] and in a structure of a SecA monomer complexed with ADP[37], whereas an open conformation where the PBD is solvent accessible exists in a crystal structure of a monomeric form[38] and in one dimer structure[31]. In the open state, the groove between the PBD and NBD II forms a polypeptide-binding clamp[24, 39]. Binding of a non-signal sequence peptide in this region results in movement of the PBD towards NBD II[37]. Interaction of SecA with SecYEG induces further movement of the PBD towards NBD II, closing the clamp[24]. NMR analysis indicates that SecA interconverts between open and closed states in solution, with the equilibrium favoring the open state[8]. Cysteine cross-linking indicates that movement of the PBD from the open state to the closed clamp state is required for initiation of translocation[40].

To resolve these conflicting interpretations, we collaborated with Prof. James Cole's group in U. Connecticut and combined three biophysical strategies to identify the dimer interface of SecA in solution and to characterize the conformational changes associated with dimerization. First, we used hydrogen/deuterium exchange mass spectrometry (HDX MS) to identify regions on SecA that are protected from exchange upon dimerization and to probe conformational changes associated with dimerization. Second, to complement our result from HDX and achieve site-specific information on the dimer interface, our collaborators from U. Connecticut selected

amino acids for alanine substitution that are predicted to have significant energetic contributions to SecA dimerization based on alternative dimer interfaces. Each alanine-substituted protein was then analyzed by sedimentation velocity analytical ultracentrifugation (AUC) to measure SecA dimerization affinity. Third, Prof. Eric May from U. Connecticut performed normal mode analysis of SecA monomer and dimer and defined conformational changes linked to dimerization. This combined strategy locates the likely SecA dimer interface and provides a link between closure of the PBD/HWD groove and SecA dimerization. In this chapter, we will focus on the discussion of our study on HDX MS. For details of the other two methods, one can refer to the publication.

6.3 Materials and Methods

6.3.1 Wild-Type SecA Expression and Purification

Wild-type, native *E. coli* SecA was expressed and purified as previously described[16]. SecA samples were exchanged by using spin columns into a working buffer containing 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH 7.5), 0.5 mM ethylenediaminetetraacetic acid (EDTA), 0.1 mM tris(2-carboxyethyl)phosphine (TCEP), and 100 to 500 mM KCl. The SecA concentration was determined by absorbance at 280 nm, using a molar absorptivity (ε_{280}) of 75,750 M^{-1} cm^{-1} calculated with Sednterp[41].

6.3.2 HDX MS Protocol

Given the high sensitivity of SecA dimerization to salt concentration (K_d increases from 0.014 µM at 100 mM KCl to 40.4 µM in 500 mM KCl[16]) , this effect was exploited to permit comparison of HDX of SecA dimer (low salt) and monomer (high salt). Stock solutions of untagged, wild-type SecA were prepared at 6 µM in either low salt (100 mM KCl) or high salt

(500 mM KCl) pH 7.5 buffer, each containing 20 mM HEPES, 0.5 mM EDTA, and 0.1 mM TCEP. Samples were equilibrated at 25 °C for 1 h before HDX analysis. Continuous labeling with deuterium was initiated by diluting 4 µL of the stock solution into 16 µL of D_2O with an identical KCl concentration. Based on the measured dissociation constants, the low salt samples were 93% dimer, and the high salt samples were 95% monomer. HDX was measured at 10 s, 30 s, 1 min, 2 min, 15 min, 1 h, and 4 h. Two proteases, pepsin and fungal XIII, were used for protein digestion to achieve high peptide coverage. For measurement, the exchange reaction was quenched at various times by mixing with 30 µL of 3 M urea, 1% trifluoroacetic acid (TFA) at ~1 °C. The mixture was then passed over a custom-packed pepsin column (2×20 mm) at 200 µL/min. For fungal XIII digestion, the exchange reaction was quenched with 30 µL of 1.2 mg/mL fungal XIII, 1% formic acid (FA) in water and set on ice for 2 min. Prior to LC/MS analysis, the digested peptides were captured on a 2×15 mm C8 trap column (Agilent, Santa Clara, CA, USA) and desalted with a 3 min flow at 200 µL/min of H_2O containing 0.1% TFA. Peptides were then separated by using a 2.1×50 mm reversed-phase C18 column (1.9 µm Hypersil Gold; Thermo Fisher Scientific, Waltham, MA, USA) with an 11.5 min linear gradient of 4%–40% CH_3CN in 0.1% FA at 50 µL/min with a nanoACQUITY UPLC (Waters, Milford, MA, USA). Protein digestion and peptide separation were carried out in water-ice bath to minimize back exchange. MS detection was performed on an LTQ-Orbitrap (Thermo Fisher Scientific, Santa Clara, CA, USA) with the following instrument parameters: spray voltage 4 kV, capillary temperature 225 °C, capillary voltage 44 V, and tube lens 100 V. Data were collected with mass resolving power of 100,000 at *m/z* 400. Each experiment was carried out in duplicate.

6.3.3 HDX Data Analysis

To identify the products of pepsin or fungal XIII digestions and to provide a list of peptides to be followed during HDX, product-ion mass spectra were acquired in a data-dependent mode, with the six most abundant ions from each scan selected for MS/MS. The MS/MS *.raw files were then converted to mzXML files and submitted to MassMatrix for peptide identification[42]. This search was also performed against a decoy (reversed) sequence, and ambiguous identifications were discarded. The final peptide list generated from MassMatrix was output as *.csv files.

The peptide list *.csv files and Thermo *.raw files were then input into HDX workbench[43] to calculate the centroid masses of isotopic envelopes (m) and deuterium level (*D%*). As described previously[44], $D\% = \{[m(P) - m(N)]/[m(F) - m(N)]\} \times 100\%$, where m(P), m(N) and m(F) are the centroid values of partially deuterated peptide, nondeuterated peptide, and fully deuterated peptide, respectively. The retention time window used for calculation of m of each peptide in each sample was manually inspected and adjusted for accurate calculation of *D%*, and peptides that showed interference by isotopic peaks from other peptides were discarded. The deuterium level was finally adjusted because the exchange media had 80% deuterium content. The data were not corrected for back exchange because two states were compared.

For details on SecA mutagenesis, analytical ultracentrifugation, and normal mode analysis, please see *Biochemistry*, **2014**, 53(19), 3248-3260.

6.4 Results

6.4.1 Experiment Condition Control for HDX

Because the oligomeric state of SecA depends on the KCl concentration, and the intrinsic exchange rate of amide protons are largely unaffected by different salt concentrations[45], we

prepared samples of SecA containing predominantly monomer or dimer—at the same protein concentration—by simply varying the KCl concentration. Any measured changes in HDX rates would then reliably reflect any dimerization-induced changes in the protein. We previously determined that the K_d for SecA dimerization at 20 °C varies from 0.014 µM at 100 mM KCl to 40.4 µM at 500 mM KCl[16]. We also found that K_d decreases monotonically by approximately a factor of two for every 5 °C increase in temperature; thus, at 25 °C, where we conducted HDX, the K_d values of SecA in 100 mM KCl and 500 mM KCl are approximately 0.007 and 20.2 µM, respectively. For each incubation in our continuous HDX experiments, the SecA stock solution was diluted from 6 µM to 1.2 µM with D$_2$O buffer. Under these conditions, 93% of the SecA is dimeric at 100 mM KCl and 95% is monomeric in 500 mM KCl.

To obtain high coverage of SecA in the digestion following HDX, we supplemented the traditional pepsin digestion with a separate one using acidic protease, fungal XIII. Pepsin digestion alone gave 86% coverage; missing were 35 residues at the N-terminus and 30 residues at the C-terminus. Fungal XIII digestion, however, allowed us to recover many peptides covering the N-terminus. Combining the peptides from the two digestions provided high (96%) coverage for the total 901 amino acid residues in this large protein.

6.4.2 HDX Analysis of SecA Monomer

We expect that regions on the protein with defined secondary structures will undergo slower HDX than unstructured regions or those in loops. This is because protein backbone amide hydrogen exchange rates are highly dependent on local fluctuations in structure and solvent accessibility[46]. We mapped for visual convenience the HDX results for the SecA monomer onto the NMR structure of *E. coli* SecA (PDB 2VDA). The time-dependent deuterium uptake for the

SecA peptides from pepsin and fungal XIII digestion, measured at seven time points, was used in the mapping (Figure 6-3).

In general, the peptides located completely in α-helical or β-sheet regions exhibit slower HDX (< 40% at short time points) and are more protected than those at loops and termini. One exception is the peptide covering residues 602-609. Although this region is an α-helix in the NBD II domain, it is dynamic with over 70% of the amide hydrogens exchanged after 10 s. Judging from the NMR structure, we see that this helix is located at the periphery of the NBD II domain and is detached from the main body by two loops, imparting greater solvent accessibility and more frequent local conformation fluctuations to this region. Noticeably, the N-terminus peptides covering residues 1-6 and 7-15 were also found to be highly dynamic, with maximal exchange within 10 s. When looking at different exchange behavior from 10 s to 4 h, most of the peptides contained within α-helical or β-sheet regions underwent increases of less than 40%. Interestingly,
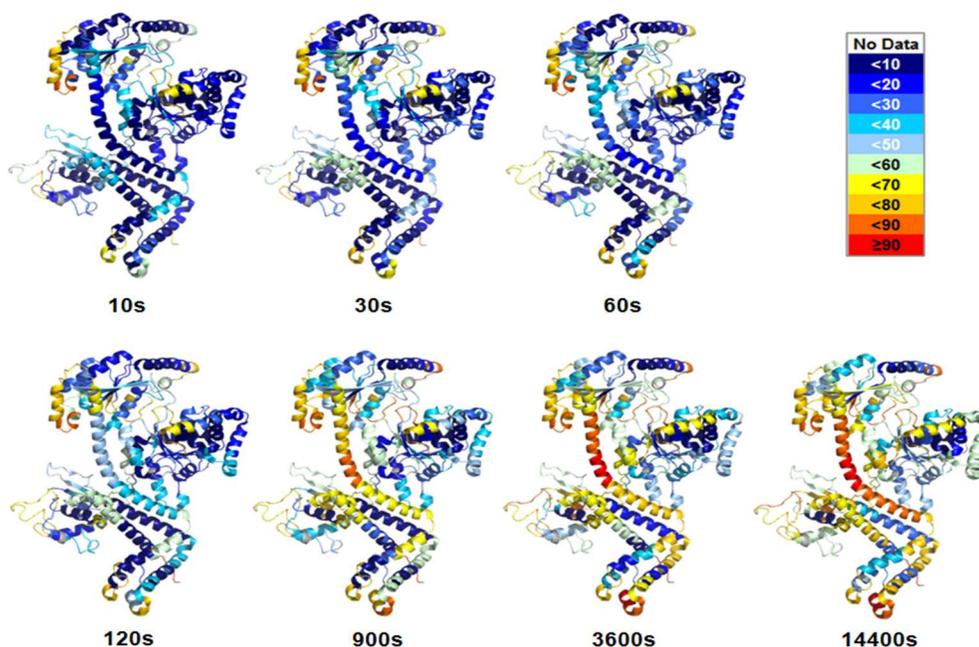
Figure 6- 3. HDX dynamics of SecA monomer. Different color indicates different amount of D uptake.

unlike most others, peptides derived from the long α-helix in HSD covering residues 627-666, showed an increase to more than 80% in deuterium uptake. This may reflect the relatively higher dynamics of this α-helical region, also named the "transducer helix[47]", which, as indicated by its name, is important for allosteric communication between NBD I , NBD II, PBD, and the C-domain[37].

6.4.3 Differential HDX Analysis of SecA Monomer and Dimer

To determine the dimerization interface of SecA, we applied comprehensive differential HDX analysis of the monomer and dimer. In this approach, the HDX kinetics of peptic/fungal peptides derived from the SecA monomer and dimer were compared (percentage-of-deuterium-incorporated, $D\%$, versus time (log scale) plots are made for all peptides). As expected, some regions in the dimer became more protected against exchange than in the monomer. The average differences in $D\%$ ($\Delta D\%$) for the duplicate analysis of seven exchange time points were calculated, and regions with significant differential deuteration levels are identified in Table 6-1. The differences between the monomer and dimer in these regions are considered significant because all peptic and fungal peptides covering the same region showed consistent HDX kinetics, smooth curves, and similar $\Delta D\%$ values. These regions are categorized into two groups: those that show a large difference ($\Delta D\% \geq 6$, red) and those that show a small difference ($3 < \Delta D\% \leq 5$, magenta). It is worth mentioning that most regions in SecA have insignificant $\Delta D\%$ values (less than 3; not listed in Table 6-1). For example, Figure 6-4D shows some HDX kinetic plots of regions that are unaffected by SecA dimerization. This consistency over most regions in SecA serves as a negative control in showing that the HDX rate is unaffected by KCl concentration and indicates there are large regions of the protein that show little change in HDX upon dimerization.

Table 6- 1. SecA peptides showing different deuterium level incorporation induced by dimerization.

Region	Structure	Pepsin Digestion		Fungal XIII Digestion	
		AA Position	Δ D%	AA Position	Δ D%
208-216	NBD	208-215 (+1)	3	208-215 (+1), 210-215 (+1)	3
				211-217 (+1)	5
313-319	PBD	314-319 (+2)	4	313-319 (+2)	3
320-338	PBD	315-320 (+2)	6	320-325 (+2)	13
		320-325 (+2), 320-326 (+1)(+2)	13	320-326 (+1)(+2)	13
		321-326 (+1)(+2)	15	327-338 (+2)	7
		327-333 (+1)(+2), 327-334 (+2)	8		
350-374	PBD	359-372 (+2), 360-372 (+2)	13	352-358 (+2), 350-362 (+2)(+3)	15
		359-368 (+2), 359-374 (+2)	13	350-358 (+2),363-368 (+1)	12
375-380	NBD	373-380 (+1)(+2)	5	369-379 (+2)	5
		375-380 (+1)	3	373-379 (+1)	5
398-408	NBD	399-404 (+1)	9	398-408 (+2), 399-407 (+2)	6
		401-409 (+2)	6	399-408 (+2)	6
		404-409 (+2)	4	402-407 (+2), 402-408 (+1)(+2)	7
427-437	NBD2	432-437 (+1)	4	427-434 (+2), 427-436 (+2)	5
		432-437 (+2)	4	430-436 (+2)	3
562-574	NBD2			562-568 (+3)	6
				569-574 (+2)	12
575-587	NBD2	572-586 (+3)(4)	5	567-587 (+4), 569-587 (+3)(+4)	5
				569-585 (+3), 569-586 (+3)(+4)	5
626-671	HSD	627-638 (+2), 627-638 (+4)	8	626-633 (+2), 626-635 (+2)(+3)	10
		627-638 (+3)	11	634-639 (+1)(+2), 654-660 (+2)	11
		627-639 (+3)	8	636-642 (+3), 661-666 (+2)	8
		639-647 (+2)(3), 648-666 (+4)	6	659-665 (+2)	10
		627-647 (+3)(+4), 649-666 (+4)	5	664-671 (+1)	6
672-690	HWD			672-677 (+1)	4
		685-690 (+1)(+2)	4	683-689 (+1)(+2), 683-690 (+2)	4
701-708	HWD	701-706 (+1), 701-708 (+1)(+2)	3		
		701-707 (+1)(+2)	4		
723-737	HWD	723-730 (+1)(+2), 731-737 (+2)	4	724-732 (+3)	3
		724-730 (+2)	7	724-733 (+2)(+3)	4
744-757	HWD	744-755 (+2)(+3), 746-758 (+2)(+3)	5	744-750 (+2)(+3)	7
		746-755 (+2)(+3), 748-755 (+2)(+3)	7	744-751 (+3)	5
		747-755 (+2)(+3), 746-757 (+2)(+3)	6		
758-774	HSD	758-767 (+3), 759-767 (+2)(+3)	3	762-768 (+2), 765-771 (+1)	4
		759-765 (+3)	4	762-771 (+2), 764-771 (+2)	3
		768-774 (+1)	3	772-778 (+2)(+3)	3
775-783	HSD	775-781 (+2), 775-782 (+2)(+3)	6	772-783 (+3)	8
		775-783 (+2)	8		
815-830	HSD	815-820 (+2)	4	812-819 (+2), 814-819 (+2)	4
		815-826 (+2), 816-826 (+2)	4	820-826 (+1)	2
		821-826 (+1)	5	820-828 (+1)(+2), 820-830 (+2)	3

The *ΔD%* values represent the average difference in percentage of deuterium incorporation across all HDX exchange points of SecA monomer and dimer. The red and magenta colors represent two groups categorized based on *ΔD%* values: large and small, respectively. The values in parentheses represent the charge state of each peptide ion; two values indicate two charge states for the same peptide.

In contrast, several regions that show significant ΔD% values (Table 6-1) must have different structures or interactions in monomer and dimer states. The largest differences are observed for peptides within the region 627-671 in the HSD (Figure 6-4A) and within the regions 313-338 and 350-374 in the PBD (Figure 6-4B). Peptides encompassing 758-774 and 815-830 in the HSD show small changes of ΔD% but a large difference is detected for 775-783 (Figure 6-4C). The kinetic curves for monomer and dimer typically come together at long times, suggesting that the dimer has a relatively large off rate, allowing the protomer to be freed from the dimer and undergo HDX as the monomer.

To help understand the interface that forms upon SecA dimerization, we mapped the regions
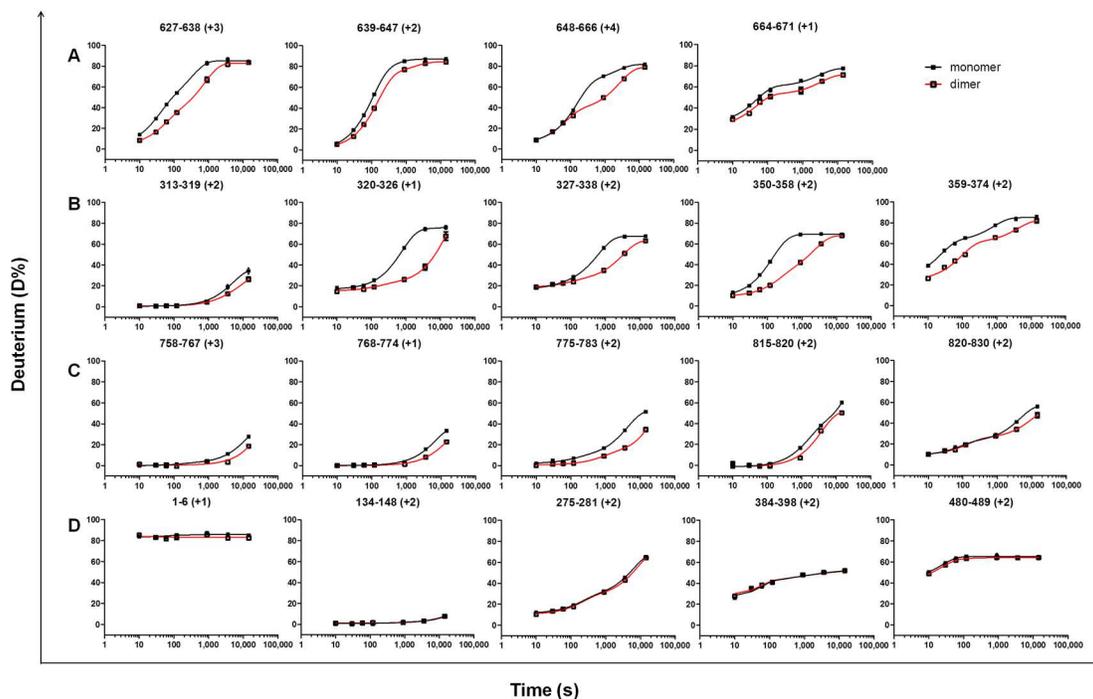
Figure 6- 4. Peptide-level HDX kinetics of SecA. A comparison between monomer (black) and dimer (red) shows significant changes in HDX at the transducer helix from HSD (A), regions in PBD (B), two helix fingers from HSD (C), and no differences in HDX at regions undergoing little structural change (D). The corresponding regions for A, B, and C are marked on Figure 6-5A.

121

identified in Table 6-1 onto the NMR structure of SecA using the same color code (Figures 6-5A and B). The three locations exhibiting the greatest difference in HDX (Figures 6-4A, B, and C), are marked with a', b' and c', respectively, on the structure shown in Figure 6-5A. Region a' corresponds to the long transducer helix in HSD. Region b' lies within the PBD and includes two antiparallel β-strands and the following α-helix. Region c' contains the two-helix finger from HSD. Other regions within HWD, NBD I, and NBD II also show some differential $\Delta D\%$ and are highlighted in Figure 6-5.

The regions of SecA that show reduced HDX upon dimerization lie on multiple faces of the
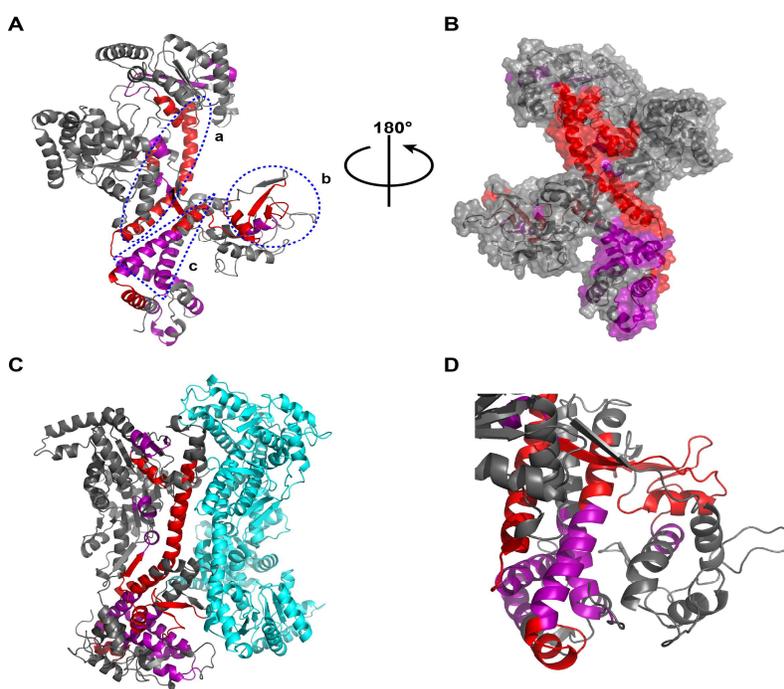
Figure 6- 5. Structural analysis of SecA HDX kinetics. A and B: exchange rates mapped onto the *E. coli* SecA (PDB 2VDA) structure. Regions exhibiting a large difference ($\Delta D\% \geq 6$) are colored red and regions exhibiting a small difference ($3 < \Delta D\% \leq 5$) are colored magenta, as in Table 6-1. A) Ribbon diagram. Regions marked as a', b', and c' correspond to peptides in Figures 6-4A, B, and C, respectively. B) Surface representation. C and D: exchange rates mapped onto the closed-form model of *E. coli* SecA. C) Dimer structure. Coloring of the left protomer corresponds to A and B. D) Interface of the PBD and the C-domain.

protomer surface, suggesting that large conformational changes accompany dimerization. Using the crystal structure 1M6N as a reference, we attribute the decreased HDX in the transducer helix (region a') to dimer-interface formation; this is also consistent with our AUC results described below. We assign the decreased HDX in the PBD (region b') to an induced allosteric effect caused by dimerization. Corresponding HDX changes are also evident within region c' and the HWD, supporting a model where SecA dimerization induces movement of the PBD into a closed conformation where it interacts with the C-terminal portion of SecA. This model is consistent with the previous observation of open and closed conformations of the PBD[14, 27, 30, 37, 38, 48, 49]. In the closed conformation, the PBD interacts extensively with the C-terminal domain and forms a compact structure, whereas in the open conformation, the PBD is more solvent exposed and would, therefore, undergo a greater extent of HDX. Owing to its vicinity to the transducer helix, the altered HDX in the two-helix finger may arise from dimerization or altered interactions with the PBD. The small $\Delta D\%$ of the short peptide 427-437 located in NBD II opposite the HSD may be due to an allosteric effect.

6.4.4 Analysis of the SecA Dimer Interface by AUC

To complement our analysis of the dimer interface by HDX, our collaborators from U. Connecticut used AUC to probe the energetic contribution of individual surface residues to SecA dimerization. Residues observed within the dimer interfaces of SecA crystal structures from multiple bacterial species[27-31] were chosen to be substituted with alanine. The program Robetta[50] was used to identify residues lying within potential dimerization interfaces and to guide selection of single alanine substitutions that are expected to strongly destabilize the SecA dimerization. AUC was then used to measure the dissociation constants of the alanine mutants. The result showed consistency with our HDX data, indicating two residue from HSD, Q662 and

E665, significantly affect SecA dimerization. Besides that, the AUC result also showed that another residue from the N-terminus, K8, is important to SecA dimerization as well, which was unrevealed in HDX results.

To analyze further the effect of N-terminus on SecA dimerization, deletion constructs, including Δ2-8, Δ2-9, and Δ2-10, were examined. And the result showed that SecA dimerization decreases with the number of residues deleted from the N-terminus. SecA constructs containing multiple alanine substitutions at the N-terminus were also generated and examined. And they found that dimerization affinity decreases as additional alanine substitutions are introduced.

For more details on this section, please see *Biochemistry*, **2014**, 53(19), 3248-3260.

6.4.5 Normal Mode Analysis

To explore the possibility of an allosteric mechanism in which SecA dimerization mediates structural changes, Prof. Eric May from U. Connecticut performed normal mode analysis on the monomer and dimer in the open and closed states. The analysis showed that the major structural difference is related to the separation between the PBD and HWD regions (Figure 6-6), and the dimer has a mode that couples much more strongly to the PBD-HWD domain separation than any mode in the monomer. The presence of this mode, which does not appear (as a single mode) in
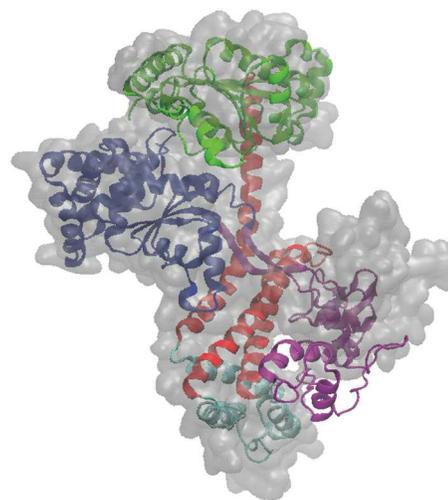
Figure 6- 6. Normal mode analysis of SecA monomer and proposed dimer configurations. The open state (PDB 2VDA) is shown as a surface representation in gray and the modeled closed state is shown in a ribbon diagram in red.

the monomer, indicates our proposed dimer interface is influencing the global dynamics of SecA, and providing flexibility to the PBD and HWD regions of the structure. These regions are distal to the proposed dimer interface and the observation is consistent with an allosteric mechanism where dimerization is linked to an open to closed transition.

For more details on this section, please see *Biochemistry*, **2014**, 53(19), 3248-3260.

6.5 Discussion

In the research described in this chapter, we utilized HDX MS in combination with site-directed mutagenesis/AUC performed by our collaborators, to probe the dimer interface of SecA in solution. HDX examines the effects of dimerization based on changes in peptide backbone solvent accessibility, with spatial resolution limited by digestion coverage and peptide size, whereas site-directed mutagenesis/AUC measures the energetic contributions of individual residues to dimerization. The HDX measurements were performed with a limited (10 s) time resolution and are, thus, insensitive to faster processes.

The HDX MS and AUC data together indicate that the HSD is a component of the dimer interface. Site-directed mutagenesis/AUC shows that residues Q662 and E665 lying at the C-terminus of the long transducer helix in the HSD are involved in dimerization. The HDX results implicate a larger region within the HSD, including the entire transducer helix and the two-helix finger. The difference in the extent of HSD involvement in the dimer interface detected by these methods may arise because some of the interfacial residues do not provide a large energetic contribution to dimerization. Alternatively, the transducer helix is known to be mobile[37], and changes in the accessibility of this helix upon dimerization may be communicated to regions lying distal from the dimer interface. With the exception of the *E. coli* structure (PDB 2FSF), each of the SecA crystal structures includes N-terminal portions of the transducer helix within

the dimer interface (Figure 6-2). However, only the *B. subtilis* 1M6N interface also encompasses the C-terminal region of this helix identified by AUC.

The AUC data indicate that the N-terminus contributes to the SecA dimer interface whereas HDX does not detect differences in this region between the monomer and dimer states. This discrepancy is likely associated with the flexible nature of the N-terminus. In this region, the deuterium uptake reaches a maximum at the shortest exchange time point (10 s), precluding detection of an interface by HDX, consistent with high on and off rates for the interaction. Our data demonstrating the involvement of the N-terminus in the SecA dimer interface disagrees with three dimer structures (PDB: 2FSF, 1NL3, and 2IBM) that do not include the N-terminus as part of the dimer interface. Note that these three structures were solved using SecA deletion mutants. Specifically, the *E. coli* SecA (PDB 2FSF) structure was solved using an N-terminus deletion mutant (Residues 9-861). In contrast, the dimer interfaces from *T. thermophilus* (PDB 2IPC) and *B. subtilis* (PDB 1M6N) structures contain the N-terminus.

Overall, the HDX and AUC data appear most consistent with the *B. subtilis* SecA 1M6N antiparallel dimer arrangement[27]. However, only 5 out of 14 alanine substitutions predicted to diminish dimerization affinity based on the 1M6N interface actually reduced dimerization of *E. coli* SecA, despite the 50% sequence identity between these orthologs. This discrepancy may indicate that the solution interface dimer is similar, but not identical, to that found in the 1M6N structure and, thus, none of the crystal structures may correspond to the interface adopted by cytoplasmic SecA. Also, another possible explanation is that different salt concentrations and buffer conditions may lead to the formation of slight different dimer interfaces. Alternatively, it was recently reported that SecA adopts multiple dimeric interfaces during translocation[51] and these may exist in equilibrium in solution. Both proposed dimer interfaces, based on the 1M6N

structure and an interface found in the *M. tuberculosis* 1NL3 crystal form (but different from the one depicted in Figure 6-2E), include the N-terminal region and portions of the HSD and HWD. However, the latter interface does not include residues Q662 and E665 that we found to affect SecA dimerization or the N-terminal portions of the transducer helix that are protected from HDX by dimerization. In agreement with our results, inter-protomer distances derived from fluorescence resonance energy transfer measurements are also most consistent with the 1M6N interface[34, 35].

The HDX data also reveal that dimerization induces movement of the PBD into a closed conformation where it interacts with the HWD. Normal mode analysis supports such an allosteric mechanism where dimerization is linked to an open to closed transition. We identified a low frequency mode in the dimer but not in the monomer that is coupled to the PBD-HWD separation. Consistent with correlation of SecA dimerization with a closing transition of the PBD, *B. subtilis* SecA adopts an open conformation in a monomeric crystal form (1TF5)[38] whereas closed forms are found in two dimeric crystal forms of the same enzyme[27, 30]. Cryo-electron microscopy analysis of dimeric *E. coli* SecA reveals a closed state[14]. However, this correlation is not always observed, as a monomeric complex of *T. maritima* SecA with ADP can be crystalized in a closed state[37] and a crystal structure of dimeric *E. coli* SecA (PDB 2FSF) reveals an open state[31]. Crystal packing forces may influence the conformation of the PBD. Consistent with a low barrier for this transition, solution NMR suggests that dimeric SecA rapidly interconverts between a closed state and a predominant open state[8]. FRET measurements indicate that signal peptide binding induces dimeric SecA to adopt a more open conformation[35]. Binding of a non-signal sequence peptide in the groove between PBD and NBD

II results in rotation of the PBD towards NBD II[37] and interaction of SecA with SecYEG induces further movement of the PBD to close the clamp[24].

Our analysis can be interpreted in the context of the closed state model of the *E. coli* SecA dimer based on the 1M6N structure (Figure 6-5C). The dimer interface is primarily contributed by the HSD and the N-terminus, two regions that lie on the same side of the protomer. The C-terminal domain regions that form an interface with the PBD become less dynamic upon dimerization (Figure 6-5D). However, the corresponding colored regions in the PBD are not in close contact with the C-terminal domain. Most of the colored residues in the PBD are also removed from the inter-protomer interface, but a short loop does contribute to dimerization (residues 338-342) and a portion of the reduced HDX in the PBD may be directly associated with dimerization. The two regions in the PBD making close contact with the C-terminal domain, residues 263-268 (loop) and 304-314 (α helix), do not show significant $\Delta D\%$ between monomer and dimer. Region 263-268 is a flexible loop that reaches 60% deuterium exchange at the shortest HDX time point, 10 s, and 80% at 4 h. In contrast, the helix in region 306-314 is rigid and shows less than 5% deuterium exchange at the longest HDX time point, 4 h. Thus, the absence of a detectable difference in monomer and dimer HDX for these regions can be ascribed to very fast and very slow exchange, respectively, that lie outside the time resolution of the present study.

Several studies show that SecA is dimeric when driving preprotein translocation across SecYEG[15, 19, 20, 37, 52, 53] whereas others report that the functional form of SecA is monomeric[18, 24, 25]. Regions within the PBD and C-domain of SecA interact with the C-terminal region of SecY[24]. When SecA binds to SecYEG, the PBD moves away from HWD and closer to the NBD II[24]. The *B. subtilis* (PDB 1M6N) dimer interface supported by the data presented here does not overlap with the SecA-SecYEG interface so that the intact dimer could interact with SecYEG.

Interaction with SecYEG and subsequent ATP hydrolysis may induce SecA dimer dissociation during the translocation cycle[1, 51]. In this context, the movement of the PBD to an open conformation upon SecA dimer dissociation may facilitate transfer of the preprotein to the SecY pore.

6.6 Acknowledgements

6.7 References

1.   Kusters, I., and Driessen, A. J. (2011) SecA, a remarkable nanomachine, *Cellular and molecular life sciences : CMLS 68*, 2053-2066.

2.   Segers, K., and Anne, J. (2011) Traffic jam at the bacterial sec translocase: targeting the SecA nanomotor by small-molecule inhibitors, *Chemistry & biology 18*, 685-698.

3.   Economou, A., and Wickner, W. (1994) SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion, *Cell 78*, 835-843.

4.   Papanikou, E., Karamanou, S., and Economou, A. (2007) Bacterial protein secretion through the translocase nanomachine, *Nature reviews. Microbiology 5*, 839-851.

5.   Tomkiewicz, D., Nouwen, N., van Leeuwen, R., Tans, S., and Driessen, A. J. (2006) SecA supports a constant rate of preprotein translocation, *The Journal of biological chemistry 281*, 15709-15713.

6.   van der Wolk, J. P., de Wit, J. G., and Driessen, A. J. (1997) The catalytic cycle of the escherichia coli SecA ATPase comprises two distinct preprotein translocation events, *The EMBO journal 16*, 7297-7304.

7.   Sardis, M. F., and Economou, A. (2010) SecA: a tale of two protomers, *Molecular microbiology 76*, 1070-1081.

8.   Gelis, I., Bonvin, A. M., Keramisanou, D., Koukaki, M., Gouridis, G., Karamanou, S., Economou, A., and Kalodimos, C. G. (2007) Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR, *Cell 131*, 756-769.

9.   Musial-Siwek, M., Rusch, S. L., and Kendall, D. A. (2007) Selective photoaffinity labeling identifies the signal peptide binding domain on SecA, *Journal of molecular biology 365*, 637-648.

10. Auclair, S. M., Moses, J. P., Musial-Siwek, M., Kendall, D. A., Oliver, D. B., and Mukerji, I. (2010) Mapping of the signal peptide-binding domain of Escherichia coli SecA using Forster resonance energy transfer, *Biochemistry 49*, 782-792.

11. Grady, L. M., Michtavy, J., and Oliver, D. B. (2012) Characterization of the Escherichia coli SecA signal peptide-binding site, *Journal of bacteriology 194*, 307-316.

12. Bhanu, M. K., Zhao, P., and Kendall, D. A. (2013) Mapping of the SecA signal peptide binding site and dimeric interface by using the substituted cysteine accessibility method, *Journal of bacteriology 195*, 4709-4715.

13. Woodbury, R. L., Hardy, S. J., and Randall, L. L. (2002) Complex behavior in solution of homodimeric SecA, *Protein science : a publication of the Protein Society 11*, 875-882.

14. Chen, Y., Pan, X., Tang, Y., Quan, S., Tai, P. C., and Sui, S. F. (2008) Full-length Escherichia coli SecA dimerizes in a closed conformation in solution as determined by cryo-electron microscopy, *The Journal of biological chemistry 283*, 28783-28787.

15. Kusters, I., van den Bogaart, G., Kedrov, A., Krasnikov, V., Fulyani, F., Poolman, B., and Driessen, A. J. (2011) Quaternary structure of SecA in solution and bound to SecYEG probed at the single molecule level, *Structure 19*, 430-439.

16. Wowor, A. J., Yu, D., Kendall, D. A., and Cole, J. L. (2011) Energetics of SecA dimerization, *Journal of molecular biology 408*, 87-98.

17. Kimura, E., Akita, M., Matsuyama, S., and Mizushima, S. (1991) Determination of a region in SecA that interacts with presecretory proteins in Escherichia coli, *The Journal of biological chemistry 266*, 6600-6606.

18.   Or, E., Navon, A., and Rapoport, T. (2002) Dissociation of the dimeric SecA ATPase during protein translocation across the bacterial membrane, *The EMBO journal 21*, 4470-4479.

19.   de Keyzer, J., van der Sluis, E. O., Spelbrink, R. E., Nijstad, N., de Kruijff, B., Nouwen, N., van der Does, C., and Driessen, A. J. (2005) Covalently dimerized SecA is functional in protein translocation, *The Journal of biological chemistry 280*, 35255-35260.

20.   Jilaveanu, L. B., and Oliver, D. (2006) SecA dimer cross-linked at its subunit interface is functional for protein translocation, *Journal of bacteriology 188*, 335-338.

21.   Wang, H., Na, B., Yang, H., and Tai, P. C. (2008) Additional in vitro and in vivo evidence for SecA functioning as dimers in the membrane: dissociation into monomers is not essential for protein translocation in Escherichia coli, *Journal of bacteriology 190*, 1413-1418.

22.   Benach, J., Chou, Y. T., Fak, J. J., Itkin, A., Nicolae, D. D., Smith, P. C., Wittrock, G., Floyd, D. L., Golsaz, C. M., Gierasch, L. M., and Hunt, J. F. (2003) Phospholipid-induced monomerization and signal-peptide-induced oligomerization of SecA, *The Journal of biological chemistry 278*, 3628-3638.

23.   Musial-Siwek, M., Rusch, S. L., and Kendall, D. A. (2005) Probing the affinity of SecA for signal peptide in different environments, *Biochemistry 44*, 13987-13996.

24.   Zimmer, J., Nam, Y., and Rapoport, T. A. (2008) Structure of a complex of the ATPase SecA and the protein-translocation channel, *Nature 455*, 936-943.

25.   Or, E., Boyd, D., Gon, S., Beckwith, J., and Rapoport, T. (2005) The bacterial ATPase SecA functions as a monomer in protein translocation, *The Journal of biological chemistry 280*, 9097-9105.

26.   Or, E., and Rapoport, T. (2007) Cross-linked SecA dimers are not functional in protein translocation, *FEBS letters 581*, 2616-2620.

27.   Hunt, J. F., Weinkauf, S., Henry, L., Fak, J. J., McNicholas, P., Oliver, D. B., and Deisenhofer, J. (2002) Nucleotide control of interdomain interactions in the conformational reaction cycle of SecA, *Science 297*, 2018-2026.

28.   Sharma, V., Arockiasamy, A., Ronning, D. R., Savva, C. G., Holzenburg, A., Braunstein, M., Jacobs, W. R., Jr., and Sacchettini, J. C. (2003) Crystal structure of Mycobacterium tuberculosis SecA, a preprotein translocating ATPase, *Proceedings of the National Academy of Sciences of the United States of America 100*, 2243-2248.

29.   Vassylyev, D. G., Mori, H., Vassylyeva, M. N., Tsukazaki, T., Kimura, Y., Tahirov, T. H., and Ito, K. (2006) Crystal structure of the translocation ATPase SecA from Thermus thermophilus reveals a parallel, head-to-head dimer, *Journal of molecular biology 364*, 248-258.

30.   Zimmer, J., Li, W., and Rapoport, T. A. (2006) A novel dimer interface and conformational changes revealed by an X-ray structure of B. subtilis SecA, *Journal of molecular biology 364*, 259-265.

31.   Papanikolau, Y., Papadovasilaki, M., Ravelli, R. B., McCarthy, A. A., Cusack, S., Economou, A., and Petratos, K. (2007) Structure of dimeric SecA, the Escherichia coli preprotein translocase motor, *Journal of molecular biology 366*, 1545-1557.

32.   Carugo, O., and Argos, P. (1997) Protein-protein crystal-packing contacts, *Protein science : a publication of the Protein Society 6*, 2261-2263.

33.   Xu, Q., Canutescu, A. A., Wang, G., Shapovalov, M., Obradovic, Z., and Dunbrack, R. L., Jr. (2008) Statistical analysis of interface similarity in crystals of homologous proteins, *Journal of molecular biology 381*, 487-507.

34.   Ding, H., Hunt, J. F., Mukerji, I., and Oliver, D. (2003) Bacillus subtilis SecA ATPase exists as an antiparallel dimer in solution, *Biochemistry 42*, 8729-8738.

35.   Auclair, S. M., Oliver, D. B., and Mukerji, I. (2013) Defining the solution state dimer structure of Escherichia coli SecA using Forster resonance energy transfer, *Biochemistry 52*, 2388-2401.

36.   Yu, D., Wowor, A. J., Cole, J. L., and Kendall, D. A. (2013) Defining the Escherichia coli SecA dimer interface residues through in vivo site-specific photo-cross-linking, *Journal of bacteriology 195*, 2817-2825.

37.   Zimmer, J., and Rapoport, T. A. (2009) Conformational flexibility and peptide interaction of the translocation ATPase SecA, *Journal of molecular biology 394*, 606-612.

38.   Osborne, A. R., Clemons, W. M., Jr., and Rapoport, T. A. (2004) A large conformational change of the translocation ATPase SecA, *Proceedings of the National Academy of Sciences of the United States of America 101*, 10937-10942.

39.   Erlandson, K. J., Miller, S. B., Nam, Y., Osborne, A. R., Zimmer, J., and Rapoport, T. A. (2008) A role for the two-helix finger of the SecA ATPase in protein translocation, *Nature 455*, 984-987.

40.   Gold, V. A., Whitehouse, S., Robson, A., and Collinson, I. (2013) The dynamic action of SecA during the initiation of protein translocation, *The Biochemical journal 449*, 695-705.

41.   Laue, T. M., Shah, B., Ridgeway, T. M., and Pelletier, S. L. (1992) Computer-aided interpretation of sedimentation data for proteins. In Analytical Ultracentrifugation in Biochemistry and Polymer Science. S. E. Harding, J. C. Horton and A. J. Rowe, editors. , *Royal Society of Chemistry, Cambridge, England*, 90-125.

42.   Xu, H., and Freitas, M. A. (2009) MassMatrix: a database search program for rapid characterization of proteins and peptides from tandem mass spectrometry data, *Proteomics 9*, 1548-1555.

43.   Pascal, B. D., Willis, S., Lauer, J. L., Landgraf, R. R., West, G. M., Marciano, D., Novick, S., Goswami, D., Chalmers, M. J., and Griffin, P. R. (2012) HDX workbench: software for the analysis of H/D exchange MS data, *Journal of the American Society for Mass Spectrometry 23*, 1512-1521.

44.   Zhang, Z., and Smith, D. L. (1993) Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation, *Protein science : a publication of the Protein Society 2*, 522-531.

45.   Tobler, S. A., Sherman, N. E., and Fernandez, E. J. (2000) Tracking lysozyme unfolding during salt-induced precipitation with hydrogen exchange and mass spectrometry, *Biotechnology and bioengineering 71*, 194-207.

46.   Wales, T. E., and Engen, J. R. (2006) Hydrogen exchange mass spectrometry for the analysis of protein dynamics, *Mass spectrometry reviews 25*, 158-170.

47.   Mori, H., and Ito, K. (2006) The long alpha-helix of SecA is important for the ATPase coupling of translocation, *The Journal of biological chemistry 281*, 36249-36256.

48.   Tang, Y., Pan, X., Chen, Y., Tai, P. C., and Sui, S. F. (2011) Dimeric SecA couples the preprotein translocation in an asymmetric manner, *PloS one 6*, e16498.

49.   Maki, J. L., Krishman, B., and Gierasch, L. M. (2012) Using a low denaturant model to explore the conformational features of translocation-active SecA, *Biochemistry 51*, 1369-1379.

50.   Kortemme, T., Kim, D. E., and Baker, D. (2004) Computational alanine scanning of protein-protein interfaces, *Science's STKE : signal transduction knowledge environment 2004*, pl2.

51.   Gouridis, G., Karamanou, S., Sardis, M. F., Scharer, M. A., Capitani, G., and Economou, A. (2013) Quaternary dynamics of the SecA motor drive translocase catalysis, *Molecular cell 52*, 655-666.

52.   Jilaveanu, L. B., Zito, C. R., and Oliver, D. (2005) Dimeric SecA is essential for protein translocation, *Proceedings of the National Academy of Sciences of the United States of America 102*, 7511-7516.

53.   Driessen, A. J. (1993) SecA, the peripheral subunit of the Escherichia coli precursor protein translocase, is functional as a dimer, *Biochemistry 32*, 13190-13197.

Chapter 7*

Fast Photochemical Oxidation of Proteins (FPOP) Maps the

Epitope of EGFR Binding to Adnectin

7.1 Abstract

Epitope mapping is an important tool for the development of monoclonal antibodies, mAbs, as therapeutic drugs. Recently, a class of therapeutic mAb alternatives, adnectins, has been developed as targeted biologics. They are derived from the tenth type III domain of human fibronectin (^{10}Fn3). A common approach to map the epitope binding of these therapeutic proteins to their binding partners is X-ray crystallography. Although the crystal structure is known for Adnectin 1 binding to human EGFR, we seek to determine complementary binding in solution and to test the efficacy of footprinting for this purpose. As a relatively new tool in structural biology and complementary to X-ray crystallography, protein footprinting coupled with mass spectrometry is promising for protein-protein interaction studies. We report in this chapter the use of fast photochemical oxidation of proteins (FPOP) coupled with MS to map the epitope of EGFR-Addnectin 1 at both the peptide and amino-acid residue levels. The data correlate well with the previously determined epitope from the crystal structure and are consistent with HDX MS data. The FPOP-determined binding interface involves various amino-acid and peptide regions near the N terminus of EGFR. The outcome adds credibility to oxidative labeling by FPOP for epitope mapping and motivates more applications in the therapeutic protein area as a stand-alone method or in conjunction with X-ray crystallography, NMR, site-directed mutagenesis, and other orthogonal methods.

7.2 Introduction

Epitope mapping is an important step in the characterization of monoclonal antibodies (mAb) for their use as therapeutic drugs. Therapeutic applications of mAbs are emerging or under development for oncology, autoimmune diseases, inflammatory disorders, and organ transplantation. Recently, a new class of biologics that mimic the binding region of mAbs, Adnectins, has been developed as therapeutic mAbs analogs[1-3].

Adnectins are a class of biologics developed from the tenth human fibronectin type III domain (^{10}Fn3), and they bind target proteins as potential therapeutic agents[1-6]. ^{10}Fn3 is a small (10 kDa), highly stable and soluble β-sandwich protein that resembles an immunoglobulin variable domain but has no disulfide bonds[7-10]. The protein fold contains three solvent-exposed loops, termed BC, DE and FG, which are structurally analogous to antibody complementarity-determining regions (CDRs) and tolerant to extensive mutations[4]. Mutations in the loops can impart different binding capacities to the ^{10}Fn3-based variants. Therefore, adnectins can have similar functions as therapeutic monoclonal antibodies by binding with high affinity to their targets. The first adnectin tested in preclinical and phase I studies, CT-322, targets vascular endothelial growth factor receptor-2, giving some desired pharmacological effects[11-14].

Epidermal growth factor receptor (EGFR), a member of the ErbB receptor family, mediates cell proliferation, differentiation, survival, angiogenesis, and migration[15]. EGFR has been implicated in many human cancers[16]. Its structure consists of an extracellular domain (exEGFR), a transmembrane domain, and an intracellular tyrosine-kinase domain. Cell signaling is initiated by binding of ligands, such as EGF, to exEGFR, followed by dimerization of EGFR and phosphorylation of specific EGFR tyrosine residues[15]. We recently studied the effects of dimerization and subsequent phosphorylation by mass spectrometry (MS)-based footprinting of

one family member attached to a membrane[17]. Tyrosine-phosphorylated receptors serve as docking sites for intracellular signaling proteins, which initiate many signaling cascades to produce a physiological outcome[18]. This EGFR signaling network is often dys-regulated in cancer and motivates a strategy for cancer therapy with EGFR-blocking agents, including mAbs targeting exEGFR and small molecules targeting the EGFR intracellular domain[19-21].

To develop mAb alternatives for cancer therapy, a representative adnectin (Adnectin 1) that specifically binds EGFR was generated using the mRNA display technique[22, 23]. As a requirement in therapeutic drug development, the EGFR-Adnectin 1 interaction was previously characterized, in this case by X-ray crystallography, revealing the binding epitope to be distinct from those of approved mAbs[22]. Although X-ray crystallography is often regarded as the "gold standard" for structure determination, it has limitations. For example, crystallization is sometimes difficult for certain antigen-antibody complexes and requires a relatively large amount of protein. Furthermore, whether contacts from crystal structures faithfully represents biologically relevant protein-protein interfaces is questionable[24, 25]. Another approach, site-directed mutagenesis[26], is labor intensive and could cause false positives because the protein is submitted to modifications, some of which could affect structure. Binding assays[27] and PEPSCAN[28] are often restricted to linear (continuous) epitopes, and are likely to fail when facing conformational (discontinuous) epitopes, characterization of which requires maintaining the native structure of both antigen and antibody. Limited proteolysis[29] followed by MS analysis is another approach but of low resolution.

There is a growing interest in mass-spectrometry-based methods for mapping epitopes and other protein-ligand interactions. Two appropriate approaches are hydrogen-deuterium exchange (HDX)[30-36] and oxidative labeling (e.g., OH radicals, FPOP)[37-40]. These methods are sensitive,

can accommodate large proteins, and can be used with the protein in native or near-native states. For example, the epitope of EGFR binding to Adnectin 1 was mapped by Iacob and Engen using HDX MS[41], and the results correlate well with the structure from X-ray crystallography. Oxidative labeling has advantages because the labeling is fast, irreversible, and residue-level information can often be generated. Fast oxidative labeling (millisec time scale) was first established by Chance and coworkers[42] by radiolytic modification. Photochemical labeling initiated by laser irradiation of H_2O_2 was later developed by different groups[43-45]. Of those methods, photochemical oxidation of proteins (FPOP), developed by Hambly and Gross[45], uses a flow system and controls the radical life to ~ 1 µs by introducing a radical scavenger to prevent excessive labeling. Recently, Jones and Gross[40] reported the use of FPOP coupled to MS to map the conformational epitope of the serine protease thrombin and showed results that are consistent with those from HDX.

In this chapter, we describe the use of FPOP to map the epitope of exEGFR-Adnectin 1 complex. To accomplish the mapping, we modified the experimental conditions from traditional FPOP settings to allow measurement in formulation-relevant conditions and to produce moderate amount of oxidation of exEGFR even in the presence of the adnectin. The FPOP data are largely consistent with the X-ray crystallography results, which show that the epitope binds at the N terminus in Domain I, as shown by decreased modifications upon Adnectin 1 binding. In addition, FPOP also revealed changes of a residue located outside the previously determined binding interface, F24, which may be due to an allosteric change in side-chain orientation that produces decreased solvent accessibility.

7.3 Experimental

7.3.1 Materials and Reagents

30% hydrogen peroxide, formic acid, trifluoroacetic acid, *L*-histidine, *L*-glutamin, *L*-methionine, catalase, phosphate-buffered saline, ammonium bicarbonate, dithiothreitol (DTT), iodo-acetamide (IAA), and HPLC-grade solvents were purchased from Sigma Aldrich (St. Louis, MO). Trypsin/Lys-C Mix was obtained from Promega Corp. (Madison, WI). PNGase F was purchased from New England Biolabs inc (Ipswich, MA). RapiGest SF was from Waters Corp. (Milford, MA). West Nile Virus E protein Domain III (WNV E DIII) was provided by D. Fremont from Department of Pathology and Immunology at Washington University, school of medicine. Human exEGFR (residues 1-642) was expressed in Sf9 cells mammalian cells with a C-terminal His tag and purified as described previously[22]. It was supplied in phosphate buffer saline (PBS:137mM NaCl, 2.7mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4 (pH 7.2). Adnectin 1 was supplied in PBS buffer (pH 7.2). Its expression and purification were described previously[22].

7.3.2 FPOP labeling of Proteins

exEGFR and Adnectin 1 were mixed together at a 1:1 molar ratio at 50 μM concentration ($K_d \sim 2$ nM[22]) and incubated for 1 h at room temperature. FPOP labeling of proteins was performed on the same day under the same conditions. Each sample started with 5 μM protein reconstituted in 1×PBS containing 350 μM His and 20 mM H_2O_2. H_2O_2 was added just prior to infusing the solution into the tubing for FPOP. FPOP was performed as described previously[45, 46]. The KrF excimer laser power (GAM Laser Inc., Orlando, FL) was adjusted to 44 mJ/pulse, and its pulse frequency set to 6.5 Hz. The width of the laser beam at the clear sample tube window was 1.8

mm. The flow rate was adjusted to 15 µL/min to ensure a 20% exclusion volume to avoid repeated HO·exposure and reaction[44]. All collections were in vials containing 10 mM catalase and 20 mM Met to reduce left-over H_2O_2. For each state, exEGFR with Adnectin 1 and exEGFR alone, three separate samples were made and submitted to FPOP labeling. In addition, three control samples for exEGFR alone were handled in the same manner, but they were not laser irradiated. All samples were stored at -80 °C after solvent evaporation if they were not directly subjected to trypsin digestion.

7.3.3 Proteolysis

A 20 µL aliquot of each FPOP-labeled exEGFR sample was mixed with 2 µL of RapiGest SF and incubated at 80 °C for 15 min. The samples were reduced with 10 mM dithiothreitol (2 µL of 10 mM solution) at 55 °C for 30 min, alkylated with 23 mM iodoacetamide (2 µL of 300 mM solution) in the dark at RT for 30 min, and deglycosylated with 2 µL PNGsae F at 37 °C for 2 h. Solutions of 40 µL of 1×PBS and 24 µL of 0.1 M ammonium bicarbonate were then added to each aliquot to bring the pH to ~8. Samples were digested overnight with the Trypsin/Lys-C Mix at an enzyme:protein ratio of 1:25. The reaction was quenched with trifluoroacetic acid (TFA) to give a final concentration of 1%. An aliquot of 30 µL of the digested samples was cleaned up with NuTip C-18 zip tips (Glygen Corp., Columbia, MD) and eluted into 30 µL of 60% acetonitrile/40% H_2O/0.1% formic acid (FA). The eluent was dried and reconstituted in 30 µL of water with 0.1% FA for autosampler loading and subsequent analysis.

7.3.4 Mass Spectrometry

For peptide and residue level analysis of labeled exEGFR, samples were analyzed on a LTQ Orbitrap XL (Thermo Fisher, San Jose, CA) operated in data-dependent acquisition mode. An

aliquot of 5 μL of each sample was loaded by the autosampler onto and eluted from a 15 cm column with a PicoFrit tip (New Objective, Inc., Woburn, MA), custom-packed with C18 reversed phase material (Magic, 0.075 mm × 150 mm, 5 μm, 100 Å; Michrom Bioresources Inc., Auburn, CA); the chromatograph was an Ultra 1D+ UPLC (Eksigent, Dublin, CA). Peptides were eluted by a 95 min, 260 nL/min gradient coupled to the nanospray source. The gradient started with a linear increase from 2% of solvent B (CH$_3$CN/0.1%FA) to 32% in 65 min, and then to 90% in 13 min, held at 90% B for 5 min, and re-equilibrated to solvent A (H$_2$O/0.1%FA) for 12 min. Mass spectra were obtained at high mass resolving power (60,000, FWHH, for ions of m/z 400), and the six most abundant ions eluting per scan were subjected to CID MS^2 in the ion trap, with charge-state rejection of +1 ions enabled. Precursor ions were added to a dynamic exclusion list for 12 s to ensure good sampling of each elution peak.

7.3.5 Adjustment of the Scavenger for FPOP

FPOP used an excimer laser to photolyze hydrogen peroxide to give two OH radicals; the H$_2$O$_2$ was added in low concentration to the protein solution, to form hydroxyl radicals[45, 47]. To control the radical lifetime, a pulsed laser and selected scavenger of 20 mM Gln were used for protein labeling at a microsecond time scale, faster than protein unfolding[46]. There were concerns, however, regarding the impact of high concentration of Gln on protein conformation and behavior and whether it faithfully mimicked the protein native state.

Thus, His, a common reagent in protein formulation buffers, was selected to replace Gln as a scavenger in the FPOP experiments. Because the reaction rate constant of ·OH with His is larger than that with Gln, His was expected to be a more efficient scavenger than Gln and to work at lower concentrations. To determine the concentration of His as an alternative to the previously used 20 mM Gln scavenger, each sample started with 10 μM WNV E DIII (a model protein), 20

mM H_2O_2 and different concentrations of His (200 µM, 350 µM, 500 µM) or 20 mM Gln in 1×PBS. The traditional FPOP condition with 20 mM Gln was included to serve as a positive control. After FPOP labeling, protein-level analysis of labeled WNV E DIII was performed on a Bruker Maxis 4G MS (Bruker Biosciences Corporation, Billerica, MA). Samples were trapped on a 2×15 mm C8 trap column (Agilent, Santa Clara, CA, USA) and desalted with a 3 min flow. A 5 min linear gradient of 4-40% CH_3CN in 0.1% formic acid was used to elute the sample for protein-level analysis. The peak corresponding to +8 charge state from the global mass spectrum was used for data analysis

Global analysis with the custom program afforded the fraction of unmodified protein after FPOP for all four conditions, as shown in Table 7-1. The table indicates that 350 µM His gave the closest scavenging ability comparing to 20 mM Gln. According to previous studies[48], one important criterion indicating the sampling of a single (native) conformation of the protein is the Poisson distribution of +0, +16, +32 species after FPOP labeling. Therefore, a more rigorous analysis was conducted, to model the product distribution for all four experiments. The goodness of the fit was indicated by the "Poisson distribution factor" in Table 7-1, for which a value closer to "1" indicated the better fit. The result also indicated the suitability of 350 µM His as a scavenger for FPOP experiments.

Table 7- 1. Analysis of FPOP modification on WNV E DIII at protein level with different scavengers

Scavenger	Unmodified WNV E DIII%	Poisson distribution factor
20 mM Gln	24.9%	0.89
500 µM His	26.4%	0.89
350 µM His	26.2%	0.91
200 µM His	17.4%	0.75

7.3.6 Data Analysis for FPOP labeling of exEGFR

The raw MS files were converted to .mgf files and profile mzXML files by using MassMatrix Mass Spec Data File Converter. The .mgf files of control samples were searched using MASCOT (Matrix Science, London, U.K.) against a custom-built database containing the sequence of exEGFR for sequence coverage and including deamidation at Asn or Gln, which results from *N*-deglycosylation in the sample handling. Oxidation at Met was also added as a variable modification. The alkylation of the samples with iodo-acetamide added a carbamidomethyl group (MW= 57.0214) to cysteines; therefore, it too was set as a fixed modification in the search. Several Asn residues were found to be entirely deamidated to form Asp or isoAsp, and they were changed in the sequence from N to D for further FPOP modification analysis. The profile mzXML files of all samples were then searched against the modified sequence for labeled and unlabeled exEGFR tryptic peptides by using ProtmapMS 2.1 (Case Western Reserve Univ., OH) with oxidative mode set to 2, meaning up to two variable modifications were allowed to occur on the same peptide. All known hydroxyl radical side-chain reaction products[49-51] were added to the modification database for searching as variable modifications. Carbamidomethyl group modification on Cys was added as a fixed modification. From the identified modifications on each tryptic peptide (Table 7-2), the intensities of signals for the modified (I_{ox}) and the unmodified species (I) for each peptide were then read from the raw data files. The extents of modification at peptide level were calculated by using the following equation[52]:

$$Extent\ of\ Modification = \frac{\sum I_{ox}}{\sum I_{ox} + I}$$

Table 7- 2. Modifications by FPOP on all peptides as identified by ProtmapMS

peptide	Modification species								
	15.995	13.979	-10.032	-22.032	31.99	-30.011	4.979	-22.032	-31.99
14-29	+	+	-	-	-	-	-	-	-
30-48	+	-	-	-	-	-	-	-	-
49-56	+	-	-	-	-	-	-	-	-
57-74	+	-	-	-	-	-	-	-	-
75-84	+	-	-	-	-	-	-	-	-
85-105	+	-	-	-	-	-	-	-	-
106-	-	-	-	-	-	-	-	-	-
110-	-	-	-	-	-	-	-	-	-
115-	+	-	+	+	-	-	-	-	-
126-	+	-	-	-	+	-	-	-	-
142-	+	-	-	-	+	-	-	-	-
166-	+	-	-	-	-	-	-	-	-
189-	+	-	-	-	-	-	-	-	-
201-	+	-	-	-	-	-	-	-	-
203-	+	-	-	-	-	+	+	-	-
221-	+	-	-	-	-	-	-	-	-
230-	+	-	-	-	+	-	-	-	-
238-	+	-	-	-	+	-	-	-	-
261-	+	-	-	-	-	+	+	-	-
261-	+	-	-	-	-	-	-	-	-
274-	+	-	-	-	+	+	+	+	-
286-	+	-	-	-	-	-	-	-	-
286-	+	+	-	-	+	-	-	-	+
311-	+	-	-	-	-	-	-	-	-
312-	+	-	-	-	-	-	-	-	-
323-	+	+	-	-	-	-	-	-	-
334-	-	-	-	-	-	-	-	-	-
337-	+	-	-	-	+	-	-	-	-
354-	+	+	+	+	+	-	-	-	-
376-	+	-	-	-	+	-	-	-	-
391-	+	+	-	-	-	-	+	-	-
408-	-	-	+	-	-	-	+	-	-
428-	+	-	-	-	-	-	-	-	-
431-	+	+	-	-	-	-	-	-	-
444-	+	-	-	-	+	-	-	-	-
444-	+	-	-	-	+	-	-	-	-
456-	+	+	-	-	-	-	-	-	-
477-	+	-	-	-	+	+	+	-	-
515-	+	-	-	-	-	-	-	-	-
524-	+	-	-	-	+	+	-	-	-

"+" indicates the modification is detected and "-" indicates the modification is not detected

Quantitative analysis of oxidative labeling at residue level was also performed for regions that showed differences in extent of modification at the peptide level between exEGFR alone and exEGFR with Adnectin 1. Modification sites on the peptide were assigned with MS^2 information. In a few cases, where location of a modification to a single residue wasn't possible owing to limited fragmentation information from MS^2 or to the presence of interference from co-elution of peptide isomers, the modification was indicated to occur on a set of possible residues.

7.4 Results and Discussion

7.4.1 FPOP labeling and analysis of exEGFR

FPOP is a chemical footprinting approach that uses short-lived $\cdot$OH radicals to "snapshot" the state of a protein or a protein-complex. Hydroxyl radicals probe the solvent accessibility of the protein because its size is similar to that of water molecules, and its high reactivity allows modification of up to 14 of 20 amino acids. In the presence of a scavenger, the radical lifetime is controlled to the μs timescale, which allows fast labeling presumably of a single state of the protein. In most FPOP footprinting experiments, the comparison is two-state; that is, of the same site of a protein existing in two states to reveal changes in solvent accessibility.

By introducing an irreversible modification, the FPOP-introduced label is maintained during post-labeling sample handling, making it adaptable to a typical "bottom-up" proteomics methodology of proteolysis and LC/MS/MS. This robustness is important for membrane and secreted proteins containing complex glycosylation and other modifications, which are important for the structural and functional roles of the protein[53]. The diversity of glycan modifications often adds to the difficulty of data processing in proteomics studies. In the case of exEGFR, we found that, without a deglycosylation step, trypsin digestion followed by typical database

searching gave only 74% of sequence coverage. Missing were several regions between AA320 to

AA550. Interestingly, these missing regions all contain a "NXT" sequence, which is the common

site for N-glycosylation. Therefore, we added a deglycosylation step before trypsin digestion to

bring the coverage up to 89%, yet still missing several short regions containing multiple Lys or

Arg residues (Figure 7-1).

```
1     LEEKK VCQGT SNKLT QLGTF EDHFL SLQRM FNNCE VVLGN LEITY VQRNY DLSFL KTIQE
              ──────────────────────────────────────>              ──────────────>    ──────>

61    VAGYV LIALN TVERI PLENL QIIRG NMYYE NSYAL AVLSN YDANK TGLKE LPMRN LQEIL
      ───────────────────>        ─────>                        ─────────>    ──────>

121   HGAVR FSNNP ALCNV ESIQW RDIVS SDFLS NMSMD FQNHL GSCQK CDPSC PNGSC WGAGE
      ──────>              ──────────────>                        ──────────────>

181   ENCQK LTKII CAQQC SGRCR GKSPS DCCHN QCAAG CTGPR ESDCL VCRKF RDEAT CKDTC
      ──────>                                                  ──────>    ─────────────

241   PPLML YNPTT YQMDV NPEGK YSFGA TCVKK CPRNY VVTDH GSCVR ACGAD SYEME EDGVR
      ══════════════════    ──────────────>    ─────>  ──────────────────>  ─────────────>

301   KCKKC EGPCR KVCNG IGIGE FKDSL SINAT NIKHF KNCTS ISGDL HILPV AFRGD SFTHT
      ─>        ───────────────────>        ─────>        ──────────────>      ─────>

361   PPLDP QELDI LKTVK EITGF LLIQA WPENR TDLHA FENLE IIRGR TKQHG QFSLA VVSLN
          ────────────────>    ────────────>        ──────>    ──────>      ────────────

421   ITSLG LRSLK EISDG DVIIS GNKNL CYANT INWKK LFGTS GQKTK IISNR GENSC KATGQ
              ──────>        ──────>        ─────>                              ─────

481   VCHAL CSPEG CWGPE PRDCV SCRNV SRGRE CVDKC NLLEG EPREF VENSE CIQCH PECLP
          ──────────────>                              ────────────>

541   QAMNI TCTGR GPDNC IQCAH YIDGP HCVKT CPAGV MGENN TLVWK YADAG HVCHL CHPNC
      ──────>                  ──────────────>                  ────────────>

601   TYGCT GPGLE GCPTN GPKHH HHHH
      ──────────────────────>
```

Figure 7- 1. Coverage map obtained for the tryptic digest of exEGFR.

7.4.2 Epitope mapping of exEGFR-Adnectin 1

We first performed FPOP labeling of exEGFR with 350 μM His as scavenger in the absence of

Adnectin 1 and analyzed the oxidation extent at the peptide level to evaluate the labeling

efficiency. This is important because different proteins contain different reactive sites, and,

therefore, may require different concentrations of scavengers to achieve a proper modification

extent. Because exEGFR is much larger in size than WNV E DIII, it is unlikely to get over-

oxidized but possibly under-oxidized under the same scavenger concentration. Therefore, we only studied the oxidation extent but didn't perform distribution analysis for the peptides after FPOP labeling. The results showed that moderate oxidation, between 0% and 25%, can be achieved for all peptides (darker bars in **Figure 7-2**). This outcome is the basis for a more detailed FPOP study of the two states.
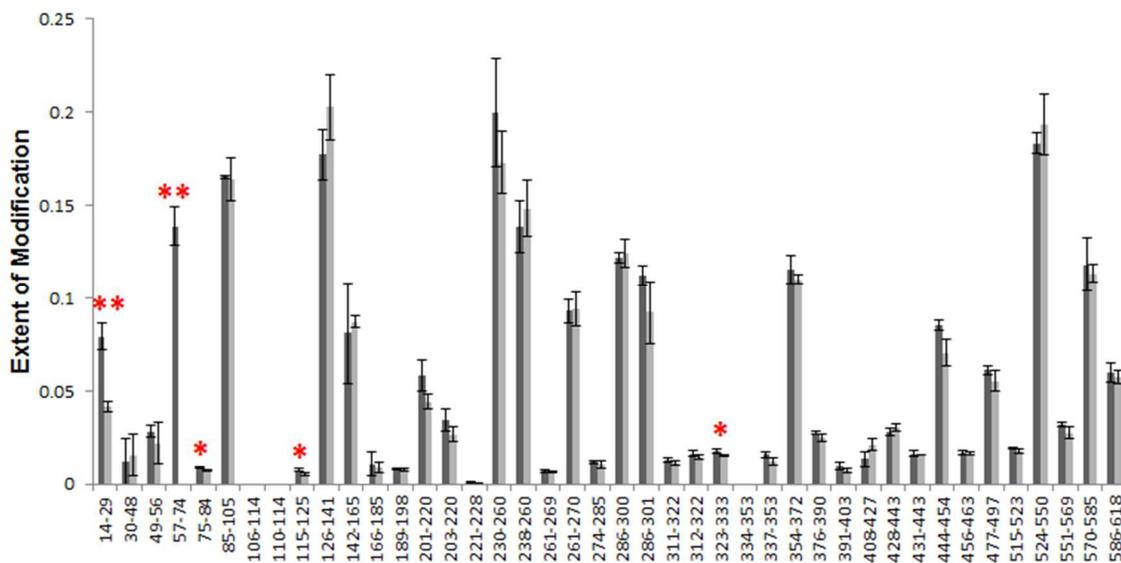
Figure 7- 2. Extent of modification of EGFR alone (dark bars) and Adnectin 1-bound EGFR (light bars) for tryptic peptides covering 89% of the sequence. The light bar in 57-74 is absent because the modification extent is less than 0.1%, and so are several other peptides.

When comparing the modification extent of the two states of exEGFR, we used the Student's t test on the sets of triplicates of both states for each peptide to determine if the two sets of data are significantly different from each other. The two-sample, two-tailed t test was applied assuming unequal variances, and the resulting P values for each peptide are listed in Table 7-3. We found a decrease in modification extent at the N-terminus for 2 of the 42 tryptic peptides of exEGFR upon Adnectin 1 binding, namely, for peptides 14-29 and 57-74 ("**" in **Figure 7-2**), of which the P values are 0.006 and 0.007, respectively. On the contrary, most of the other peptides have larger P values (Table 7-3), indicating no significant difference in oxidation extent between the

free and bound states of exEGFR. Three other peptides, namely, 75-84, 115-125 and 323-333, also have relatively small P values ($0.03 < P < 0.05$) compared to most of the data ("*" in Figure 7-2), showing a similar trend of increased protection in the bound state. These differences are too small, however, and might be caused by structure fluctuations upon Adnectin binding rather than by interactions at the direct binding site. Noteworthy is that the modification extents for both states presented here are already adjusted by deducting the modification percentage in the control

Table 7- 3. Student's t-tests for all triplicates of both EGFR free and EGFR-Adnectin1 bound states.

peptide	P value	Peptide	P value
14-29	**0.006**	286-300	0.641
30-48	0.822	286-301	0.157
49-56	0.421	311-322	0.143
57-74	**0.007**	312-322	0.177
75-84	0.044	323-333	0.048
85-105	0.449	334-353	NA
106-114	NA	337-353	0.069
110-114	NA	354-372	0.352
115-125	0.032	376-390	0.517
126-141	0.123	391-403	0.122
142-165	0.807	408-427	0.063
166-185	0.617	428-443	0.186
189-198	0.713	431-443	0.686
201-220	0.087	444-454	0.052
203-220	0.147	456-463	0.403
221-228	0.598	477-497	0.231
230-260	0.253	515-523	0.104
238-260	0.453	524-550	0.541
261-269	0.446	551-569	0.111
261-270	0.902	570-585	0.635
274-285	0.425	586-618	0.570

samples from them, in an effort to exclude oxidation introduced during sample handling. The reduced reactivity at those two regions indicates reduced solvent accessibility upon binding, suggesting that they are at or near the location of the binding site with Adnectin 1, assuming no allosteric effects are occurring.

To obtain residue-specific information of the binding site, we performed residue-level analysis on the two regions that showed different modification extents for the two states. For example, the extracted ion chromatograms (EIC) and the corresponding mass spectra of peptide 14-29 (unmodified and +16 modified) are shown in Figure 7-3a and Figure 7-3b, respectively,

indicating a mass shift of 15.9936 Da after modification. To assign residue-specific information to peptides eluting at each peak in the EIC, a single MS/MS scan of the +16 species at one elution time is used. An example product-ion spectrum of peptide 14-29 with a +16 modification from a single sampling (peak 1 in Figure 7-3a) is shown in Figure 7-4a. Two product ions at *m/z* 725.5 and 790.0, assigned to y_{12} and y_{13} +16, respectively, indicated the +16 modification is on L17. This assignment is also supported by other fragment ions, such as b_3 and b_4 +16. Similarly, Figure 7-4b shows an isomeric peptide with a +16 modification eluting at a different time (peak 2 in Figure 7-3a) and demonstrating that another oxidative modification occurs at a different location, here shown to be F20. In some cases, the product-ion spectrum was insufficiently informative for the assignment of the modification to a single residue. In that case, the modification was assigned to the narrowest region determined by fragment ions (e.g., region L17-H23). Worth mentioning is that positional isomers of the same residue modification are also separated in the EIC, and such peaks corresponding to the same site modification are added together for calculation.
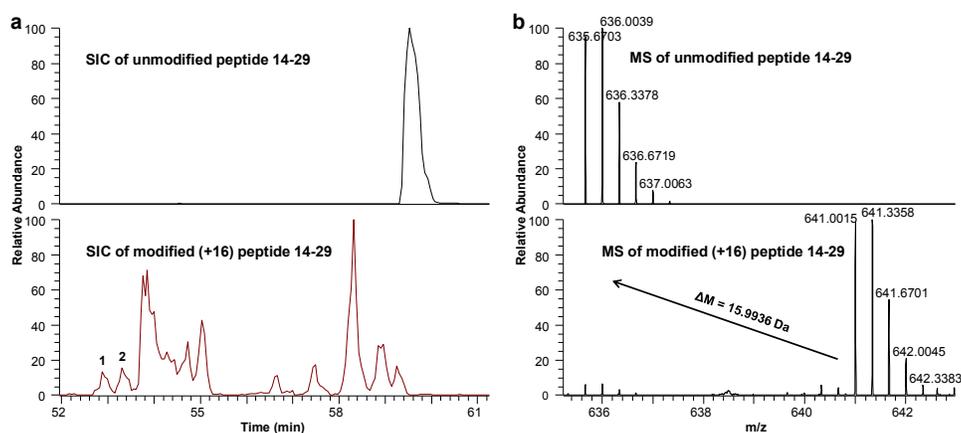
Figure 7- 3. (a) Extracted ion chromatograms of peptide 14-29 unmodified (top) and +16 modified (bottom). (b) Mass spectra of peptide 14-29 normalized to 100% unmodified (top) and modified (bottom).
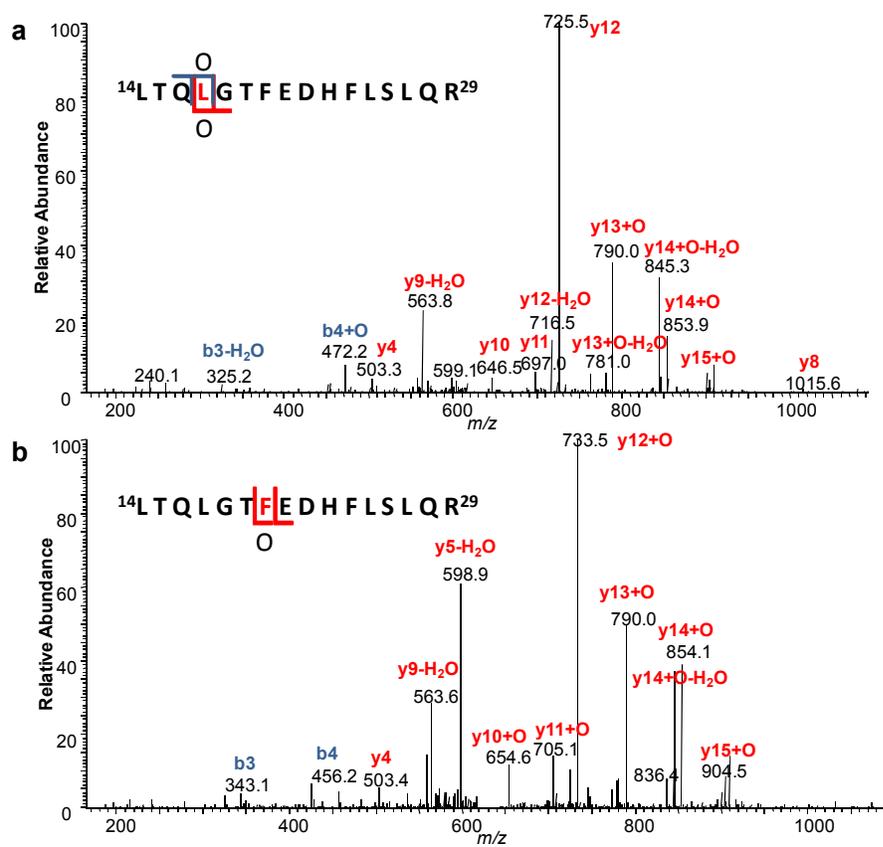
Figure 7- 4. Product-ion (MS/MS) spectrum of peptide 14-29 with +16 modification on (a) L17, colored in red and (b) F20, colored in red.

A total of seven specific residues or short regions were identified to bear the oxidations in the regions represented by peptides 14-29 and 57-74; all but one showed protection upon Adnectin 1 binding (**Figure 7-5**). A similar Student's t-test also performed on the seven residues evaluated the significance of the observed differences between the two states (Table 7-4). For peptide 14-29, residues L14, L17, F20, F24 were modified and showed different modification extents between the bound and unbound states ($P \leq 0.02$), whereas another modification assigned to region L17-H23 showed no change between the two states ($P = 0.2$). In addition to the N-terminal residues, L69 from peptide 57-74, assigned with MS/MS, was also found to undergo decreased solvent accessibility and labeling extent upon Adnectin 1 binding ($P = 0.03$). There is

another +16 modification of peptide 57-74, but we were unable to determine its location from the

MS/MS information.
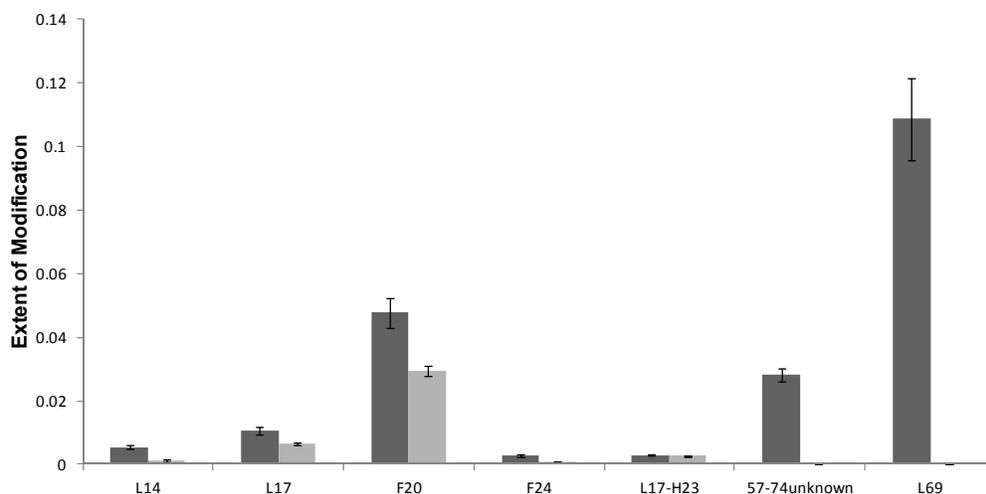
Figure 7- 5. Extent of modification of EGFR alone (dark bars) compared to Adnectin 1-bound EGFR (light bars) at the residue level for regions that showed labeling differences between the two states at peptide level.

Table 7- 4. Student's t-tests for all seven residues or short regions labeled in peptide 14-29 and 57-74 of both EGFR free and EGFR-Adnectin1 bound states.

residue	L14	L17	F20	F24	L17-H23	57-74unknown	L69
p value	0.003	0.017	0.014	0.021	0.204	0.03	0.032

7.4.3 Epitope detected by FPOP labeling vs. by other methods

To compare the epitope of EGFR bound to Adnectin 1 as determined by FPOP labeling and crystallography, we mapped the five residues identified to undergo decreased labeling upon Adnectin 1 binding on the crystal structure of exEGFR-Adnectin 1 (PDB: 3QWQ, Figure 7-6). Four of the five residues, L14, L17, F20 and L69, lie at the interface of the two species when mapped onto the exEGFR-Adnectin 1 crystal structure (Figure 7-6). Specifically, the differential modification of L17 from FPOP labeling is consistent with the crystallography study[22], which

shows that L17 interacts with K79 of Adnectin 1 via H-bonding. The crystallography also pinpoints several neighboring residues (T15, Q16 and G18) that interact with Adnectin 1 via H-bonding, and others (L14, L69, S99 and Y101) that interact via van der Waals forces. Information on T15, Q16 and G18 was not obtained, however, by FPOP labeling owing to a lack of reactivity of these residues. Although FPOP resolved interactions involving L14 and L69, it was silent concerning S99 and Y101 because these residues showed little reactivity; evidence for oxidation on those two residues was unclear because of interference from the highly oxidized M87 (located in an exposed loop) on the same peptide in the LC chromatogram. F20, although not thought to interact with ligands of EGFR, is at the buried surface of Adnectin 1 binding determined by crystallography, and, therefore, shows decreased solvent accessibility in the complex. The last residue, F24, changing its solvent accessibility as identified by FPOP, however, is distant to the putative EGFR-Adnectin 1 interface according to the crystal structure. The decrease in labeling extent on F24 may be due to a change of side-chain orientation that leads to decreased solvent accessibility or a slight change in solvent accessibly (from small to smaller) allosterically controlled by Adnectin binding.

When comparing the epitope mapped by HDX[41] and by FPOP labeling, we also found good consistency. HDX MS
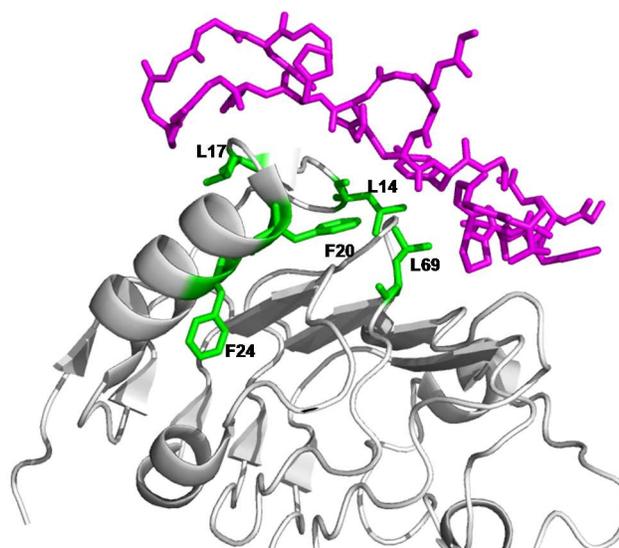

Figure 7- 6. Structural model of EGFR-Adnectin 1 (PDB file 3QWQ) with the interacting residues from Adnectin 1 colored in magenta, and domain I of exEGFR, in other colors. Residues from regions 14-29 and 57-74 that showed different modification extents are in green.

coupled with ETD achieved epitope mapping at residue level and the most important interaction site was assigned as K13 to Q16. Noteworthy is that, unlike HDX, FPOP labeling is a measure of solvent accessibility of a side chain whereas HDX responds to changes in H-bonding of the protein backbone. Therefore, the two methods provide orthogonal information in epitope mapping, and both can complement other epitope mapping methods, such as X-ray crystallography, but reveal more relevant solution binding.

7.5 Conclusions

As a relatively new protein footprinting method, FPOP is less studied than HDX in applications involving mapping epitopes of protein antigens. In this application, we make two-state (bound and unbound) comparisons of labeling extents at identical sites. The comparisons reflect changes in solvent accessibility accompanying intermolecular interactions or allosteric changes[48]. The irreversible nature of FPOP labeling allows for complicated sample handling before MS detection, including deglycosylation, digestion, and long chromatography. This attribute is particularly important when dealing with complex samples and achieving good coverage and detailed CID analysis for residue-specific information. Here, we demonstrated the successful application of FPOP in mapping the epitope of EGFR-Adnectin 1 complex, showing largely consistent results with that from X-ray crystallography and HDX studies but revealing an additional interaction site. Five residues or short regions show decreased modification extent upon Adnectin 1 binding; of these, four lie in the binding interface determined by the crystal structure. The other (F24) is possibly caused by a change in side-chain orientation or a decrease in solvent accessibility of Adnectin 1 when bound. One advantage of MS-based protein footprinting is that it can sample physiologic mixtures. Therefore, FPOP and/or HDX can be

used as stand-alone methods or in conjunction with crystallography to elucidate confidently specific binding sites of antigens in a physiological-relevant solution state.

7.6 Acknowledgements

7.7 References

1. Xu, L., Aha, P., Gu, K., Kuimelis, R. G., Kurz, M., Lam, T., Lim, A. C., Liu, H., Lohse, P. A., Sun, L., Weng, S., Wagner, R. W., and Lipovsek, D. (2002) Directed evolution of high-affinity antibody mimics using mRNA display, *Chemistry & biology 9*, 933-942.

2. Hackel, B. J., Kapila, A., and Wittrup, K. D. (2008) Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling, *Journal of molecular biology 381*, 1238-1252.

3. Wojcik, J., Hantschel, O., Grebien, F., Kaupe, I., Bennett, K. L., Barkinge, J., Jones, R. B., Koide, A., Superti-Furga, G., and Koide, S. (2010) A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain, *Nature structural & molecular biology 17*, 519-527.

4. Koide, A., Bailey, C. W., Huang, X., and Koide, S. (1998) The fibronectin type III domain as a scaffold for novel binding proteins, *Journal of molecular biology 284*, 1141-1151.

5. Koide, A., Abbatiello, S., Rothgery, L., and Koide, S. (2002) Probing protein conformational changes in living cells by using designer binding proteins: application to the estrogen receptor, *Proceedings of the National Academy of Sciences of the United States of America 99*, 1253-1258.

6. Karatan, E., Merguerian, M., Han, Z., Scholle, M. D., Koide, S., and Kay, B. K. (2004) Molecular recognition properties of FN3 monobodies that bind the Src SH3 domain, *Chemistry & biology 11*, 835-844.

7.   Main, A. L., Harvey, T. S., Baron, M., Boyd, J., and Campbell, I. D. (1992) The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions, *Cell 71*, 671-678.

8.   Dickinson, C. D., Veerapandian, B., Dai, X. P., Hamlin, R. C., Xuong, N. H., Ruoslahti, E., and Ely, K. R. (1994) Crystal structure of the tenth type III cell adhesion module of human fibronectin, *Journal of molecular biology 236*, 1079-1092.

9.   Leahy, D. J., Erickson, H. P., Aukhil, I., Joshi, P., and Hendrickson, W. A. (1994) Crystallization of a fragment of human fibronectin: introduction of methionine by site-directed mutagenesis to allow phasing via selenomethionine, *Proteins 19*, 48-54.

10.   Leahy, D. J., Aukhil, I., and Erickson, H. P. (1996) 2.0 angstrom crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region, *Cell 84*, 155-164.

11.   Mamluk, R., Carvajal, I. M., Morse, B. A., Wong, H., Abramowitz, J., Aslanian, S., Lim, A. C., Gokemeijer, J., Storek, M. J., Lee, J., Gosselin, M., Wright, M. C., Camphausen, R. T., Wang, J., Chen, Y., Miller, K., Sanders, K., Short, S., Sperinde, J., Prasad, G., Williams, S., Kerbel, R., Ebos, J., Mutsaers, A., Mendlein, J. D., Harris, A. S., and Furfine, E. S. (2010) Anti-tumor effect of CT-322 as an adnectin inhibitor of vascular endothelial growth factor receptor-2, *mAbs 2*, 199-208.

12.   Bloom, L., and Calabro, V. (2009) FN3: a new protein scaffold reaches the clinic, *Drug discovery today 14*, 949-955.

13.   Molckovsky, A., and Siu, L. L. (2008) First-in-class, first-in-human phase I results of targeted agents: highlights of the 2008 American society of clinical oncology meeting, *Journal of hematology & oncology 1*, 20.

14.   Tolcher, A. W., Sweeney, C. J., Papadopoulos, K., Patnaik, A., Chiorean, E. G., Mita, A. C., Sankhala, K., Furfine, E., Gokemeijer, J., Iacono, L., Eaton, C., Silver, B. A., and Mita, M. (2011) Phase I and pharmacokinetic study of CT-322 (BMS-844203), a targeted Adnectin inhibitor of VEGFR-2 based on a domain of human fibronectin, *Clinical cancer research : an official journal of the American Association for Cancer Research 17*, 363-371.

15.   Janmaat, M. L., and Giaccone, G. (2003) The epidermal growth factor receptor pathway and its inhibition as anticancer therapy, *Drugs of today 39 Suppl C*, 61-80.

16.   Yarden, Y., and Sliwkowski, M. X. (2001) Untangling the ErbB signalling network, *Nature reviews. Molecular cell biology 2*, 127-137.

17.   Zhang, H., Shen, W., Rempel, D., Monsey, J., Vidavsky, I., Gross, M. L., and Bose, R. (2011) Carboxyl-group footprinting maps the dimerization interface and phosphorylation-induced conformational changes of a membrane-associated tyrosine kinase, *Mol Cell Proteomics 10*, M110 005678.

18.   Schlessinger, J. (2000) Cell signaling by receptor tyrosine kinases, *Cell 103*, 211-225.

19.   Masui, H., Kawamoto, T., Sato, J. D., Wolf, B., Sato, G., and Mendelsohn, J. (1984) Growth inhibition of human tumor cells in athymic mice by anti-epidermal growth factor receptor monoclonal antibodies, *Cancer research 44*, 1002-1007.

20.   Sato, J. D., Kawamoto, T., Le, A. D., Mendelsohn, J., Polikoff, J., and Sato, G. H. (1983) Biological effects in vitro of monoclonal antibodies to human epidermal growth factor receptors, *Molecular biology & medicine 1*, 511-529.

21.   Flynn, J. F., Wong, C., and Wu, J. M. (2009) Anti-EGFR therapy: mechanism and advances in clinical efficacy in breast cancer., *J. Oncol. 10.1155/2009/526963*.

22. Ramamurthy, V., Krystek, S. R., Jr., Bush, A., Wei, A., Emanuel, S. L., Das Gupta, R., Janjua, A., Cheng, L., Murdock, M., Abramczyk, B., Cohen, D., Lin, Z., Morin, P., Davis, J. H., Dabritz, M., McLaughlin, D. C., Russo, K. A., Chao, G., Wright, M. C., Jenny, V. A., Engle, L. J., Furfine, E., and Sheriff, S. (2012) Structures of adnectin/protein complexes reveal an expanded binding footprint, *Structure 20*, 259-269.

23. Emanuel, S. L., Engle, L. J., Chao, G., Zhu, R. R., Cao, C., Lin, Z., Yamniuk, A. P., Hosbach, J., Brown, J., Fitzpatrick, E., Gokemeijer, J., Morin, P., Morse, B. A., Carvajal, I. M., Fabrizio, D., Wright, M. C., Das Gupta, R., Gosselin, M., Cataldo, D., Ryseck, R. P., Doyle, M. L., Wong, T. W., Camphausen, R. T., Cload, S. T., Marsh, H. N., Gottardis, M. M., and Furfine, E. S. (2011) A fibronectin scaffold approach to bispecific inhibitors of epidermal growth factor receptor and insulin-like growth factor-I receptor, *mAbs 3*, 38-48.

24. Carugo, O., and Argos, P. (1997) Protein-protein crystal-packing contacts, *Protein science : a publication of the Protein Society 6*, 2261-2263.

25. Xu, Q., Canutescu, A. A., Wang, G., Shapovalov, M., Obradovic, Z., and Dunbrack, R. L., Jr. (2008) Statistical analysis of interface similarity in crystals of homologous proteins, *Journal of molecular biology 381*, 487-507.

26. Benjamin, D. C., Berzofsky, J. A., East, I. J., Gurd, F. R., Hannum, C., Leach, S. J., Margoliash, E., Michael, J. G., Miller, A., Prager, E. M., and et al. (1984) The antigenic structure of proteins: a reappraisal, *Annual review of immunology 2*, 67-101.

27. Scott, J. K., and Smith, G. P. (1990) Searching for peptide ligands with an epitope library, *Science 249*, 386-390.

28.   Geysen, H. M., Meloen, R. H., and Barteling, S. J. (1984) Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid, *Proceedings of the National Academy of Sciences of the United States of America 81*, 3998-4002.

29.   Jemmerson, R., and Paterson, Y. (1986) Mapping epitopes on a protein antigen by the proteolysis of antigen-antibody complexes, *Science 232*, 1001-1004.

30.   Yamada, N., Suzuki, E., and Hirayama, K. (2002) Identification of the interface of a large protein-protein complex using H/D exchange and Fourier transform ion cyclotron resonance mass spectrometry, *Rapid communications in mass spectrometry : RCM 16*, 293-299.

31.   Baerga-Ortiz, A., Hughes, C. A., Mandell, J. G., and Komives, E. A. (2002) Epitope mapping of a monoclonal antibody against human thrombin by H/D-exchange mass spectrometry reveals selection of a diverse sequence in a highly conserved protein, *Protein science : a publication of the Protein Society 11*, 1300-1308.

32.   Lu, J., Witcher, D. R., White, M. A., Wang, X., Huang, L., Rathnachalam, R., Beals, J. M., and Kuhstoss, S. (2005) IL-1beta epitope mapping using site-directed mutagenesis and hydrogen-deuterium exchange mass spectrometry analysis, *Biochemistry 44*, 11106-11114.

33.   Coales, S. J., Tuske, S. J., Tomasso, J. C., and Hamuro, Y. (2009) Epitope mapping by amide hydrogen/deuterium exchange coupled with immobilization of antibody, on-line proteolysis, liquid chromatography and mass spectrometry, *Rapid communications in mass spectrometry : RCM 23*, 639-647.

34.   Zhang, Q., Willison, L. N., Tripathi, P., Sathe, S. K., Roux, K. H., Emmett, M. R., Blakney, G. T., Zhang, H. M., and Marshall, A. G. (2011) Epitope mapping of a 95 kDa

antigen in complex with antibody by solution-phase amide backbone hydrogen/deuterium exchange monitored by Fourier transform ion cyclotron resonance mass spectrometry, *Analytical chemistry 83*, 7129-7136.

35.   Pandit, D., Tuske, S. J., Coales, S. J., E, S. Y., Liu, A., Lee, J. E., Morrow, J. A., Nemeth, J. F., and Hamuro, Y. (2012) Mapping of discontinuous conformational epitopes by amide hydrogen/deuterium exchange mass spectrometry and computational docking, *Journal of molecular recognition : JMR 25*, 114-124.

36.   Jensen, P. F., Jorgensen, T. J., Koefoed, K., Nygaard, F., and Sen, J. W. (2013) Affinity capture of biotinylated proteins at acidic conditions to facilitate hydrogen/deuterium exchange mass spectrometry analysis of multimeric protein complexes, *Analytical chemistry 85*, 7052-7059.

37.   Takamoto, K., and Chance, M. R. (2006) Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes, *Annual review of biophysics and biomolecular structure 35*, 251-276.

38.   Pan, Y., Brown, L., and Konermann, L. (2009) Mapping the structure of an integral membrane protein under semi-denaturing conditions by laser-induced oxidative labeling and mass spectrometry, *Journal of molecular biology 394*, 968-981.

39.   Pan, Y., Stocks, B. B., Brown, L., and Konermann, L. (2009) Structural characterization of an integral membrane protein in its natural lipid environment by oxidative methionine labeling and mass spectrometry, *Analytical chemistry 81*, 28-35.

40.   Jones, L. M., J, B. S., J, A. C., and Gross, M. L. (2011) Fast photochemical oxidation of proteins for epitope mapping, *Analytical chemistry 83*, 7657-7661.

41.   Iacob, R. E., Chen, G., Ahn, J., Houel, S., Wei, H., Mo, J., Tao, L., Cohen, D., Xie, D., Lin, Z., Morin, P. E., Doyle, M. L., Tymiak, A. A., and Engen, J. R. (2014) The influence of adnectin binding on the extracellular domain of epidermal growth factor receptor, *Journal of the American Society for Mass Spectrometry 25*, 2093-2102.

42.   Maleknia, S. D., Brenowitz, M., and Chance, M. R. (1999) Millisecond radiolytic modification of peptides by synchrotron X-rays identified by mass spectrometry, *Analytical chemistry 71*, 3965-3973.

43.   Sharp, J. S., Becker, J. M., and Hettich, R. L. (2004) Analysis of protein solvent accessible surfaces by photochemical oxidation and mass spectrometry, *Analytical chemistry 76*, 672-683.

44.   Aye, T. T., Low, T. Y., and Sze, S. K. (2005) Nanosecond laser-induced photochemical oxidation method for protein surface mapping with mass spectrometry, *Analytical chemistry 77*, 5814-5822.

45.   Hambly, D. M., and Gross, M. L. (2005) Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale, *Journal of the American Society for Mass Spectrometry 16*, 2057-2063.

46.   Gau, B. C., Sharp, J. S., Rempel, D. L., and Gross, M. L. (2009) Fast photochemical oxidation of protein footprints faster than protein unfolding, *Analytical chemistry 81*, 6563-6571.

47.   Hambly, D., and Gross, M. (2007) Laser flash photochemical oxidation to locate heme binding and conformational changes in myroglobin, *Int J Mass Spectrom 259*, 124-129.

48.   Gau, B. C., Chen, H., Zhang, Y., and Gross, M. L. (2010) Sulfate radical anion as a new reagent for fast photochemical oxidation of proteins, *Analytical chemistry 82*, 7821-7827.

49. Xu, G., and Chance, M. R. (2004) Radiolytic modification of acidic amino acid residues in peptides: probes for examining protein-protein interactions, *Analytical chemistry 76*, 1213-1221.

50. Xu, G., and Chance, M. R. (2007) Hydroxyl radical-mediated modification of proteins as probes for structural proteomics, *Chemical reviews 107*, 3514-3543.

51. Zhu, W., Smith, J. W., and Huang, C. M. (2010) Mass spectrometry-based label-free quantitative proteomics, *Journal of biomedicine & biotechnology 2010*, 840518.

52. Zhang, H., Gau, B. C., Jones, L. M., Vidavsky, I., and Gross, M. L. (2011) Fast photochemical oxidation of proteins for comparing structures of protein-ligand complexes: the calmodulin-peptide model system, *Analytical chemistry 83*, 311-318.

53. Edited by Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., and Etzler, M. E. (2009) Essentials of Glycobiology (2nd edition), *Cold Spring Harbor Laboratories Press (NY)*.

Yuetian Yan's Resume

665 S Skinker Blvd., APT 15H   314-603-7681 (Cell)
Saint Louis, MO 63105   karyyan@gmail.com

Education:
Ph.D.: Washington University in St. Louis, United States
Analytical Chemistry (Mass Spectrometry)   Cumulative GPA: 3.92/4.00   January 2015
BS: Nanjing University, China
Department for Intensive Instruction (major: chemistry)   Cumulative GPA: 90.1 /100   May 2008

Industry Experience:
Bristol-Myers Squibb
Biologics Development and Operations/PCO BDAS   Princeton, New Jersey
Research Co-op, Groups of Li Tao and Guodong Chen   July 2014 – December 2014
● Utilized HDX MS to study the local and global conformational impact of non-enzymatic degradation in the CDRs of a therapeutic antibody, including isomerization and oxidation.
● Utilized HDX MS to study the relationship between structure and the degradation pattern of a PEGylated Adnectin.
● Utilized native ESI MS and HDX MS to study an intermediate state of an antibody, which is composed of an extra light chain and a monomer. This study was performed on a Waters Synapt G2 instrument. Native ESI MS was used to confirm the non-covalent interaction between the light chain and the monomer. Separation of Fab and Fc followed by Native ESI MS indicated that the extra light chain binds to the Fab. HDX MS then reported on the interaction site of Fab and the light chain, with support from structural modeling.

Reliable Biopharmaceutical Corporation   Saint Louis, Missouri
Collaboration service   December 2011 – May 2014
● Provided MS analysis in carbohydrates drug and small molecule drug development, and generated scientific reports and communicate with customers.

Academic Experience:
Washington University in St. Louis   Saint Louis, Missouri
Research Assistant, Laboratory of Michael L. Gross   June 2010 – Present
● Studied the dimerization of SecA (a 102 kDa protein) by HDX MS. SecA is motor protein in bacteria preprotein translocation, and both its dimeric interface and active state remain controversial. We controlled the oligomeric state of SecA by varying salt concentrations, and used HDX MS to characterize the dimer interface. The determined interface is consistent with Alanine mutation/Analytical Ultracentrifugation results, and majorly agrees with one of the five controversial crystal structures. A conformational change from open to closed forms was also identified at the pre-protein binding domain upon dimerization. This study also provided leads for the active form of SecA**.**
● Mapped the epitope of Adnectin1-EGFR interaction at the amino-acid residue level by fast photochemical oxidation of proteins (FPOP) Mass Spectrometry, in collaboration with BMS. The epitope identified agrees with that from both HDX study and crystallography results, presenting more evidence of the capability of FPOP in epitope mapping.

166

● Built a tool box with LC/MS and GC/MS for structural characterization of steroid metabolites that are active pheromones in mice.

● Implemented charge-remote fragmentation using high-energy CID on a MALDI TOF/TOF instrument to characterize the structures of steroid metabolites.

● Developed a method involving MS to characterized the structure of an unknown steroid metabolite that activates mouse pheromone-sensing neurons from mouse urine. The proposed structure was confirmed by customer-synthesis.

● Performed MS/LC-MS analysis of small molecules, polymers, proteins and peptides for internal and external customers, as a requirement on members of an NIH Research Resource.

Technical Experience:

● **Mass spectrometers:** Thermo LTQ-Orbitrap/FT; Waters Synapt G2/G2-Si; Brukers Maxis Q-TOF; and Agilent 7200 Q-TOF GC/MS, etc.)

● **LC systems:** Waters ACQUITY/ nanoACQUITY systems, Agilent 1100/1200 systems, Eksigent 1D/1D+ systems, and Dionex Ultimate 3000

● **Waters HDX system**.

● **Separation/biochemical techniques:** Reverse-phase chromatography; 1D-Electrophoresis; Solid phase extraction; Buffer exchange; Protein enzyme digestion, etc.

● **Bioinformatics tools:** Xcalibur, MassLynX, BioPharmaLynx, ProteinLynx Global Server, DynamX, HDX Workbench, HDX Examiner, MASCOT, ProtMapMS, etc..

Publications:

● <u>Yan Y.</u>, Wowor A.J., Auclari S.M., Yu D., Zhang J., May E.R., Gross M.L., Kendall D.A., Cole J.L., *Analysis of SecA Dimerization in Solution, Biochemistry, 53(19), 3248-3260 **(2014)***

● <u>Yan Y.</u>, Chen G., Wei H., Huang R.Y., Mo J., Rempel D.L., Tymiak A.A., and Gross M.L., *Fast Photochemical Oxidation of Proteins (FPOP) Maps the Epitope of EGFR Binding to Adnectin, JASMS, 25(12), 2084-2092 **(2014)***

● <u>Yan Y.</u>, Rempel, D.L., Holy T.E., Gross M.L., *Mass Spectrometry Combinations for Structural Characterization of Sulfated-Steroid Metabolites, JASMS, 25(5), 869-879 **(2014)***

● <u>Yan Y.</u>, Ubukata M., Cody R.B., Holy T.E., and Gross M.L., *High-energy Collision-induced-dissociation Causes Charge-remote-fragmentation of Steroid Sulfates on MALDI TOF/TOF, JASMS, 25(8), 1404-1411 **(2014)***

● Fu X., <u>Yan Y.</u>, Xu S.P., Geerlof-vidavsky I., Chong W., Gross M.L. and Holy T.E., *A Natural Ligand for Mouse Vomeronasal Sensory Neurons Revealed by Systematic Correlation Analyses.* **Reviewed by Cell.**

● Yefremova Y., Al-Majdoub M., Opuni K. F., Koy C., Cui W., <u>Yan Y.</u>, Gross M. L., and Glocker M. O., *"De-novo" amino acid sequence elucidation of protein G'e by combined "Top-Down" and "Bottom-Up" mass spectrometry*, JASMS, DOI: 10.1007/s13361-014-1053-2

Presentations:

● <u>Yan Y.</u>, Chen G.., Wei H., Huang R., Mo J., Rempel D.L., Tymiak A.A., Gross M.L., *Epitope Mapping of EGFR Binding to an Adnectin by Fast Photochemical Oxidation of Proteins,* **ASMS 2014 Annual Conference**

● <u>Yan Y.</u>, Chen G.., Wei H., Huang R., Mo J., Rempel D.L., Tymiak A.A., Gross M.L., *Epitope Mapping of EGFR Binding to an Adnectin by Fast Photochemical Oxidation of Proteins,* ***3rd** International Symposium on Higher order structure of Protein Therapeutics (2014).* **Oral presentation.**
● <u>Yan Y.</u>, Wowor A.J., Zhang J., Cole J.L., Kendall D.A., Gross M.L., *H/DX MS Reveals Features of SecA Dimeric Interface and a Conformational Change from Open to Closed Forms upon Dimerization,* ***ASMS 2013 Annual Conference***
● <u>Yan Y.</u>, Fu X., Holy T.E., Gross M.L., *Mass Spectrometric Structural characterization of Sulfated Steroids: Implications in Animal Communication,* ***ASMS 2012 Annual Conference***
● <u>Yan Y.</u>, Holy T.E., Gross M.L., *Method Development for Structural Characterization of Sulfated Steroids with Mass Spectrometry: Applications in Animal Communication,* ***ACS 2011 Regional Meeting***

Honors and Awards:
● Dean Dissertation Fellowship, Washington University, United States   2014
● Graduate Research Fellowships, Washington University, United States   2011 – 2013
● Graduate Teaching Fellowships, Washington University, United States   2009 – 2011
● Endeavour Australia Cheung Kong Scholarship, Melbourne University, Australia   2007
● National People's Scholarship, 2nd class (top 5%), Nanjing University, China   2005, 2007
● National People's Scholarship, 3rd class (top 10%), Nanjing University, China   2006

Professional Reference:
Michael L. Gross, Ph.D.
Professor of Chemistry, Washington University in St. Louis
Department of Chemistry, Campus Box 1134, Washington University in St. Louis
One Brookings Drive, Saint Louis, MO 63130
314-935-4814 (Office)
mgross@wustl.edu

Timothy E. Holy, Ph. D.
Associate Professor of Anatomy and Neurobiology, Washington University School of Medicine
Department of Anatomy & Neurobiology, Campus Box 8108, Washington University in St. Louis
660 S. Euclid Avenue, Saint Louis, MO 63110-1093
314-362-0086 (Office)
holy@wustl.edu

Li Tao, Ph.D.
Sr. Principal Scientist
Global Manufacturing & Supply / Biologics Development & Operations
Bristol-Myers Squibb
Building 9, Room 123
311 Pennington Rockey Hill Rd, Pennington, NJ, 08534
609-818-5527 (Office)
Li.Tao@bms.com