Development and Application of Mass Spectrometry-Based Biophysical Approaches

Ying Zhang
Washington University in St. Louis

Follow this and additional works at: http://openscholarship.wustl.edu/art_sci_etds
Part of the Chemistry Commons

Recommended Citation
Development and Application of Mass Spectrometry-Based Biophysical Approaches

by

Ying Zhang

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 .......................................................................................................................... vi
List of Tables .............................................................................................................................. viii
Acknowledgments ................................................................................................................... ix

ABSTRACT OF THE DISSERTATION .................................................................................. x

Chapter 1: Introduction ........................................................................................................... 1

1.1 Hydrogen Deuterium Exchange (HDX) ........................................................................ 2

1.1.1 Types of Hydrogens in Proteins ............................................................................. 3

1.1.2 pH and Temperature Effect .................................................................................... 3

1.1.3 HDX-MS ................................................................................................................. 4

1.1.4 EX1 and EX2 ........................................................................................................... 6

1.2 Protein-Ligand Interactions .......................................................................................... 6

1.3 Protein-Ligand Affinity Measurements ......................................................................... 8

1.4 Conventional Methods for Ligand Binding Characterization ......................................... 10

1.4.1 Isothermal Titration Calorimetry (ITC) ................................................................. 10

1.4.2 Fluorescence ........................................................................................................... 11

1.4.3 Fourier Transform Infrared spectroscopy (FTIR) ................................................... 11

1.4.4 Surface Plasmon Resonance (SPR) .................................................................... 12

1.4.5 Enzyme-Linked Immunosorbent Assay (ELISA) .................................................... 12

1.4.6 Circular Dichroism (CD) ....................................................................................... 13

1.4.7 Nuclear Magnetic Resonance (NMR) .................................................................. 13

1.4.8 X-ray Crystallography ......................................................................................... 14

1.5 Direct Mass Spectrometry Method ................................................................................ 14

1.6 HDX-MS .................................................................................................................... 15

1.6.1 HDX-MS for Binding Regions .............................................................................. 15

1.6.2 HDX-MS for Binding Affinity ................................................................................ 16

1.6.3 HDX-MS for Binding Order .................................................................................. 34

1.6.3 HDX-MS for Protein-Protein Interactions ............................................................ 35

1.7 Conclusion .................................................................................................................... 39

1.8 Acknowledgements ....................................................................................................... 39
### Chapter 4: Differential Hydrogen Deuterium Exchange Mass Spectrometry Studies of Protein Arginine Deiminase 2 (PAD2)

- **Abstract** .......................................................... 105
- **Introduction** ...................................................... 105
- **Methods and Materials** ....................................... 109
  - 4.3.1 Chemicals and Proteins .................................. 109
  - 4.3.2 HDX-MS ...................................................... 109
  - 4.3.3 Peptide Identification and Data Processing ........... 110
- **Results and Discussion** ....................................... 111
  - 4.4.1 HDX Studies of Apo and Holo Wild-Type (WT) PAD2 at 4°C ........................................ 111
  - 4.4.2 HDX Studies of Apo and Holo WT PAD2 at 25°C ..................................................... 134
  - 4.4.3 Comparing HDX Studies (4 °C vs. 25 °C) of Apo WT PAD2 ........................................ 150
  - 4.4.4 HDX Studies of Mutant PAD2 ............................. 163
- **Conclusion** ....................................................... 180
- **Acknowledgements** ............................................. 181
- **References** ....................................................... 182

### Chapter 5: PLIMSTEX studies of calcium binding of protein arginine deiminase 2 (PAD2)

- **Abstract** .......................................................... 185
- **Introduction** ...................................................... 185
- **Materials and Methods** ....................................... 187
  - 5.3.1 Expression of WT-PAD2; Peptide Level HDX; Peptide Identification and HDX Data Processing .................................................. 187
  - 5.3.2 PLIMSTEX ...................................................... 187
  - 5.3.3 PLIMSTEX Curve Fitting ................................. 188
- **Results and Discussion** ....................................... 188
- **Conclusion** ....................................................... 198
- **Acknowledgements** ............................................. 199
- **References** ....................................................... 200
# List of Figures

## Chapter 1

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>A typical HDX-MS experiment workflow</td>
<td>5</td>
</tr>
<tr>
<td>1.2</td>
<td>Outcome of HDX under EX1 (A) and EX2 (B) regimes.</td>
<td>7</td>
</tr>
<tr>
<td>1.3</td>
<td>A typical workflow for PLIMSTEX</td>
<td>17</td>
</tr>
<tr>
<td>1.4</td>
<td>A typical PLIMSTEX curve</td>
<td>19</td>
</tr>
<tr>
<td>1.5</td>
<td>Examples of ‘sharp break’ PLIMSTEX curves</td>
<td>23</td>
</tr>
<tr>
<td>1.6</td>
<td>On- and off-fluxes among various species in PLIMSTEX experiments</td>
<td>24</td>
</tr>
<tr>
<td>1.7</td>
<td>A typical work flow for the dPLIMSTEX</td>
<td>26</td>
</tr>
<tr>
<td>1.8</td>
<td>A typical workflow for SUPREX</td>
<td>28</td>
</tr>
<tr>
<td>1.9</td>
<td>A typical workflow for SUPREX with digestion</td>
<td>31</td>
</tr>
<tr>
<td>1.10</td>
<td>HDX patterns for various Ca(^{2+}) binding sites</td>
<td>36</td>
</tr>
</tbody>
</table>

## Chapter 2

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Western blotting for both A(\beta_{40}) and A(\beta_{42}).</td>
<td>54</td>
</tr>
<tr>
<td>2.2</td>
<td>Pulsed HDX results for three peptides from A(\beta_{40}).</td>
<td>56</td>
</tr>
<tr>
<td>2.3</td>
<td>Raw data for HDX experiments</td>
<td>61</td>
</tr>
<tr>
<td>2.4</td>
<td>Pulsed HDX and Western blotting results for three peptic peptides from A(\beta_{42}).</td>
<td>62</td>
</tr>
<tr>
<td>2.5</td>
<td>A flow chart for the modified F-W modeling and bootstrap strategy</td>
<td>65</td>
</tr>
<tr>
<td>2.6</td>
<td>A zoom in view for the outcome of application of the bootstrap strategy.</td>
<td>66</td>
</tr>
<tr>
<td>2.7</td>
<td>t(_{1/2}) distribution from bootstrap analysis for different peptide.</td>
<td>67</td>
</tr>
<tr>
<td>2.8</td>
<td>Pulsed HDX results for undigested species from A(\beta_{42}).</td>
<td>70</td>
</tr>
<tr>
<td>2.9</td>
<td>TEM images for effect of quenching solution on A(\beta_{42}) fibrils.</td>
<td>72</td>
</tr>
<tr>
<td>2.10</td>
<td>Pulsed HDX results for three peptides from A(\beta_{42}).</td>
<td>75</td>
</tr>
</tbody>
</table>

## Chapter 3

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Molecular structures of the BFCs L1 and L2 and amentoflavone (AMF)</td>
<td>90</td>
</tr>
<tr>
<td>3.2</td>
<td>Pulsed HDX results for three peptic peptides from A(\beta_{42}) interacting with L1.</td>
<td>92</td>
</tr>
<tr>
<td>3.3</td>
<td>Pulsed HDX results for three peptic peptides from A(\beta_{42}) interacting with L2.</td>
<td>94</td>
</tr>
<tr>
<td>3.4</td>
<td>Pulsed HDX results for three peptic peptides from A(\beta_{42}) interacting with AMF.</td>
<td>97</td>
</tr>
</tbody>
</table>

## Chapter 4

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Function of PAD</td>
<td>106</td>
</tr>
<tr>
<td>4.2</td>
<td>HDX peptide coverage map for WT PAD2</td>
<td>112</td>
</tr>
<tr>
<td>4.3</td>
<td>HDX kinetic curves for WT PAD2 (4 °C)</td>
<td>127</td>
</tr>
<tr>
<td>4.4</td>
<td>Dynamics map for HDX of PAD2 (4 °C).</td>
<td>129</td>
</tr>
<tr>
<td>4.5</td>
<td>Sequence alignment of PAD2 and PAD4.</td>
<td>132</td>
</tr>
<tr>
<td>4.6</td>
<td>HDX kinetic curves for WT PAD2 (25 °C).</td>
<td>148</td>
</tr>
<tr>
<td>4.7</td>
<td>Dynamics map for HDX of PAD2 (25 °C).</td>
<td>149</td>
</tr>
<tr>
<td>4.8</td>
<td>HDX kinetic curves for WT PAD2 apo state comparison between two temperatures.</td>
<td>162</td>
</tr>
<tr>
<td>4.9</td>
<td>HDX kinetic curves for PAD2 C647A mutant</td>
<td>177</td>
</tr>
</tbody>
</table>
Figure 4.10 Dynamics map for HDX of C647A PAD2 (4 °C)................................................................. 178
Chapter 5
Figure 5.1 PLIMSTEX data for regions that do not show deuterium uptake decrease. .......... 190
Figure 5.2 PLIMSTEX data for five Ca\(^{2+}\) binding regions.............................................................. 192
Chapter 6
Figure 6.1 HDX peptide coverage map ............................................................................................. 209
Figure 6.2 Kinetic curves of all the peptides used for HDX.............................................................. 215
Figure 6.3 Selected HDX kinetic curves for demonstration of the precision for each experiment.
......................................................................................................................................................... 216
Figure 6.4 Homology models A to D.................................................................................................. 219
Figure 6.5 Model D with disorder prediction data............................................................................. 225
Figure 6.6 Correlation of HDX and SASA data for all models .......................................................... 227
Figure 6.7 The deuterium uptake differences between the oxidized and reduced states of DHCC are mapped onto Model D........................................................................................................ 228
Figure 6.8 The deuterium uptake levels for all peptides mapped onto Model D......................... 230
Figure 6.9 Proposed function of diheme cytochrome c. .................................................................. 235
List of Tables

Chapter 4
Table 4.1 Calcium binding regions in PAD2 .......................................................... 130

Chapter 5
Table 5.1 Examples of the Ca binding cooperativity .............................................. 194
Table 5.2 All tested binding patterns for Ca binding cooperativity ....................... 195
Table 5.3 Comparison of the data between PLIMSTEX and X-ray crystallography .. 197

Chapter 6
Table 6.1 Homology modeling data ........................................................................ 220
Table 6.2 Comparison of HDX data and Homology models ................................... 223
Acknowledgments

To my advisors, Professors Michael Gross and Liviu Mirica: thank you for the long-lasting support. Thank you for providing useful advice, patient guidance, free research environment and great instruments during these five years. You taught me how to be a good scientist, and more importantly, a good person. I feel very proud to be a member of your groups.

To all lab members: thank you all for your selfless help in my work and life; I especially thank staff scientists and postdocs Drs. Henry Rohrs, Don Rempel, Jun Zhang, Weidong Cui, Anuj Sharma, Daryl Giblin, and Hao Zhang, for help in modeling and instrumentation.

To all my collaborators: thank you for providing me with all the great projects. The collaborators include Dr. Yuanyuan Wang, Prof. William Burho, Dr. Hai Yue, Erica Majumder, Prof. Robert Blankenship, Dr. Kathrin Lindner, Prof. Michael Przybylski, Dr. Daniel Slade, Prof. Paul Thompson, Jiaxi Fang, and Prof. Pratim Biswas. I want to express my special thanks to my collaborators at Genentech, Drs. Aaron Wecksler, Patricia Molina, Galahad Deperalta, and Victor Ling, who provided not only interesting projects for our collaboration but also important help to understand industry better and advice for my career development.

To my friends and family: thank you all for being there for me.

Ying Zhang

Washington University in St. Louis

December 2014
ABSTRACT OF THE DISSERTATION

Development and Application of Mass Spectrometry-Based Biophysical Approaches
by
Ying Zhang
Doctor of Philosophy in Chemistry
Washington University in St. Louis, 2014
Professor Michael Gross, Chair
Professor Liviu Mirica, Co-Chair

Mass spectrometry (MS)-based biophysical approaches are new “tools” for protein characterization owing to its capability to analyze proteins and protein complexes that range in molecular weight from kDa to MDa. Protein characterization requires more than identifying the primary structure. More importantly, protein high order structures (i.e., secondary, tertiary and quaternary structures) are needed for biological studies. MS has become the major tool in studies of protein primary structure and post translational modifications (PTMs) over the past two decades. Because MS has high sensitivity and fast turnaround, more and more biophysical approaches rely on MS to generate information for protein higher order structures.

One of the emerging biophysical approaches is MS-based protein footprinting. Protein surface regions can be covalently labeled by chemical reagents in a biologically relevant environment. These chemical labels can be read out by MS through either bottom-up or top-down MS proteomics analysis. The outcome provides protein conformational information. Among various chemical labeling strategies, hydrogen deuterium exchange (HDX) is one of the most commonly used approaches in MS-based protein biophysical studies.

HDX-MS is introduced in Chapter 1 by covering the early developments and new applications especially in measuring interaction affinities. Although HDX-MS has been developed for decades, there are still many challenges in protein characterization that require new or improved
HDX method development. One such challenge is characterization of protein aggregation. Protein aggregation leads to loss of protein function, and protein aggregates are implicated in several neurodegenerative diseases like Alzheimer’s and Parkinson’s diseases. A key issue in studies of protein aggregation is real-time monitoring under biologically relevant condition. We developed an HDX-MS-based approach by studying Alzheimer’s disease related Aβ aggregation, and we described this development in Chapter 2. Aβ proteins are labeled by deuterium in a pulsed way during Aβ aggregation. The extents of aggregations are monitored by MS as deuterium uptake. This pulsed HDX platform provides peptide-level information about Aβ aggregation. Ligands (drug candidates) were also evaluated with this platform to determine how the drug candidates affect oligomerization (Chapter 3).

Ligand interactions can induce protein conformational changes, which are required in various protein functions like signaling, enzyme activity. Such interactions are fundamental to all biological processes. One of the often used ligands in cells is calcium. Calcium interacts with a variety of calcium-binding proteins, most of which have conserved sequence that form EF-hand motifs to bind calcium. MS-HDX has been an important tool in studies of these typical calcium-binding proteins. Many proteins without an EF-hand motif also require calcium for their function. For example, protein-arginine deiminase (PAD) is an enzyme for arginine citrullination and binds calcium without EF-hand motif. We conducted differential HDX studies on PAD2 protein (Chapter 4). Multiple and cooperative calcium binding of PAD2 are detected by HDX. HDX was further extended by applying protein-ligand titration in an HDX experiment (i.e., Protein-ligand interactions by mass spectrometry, titration and H/D exchange, PLIMSTEX). The calcium binding affinity of each binding site can be elucidated by PLIMSTEX (Chapter 5).
Protein aggregation or ligand-binding induced conformational changes can also be detected by MS-HDX. One significant question in MS-based biophysical studies is how to generate structural information for proteins in the absence of a high resolution structure. In a newly developed platform, we combined a traditional structural biology approach, homology modeling, and MS-HDX to generate a structural model for diheme cytochrome c (DHCC) from *Heliobacterium* (Chapter 6), a protein for which solvent accessibility information from HDX experiment was used as the guide for homology modeling and used to generate a refined structural model of DHCC by using various computational approaches.

In summary, we describe in this thesis development and application of several new, refined approaches of HDX and analyze protein aggregation, protein-ligand binding and unknown protein structures. We hope other scientists can apply these approaches to solve complicated and demanding biological problems that are difficult to investigate using traditional biophysical methods.
Chapter 1: Introduction

This work is part of a book chapter in “Hydrogen-Deuterium Exchange Mass Spectrometry: Fundamentals, Techniques, and Applications” that has been submitted:

Chapter 11: “Hydrogen Deuterium Exchange for Analysis of Ligand Binding and Protein Aggregation”, Zhang, Y.; Rempel, D.; and Gross, M.
Different biological functions are carried out by different proteins, protein-ligand and protein-protein complexes. Any failure in these processes, no matter how small, may lead to a malfunction of the whole system. To understand fully the biological processes, we first need to know how proteins interact with each other. There are several biophysical methods for uncovering these interactions, and they will be described in this chapter. One of the recent developments, hydrogen deuterium exchange (HDX) coupled with mass spectrometry (MS) approach, has become a powerful technique and is of high interest to us. In this thesis, we are using HDX-MS to explore different proteins and their interactions by following protein high-order structure (HOS) changes induced by perturbations, including protein/ligand binding (Chapter 2-5) and temperature (Chapter 4). We also demonstrated that for homology modeling (HM), HDX-MS can be used to adjudicate protein structures suggested by HM and that are unknown so far (Chapter 6).

In the first part of this introduction, we describe the basic principles of HDX and several factors that affect HDX. This section is purposefully brief as these ideas are now well-known. In the second part, we review in more detail the use of MS and HDX for characterization of protein-ligand and protein-protein binding.

1.1 Hydrogen Deuterium Exchange (HDX)

HDX, as a powerful biophysics tool, been under development since the 1950’s. The occurrence of hydrogen and deuterium exchange induces a mass shift that can be monitored by MS. When coupled with MS detection, its use increases profoundly because it can follow protein conformational changes, dynamics and folding with relative ease and high sensitivity at both the global or protein level, at the peptide level by using peptides released from proteolytic digestion,
and even at the amino-acid level. More papers covering protein HDX now report the use of MS as the monitoring approach than any other method.

1.1.1 Types of Hydrogens in Proteins
There are three types of hydrogens in a protein: the hydrogens in carbon-hydrogen bonds, those in side chain groups, and those that are located on the amides of peptide bonds (also called the backbone hydrogens). The intrinsic exchange rate of hydrogens of carbon-hydrogen bond is too slow to observe, and that of the hydrogen on side chain group is so fast that back-exchange (i.e., when the protein is returned to H₂O solution) is significant. As a result, only the backbone hydrogens possess intermediate intrinsic rate that are readily measured, and they are excellent in as reporters for protein structure and dynamics because their exchange rates greatly depend on both hydrogen bonding and solvent accessibility.

1.1.2 pH and Temperature Effect
The intrinsic exchange rate of each residue is determined by primary structure of the protein (i.e., the nature of both the residue itself and its neighboring residues). In addition, this rate is also determined by other factors including temperature and pH. This is because the HDX reaction is controlled by both intrinsic rates of exchange and by protein dynamics. The latter is dependent on pH and activation energies or temperature. The higher the temperature, the lower the activation energy; thus, the faster the reaction occurs. As a result, one needs to be careful in controlling these two factors to perform proper HDX experiments. To obtain information from the backbone hydrogens when using MS, the HDX must be quenched so that the protein can be isolated, digested, and the peptides analyzed in H₂O solution. This is accomplished by controlling temperature at 0 °C and pH at ~2.5 so that the sample can be handled without causing back-exchange. This process is called “quench”, and plays an important role in HDX.
Furthermore, one can intentionally change the temperature and pH during the HDX process to obtain different information. We will give a detailed example on changing temperature in HDX experiments in Chapter 4.

1.1.3 HDX-MS

Because the absolute rate of HDX at a specific peptide bond is difficult to measure, we compare the relative deuterium uptake rates for proteins under different conditions and report the HDX deuterium uptake level as a function of HDX exchange time (Figure 1.1). In the 1990s, Roder and coworkers applied HDX and two-dimensional nuclear magnetic resonance (NMR) to protein-ligand complexes and showed that the binding interface can be located because the extent of HDX decreases in the region of binding. HDX coupled with MS began to emerge and proved to be up to the task of following protein exchange. Compared to NMR, MS has the advantages of higher sensitivity, permitting higher upper masses of the protein, and requiring no labeling or modification of a protein. Besides, HDX-MS should work with proteins under complex, biologically-relevant conditions. Sometimes proteins in the presence of impurities can be directly studied by HDX, as long as the impurities are not involved in the binding event. This advantage frees users from tedious and time-consuming protein purification. Unlike circular dichroism (CD) or fluorescence (we will discuss both of the methods in this chapter), HDX-MS can provide information at the peptide level through on-line pepsin digestion, and occasionally at the residue level by using electron-transfer dissociation (ETD) or electron-capture dissociation (ECD). Pepsin is widely used because it worked well with low-pH quenching condition. ETD or ECD experiments, however, are not routine and require careful control of the dissociation conditions.
Figure 1.1 A typical HDX-MS experiment workflow comparing between the apo (A) and holo (B) states of Calmodulin. In the holo state, Calmodulin complexes with Ca$^{2+}$ (shown in yellow spheres).
1.1.4 EX1 and EX2
HDX occurs at two distinct regimes, EX1 and EX2\textsuperscript{3,12-19}. Most proteins exhibit EX2 behavior, where the protein folding/unfolding rate is higher than the deuterium exchange rate. On the other hand, EX1 behavior occurs when protein undergoes a local unfolding event, exposing all the backbone hydrogens in this region and allowing exchange with D\textsubscript{2}O to occur rapidly before the structure folds back to the native state. Under the EX2 kinetics regime, the mass of the peptide or protein shows a continuous increase as a single population in isotopic distribution as the exchanging time increases. On the contrary, in the EX1 regime, there are always two mass populations; the low-mass population decreases whereas high mass one increases (Figure 1.2).

1.2 Protein-Ligand Interactions
In this section of the introduction, we will review in depth the use of HDX and MS for characterization of protein-ligand/protein interaction. We will also describe briefly other non-MS methods that utilize indirect or direct measurements and report information from protein to residue level.

All biological processes require intra- or inter-molecule recognition, interaction, and organization that drive the various functions of macromolecules. Proteins providing these functions include, for example, enzymes; molecular motors, whose conformational changes are involved in movement; transporters, which utilize conformational change to import or export ions, small molecules, and even proteins across cellular membranes; and detectors of cell signals, which often initiate or play an important role in signaling. All of these macromolecular functions are triggered by binding of various ligands, for example, metal ions, lipids, fatty acids, and nucleotides. Although most of these interactions are non-covalent, the affinities are usually strong, and, more importantly, the selectivity of these interactions is high. Most biological
Figure 1.2 Outcome of HDX under EX1 (A) and EX2 (B) regimes.
processes are performed by protein-ligand complexes that act in a coordinated manner. Any interruption in protein-ligand interaction can cause failure of the whole system. Thus, it is important to understand in detail how proteins and ligands interact.

1.3 Protein-Ligand Affinity Measurements

The key questions in studies of protein-ligand interactions are always where do the ligands bind and how tightly (i.e., binding affinity). In almost all protein-ligand interactions, there is equilibrium among the relevant species in solution (Equation 1.1), namely the free protein (P), free ligand (L), and protein-ligand complex (PL). Some protein-ligand complexes made up of non-specific interactions may also form transiently; however, usually their population is not significant.\(^\text{20}\)

\[ \text{P + L} \xrightleftharpoons[k_{\text{off}}]{k_{\text{on}}} \text{PL} \quad 1.1 \]

The on-rate constant, \(k_{\text{on}}\) (with units of M\(^{-1}\)s\(^{-1}\)), depends on the concentration of both free protein [P] and free ligand [L]; the higher the concentration of the protein and/or the ligand, the more likely they will encounter each other and bind. In an ideal case (sufficient [P] and [L]), \(k_{\text{on}}\) is controlled in part by the diffusion rate, the size of the protein (essentially the size of the binding site), and the nature of the ligand. The larger the protein, the faster the binding can occur to take advantage of the large surface area.\(^\text{21, 22}\) A charged binding site on the protein can attract oppositely charged ligands.\(^\text{23-25}\) In contrast, the off-rate constant, \(k_{\text{off}}\) (with a unit of s\(^{-1}\)), does not depend on concentration, and it represents only the probability of a protein-ligand dissociating to reform starting materials.
The rates \((r, \text{ in Equations 1.2 and 1.3})\) are the corresponding rate constants multiplied by the concentration of appropriate species, and \([\text{PL}]\) represents the concentration of the protein-ligand complex.

\[
    r_{\text{on}} = k_{\text{on}}[\text{P}][\text{L}] \quad 1.2
\]

\[
    r_{\text{off}} = k_{\text{off}}[\text{PL}] \quad 1.3
\]

As is usual, the association rate of free protein and free ligand to give the complex is equal to the dissociation rate of the reverse reaction (Equation 1.4) when equilibrium is established.

\[
    k_{\text{on}}[\text{P}][\text{L}] = k_{\text{off}}[\text{PL}] \quad 1.4
\]

The binding constant or association constant \(K_a\), by definition, is the ratio of \(k_{\text{on}}\) and \(k_{\text{off}}\) (Equation 1.5).

\[
    K_a = \frac{k_{\text{on}}}{k_{\text{off}}} = \frac{[\text{PL}]}{[\text{P}][\text{L}]} \quad 1.5
\]

The dissociation constant or binding affinity, \(K_d\) (with the unit of M) is the reciprocal of \(K_a\) and is more frequently used (Equation 1.6), whereas the binding affinity is a measure of the strength of the protein-ligand interaction. When the free protein and ligand reach the same value as the dissociation constant, 50% of the complex is formed.

\[
    K_d = \frac{1}{K_a} = \frac{k_{\text{off}}}{k_{\text{on}}} = \frac{[\text{P}][\text{L}]}{[\text{PL}]} \quad 1.6
\]
1.4 Conventional Methods for Ligand Binding Characterization

Measuring protein-ligand interactions is ultimately needed for most biological studies, and numerous analytical and computational approaches have been developed to characterize the interaction. The methods make use of thermodynamics, kinetics, stoichiometry, and perturbation related to ligand binding. Each approach can generate specific information about the interaction. More importantly, information from each approach can be combined, and a comprehensive view can be achieved. New additions to this “tool box” are MS-based methods. A brief overview of existing methods is provided for perspective before we focus on the MS-based methods.

1.4.1 Isothermal Titration Calorimetry (ITC)

The binding events in a protein-ligand interactions always involve an enthalpy change. One direct measurement of an enthalpy change during ligand binding is ITC\(^{26, 27}\). In an ITC experiment, a certain amount of ligand is titrated into the protein solution. The mixture is stirred to reach equilibrium quickly. The difference in the amount of heat that required to maintain a constant temperature between a reference and sample cell is related to the heat released or absorbed in the binding event. ITC can be applied to all biochemical processes that show an enthalpy change. ITC is a general approach for characterizing protein-ligand interactions without requiring any special labeling or buffer system and can afford stoichiometry \(n\), binding affinity \(K_a\), and enthalpy \(\Delta H^0\). ITC has limited sensitivity, however, requiring relative large quantities of purified and water-soluble proteins. The approach does not directly provide any information on the location of the interaction.
1.4.2 Fluorescence
Fluorescence-based methods have been widely applied in studies of protein-ligand interactions\textsuperscript{\textcopyright 28-30}. Most commonly, binding affinity can be elucidated in a titration-type experiment in which fluorescence is used to measure equilibrium concentrations. Another approach is fluorescence anisotropy that utilizes the degree of decorrelation in the polarization of emitted and incident light. Such differences are induced by the free tumbling of the molecule that is reduced significantly upon binding to the protein.

Another popular fluorescence-based method is Förster Resonance Energy Transfer (FRET) for which energy movement between an excited fluorophore, the donor, and a neighboring, ground-state fluorophore, the acceptor, occurs and is monitored. FRET can provide the distance between the donor and acceptor\textsuperscript{31}. With different labeling positions, FRET can provide some view of the conformational changes occurring upon ligand binding or protein-protein interactions. FRET requires that the protein be modified by introduction of fluorophores, which are often attached to reactive amino acid side chains, most frequently cysteine and lysine. One potential problem is that the modification could alter binding behavior. Because this is not known a priori, effort is often needed to ensure that the target protein is not affected. The interpretation of fluorescence data for systems with multiple protein-ligand interaction states could be problematic as well.

1.4.3 Fourier Transform Infrared spectroscopy (FTIR)
Protein-ligand interactions involve conformational changes of proteins. FTIR is able to monitor these changes by measuring secondary structure content (e.g., α-helix and β-sheet content) of the protein\textsuperscript{32,33}. Different bond types (e.g., amide I and II) have distinct infrared absorption, and each absorption band corresponds to different molecular vibrations. For example, amide I absorption contains mainly a C=O stretch, whereas amide II absorption arises from both N-H
bending and C-N stretching. The FTIR-based methods are accurate and sensitive. The spectra are information-rich, recording nearly many types of molecular vibrations. To obtain spatially-resolved information, the bands must be assigned to individual groups, usually requiring isotopic labeling by site-directed mutagenesis.

1.4.4 Surface Plasmon Resonance (SPR)
Optical biosensors, like SPR, have become important tools for basic research and drug discovery. SPR, with its exceptional sensitivity, can be viewed as a biosensor for specific biological interactions (e.g., antigen-antibody binding in biopharmaceutical research). The binding kinetics as well as the binding affinities of protein-ligand interactions can be obtained. In a typical SPR experiment, the ligand is immobilized on a gold surface of an SPR chip, and the protein solution flows over the surface. The sensor detects changes of electromagnetic waves formed by electrons (i.e., surface plasmons of the gold layer) when their frequency matches the frequency of the incident light at a specific angle. The surface plasmon is influenced by its environment and will change upon interaction between an immobilized ligand and a protein. One limitation of this surface-based measurement is the need for immobilization of the ligand on the surface. A specific problem is that immobilizing the analyte may restrict its rotational freedom and diffusional properties, and these alter the thermodynamics and the kinetics of binding.

1.4.5 Enzyme-Linked Immunosorbent Assay (ELISA)
ELISA is a popular bio-analytical assay and another spectroscopic approach to examine protein-ligand binding. In general, the key step in ELISA is the immobilization of the antigen, which can be accomplished either by direct adsorption of the assay plate or indirect capture via an antibody that has already been immobilized on the plate. ELISA takes advantage of specific binding between antigen and antibody when a detectable signal from another enzyme-substrate reaction
(normally an absorbance change) can be observed\textsuperscript{38, 39}. ELISA is convenient and sensitive. The observation must be made sufficiently fast, however, to avoid misleading results from over-reaction between the enzyme and substrate. It can provide some information on the site of the interaction when the fluorophore is moved from site to site of the protein, but requires longer time.

1.4.6 Circular Dichroism (CD)
Traditional approaches in structural biology can also be used to determine protein-ligand interactions. CD detects the difference in absorption of left and right circularly polarized light, usually in protein solutions comparing between in presence and absence of the ligand. A CD spectrum responds to the bonds and sub structures responsible for the chirality\textsuperscript{40}. When a ligand binds to a protein, the binding perturbs the protein’s chirality\textsuperscript{41, 42}. CD is fast and solution-based. Disadvantages are low structural resolution and requirement of a pure protein sample.

1.4.7 Nuclear Magnetic Resonance (NMR)
NMR can provide high resolution for protein-ligand interaction in solution under near physiological conditions. Binding affinity, interface, and structural rearrangements accompanying protein-ligand interaction can be elucidated. NMR targets changes of the electromagnetic energy that nuclei absorb and release in a magnetic field\textsuperscript{43, 44}. In contrast to all methods mentioned above, NMR is the one of few methods that can generate a high-resolution (atomic level) structural model of protein-ligand complexes. Another advantage is its application to systems with weak interactions\textsuperscript{45}. Although NMR has low sensitivity (requiring high-concentration of pure protein sample) and low throughput, many applications of NMR have been reported in studies of protein-ligand interactions\textsuperscript{46}. 

- 13 -
1.4.8 X-ray Crystallography
Similar to NMR, X-ray reports protein structure at a molecular level. Ligand binding can perturb the electron density around the binding sites, which is reflected by the method\textsuperscript{47}. One can titrate a protein with a metal ion, for example, and use electron density to measure metal-ion occupancy. The major drawback of crystallography-based approaches is that the results pertain to a solid-state structure, and dynamics involved in protein-ligand interactions is missing. Proteins (e.g., apo forms) may be so dynamic that they are difficult to study by crystallography-based approaches. In addition, crystallography requires, of course, that the protein-ligand complex can be crystallized.

1.5 Direct Mass Spectrometry Method
All of the above approaches have limitations. Many have good sensitivity but poor or no structural resolution. Others have high structural resolution (e.g., NMR, X-ray crystallography) but low sensitivity or inability to work with large proteins or those that do not crystallize. Mass spectrometry has become a powerful analytical approach for proteins, thanks to the availability of appropriate ionization methods (e.g. Electrospray Ionization (ESI) and Matrix-Assistant Laser Desorption/Ionization (MALDI)). An accurate and sensitive protein analysis can be done with relatively small samples. More importantly, a modification of ESI (called “native” MS) can introduce proteins directly from solutions that are similar to a protein’s native environment. In that case, ESI solvents are aqueous (e.g., ammonium acetate solution at pH 7), making it possible to observe gas-phase, protein-ligand complexes that are held together by non-covalent interactions. This sets up direct methods whereby MS measures equilibrium concentrations at both high sensitivity and perhaps with some structural resolution from MS/MS methods\textsuperscript{48}.
Klassen and coworkers\(^4\) have played a major role in the development of a direct MS method in which key concentrations in Equations 1.5 and 1.6 are measured by the mass spectrometer. Protein-ligand complexes, as well as free proteins, are introduced by ESI into the instrument. The intensities of signals representing different species can be recorded at different protein-ligand concentrations, so that information about protein-ligand interaction (e.g., binding affinity) can be elucidated. Various states of a complex can be isolated and studied separately. A “catch and release” approach, which includes isolating the complex ions, dissociating them by applying energy (e.g., collision-induced dissociation (CID)) and detecting the product ions by high resolution instrument, has also been developed and applied in drug screening and for estimating affinities.

### 1.6 HDX-MS

As mentioned above, HDX-MS has become an attractive complement for biophysical studies of protein-ligand interaction. The application of HDX-MS to protein-ligand systems can afford detailed information on binding regions, binding affinity, and binding order. There are several review articles on HDX-MS\(^5\)\(^-\)\(^4\). In the remainder of this chapter, we will focus on methodology and analysis for HDX-based studies of protein-ligand and protein-protein interactions.

#### 1.6.1 HDX-MS for Binding Regions

Often, the first question about a protein-ligand interaction is where the binding interface lies. The answer can be provided by HDX approaches (e.g., continuous labeling HDX). The relative deuterium uptake rates (in form of an HDX kinetics curve, where deuterium uptake is plotted against time of exchange) are compared between ligand-free (apo) and ligand-bound (holo) states. Ligands can be metal ions\(^5\)\(^5\), inhibitors\(^5\)\(^6\), glycans\(^5\)\(^7\), lipids\(^5\)\(^8\), or nucleotides\(^5\)\(^9\). Regions of proteins affected by ligand binding usually exhibit slower deuterium uptake because hydrogen...
bonds along certain protein back-bone amides are affected by ligand binding and the protein is stabilized as a complex. Those regions that exhibit differences are likely to be the binding interface or remote regions that change as an allosteric interaction (an application will be described in Chapter 4).

1.6.2 HDX-MS for Binding Affinity

There are two approaches based on HDX to determine protein-ligand binding affinity: Protein-Ligand Interactions by Mass Spectrometry, Titration, and H/D Exchange (PLIMSTEX), and Stability of Unpurified Proteins from Rates of H/D Exchange (SUPREX). Both methods measure changes in deuterium uptake as a function of either ligand or denaturant concentration and can report quantitatively thermodynamic properties of the protein-ligand complex. The methods are applicable when the target protein undergoes HDX via an EX2 mechanism (i.e., the intrinsic rates of HDX are slower than the off rate of the ligand). Additionally, the methods require a predetermined incubation time when using continuous HDX labeling such that all points are measured at a constant time.

1.6.2.1 PLIMSTEX

PLIMSTEX was developed by Gross and coworkers\textsuperscript{60} as a means to measure binding affinities (Figure 1.3). It compares deuterium uptake level of the apo and holo protein, not as a function of HDX time, but of the total protein ligand ratios (i.e., $\frac{[L]_T}{[P]_T}$, where $[L]_T$ represents total ligand concentration, and $[P]_T$ is total protein concentration) to afford protein-ligand binding affinities. Usually, $[P]_T$ is constant, and the only experimental variable in PLIMSTEX is $[L]_T$.\textsuperscript{61} The underlying principle of PLIMSTEX is similar to that of fluorescence but without the need to measure the free-ligand concentration.
Figure 1.3 A typical workflow for PLIMSTEX.
The first step of PLIMSTEX is a standard continuous labeling HDX experiment, which can afford kinetic curves, either global or at the peptide level. By comparing the kinetic curves between apo and ‘hard’ holo (i.e., the protein is completely ligand-bound) states, one can determine an HDX time point for which the deuterium uptake is nearly constant and there are relatively large differences between the apo and holo states. Under these conditions, HDX of the protein and protein-ligand complex are nearly steady-state so that small errors in time have little effect on the extent of HDX. Furthermore, the large difference in HDX between apo and holo also adds to the accuracy and facilitates analysis of complicated systems (i.e., proteins with multiple binding sites for same ligands).

After determining the titration time for PLIMSTEX, a set of mixing experiments is conducted by incubating proteins with increasing \([L]\) (i.e., titrating), from zero (i.e., apo state) to excess ligand (i.e., holo state) (Figure 1.4). When the protein-ligand complex has reached equilibration with free protein and ligand, HDX is initiated by adding a buffer in \(\text{D}_2\text{O}\). The steps for MS analysis are the same as those for continuous labeling HDX experiments: quench with acidic solution, desalt by loading onto a reversed-phase column, elute the trapped protein/peptides into the mass spectrometer, and measure the mass shift (\(\Delta D\)). After obtaining \(\Delta D\), the extent of HDX as a function of \([L]\) is plotted giving a “PLIMSTEX curve”. Typical results show that the deuterium uptake level decreases with increasing \([L]\), reflecting increasing protection of the backbone amide hydrogens in the binding regions or other regions involved in allosteric interactions. Other binding intermediates can be monitored as well. The PLIMSTEX curve is then fit by a mathematical model to afford the \(K_d\) value$^{62}$. 
Figure 1.4 A typical PLIMSTEX curve obtained for 15 µM porcine calmodulin titrated with Ca\(^{2+}\) in 50 mM HEPES. Redrawn from data in ref.\(^{61}\)
A nonlinear least squares (NLLS) regression is performed by procedures implemented with MathCAD 14 (PTC Inc., Needham, MA). As mentioned above, the titration data are fitted using a 1:n = protein:ligand binding system, where n is the number of binding sites of the same ligand to the target protein. In PLIMSTEX modeling, $\Delta D$ is a function of $[L]_T$, and the cumulative binding constants $\beta_i$ (i.e., product of the stepwise binding constants $K_i$, where $i$ varies from 1 to n (Equation 1.7).

$$\beta_i = K_1 \times \cdots \times K_i \quad 1.7$$

$D_0$ is the deuterium uptake of the apo state protein, and $\Delta D_i$ is the deuterium uptake differences between the apo state and the intermediates $i$ (i.e., $\Delta D_i = D_0 - D_i$). Typically, $\Delta D_n$ is the largest $\Delta D$, as ligand binding induces protection of the binding sites, so deuterium uptake decreases as ligand is added. $D_0$ is treated in the modeling as a variable to minimize experimental errors. The best fit is obtained by searches, changing all the variable parameters (i.e., $\beta_i$, $D_0$, and $\Delta D_i$) to minimize the error between the fitted curve and the experimental data by iterating through many trials.

1.6.2.11 Examples of PLIMSTEX
PLIMSTEX was first demonstrated by using four model proteins: rat intestinal fatty acid binding protein (I-FABP) interacting with potassium oleate (1:1), GDP-bound human p21$^{H-ras}$ protein (Ras-GDP) interacting with Mg$^{2+}$ (1:1), Ca$^{2+}$-saturated porcine calmodulin (holo-CaM) interacting with the peptide melittin (1:1), and apo-CaM interacting with Ca$^{2+}$ (1:4)$^{60}$. PLIMSTEX was able to provide accurate $K_d$‘s of the four systems ranging from $10^4$ to $10^8$ M$^{-1}$. These early experiments also suggested that PLIMSTEX can be used for quick determination of binding stoichiometry and purity of the proteins$^{63}$. 

- 20 -
In an early collaborative application, we applied PLIMSTEX to a protein-DNA binding system, human telomeric repeat binding factor 2 (hTRF2) interacting with double-stranded telomeric DNA (repeats of TTAGGG). When accompanied by protease digestion, PLIMSTEX curves can provide binding information at the peptide level. For this system, the PLIMSTEX-determined binding affinity is within a factor of three of a previously reported value. Sequel studies showed applicability to other systems as well. Another application of PLISMTEX on a Ca\(^{2+}\) binding protein will be described in Chapter 5.

### 1.6.2.1.2 Advantages of PLIMSTEX

One advantage of this approach is that affinities determined are for a protein in a native, or at least near-native, condition during the HDX. The amount of protein required for the titration is small as is the concentration (µM and lower in favorable cases). Unlike SUPREX, which will be introduced later, PLIMSTEX does not require a denaturant, which may affect protein binding. In addition, PLIMSTEX does not need any tagging reaction to measure the free-ligand concentration \([L]\) during the titration process. It only relies on the measurement of \(\Delta D_i\). Unlike gas-phase direct measurements, the titration is done in solution, and the extent of HDX, although measured in the gas phase of a mass spectrometer, simply reports on the status of the binding. The approach can deal with 1:n binding systems, affording analysis of multiple macroscopic binding constants \(\beta_i\), providing each binding events induces a measurable change in the deuterium uptake. The stoichiometry of protein-ligand complexes can also be elucidated. As such, PLIMSTEX complements and validates direct MS measurements where non-covalent protein-ligand complexes can be introduced into the gas-phase and their molecular weights measured to reveal stoichiometry. The direct measurement, based on the size of the complex, can be misleading, however, if the protein-ligand complex is not faithfully transferred from solution...
to the gas phase. When [P] is sufficiently large (e.g., 100 times the $K_d$), reliable $K_d$ values cannot be measured. Instead, a “sharp-break” curve is observed, reflecting the binding stoichiometry (Figure 1.5). In addition, these curves may also be useful to determine the purity of the protein by titrating with a ligand of known purity. The throughput of PLIMSTEX (e.g., in drug discovery) may be increased by using MALDI instead of LC/ESI-MS, but this remains to be demonstrated. PLIMSTEX has the potential to give peptide (regional) resolution by applying protease digestion and following HDX for peptides, thus locating the binding site and the $K_d$. This will be an important advance especially when a protein binds multiple ligands. We are now testing this prospect.

1.6.2.1.3 Disadvantages of PLIMSTEX

One concern of PLIMSTEX is that protein or protein/ligand dynamics may distort the outcome. Intermediates (especially for a 1:n system where $n > 1$) may introduce problems, as multiple equilibria occur simultaneously (Figure 1.6). Apo (H) and apo (D) represent apo state protein, in the absence or presence of deuterium, respectively, whereas holo (H) and holo (D) have similar meanings for the complex. Based on Equation 1-2 and 1-3, both on and off rates can be calculated; the former is a bimolecular reaction whereas the latter is a unimolecular process. At the beginning of the titration, [L] is small, and the $r_{on}$ is relatively smaller than $r_{off}$. As a result, much of the protein, even as complexed, will exist in the apo state if the off rate is high and undergo exchange characteristic of the apo (D) state. Of course, it is not possible for holo(D) to become holo(H). Thus, when [L] is small, there is a higher possibility of obtaining more deuterium than expected, introducing a distortion of the PLIMSTEX curve at the beginning of the titration. Therefore, [L] has to be sufficiently large to shut down the other fluxes. The outcome can be that the $K_d$ calculated with the PLIMSTEX platform is too large.
Figure 1.5 Examples of ‘sharp break’ PLIMSTEX curves for titration of (a) melittin and (b) mastoparan of 15 μM Ca²⁺-saturated porcine calmodulin (CaM-4Ca) in 50 mM HEPES, 100 mM KCl, 0.49 mM Ca²⁺, 99% D₂O, apparent pH 7.4. Redrawn from data in ref. 63.
**Figure 1.6** On- and off-fluxes among various species in PLIMSTEX experiments.
In addition, PLIMSTEX is unable to measure tight binding lower than nM. Nevertheless, for tight binding, it should be possible to determine the stoichiometry, as mentioned above, from a sharp-break curve, and put limits on the affinity.

1.6.2.1.4 dPLIMSTEX

dPLIMSTEX is an adaption of PLIMSTEX incorporating a dilution strategy. It was developed for protein-peptide systems (e.g., antibody-antigen system) to minimize sample consumption, which is an issue with standard PLIMSTEX and valuable proteins. The “ligand” peptide (antigen) is used as readout, unlike signals from proteins in the standard PLIMSTEX, taking advantage of the improved capacity of mass spectrometers to measure the mass of a peptide ligand with higher accuracy and precision than that of a large protein.

The work flow for dPLIMSTEX starts with half volume of an equilibrated protein-peptide complex for measurement. The other half is diluted in aqueous buffer before incubation (Figure 1.7). The dilution step is continued until the concentration of the peptide is too low to be detected. dPLIMSTEX was first demonstrated by using a model system, calcium-saturated calmodulin with the opioid peptide β-endorphin, and it yielded a similar binding constant as compared to that determined by standard PLIMSTEX and other methods. It was then applied to a monoclonal anti-nitrotyrosine antibody, in complex with a 3-nitrotyrosine-modified peptide system. A binding stoichiometry of 1:2 was confirmed. In addition, a $K_d$ in the low nM range and a minimum of five amino acid constituting the epitope were determined. Compared to the standard PLIMSTEX protocol, dPLIMSTEX has the advantage of consuming less material, and being less subject to error on the mass shift because it monitors, in this case, a peptide rather than a large protein.
Figure 1.7 A typical work flow for the dPLIMSTEX. Terms x and y are the initial concentrations of protein and peptide, respectively, and z is the dilution factor. Redrawn from data in ref. 67.
1.6.2.2 SUPREX
Another HDX-based method, SUPREX, was developed by the Fitzgerald and coworkers\textsuperscript{68}. It is analogous to chemical denaturation methods employing CD or fluorescence, which were previously known.

It begins with incubating a protein (apo or holo state) in D\textsubscript{2}O buffers containing different concentrations of denaturant ([denaturant]), usually urea or guanidinium chloride (GdmCl), for a predetermined time, which is constant for all the measurements. Approximately ten different [denaturant] concentrations are commonly used to plot a complete SUPREX curve. After quenching the HDX, a measurement is made, often with MALDI-MS, although ESI can also be used. As [denaturant] increases, the deuterium uptake also increases, reflecting the loss of stability of the protein (Figure 1.8). In the presence of the ligand, the protein is stabilized, and the curve shifts to larger concentrations of denaturants. The difference between the apo and holo states is a measure of the binding constant. SUPREX reports mostly on the globally protected and slow exchanging regions, providing that these regions are involved in the denaturation.

Similar to PLIMSTEX, SUPREX data are plotted as $\Delta D$ (deuterated – undeuterated) of an apo/holo state protein as a function of [denaturant], and fitted using NLLS analysis. The fitted curve affords a transition midpoint ($C^2_{\text{SUPREX}}$, as a [denaturant] at this point), which can be used to calculate $\Delta G_f$ and an $m$-value (defined as the sharpness of the transition in the fitted curve (i.e.,

$$\frac{\delta \Delta G_f}{\delta [\text{denaturant}]}$$

(Equation 1.8).

$$K_{\text{fold}} = e^{\frac{\Delta G_f + m[\text{denaturant}]}{RT}} \frac{1}{1.8}$$
Figure 1.8 A typical workflow for SUPREX. Redrawn from data in ref. 69
\[ K_{\text{fold}} = \frac{k_{\text{close}}}{k_{\text{open}}} \]

In Equation 1.8, \( K_{\text{fold}} \) is the equilibrium constant between the exchange incompetent and competent states, which are related to protein-folding dynamics (Equation 1.9). \( R \) is the gas constant, and \( T \) is the temperature (in Kelvin). By obtaining differences between \( \Delta G_f \) (holo) and \( \Delta G_f \) (apo), together with the \( m \)-value, \( \Delta \Delta G_f \) can be calculated\(^{70-72} \) and converted into a dissociation constant \( K_d \) (Equation 1.10). For the apo state of a protein, the result relates to protein dynamics, and that of a holo-state protein gives information of the stability of protein-ligand complex.

\[ \Delta \Delta G_f = -nRT \ln(1 + \frac{[L]}{K_d}) \]

All ligand-bound experiments are performed in a 50-100 fold excess of ligands, so that most of the protein is in a “hard”-holo state\(^73 \). The changes of \( m \)-value, on the other hand, can be used to evaluate the surface area that is buried upon protein folding/unfolding or ligand binding\(^74 \). The differences, if any, of the baselines before the transition between two states in SUPREX curves also indicate the protection of the protein when the ligand binds. The lower the baseline compared to that of the apo state, the less solvent-accessible the protein. This difference, which is crucial for the success of PLIMSTEX, does not impact SUPREX.

For a strict two-state model (i.e., partially folded state is transient and not significantly populated), SUPREX provides reliable stability data (i.e., \( \Delta G_f \) and an \( m \)-value) compared to traditional methods (e.g., CD and fluorescence)\(^{70-72,75} \). The difference (\( \Delta \Delta G_f \)) between apo and holo states of the target protein is the binding free energy of the ligand. In contrast, if the system
has more than two states in equilibrium, with folding/unfolding events happening cooperatively, the absolute values of such stability data (i.e., $\Delta G_i$) are less useful. Relative differences (i.e., $\Delta \Delta G_i$), however, can still accurately reflect the binding free energy. In the case of systems that fold/unfold uncooperatively (i.e., different domains of the same protein fold/unfold independently), global SUPREX cannot produce useful information. An alternative is to turn to the peptide level by using protease digestion.

Most SUPREX experiments have been performed on small proteins (less than 15 kDa), as MALDI always produce ions of +1 charge state. Instead of increasing the dynamic range of a MALDI instrument, Fitzgerald and coworkers developed a strategy where protease digestion is coupled with standard SUPREX, after the exchange and quenching steps (Figure 1.9). The purpose of a rapid digestion is to generate peptides that can cover different individual domains. Such a protocol significantly expands the application of SUPREX to larger proteins. As for all HDX experiments, the footprints (i.e., number of deuteriums) of each proteolytic product are already fixed before analysis, and this information is largely preserved until analysis. For success, the peptides must undergo a sufficiently large conformational change so that changes can be followed.

1.6.2.2.1 Examples of SUPREX
The thermo-stability of monomeric $\lambda$ repressor variants and maltose-binding proteins were used in 2000 to demonstrate SUPREX. The stability measurements were compared with CD denaturation curves of same purified samples. The experimental requirements of SUPREX, the back exchange rates in MALDI matrix and the deuterium uptake as function of time in different [denaturant] were also determined. The addition of the denaturant, in this case, GdmCl, dramatically increases the exchange rate through protein global unfolding. More importantly,
Figure 1.9 A typical workflow for SUPREX with digestion. Redrawn from data in ref. 77.
the SUPREX curves of maltose-binding protein shifts between apo and holo proteins (i.e., in presence and absence of maltose). The presence of maltose increases the stability by binding to the native state, and the protein unfolds at higher [GdmCl]. The thermodynamic stability information can be extracted from the deuterium uptake curves as well.

The original SUPREX approach requires accurate MS measurement of an intact protein, especially in the proposed single-point SUPREX for drug candidate screening. The protease digestion from traditional HDX experiment can be integrated into the SUPREX protocol to improve the accuracy of deuterium-uptake measurement. The strategy utilizes the SUPREX behavior of the detected peptides to report on behavior of different regions from the protein. The modern MS instrument can deliver high resolving power that will greatly improve the accuracy of SUPREX and avoid the difficulty of using direct measurements of large, intact proteins. Two model proteins, CypA and glyoxylate aminotransferase (ACTmi) were used as model to demonstrate the protease-assisted, single-point SUPREX.

1.6.2.2.2 Advantage of SUPREX
One advantage of SUPREX, like continuous labeling HDX and PLIMSTEX, is that it can be applied to proteins under biologically relevant or near relevant conditions. Competitive methodology, like fluorescence, often needs modifications of the target protein, to incorporate a fluorescent label, as discussed earlier. SUPREX can also work with unpurified samples as long as the other components do not interact with the protein and do not significantly suppress its signal in the mass spectrometer. SUPREX is a particularly appropriate choice when salt contamination is important because MALDI has a high tolerance of salt.

Another advantage of SUPREX is that it is a good candidate for high throughput screening (HTS). As mentioned earlier in this chapter, HTS is a problem for standard HDX. With the
growing availability of fully automated instrumentation optimized for HTS, one can potentially screen up to 100,000 ligands per day, using a single-point SUPREX protocol\textsuperscript{78}, in which one compares only a single $\Delta D$ between two states: protein-target ligand complex and protein-reference ligand complex. The greater $\Delta D$ is, the tighter the ligand binds to the protein. The limitation in the data rate is the instrumentation. Nevertheless, SUPREX is well-suited for projects aiming at the discovery of new ligands that bind to a specific protein in drug discovery.

1.6.2.2.3 Disadvantage of SUPREX
The advantage of HTS comes at the expense of losing detailed structural information, which is often provided with continuous labeling HDX or PLIMSTEX\textsuperscript{69}. Even with application of a protease-digestion protocol, the resolution can be at best a single domain of the protein\textsuperscript{77}, instead of resolution at the peptide or even residue level. In addition, the requirement of denaturant is problematic because the denaturant may alter or even prevent protein-ligand binding; thus, protein folding/unfolding is not the only parameter that affects the results. SUPREX can only be used for a 1:1 binding system because a higher ligand binding system cannot satisfy the two-state assumption that is required in the $\Delta G_f$ calculation. Nevertheless, SUPREX is an important method for deciphering biophysical properties of protein-ligand complexes, and it can have reasonable throughput.

All of the above mentioned HDX-based methods are limited, because they require a conformational change or a relatively large shielding or protection to be introduced by the ligand binding. For the systems that do not undergo dramatic changes during complex formation, both PLIMSTEX and SUPREX may be useful if coupled with a competition experiment (i.e., performing experiments in which another ligand of known affinity competes) or with a pulsed-
labeling strategy that allows the analyst to use short HDX times and focus on the fast-exchanging amide hydrogens.

1.6.3 HDX-MS for Binding Order

Once information is obtained on where and how tightly ligands bind to target proteins, one may need to follow the order of ligand binding when a protein binds more than one ligand. Here the question is which binding site is used first and which is last. This binding order is of importance because such information can guide research on drug development, particularly when the protein-ligand system is asymmetrical (i.e., different ligand binding regions are not identical). An example is aminoglycoside-N3-acetyltransferase-IIIb (AAC)\textsuperscript{79}. Its order of substrate binding (either coenzyme A or antibiotic) yielded ternary complexes with different dynamic properties.

To measure the order of binding, Gross and coworkers\textsuperscript{65} developed a PLIMSTEX-related approach. Troponin C (TnC), a Ca\textsuperscript{2+}-bound protein, is an excellent model, because four Ca\textsuperscript{2+} ions bind to four EF hands of the protein. Because Ca\textsuperscript{2+} binding to EF-III and EF-IV are known to be high affinity, a two-step dialysis procedure was carried out to ensure a fully apo state. In the work flow, a “sharp break” PLIMSTEX curve is first obtained (i.e., at high protein concentration) to reveal the binding stoichiometry, and give an estimation of the $\Delta D$ values to guide the curve fitting to obtain equilibrium constants. The binding constants for each Ca\textsuperscript{2+} are then used to calculate fractional species curves. Based on these curves, one can calculate the concentration of the various binding species (in this case, apo-TnC, TnC-Ca\textsuperscript{2+}, TnC-2Ca\textsuperscript{2+}, TnC-3Ca\textsuperscript{2+}, and holo-TnC) as a function of [Ca\textsuperscript{2+}]. By applying continuous labeling HDX under various predetermined concentrations of Ca\textsuperscript{2+}, one can determine when different regions bind Ca\textsuperscript{2+} as HDX reports changes caused by ligand binding at that site. In the case of troponin, the
determined binding order to four EF hands is III, IV, II and I (Figure 1.10), which would be difficult if not impossible to determine in any single experiment.

The advantage of identifying binding order using an HDX-based method is the approach does not depend on accurately detecting abundances of the various binding species, but on their HD exchanged mass. The advantages of PLIMSTEX also apply. The equilibrium information obtained from the PLIMSTEX experiment is essential for the design of the subsequent kinetics experiments that extract deuterium distribution for different binding regions. For a tight-binding system, no binding order information can be obtained. Compared to traditional approaches, this method is time-consuming. Nevertheless, this method can be applied to proteins that bind multiple ligands, without mutating the protein, which may disturb the system.

1.6.3 HDX-MS for Protein-Protein Interactions

1.6.3.1 SIMSTEX for Protein Association

Thus far, we have emphasized protein/ligand binding. Proteins also self-associate; sometimes, self-association is simple, producing small oligomers that readily dissociate. Other times, the protein undergoes uncontrolled aggregation and ultimately forms fibrils (amyloids) that have serious consequences in human health (e.g., as in Alzheimer’s disease, Parkinson’s disease).

Insulin is an excellent model to test whether HDX can be used to assess self-association. It is a biologically relevant protein of modest size, but it still possesses all of the structural features of a large protein. It is stored in the pancreas as a hexamer, but it functions as a monomer\textsuperscript{80, 81}. Gross and coworkers\textsuperscript{82} adapted PLIMSTEX to determine self-association equilibrium constants for proteins, using insulin as a model, and called the approach SIMSTEX (Self-association Interactions using Mass Spectrometry, self-Titration and H/D Exchange). The work flow is, in effect, a titration, but one in which the protein is titrated with itself. This is accomplished by
Figure 1.10 HDX patterns for various Ca\(^{2+}\) binding sites (obtained as peptides by peptic digestion) indicated the binding orders. Each column represents deuterium distribution of the peptic peptides for each EF hand: III, IV, II, and I (left to right). Each row represents various Ca\(^{2+}\) bound states: 0, 1, 2, 3, and 4 Ca\(^{2+}\) bound (top to bottom). Redrawn from data in ref.\(^{65}\)
simply changing its concentration. As the concentration of the protein increases, more oligomerization occurs, resulting in more amide hydrogens being protected, and less deuterium uptake. To model the HDX data, we assumed that the oligomerization of r-human insulin occurs as monomer ⇌ dimer ⇌ tetramer ⇌ hexamer. Furthermore, different mutants of insulin were also compared to show that substituting amino acids B9, B13, B21 and B27 significantly impacts the self-association of insulin. Some of them (e.g., B13Gln) show increased tendency to self-associate, whereas others (e.g., lispro and B9Asp) show less tendency. One concern of SIMSTEX is that the dissociation/association kinetics can distort the shape of the curve. In this example, however, the distortion is reduced, possibly owing to the slower kinetics of HD exchange, compared to the dissociation/association. The resulting association constants agree with literature values, suggesting that the SIMSTEX approach can be useful.

The advantage of SIMSTEX is that it is not time-consuming, and it provides an accurate comparison between different mutants. It has potential to improve the structural resolution to the peptide or even single amino-acid residue level. Its limitation is the requirement that self-association increases H bonding and decreases solvent accessibility such that the deuterium uptake level change can be detected. Furthermore, the approach likely cannot handle the complexity of aggregation of proteins like Amyloid beta. Nevertheless, the approach motivates continued development, not only for drug screening, but also for the study large-protein self-aggregation.

1.6.3.2 Pulsed HDX for Protein Aggregation
This protein aggregation is the immediate causes of Alzheimer’s, Parkinson, and Huntington disease, motivating development of new biophysical approaches. Although uncontrolled aggregation of amyloid proteins is also a deleterious example of protein self-interaction causing
disease, the process is far too complex for application of SIMSTEX. Amyloid beta (Aβ) (Alzheimer’s disease, AD) and α-synuclein (Parkinson’s disease) are of importance because their propensity to aggregate affects brain function. In AD, the aggregation of the 40- and 42-aa-long Aβ peptide, generally called Aβ40 or Aβ42, as well as related peptides is proposed to be involved in the onset of the disease. Aβ42 appears to be the most neurotoxic and amyloidogenic member of the family.

The structures regarding the first and last stage in Aβ42 aggregation have been studied extensively by NMR, X-ray and fluorescence. Nevertheless, there is still little understanding of the soluble aggregates of Aβ42 because the aggregation gives intrinsically high heterogeneity of oligomers and of time-dependent behavior. Thus, a continuous labeling HDX or SIMSTEX strategy is of little use. Furthermore, the application of HDX in a time-dependent manner is challenging because both HDX and aggregation occur simultaneously. To minimize aggregation during the time for HDX and thereby separate the two processes, we developed a pulsed HDX strategy, whereby Aβ42 is allowed to incubate for various times and then submitted to pulsed HDX (1 min) followed by the standard quenching, proteolysis and analysis. More details will be described in Chapter 2 and 3. The advantage of the pulsed HDX approach is that it is applicable to studies under various experimental conditions of aggregation. Thus, it can be a general tool to support future efforts to characterize protein aggregation as a function of different parameters (e.g., concentration, presence of different ligands or protein, and pH). With the help of ECD/ETD, site-specific information can be obtained as well.

Although this approach is convenient and easy to understand, one needs to predetermine the aggregation times carefully, as the samples are incubated separately. Different time points need
to be chosen for different aggregation conditions. For example, higher temperature and faster agitation accelerates the aggregation process, thus necessitating shorter incubation times. Additional efforts are needed to determine the appropriate aggregation times under conditions that have unknown effects.

1.7 Conclusion
There are three essential questions that need to be addressed in protein-ligand binding: binding region, affinity, and order. Various non-MS approaches are now well developed; however, they suffer from either low structural resolution or low sensitivity. Although direct MS can be utilized to study binding affinity, HDX-based approaches can be applied to answer many questions. A standard HDX-MS can report binding regions because it is sensitive to the slower kinetics of the Holo vs. the Apo state. PLIMSTEX/SUPREX experiments can be applied to measure binding affinity. Specifically, by applying PLIMSTEX, one can extract binding affinities for multi-binding systems (i.e., various ligands bind to different domain of the protein), and binding orders if several kinetic curves are performed at various ligand concentration. Furthermore, an adaptation of PLIMSTEX, called SIMSTEX, can be applied for investigating protein association. Even for complex and fast protein aggregation (e.g., for amyloid beta aggregation), a pulsed HDX strategy, which reports aggregation kinetics under various conditions, appears to have high potential.

1.8 Acknowledgements
I acknowledge Don Rempel for help in PLIMSTEX modeling.
1.9 References


Chapter 2: Pulsed Hydrogen/Deuterium Exchange Mass Spectrometry Probes Conformational Changes in Amyloid Beta (Aβ) Aggregation

This work is part of the following publication:

2.1 Abstract
Probing the conformational changes of amyloid beta (Aβ) aggregation is challenging owing to the vast heterogeneity of the resulting soluble aggregates. To investigate the formation of these aggregates in solution, we designed a mass-spectrometry (MS)-based biophysical approach and applied it to the formation of soluble aggregates of the Aβ42 peptide, the proposed causative agent in Alzheimer’s disease (AD). The approach incorporates pulsed hydrogen deuterium exchange (HDX) coupled with MS analysis. The combined approach provides evidence for a self-catalyzed aggregation with a lag phase, as observed previously by fluorescence methods. Unlike those approaches, pulsed HDX does not require modified Aβ42 (e.g., labeling with a fluorophore). Furthermore, the approach reveals that the center region of Aβ42 is first to aggregate, followed by the C- and N-termini. We also found that the lag phase in the aggregation of soluble species is affected by temperature and Cu^{2+} ions. This MS approach has sufficient structural resolution to allow interrogation of Aβ aggregation in physiologically relevant environments. This platform should be generally useful for investigating the aggregation of other amyloid-forming proteins and neurotoxic soluble peptide aggregates.

2.2 Introduction
Protein aggregation is one of the immediate causes of Alzheimer’s disease, Parkinson’s, and Huntington’s disease, motivating biophysical studies of the responsible proteins. More than twenty small proteins undergo amyloidosis in humans. In Alzheimer’s disease (AD), the aggregation of the 40 or 42 amino-acid long amyloid beta (Aβ) peptide, generally called Aβ40 or Aβ42, respectively, is proposed to be involved in the onset of the disease\textsuperscript{1, 2}. Aβ42 is more amyloidogenic and more neurotoxic than Aβ40. Although the amyloid-cascade hypothesis
suggests that the Aβ-containing amyloid plaques are responsible for neurodegeneration\textsuperscript{3-7}, other studies suggest that soluble aggregates of Aβ\textsubscript{42} are more neurotoxic than the amyloid plaques\textsuperscript{8-13}. The amyloid plaques in AD-affected brains contain high levels of copper, zinc and iron\textsuperscript{14-20}. Among these, Cu has drawn the most attention because the Aβ precursor protein (APP) is likely a Cu-chaperone protein\textsuperscript{21}. Several studies of Cu\textsuperscript{2+}-Aβ\textsubscript{40} interactions show that Cu\textsuperscript{2+} can promote Aβ\textsubscript{40} aggregation\textsuperscript{14,18,19}.

The structure of Aβ\textsubscript{42} and its aggregates, although studied extensively, remains of high interest. Studies of amyloid fibrils invoke X-ray crystallography\textsuperscript{22-24}, electron microscopy (EM)\textsuperscript{19,25,26}, and thioflavin T (ThT) fluorescence\textsuperscript{19,27}, revealing the polypeptide’s global behavior, whereas NMR studies provide residue-level information for the fibrils\textsuperscript{28-30}. Nevertheless, we know little about soluble Aβ aggregates owing to their intrinsically high heterogeneity.

MS should offer an opportunity for investigating soluble aggregates of Aβ\textsubscript{42}. Thus far, there are no MS-based, time-dependent studies of the formation of soluble aggregates. Moreover, there are no other biophysical studies of Aβ\textsubscript{42} aggregation at the peptide (regional) level. MS, however, was used for analyzing the aggregated Aβ fibrils\textsuperscript{31-33} and, with ion mobility\textsuperscript{34-36}, for soluble Aβ aggregates. Hydrogen deuterium exchange (HDX)\textsuperscript{37-42}, even with top-down sequencing, can afford residue-level information\textsuperscript{43,44} and provide insight on Aβ\textsubscript{42} fibril core structure\textsuperscript{31,32} and its recycling\textsuperscript{33,45}.

In light of the dearth of aggregation studies at the peptide level, we have employed herein pulsed HDX to study the aggregation of the Aβ\textsubscript{40} and Aβ\textsubscript{42} peptides. Our platform is suitable for confirming the effect of temperature, agitation, and presence of Cu\textsuperscript{2+} ions on Aβ aggregation.
Pulsed HDX, developed by Englander\textsuperscript{46}, can be used for detecting protein folding intermediates\textsuperscript{47} and membrane protein behavior\textsuperscript{48}. Others have also applied pulsed HDX to analyze fibril\textsuperscript{43} and oligomer\textsuperscript{43, 44} Aβ structures. To explain the experimental data, we used a self-catalyzed aggregation model, and its success suggests utility for other amyloid-forming proteins.

2.3 Materials and Methods

2.3.1 Chemicals
All chemicals and reagents were purchased from Sigma Aldrich (St. Louis, MO) unless stated otherwise. All gels, membranes, and reagents for Western blotting were purchased from Invitrogen (Grand Island, NY). D\textsubscript{2}O buffers were prepared by dissolving phosphate buffer (PBS) in D\textsubscript{2}O (Cambridge Isotope Laboratories Inc., Andover, MA).

2.3.2 Protein Preparation
A synthetic, wild-type human form of Aβ\textsubscript{40} (Keck Biotechnology, Yale University) and a recombinant, wild-type human form of Aβ\textsubscript{42} (rPeptide, Bogart, GA) were used. Sample preparation followed a reported procedure\textsuperscript{49}. Both Aβ\textsubscript{40} and Aβ\textsubscript{42} were treated to remove any aggregated species prior to storage by dissolving the material in hexafluoro-isopropyl alcohol (HFIP) at 0.1 mM. The protein solution was aliquoted in low-binding tubes (Eppendorf, Hauppauge, NY), and the solvent was evaporated overnight. The samples were vacuum-centrifuged (Eppendorf, Hauppauge, NY) to remove any existing HFIP, leaving an Aβ film in the tubes. This Aβ film was then stored at -80 °C for future use.

Aβ films were thawed and dissolved in dry dimethyl sulfoxide (DMSO) at 1 mM prior to aggregation. Aβ aggregation studies were initiated by diluting 1:19 monomeric Aβ (either Aβ\textsubscript{40} or Aβ\textsubscript{42}) into PBS buffer (pH 7.4) at 25 °C. Aβ aggregation experiments were investigated at 9
time points for Aβ_{40} and 20 time points for Aβ_{42} ranging from 1 min to 48 h. Soluble aggregates were enriched by centrifugation (16,000 g, 5 min at 4 °C). The centrifuged sample was then carefully divided into two equal-volume parts (i.e., 10 µL each), referred to in all contexts as “top half” and “bottom half”. Only the “bottom half” solution was submitted to MS analysis to obtain the plots shown in the various figures.

To study the effect of various parameters on Aβ_{42} aggregation, a procedure similar to that described above was followed except for those involving Cu^{2+}. CuCl₂ was dissolved in a PBS buffer to give a Cu^{2+} concentration equal to the protein concentration. The stages of aggregation were interrogated at 13 time points by varying the incubation time from 1 min to 48 h. Prior to MS analysis, the sample was submitted to centrifugation under the same conditions as described above, and only the “bottom half” solution was collected.

For a comparison, the process of aggregation was also monitored by varying the temperature (i.e., 37 vs 25 °C) and agitation (i.e., incubator rotating horizontally at 150 rpm). Experiments were investigated for an average of 15 time points, in absence or presence of Cu^{2+}.

Aβ_{42} fibrils were formed following a similar procedure, but incubating at 37 °C for 3 weeks. After a white precipitate formed, the solution was centrifuged, and the pellet was re-dissolved in PBS buffer and analyzed by pulsed HDX.

Native gel electrophoresis and Western blotting were conducted as previously described.

### 2.3.3 Pulsed Hydrogen Deuterium Exchange
HDX experiments were carried out by mixing the incubated protein sample with D₂O buffer in a 1:1 ratio (pulsed labeling for 1 min at 0 °C). We chose a 1:1 dilution over, for example, 1:10, to minimize dilution and consequent shifts in the equilibrium of aggregation/disaggregation.
Importantly, the HDX time was fixed for each experiment, so the term “time” hereafter means aggregation time unless stated otherwise. HDX was then quenched by mixing the exchanging solution with 30 μL 3 M urea and 1% trifluoroacetic acid (TFA) to give a pH of 2.5. The quenched exchanged sample was then digested by passing through an custom-packed pepsin column at 200 μL/min, and peptic peptides were captured on a C8 trap column (2 mm x 1 cm, Agilent Inc., Santa Clara, CA) and desalted (total time for digestion and desalting was 3 min). Peptides were then separated with a linear gradient of 4% - 40% CH3CN, 0.1% formic acid, over 5 min. Both protein digestion and peptide separation were performed at 0 °C to minimize back exchange. The eluted peptides were analyzed by a MaXis quadrupole time-of-flight (Q-TOF) (Bruker Daltonics Inc., Bremen, Germany) in the positive-ion ESI mode. All analyses were done in triplicate.

2.3.4 Aβ40 Back Exchange
We evaluated the extent of back exchange by incubating Aβ40 (same concentration as described above, starting by diluting from stock solution in d6-DMSO) in D2O buffer, with 4.5 M GdnDCl (GdnHCl already exchanged with D2O) to unfold fully the protein. The solution was allowed to stand for 2 days at 25 °C without agitation, based on a published procedure50. After incubation, the sample was treated exactly the same as described above.

2.3.5 Data Analysis and Modeling
Peptic peptides identification was performed as described previously51. The centroid mass of the peptides was converted by MagTran v1.03. The percent protection was calculated by using Equation 2.1,

\[
\text{% protection} = 100\% - \%D = (1 - \frac{m_{\text{HDX}} - m_{\text{control}}}{(N-2) \times 0.5}) \times 100\%
\]

2.1
where \( m_{\text{HDX}} \) is the centroid mass of the deuterated peptides; \( m_{\text{control}} \) was the centroid mass of non-deuterated peptides; \((N - 2)\) is the number of exchangeable amide hydrogens; and 0.5 is the final D\(_2\)O content of the buffer system.

The experimentally determined time-dependent data (mass shift versus incubation time) were characterized phenomenologically by following the recommendation of Finke and coworkers\(^{52}\) to use the simplest model consistent with the data. The modeling was carried out by Mathcad 14.0 M020 (Parametric Technology Corporation, Needham, MA). The least squares fit was implemented with the "Minimize" function in the "Nonlinear Quasi-Newton" mode. In each trial of the minimization process, the postulated normalized rate constants were converted to their equivalent physical values by multiplication with their respective initial physical rate constants.

The physical rate constants specified an ordinary differential equation, which was solved with the adaptive step-size fourth order Runge-Kutta "Rkadapt" function in Mathcad. Each reaction species accounted for a time-varying fraction of the total original monomer molecules in solution. Each fraction was weighted by a postulated percentage of protection for that species. The postulated rate constants and protection weights were changed for each trial of the search to minimize the difference between the data curve and the sum of weighted species fractions. The modified F-W modeling was applied to all experiments except experiments done at 37 °C, with agitation at 150 rpm and in absence of Cu\(^{2+}\), which shows constant increase in the curve and does not need modified F-W modeling.

A bootstrap resampling method\(^{53-55}\) was then used to evaluate the uncertainty of \( t_{1/2} \). We constructed trial data sets by randomly resampling with replacement at each time point. The number of times for resampling equals the number of replicates at that time point. The model
was refitted to each of the 104 trial data sets. Analysis of the accumulated results gave a mean and standard deviation.

2.4 Results and Discussion
Our goal is to implement a general MS approach to monitor Aβ aggregation under different incubation conditions. Both HDX and aggregation are time-dependent phenomena. To minimize aggregation during HDX and thereby separate aggregation and HDX, we used pulsed HDX whereby we allowed Aβ to incubate for various times and then submitted the mixture to a rapid pulsed HDX (1 min) followed by the usual quench procedure. Prior to MS analysis, we centrifuged the sample, split it into two equal parts representing the top and bottom parts, respectively, and analyzed the latter aliquot that is enriched in Aβ oligomers and aggregates.

2.4.1 Comparison of Aβ₄₂ and Aβ₄₀ Aggregation by Native Gel and Western Blotting
After various aggregation times, we analyzed the Aβ₄₀ and Aβ₄₂ samples by native gel and Western blotting to permit an accurate comparison of the aggregated species (Figure 2.1). We found that the top half of the solution after centrifugation showed no detectable aggregates but only Aβ monomer, indicating that most of the monomer is homogeneously distributed, whereas larger soluble species move to the bottom half. We used the Aβ₄₀ peptide as a control, which showed only the presence of Aβ₄₀ monomer in all the “bottom half” samples (Figure 2.1a-c), in agreement with previous work. Aβ₄₂, however, showed a quite different behavior. After only 1 min of incubation, formation of low molecular-weight (LMW) Aβ₄₂ oligomers (i.e., mostly trimer and tetramer) was observed. With longer incubation time, high molecular-weight (HMW) soluble Aβ₄₂ oligomers in the 50-110 kDa range were observed. The amyloid fibrils, which are too large to enter the gel, were observed at the top of gel lanes. All these results are consistent
Figure 2.1 Western blotting for both Aβ_{40} and Aβ_{42}. Aβ_{40} samples are shown in a-c and Aβ_{42} in (d-f). The sample film was dissolved in DMSO to maintain monomeric form (1 mM). The sample was diluted by 20-fold in PBS buffer to initiate aggregation. Aggregation was at 25 °C, no agitation, and in absence of Cu^{2+}. The sample was then centrifuged at 16,000 g for 5 min at 4 °C. Aggregation time points were 1 min (a and d), 24 h (b and e), and 48 h (c and f), respectively.
with previously published work showing that Aβ42 has a greater tendency to form aggregates than Aβ40.34,56.

2.4.2 Pulsed HDX Analysis of Soluble Aβ Species
To conduct a more detailed characterization of various regions of Aβ, we applied pepsin proteolysis following the pulsed HDX treatment described above. As most of the soluble Aβ species were on the bottom half fraction upon centrifugation, we performed all analyses on the bottom half solution. Seven peptides were observed: the N-terminus region (1-19 and 4-19), middle region (20-33, 20-34 and 20-35) and C-terminus region (35-40/42 and 36-40/42). We used three of these peptides (i.e., 1-19, 20-35, and 36-40/42) to analyze the HDX results because these fragments provide full Aβ coverage.

The pulsed HDX approach used to probe the Aβ aggregation states as a function of incubation time is unlike most HDX protocols, in which the protein is incubated in D2O buffer for various times, quenched, and analyzed. Here the HDX time is a short (1 min) “pulse” after various incubation times. This pulse time is sufficient to allow the maximum exchange for this largely intrinsically disordered protein, yet is negligible when compared to the incubation times. We view this pulsed HDX as a “recording” tool to monitor aggregation without competing with it, allowing us to separate experimentally the aggregation from HDX.

Initially, we followed the aggregation of Aβ40 at 25 °C by pulsed HDX as a control experiment (Figure 2.2). All three peptic peptides monitored show a constant HDX extent as a function of incubation time. This indicates that no detectable conformational changes or self-association occurred from 1 min to 48 h at 25 °C. The apparent protection is likely due to two factors: (1) the exchange time is sufficiently short for the pulse (1 min) that HDX is not complete and (2) there
Figure 2.2 Pulsed HDX results for three peptides from Aβ40. Aggregation was done at 25 °C and in the absence of Cu²⁺. Peptic peptides 1-19 (a), 20-35 (b), and 36-40 (c) are represented by green triangles, orange circles, and blue diamonds, respectively. Lines were fit by using “linear fit” in OriginPro 8.5.
is significant back exchange for the highly disordered Aβ_{40}. To evaluate the back exchange, after quenching the “HDX” for the fully deuterated Aβ_{40}, the solution was subjected to online-pepsin digestion, column trapping and eluting. During these processes, all the solvents are H\textsubscript{2}O based, which means the protein undergoes back exchange. We found only ~60% of deuterium uptake (an average based on the peptide level). This result shows that there is ~40% back exchange for this highly disorder protein on our HDX platform. Thus, the extent of HDX for Aβ_{40} in the pulsed HDX experiment is consistent with not only our native gel and Western blotting experiments, but also with the common perception that Aβ_{40} forms fewer aggregates than Aβ_{42}^{57, 58}\textsuperscript{.}

By contrast, HDX of Aβ_{42} showed an appreciable increase in protection by a modified sigmoidal behavior (Figure 2.4, solid line, raw HDX data shown in Figure 2.3a-d). The first stage is rapid and exponential-like, showing a rapid increase in protection. A first plateau follows wherein no significant change occurs in protection. A second increase in protection appears as a sigmoidal upward break in the curve, followed by a second plateau, indicating that the system reaches equilibrium among all the species. The HMW species correspond to Aβ_{42} soluble aggregates that concentrate upon centrifugation to the bottom half solution. Importantly, these results are consistent with native gel and Western blotting that show formation of HMW Aβ_{42} soluble species on the same timescale as the pulse HDX experiments (Figure 2.4, right panel, lane 1).

We attribute the first rapid increase in the protection level to formation of small Aβ_{42} oligomers (dimer, trimer, and tetramer, etc), in line with previous work\textsuperscript{34, 59}. The plateau that follows indicates no additional significant hydrogen-bond formation or conformational change. During this stage, larger oligomers are forming with little increase in protection as Aβ_{42} molecules adds.
No HDX control

1 min

1 h

5 h

48 h

m/z
Figure 2.3 Raw data for HDX experiments shown only for the peptide 1-19 (+4 charge). Experimental conditions: a) 25 °C, no agitation, in the absence of Cu$^{2+}$; b) 25 °C, no agitation, in the presence of Cu$^{2+}$; c) 37 °C, 150 rpm agitation, in the absence of Cu$^{2+}$; d) 37 °C, 150 rpm agitation, in the presence of Cu$^{2+}$. The black traces in each panel are the spectra showing isotopic distribution with no deuterium uptake, while the rest of the spectra show isotopic distribution with deuterium uptake at different incubation times.
Figure 2.4 Pulsed HDX and Western blotting results for three peptic peptides from Aβ_{42}. Aggregation was done at 25 °C, and in the absence (solid line) or presence of Cu^{2+} (dashed line). Peptic peptides 1-19 (a), 20-35 (b), and 36-42 (c) are represented by green triangles, orange circles, and blue diamonds, respectively. Straight lines were fit by using “linear fit” in OriginPro 8.5. Curve lines were fit by using modified F-W modeling in MathCAD. Native gel and Western blotting of the aggregated Aβ_{42} samples (incubated at 25 °C for 48 h), in the absence (lane 1) or presence of Cu^{2+} (lane 2).
to the large soluble aggregates. After some time, the oligomers do reorganize, in an autocatalytic fashion, to a structure showing higher protection. The upward-breaking sigmoidal curve represents this reorganization and is consistent with the behavior of other amyloid-like proteins (i.e., they undergo nucleation, growth, and stabilization)\textsuperscript{52}. Whereas the rate of nucleation is slow, giving nearly constant HDX, the growth rate is much faster, giving the sigmoidal behavior of HDX. Similar sigmoidal behavior was reported recently for A\textsubscript{β\textsubscript{42}} aggregation followed by a fluorescent tetramethylrhodamine (TMR) labeled A\textsubscript{β\textsubscript{42}} derivative\textsuperscript{59}. Our approach has an advantage that it can be employed directly without the need for adding amino acids or fluorophores to the protein.

### 2.4.3 Application of Finke-Watzky (F-W) Modeling and Bootstrap Strategy

Amyloid-like proteins usually undergo a self-catalyzed aggregation process, which involves one nucleation step and one growth step. Although there are many models to describe a slow, continuous nucleation, we chose the simplest one to fit the time-dependent aggregation of A\textsubscript{β\textsubscript{42}} (Figure 2.4), following the recommendation of Finke and coworkers\textsuperscript{52}. We modified the F-W model by adding a “dimerization” process (Equation 2.2). The model describes a reaction that proceeds from a monomer A\textsubscript{β} in state “A” to a “dimer” in state "B" (Equations 2.2-2.4). Admittedly, state “B”, consisting of a heterogeneous mix of small oligomers, is more complicated than can be addressed in this modified F-W model. Nevertheless, considering it as “dimer”, we subsequently allowed its transformation to state "C" by nucleation of large, soluble oligomers, heterogeneous in structure and number. The model includes a forward rate constant $k_1$ of oligomerization, a rate constant $k_1$ for the reverse reaction, a forward nucleation rate constant $k_2$ from "B" to "C", and a forward self-catalysis rate constant $k_3$ involving "B" and "C". A useful
parameter that can allow for a direct comparison of different aggregation curves is $t_{1/2}$, which is the time point at which the number of Aβ molecules in state "C" equals to the remaining Aβ molecules.

$$A + A \xrightarrow{k_1/k_{-1}} B \quad 2.2$$

$$B \xrightarrow{k_2} C \quad 2.3$$

$$B + C \xrightarrow{k_3} 2C \quad 2.4$$

In fitting the data to the F-W model, we sought a single parameter to characterize the outcome. The $t_{1/2}$ value (the time to reach one-half of the step height in the sigmoidal plot) for the transitions is likely that parameter, and it could be obtained by averaging the $t_{1/2}$ values from three determinations. However, to obtain a better measure and utilize the full statistical value of the data, we evaluated the statistics by using a bootstrap resampling strategy. This gives a more reliable $t_{1/2}$ along with its precision, permitting the use of a simple t-test for establishing differences. By applying the bootstrap strategy to the F-W model\textsuperscript{53-55} (Figures 2.5 and 2.6), we retrieved not only $t_{1/2}$ values, but also rate-constant information and precision. Although there are many families of $k_{-1}, k_1, k_2, \text{and } k_3$ values that fit the experimental data, they are highly correlated (i.e., an increase in $k_2$ and a commensurate decrease in $k_3$ gives a good fit to the model). There is, however, a nearly constant $t_{1/2}$ for all the fits for one peptide. More importantly, the $t_{1/2}$ values corresponding to the three peptides are different. The middle region (i.e., 20-35) has the shortest $t_{1/2}$ ($1070 \pm 30$ min), whereas C-terminus (i.e., 36-42) and N-terminus (i.e., 1-19) regions have longer $t_{1/2}$ values ($1230 \pm 30$ min and $1420 \pm 20$ min, respectively, Figure 2.7). That these
Figure 2.5 A flow chart for the modified F-W modeling and bootstrap strategy.
Figure 2.6 A zoom in view for the outcome of application of the bootstrap strategy. Three different lines (overlapping in three different colors) were randomly selected from 104 fitting results following fitting by the modified F-W modeling on peptide 1-19 (+4 charge). Experimental conditions: 25 °C, no agitation, in absence of Cu^2+. 
Figure 2.7 t_{1/2} distribution from bootstrap analysis for different peptide. Experimental condition: 25 °C, no agitation, in absence of Cu^{2+}. The frequency of each peptide was calculated using the “histogram” function in Excel data analysis. Bin range is from 1000 to 1500, using 10 as increase interval.
differences among the three peptides are statistically significant is supported by the t-test at ≥ 99.99% confidence level, made possible by employing the bootstrap strategy. Although it may be surprising that the outcome treats the three peptides as separate entities, we note that proteins are complex species and give, for example, different HDX rate constants for different regions. These results now establish that the middle region (i.e., residues 20-35) is the first to report entering an organized state. This is consistent with most molecular dynamics simulations that point to the middle region as either a folding nucleus or a dimerization interface. The hinge region identified in the solid-state NMR results for Aβ₄₂ fibrils is from S26 to I31, which lies in the middle of the region we identified here. This smaller region may also be important in initiating aggregation, as suggested by aggregation studies of myriad mutation forms. For example, Flemish (A21G), Arctic (E22G), Dutch (E22Q), Iowa (D23N) and others exhibit different aggregation behavior compared to the wild-type Aβ₄₂ (WT-Aβ₄₂), supporting that the middle region of Aβ is important in the aggregation process.

The C-terminus (residues 36-42) aggregates second, confirming that it plays a smaller but still important role in aggregation. All the residues in this region (i.e., VGGVVIA) are hydrophobic, serving as an interface for Aβ₄₂ aggregation. Besides, the only differences between Aβ₄₀ and Aβ₄₂ are the last two residues of this region (i.e., IA) of Aβ₄₂. The strong differences in aggregation between Aβ₄₀ and Aβ₄₂ hint at the importance of the hydrophobic interaction within this region. As a result, we propose that without the last two residues, the hydrophobic interaction would be insufficient to “pull” small aggregates together and form larger aggregates.

The N-terminus (residues 1-19) aggregates last, consistent with its hydrophilic nature. Unlike other approaches showing that the N-terminus is flexible, our data show some ordering in this
region, and as expected, it is the last involved in the aggregation process. The observation that mice never develop AD, however, may be consistent with the importance of the N-terminus, as the three mutations between mouse vs human Aβ represent amino-acid substitutions in this region (i.e., R5G, Y10F, and H13R). Nevertheless, the N-terminus is involved in aggregation, and this might be the cause for the structural difference between soluble Aβ aggregates and insoluble Aβ fibrils. In addition, the structural difference may be relevant to the different levels of neurotoxicity exhibited by soluble Aβ aggregates vs insoluble fibrils. Thus, these pulsed HDX studies should have an important impact in the design of drugs that can alter the aggregation of Aβ species and ultimately their neurotoxicity.

### 2.4.4 Pulsed HDX Analysis of Aβ₄₂ Fibrils

Even though the LC-MS/MS identification showed full coverage for the peptides, there was some un-digested Aβ₄₂ observed as monomer mass in the above experiments; and its corresponding relative peak intensity increased with increasing incubation time. Our online pepsin digestion has been tested in many other projects⁶⁸, ⁶⁹, and the digestion efficiency approaches 100% for proteins, most of which are larger than Aβ₄₂. The Aβ₄₂ peptide showed the lowest digestion efficiency, suggesting presence of amyloid fibrils or other high MW aggregates that are slow to proteolyze. The protection levels of these undigested species (Figure 2.8a) are higher than that of Aβ₄₂ species that can undergo pepsin proteolysis. Given that we observed no significant peaks corresponding to the monomeric mass of Aβ₄₀ in its aggregation experiments, the result for Aβ₄₂ is further evidence that the amyloid fibrils are the source of peaks representing undigested Aβ₄₂.
Figure 2.8 Pulsed HDX results for undigested species from Aβ42. Experimental conditions: a) 25 °C, no agitation, in the absence of Cu²⁺; b) 25 °C, no agitation, in the presence of Cu²⁺; c) 37 °C, 150 rpm agitation, in the absence of Cu²⁺; d) 37 °C, 150 rpm agitation, in the presence of Cu²⁺.
To study directly the amyloid fibrils, we applied pulsed HDX to the pre-formed Aβ42 fibrils, which represent the final stage of Aβ aggregation. Digestion of amyloid fibrils is difficult\textsuperscript{31, 70} and usually requires the use of “strong” chemicals including hexafluoroisopropanol (HFIP) or DMSO to dissociate all aggregates into Aβ42 monomer\textsuperscript{49}. This time-consuming and MS-unfriendly digestion protocol is not suitable for an HDX platform. We tested the effectiveness of the HDX quenching procedure (3 M urea with 1% TFA) by using TEM, and found the protocol was able to dissociate to some extent the amyloid fibrils, allowing us to measure the HDX of the fibrils (Figure 2.9).

Upon pulsed HDX to the pre-formed fibrils we observed both the peptides formed by pepsin digestion (i.e., 1-19, 20-35, and 36-42) and the undigested Aβ42. The peptides and the full Aβ42 are similarly protected (89 ± 1% for the undigested Aβ42, 85 ± 1%, 84 ± 1%, and 91 ± 2% for 1-19, 20-35 and 36-42, respectively), consistent with the peptides being proteolytic fragments from the amyloid fibrils. This conclusion is consistent with a study that shows the recycling mechanism of Aβ42 fibrils with the Aβ42 monomer\textsuperscript{45}. The monomer thus formed carries information of the “imprinted” fibrils because quenching HDX preserves information prior to MS analysis. These monomers digest well with pepsin owing to a lack of the highly packed structure characteristic of the amyloid fibrils. Thus, the peptide fragments are useful “fibril reporters” because they contain the same protection as their precursor fibrils. These amyloid fibrils, which become ultimately the main component in the incubation, are considerably more difficult to interrogate because they have a highly compact structure. Nevertheless, the highly organic mobile phase we used is sufficient to denature some of the fibrils and release detectable amounts of Aβ42 monomer\textsuperscript{31}. The protection level of undigested Aβ42 seen in the previously
Figure 2.9 TEM images for effect of quenching solution on \( \text{A}\beta_{42} \) fibrils. (a) \( \text{A}\beta_{42} \) fibrils; (b) \( \text{A}\beta_{42} \) fibrils with 3 M urea; (c) \( \text{A}\beta_{42} \) fibrils with 3 M urea and 1% TFA (bar scale is 500 nm).
mentioned experiments (Figure 2.8a) is similar to that of the pre-formed fibrils, both digested and undigested, indicating that the full Aβ42 peptide originates from the Aβ42 fibrils whereas the peptides arise from soluble species.

To rule out a significant contribution of undigested Aβ42 to our time-dependent studies described earlier, we measured directly the digestion of amyloid fibrils by examining the relative peak-area ratios in extracted ion chromatograms (EIC) of the peptide fragments and the undigested Aβ42. We used the N-terminal region (i.e., 1-19) as a measure of fibril digestion, and the ratio of its EIC peak area compared to that of the undigested Aβ42 is 0.078, indicating that only a small fraction of fibrils undergo digestion. Thus, the contribution of fibril digestion to the peptide-fragment signals is likely small because the peptide 1-19 signal area in the EIC was considerably greater than the EIC area corresponding to the undigested protein (see below). Because HDX reports an ensemble average, the protection levels seen in the digested peptides are a measure of soluble Aβ42 aggregates, rather than of the insoluble Aβ species.

2.4.5 Pulsed HDX Analysis of Factors Affecting Aβ42 Aggregation
One motivation for the HDX pulsed platform is to develop a tool to investigate the many factors that affect Aβ42 oligomerization and aggregation (and of other amyloid-forming proteins). Such a platform should rapidly provide insight on those factors that affect Aβ42 aggregation. Shifts in or losses of the sigmoidal behavior are a measure of the effect. One factor of high interest to us is that of Cu2+, which has been previously implicated in Aβ42 aggregation14, 18, 19. Thus, we added Cu2+ to the PBS buffer before mixing with Aβ42 such that the final stoichiometry of Cu2+ and Aβ42 is 1:1 (molar ratio). The pulsed HDX results of Aβ42 incubation in absence of Cu2+ provide the reference point for this experiment. The protection level of Aβ42 incubated with Cu2+ is
remarkably linear and unchanged for the three peptic peptides (i.e., 1-19, 20-35, and 36-42, Figure 2.4, dashed lines, raw data shown in Figure 2.3b), in contrast to their sigmoidal behavior in the absence of Cu\(^{2+}\) (Figure 2.4, solid lines, raw data shown in Figure 2.3a). The plateau stage of protection in the presence of Cu\(^{2+}\) is similar to that represented by the first plateau for 1-19 in the absence of Cu\(^{2+}\) (55%). The protection is lower than that of both 20-35 and 36-42 (56% vs 60% and 60% vs 75%, respectively).

We hypothesize that the differences of protection in the presence and absence of Cu\(^{2+}\) is that Cu\(^{2+}\) stabilizes different forms of Aβ aggregates by interacting with them. The coordination sites for Cu\(^{2+}\) are likely on the N-terminus (i.e., H6, H13, and H14)\(^{14, 17}\), and coordination to these sites would lead to more protection of that region of Aβ\(_{42}\). By contrast, the C-terminus, which can principally interact via hydrophobic forces, may be diminished to compensate for the Cu\(^{2+}\) coordination. In addition, native gel and Western blotting of the samples corresponding to the largest difference on the pulsed HDX curve show much less high MW soluble aggregates (50-110 kDa) and more low MW aggregates (i.e., mainly trimer and tetramer) in the presence vs absence of Cu\(^{2+}\) (Figure 2.4, right panel). This result reinforces our previous report showing that Cu\(^{2+}\) stabilizes the soluble Aβ\(_{42}\) aggregates and inhibits formation of amyloid fibrils\(^{14, 19}\).

We also tested the ability of our pulsed HDX approach under rapidly aggregating conditions such as higher temperature and with agitation. We chose a higher and more physiologically relevant temperature (37 °C) and agitated the solution (at 150 rpm, Figure 2.10, solid line, raw data shown in Figure 2.3c). Compared to the results at 25 °C in absence of Cu\(^{2+}\) and without agitation, the aggregation behavior speeded up considerably. The lag phase disappears, indicating that the transformation from low MW aggregates to high MW aggregates is expedited.
Figure 2.10 Pulsed HDX results for three peptides from Aβ_{42}. Aggregation was done at 37 °C, with agitation at 150 rpm and in absence (solid line) or presence of Cu^{2+} (dashed line). Peptic peptides 1-19 (a), 20-35 (b), and 36-42 (c) are represented by green triangles, orange circles, and blue diamonds, respectively. Solid lines were fit by using modified FW modeling while dashed lines by simple F-W modeling in MathCAD.
Next, we repeated the experiment at higher temperature and subjected the solution to agitation, as well as added Cu\textsuperscript{2+} to the solution. According to the results just discussed, applying higher temperature and agitation should decrease the lag phase of Aβ\textsubscript{42} aggregation, whereas Cu\textsuperscript{2+} should slow down the aggregation. Indeed, the results (Figure 2.10, dashed line, raw data shown in Figure 2.3d) indicate that the Aβ\textsubscript{42} aggregation is faster than that at 25 °C, yet a lag phase is observed in presence of Cu\textsuperscript{2+}. This agrees with the hypothesis that higher temperature and agitation accelerate the aggregation. On the other hand, the results showed slower kinetics than that for Aβ\textsubscript{42} aggregation at 37 °C, 150 rpm, in absence of Cu\textsuperscript{2+}, providing evidence that Cu\textsuperscript{2+} slows down the Aβ\textsubscript{42} aggregation, even at 37 °C, yet without completely preventing it.

As we observed in experiments of Aβ\textsubscript{42} incubated at 25 °C, the peaks corresponding to the Aβ\textsubscript{42} monomer (undigested species) were observed in all experiments mentioned in this section. The extent of protection of the corresponding undigested species is higher than observed for peptide fragments in the same experiments, and similar to those of both digested and undigested pre-formed fibrils (Figure 2.8b-d).

The ratios of peak intensities corresponding to Aβ\textsubscript{42} that can be digested relative to that resistant to digestion now range from 17 to 1.1. After longer incubation times, more and more Aβ\textsubscript{42} becomes resistant to digestion. These ratios are 210 to 14 times higher than the ratio calculated from pre-formed fibrils described above (i.e., 0.078), indicating that only a very small amount of the fibrils was digested and thus is not expected to interfere to a large extent with the protection level calculated from these experiments. Importantly, this result indicates that our platform reports only on the soluble Aβ species and not insoluble aggregates.
2.5 Conclusion
A new implementation of the pulsed HDX technique interfaced with LC/MS provides an opportunity to examine the details of aggregation of the Aβ peptide at various stages. By using this method, we successfully decoupled the aggregation from the HDX process. Importantly, we extracted kinetic information on the Aβ42 aggregation at 25 °C, indicating that the middle region of the Aβ42 peptide (i.e., 20-35) was the “seeding” region in aggregation, followed by the C-terminus hydrophobic region (i.e., 36-42) and then the N-terminus hydrophilic region (i.e., 1-19). Finally, we showed that this approach allowed us to examine directly the factors that affect the oligomerization of Aβ42. For example, at 37 °C and with agitation, Aβ42 aggregated faster than at 25 °C. On the contrary, the presence of Cu²⁺ slowed down the Aβ42 aggregation, presumably by complexing the polypeptide in the N-terminal region and stabilizing the soluble Aβ42 species.

We envision this approach as a general tool to support future efforts to measure rates of Aβ42 oligomerization and aggregation as a function of various parameters (e.g., concentration, presence of different ligands or proteins, and pH). We will discuss in detail on effect of ligand binding in Chapter 3. We also see the need for more effort in using MS site-specific ion activation (i.e., electron-transfer dissociation, ETD) to obtain information at the amino-acid level and probe in more detail the aggregation interface and the Cu²⁺ binding sites. Alternative footprinting approaches including fast photochemical oxidation may also offer complementary views to understand amyloid formation. Moreover, the method described herein is applicable to the study of various experimental conditions on the oligomerization and aggregation of other amyloid-forming proteins involved in neurodegenerative diseases.
2.6 Acknowledgments
I acknowledge Dr. Anuj Sharma for performing native gel and TEM experiments, and Don Rempel for performing modeling.
2.7 References


Chapter 3: Structural Analysis of Amyloid Beta Peptides Interacting with Different Ligands by Pulsed Hydrogen Deuterium Exchange Mass Spectrometry
3.1 Abstract
The study of amyloid β (Aβ) peptide aggregation in Alzheimer’s Disease (AD) has been hampered by the high heterogeneity of soluble Aβ aggregates. To extend the ideas of Chapter 2, we further applied the pulsed HDX platform to investigate the effect of different compounds on the Aβ42 aggregation process. Three ligands are included in this study, L1, L2, and amentoflavone (AMF). The pulsed HDX technique reveals that the three ligands interfere with Aβ42 aggregation process differently. L1 and L2 stabilize small oligomers when Cu²⁺ is absent. When Cu²⁺ is present, L2 accelerates the structural reorganization and formation of large Aβ42 oligomers more than does L1. On the contrary, when AMF interacts with Aβ42, the presence of Cu²⁺ does not have a significant impact on the Aβ42 aggregation behavior. Overall, these studies demonstrate that the pulsed HDX platform described in the previous chapter is applicable to study the effect of different conditions and additives that affect amyloid protein aggregation.

3.2 Introduction
There are more than twenty small proteins that undergo amyloidosis in humans. Such protein aggregation events play a role in neurodegenerative diseases including Alzheimer’s, Parkinson’s and Huntington’s, and has gained the attention of many researchers in biochemistry and biophysics. In general, only one protein is responsible for each type of amyloidosis; for example, amyloid beta (Aβ) peptide aggregation is involved in Alzheimer’s disease (AD). There are two main forms of the Aβ peptide, Aβ40 and Aβ42, which are 40 and 42 amino acids long, respectively. Aβ40 is dominant, whereas Aβ42 is more neurotoxic and amyloidogenic. Previous studies suggest that mature Aβ fibrils or Aβ plaques are responsible for memory loss and neuro
degeneration\textsuperscript{3-8}. Recent in vitro and in vivo studies, however, suggest that the soluble aggregates of Aβ\textsubscript{42} are more neurotoxic than the amyloid plaques\textsuperscript{9,14}.

The AD-affected brain contains remarkably high levels of metal ions including copper (Cu), zinc (Zn), and iron (Fe)\textsuperscript{15-20}. Among these metal ions, Cu has drawn the most attention because it has a putative role of the amyloid-beta precursor protein (APP) as a Cu chaperone protein\textsuperscript{21}. The Cu\textsuperscript{2+}-Aβ interaction can promote formation of small Aβ\textsubscript{42} oligomers\textsuperscript{19,20,22}.

Recent efforts to control Aβ aggregation via metal ion-Aβ interactions were focused on small molecules (e.g., bifunctional chelators (BFCs) that bind to both metal ions and Aβ peptides\textsuperscript{20,23-28}). Most of these studies investigated the interaction of Aβ\textsubscript{40} with BFCs. That interaction causes significantly different aggregation pathways to be followed than those of the more pathological-relevant Aβ\textsubscript{42}. In addition, natural compounds, for example, amentoflavone (AMF) and epigallocatechin gallate (EGCG), have drawn serious attention recently as affecting the occurrence of AD\textsuperscript{20,30}.

The structures of Aβ\textsubscript{42} and its aggregates have been studied extensively. Electron microscopy (EM)\textsuperscript{20,31,32} and thioflavin-T (ThT) fluorescence\textsuperscript{20,33} can be employed to investigate the global aggregation behavior, whereas X-ray crystallography\textsuperscript{34-36} and solid state NMR\textsuperscript{37-39} studies provide residue-level information on the amyloid fibrils. Not much is known about the structure of the soluble Aβ aggregates, however, owing to their low concentration and vast heterogeneity.

Most mass spectrometry (MS) measurements of Aβ have been performed on the complete aggregated stage (Aβ fibrils)\textsuperscript{40-42}. We are motivated to study soluble Aβ aggregates by MS methods as they require very little sample and can afford both sensitivity and resolution. Hydrogen deuterium exchange (HDX)-MS can afford information on hydrogen-bond formation and conformational changes due to protein/ligand or protein/protein interaction\textsuperscript{43-48}. MS has
already been utilized in determining Aβ42 fibril core structure\textsuperscript{40, 41} and recycling mechanism\textsuperscript{42, 49}. Studies on the Aβ oligomers are scarce, surprising because residue-level structural information of aggregates can be achieved by using both HDX and top down sequencing\textsuperscript{50, 51}. Pulsed HDX, in which the deuterium labeling time is constant, was first developed by Englander\textsuperscript{52} and used by Konermann\textsuperscript{53} for detecting protein-folding intermediates and membrane-protein behavior. Our groups have used this approach to elucidate the aggregation kinetics of Aβ42 under different conditions, including the presence of Cu\textsuperscript{2+} and different temperature\textsuperscript{54}. Other MS studies have utilized ion mobility MS (IM-MS)\textsuperscript{55-57} and provided size information of soluble Aβ42 oligomers, which cannot be achieved by HDX. Herein, we describe the pulsed HDX platform and its applicability to the aggregation behavior of Aβ42 with different metal-binding BFCs, in the presence and absence of Cu\textsuperscript{2+}. We were able to probe the structural interference of different ligands on Aβ42 aggregation behavior at the peptide level. We observed different effects from the three ligands tested, further validating the sensitivity of the platform. Overall, these studies demonstrate that the pulsed HDX platform that we described in the previous chapter is applicable to different conditions and additives that affect amyloid protein aggregation.

3.3 \textbf{Materials and Methods}

3.3.1 \textbf{Chemicals and Protein Preparation}
All of this information is covered in the previous chapter (Chapter 2).

3.3.2 \textbf{Pulsed HDX}
The pulsed HDX was as described in the previous chapter (Chapter 2). In brief, Aβ films were thawed and dissolved in dry dimethyl sulfoxide (DMSO) at 1 mM before aggregation. Aβ42

- 88 -
aggregation studies were initiated by diluting 1:19 (vol/vol) monomeric Aβ42 into PBS buffer (pH 7.4) at 37 °C with agitation (i.e., incubator rotating horizontally at 150 rpm). Aggregation was allowed to occur over various incubation times ranging from 1 min to 48 h (13 time points) under various experimental conditions (i.e., complexing with different ligands). Soluble aggregates were enriched by centrifugation (16,000 × g, 5 min at 4 °C). The centrifuged sample was then carefully divided into two equal-volume parts (i.e., 10 μL each). Only the lower half of the sample was submitted to MS analysis to obtain the plots shown in various figures, following the general approach described in the previous chapter.

To study the effect of different ligands on Aβ42 aggregation, a similar procedure was followed except that the protein was incubated in the presence of ligand (i.e., L1, L2, and AMF). For experiments in the presence of ligand and the absence of Cu^{2+}, Aβ42 was co-incubated with each one of the ligands (pre-diluted in PBS buffer) in a 1:1 ratio. For the experiments in the presence of both ligand and Cu^{2+}, Aβ42 was incubated with the ligand and Cu^{2+} in 1:1:1 ratio.

### 3.3.3 LC/MS, Data Analysis, and Finke-Watzky (F-W) modeling
All of this information is covered in the previous chapter (Chapter 2).

### 3.4 Results and Discussion
We chose three ligands to study their effect on Aβ42 aggregation and the formation of soluble Aβ42 oligomers (Figure 3.1). Both L1 and L2 are metal-binding BFCs, and AMF is a natural product that is also expected to interact with metal ions. L1 contains a 2-phenylbenzothiazole/vanillin group for Aβ binding and a 1,4-dimethyl-1,4,7-triazacyclononane molecular fragment for metal chelation, whereas L2 has metal-binding donor atoms integrated within a molecular framework modeled on the amyloid-binding fluorescent dye thioflavin T (ThT)²⁷.
Figure 3.1 Molecular structures of the BFCs L1 and L2 and amentoflavone (AMF) described in this work. For L1 and L2, the Aβ binding frame and metal binding parts are highlighted in red and blue, respectively.
The pulsed HDX approach employed in this study probes the samples as a function of incubation time and, unlike other typical HDX protocols in which the protein is incubated in D$_2$O buffer for various times, quenched, and analyzed. Here the HDX time is a short “pulse” (1 min) taking place after various incubation times. The pulse time is sufficient to allow the maximum exchange for the largely intrinsically disordered Aβ peptide, yet is negligible when compared to the incubation times. We view this pulsed HDX as a “recording” tool to monitor aggregation without competing with it, thus allowing us to separate the effects of aggregation from those of HDX.

To investigate Aβ$_{42}$ aggregation at higher structural resolution, we applied pepsin digestion following pulsed HDX labeling and preceding mass spectrometry analysis, so that all the HDX information is retained, yet available at the peptide level. As most of the soluble Aβ$_{42}$ aggregates were on the bottom half fraction upon centrifugation (see Materials and Methods sections), we performed all analyses following the reported protocol$^{54}$ on the bottom half of the solution. Seven peptides were observed: the N-terminus region (1-19 and 4-19), middle region (20-33, 20-34 and 20-35) and the C-terminus region (35-42 and 36-42). We used three of these peptides (i.e., 1-19, 20-35, and 36-42) to analyze the HDX results because these fragments provide full coverage. We already observed that Aβ$_{42}$ aggregation has three stages from Chapter 2: formation of small Aβ oligomers (~40% protection) from monomeric Aβ, formation of larger Aβ oligomers to afford ~50-60% protection, and a structural reorganization leading to fibrillar species (~80% or more protection)$^{54}$.

### 3.4.1 Pulsed HDX for L1

Addition of L1 to an Aβ$_{42}$ solution causes the Aβ$_{42}$ aggregation kinetics to change significantly (Figure 3.2). For example, all three regions of the protein, as seen in the peptic peptides, show significant longer lag phases in the presence of L1 and absence of Cu$^{2+}$ (Figures 3.2A, 3.2B, and
Figure 3.2 Pulsed HDX results for three peptic peptides from Aβ42 interacting with L1. Aggregation was done at 37°C, with agitation at 150 rpm, in presence of L1, either in the absence (upper panel) or presence of Cu^{2+} (lower panel). Peptic peptides 1-19 are represented in A and D, 20-35 in B and E, and 36-42 in C and F, respectively. All lines are fits using a modified F-W model in MathCAD.
3.2C) compared to the previously reported results in the absence of both L1 and Cu$^{2+}$ (see Chapter 2). This result helps establish the sensitivity of the pulsed HDX platform on reporting the Aβ$_{42}$ aggregation behavior in presence of small organic compound and other additives. In addition, the presence of an extended lag phase indicates that L1 interferes with Aβ$_{42}$ aggregation by stabilizing small, soluble Aβ$_{42}$ oligomers that show a low protection level (i.e., 40% to 50%). The $t_{1/2}$ values from all peptides (~250 min), however, are similar to those when comparing between the two conditions, indicating that L1 can also accelerate the structural reorganization to form large aggregates with a high protection level (i.e., ~80%).

When both L1 and Cu$^{2+}$ are present (Figures 3.2D, 3.2E, and 3.2F), the Aβ$_{42}$ aggregation kinetic curves are similar compared to the condition where only L1 is present (Figures 3.2A, 3.2B, and 3.2C). The only difference is a slightly longer lag phase in the presence of Cu$^{2+}$ (~300 vs. ~250 min); this is consistent with the hypothesis that Cu$^{2+}$ decelerates the aggregation process by stabilizing small oligomers$^{20, 27, 54}$. Comparing the results obtained for the presence of L1 and Cu$^{2+}$ (Figures 3.2D, 3.2E, and 3.2F) compared to those with only Cu$^{2+}$ (Chapter 2), we observe that the lag phase is significantly shorter (~300 vs. ~700 min). Whereas the presence of L1 alone increases the lag phase by interacting with Aβ$_{42}$, L1 appears to interact also with Cu$^{2+}$ (through the metal-binding portion of L1 highlighted in blue in Figure 3.1) and interfere with the Cu$^{2+}$ coordination to Aβ$_{42}$, thus leading to a different Aβ aggregation structure. This structure is likely a small oligomer, because it exhibits a lower protection level (~45%, see Figures 3.2D, 3.2E, and 3.2F) compared to the larger oligomers formed in the absence of L1 (~60%-70% Chapter 2). These small oligomers are slightly more favored vs. the larger oligomers when L1 is present, thus leading to the shorter lag phase compared to that obtained in the absence of L1.
Figure 3.3 Pulsed HDX results for three peptic peptides from Aβ42 interacting with L2. Aggregation was done at 37 °C, with agitation at 150 rpm, in presence of L2, either in the absence (upper panel) or presence of Cu²⁺ (lower panel). Peptic peptides 1-19 are represented in A and D, 20-35 in B and E, and 36-42 in C and F, respectively. All lines are fits using an F-W model in MathCAD.
3.4.2 Pulsed HDX for L2
The Aβ₄₂ binding part of L2 is similar to that of L1. Thus, we expect to see L2 exert a similar effect on the Aβ₄₂ aggregation in the absence of metal ions. Indeed, when L2 is added to an Aβ₄₂ solution, significantly longer lag phases for aggregation are observed when compared to the absence of any ligand (Figures 3.3A, 3.3B, and 3.3C vs. Chapter 2) in all regions. The $t_{1/2}$ values of the peptides shown in the presence of L2 are longer than under the other conditions of no ligands and in the presence of L1 (~500 vs. ~250 min), most likely due to the presence of the extended lag phase. During this lag time, Aβ₄₂ stays mostly as small oligomers corresponding to the section of the curve showing protection levels of ~40%. This result thus indicates that L2 can stabilize small Aβ₄₂ oligomers, similar to L1.

By comparison, the structural reorganization of the large aggregates corresponding to the rapid increase in the protection level from 40% to 80% is accelerated in the presence vs. the absence of L2, similar to L1. We suggest that in the presence of ligands (L1 or L2), the small Aβ₄₂ oligomers are slightly more stabilized, and this leads to a slower reorganization to form large aggregates.

The metal-binding fragment of L2, however, is significantly different than that of L1, leading to different Aβ₄₂ aggregation kinetic curves in presence of Cu$^{2+}$. Unlike L1, no lag phase is observed for Aβ₄₂ aggregation in the presence of L2 and Cu$^{2+}$ (Figures 3.3D, 3.3E, and 3.3F). This result further indicates that the BFCs bind to both Aβ₄₂ and Cu$^{2+}$. In this case, L2 does not stabilize the small Aβ₄₂ oligomers vs. the larger aggregates, affording significantly smaller $t_{1/2}$ values for all regions than those in the absence of L2 and presence of Cu$^{2+}$ (~100 vs. ~700 min)$^{54}$. 

- 95 -
3.4.3 Pulsed HDX for AMF
AMF is a natural product and does not have any structural similarity to L1 or L2, and we expected that the aggregation kinetic curves should reflect this feature. Indeed, the curves in all regions show a rapid increase and a leveling in protection level in the presence of AMF alone (Figures 3.4A, 3.4B, and 3.4C). Unlike in the presence of L1 or L2, the aggregation curves in the presence of AMF do not show a significant lag phase, indicating that the small Aβ42 oligomers readily reorganize themselves into large aggregates with a high protection level, and thus AMF does not stabilize to any extent these small oligomers. In addition, the protection levels at the final stage in the presence of AMF (~80-90%) is higher than those in the presence of L1 or L2 (~80%), indicating that large, fibrillar Aβ42 aggregates ultimately form in the presence of AMF54. In addition, compared to the aggregation curves obtained when no ligands were present, the aggregation curves corresponding to the presence of AMF show similar t1/2 values yet with different curvatures, further suggesting that AMF directly affects the Aβ42 aggregation. Similarly, in the presence of both AFM and Cu²⁺, the Aβ42 aggregation kinetic curves do not show a significant lag phase (Figures 3.4D, 3.4E, and 3.4F), nor does the presence of Cu²⁺ change the t1/2 values for any of the peptides. Importantly, these results strongly suggest that AMF eliminates the “Cu effect” (i.e., the Cu-mediated stabilization of small Aβ42 oligomers20,22) observed in the absence of any ligand. Overall, these results indicate that AMF efficiently binds to Aβ42, even in the presence of Cu²⁺, and promotes the Aβ42 structural reorganization to form large, fibrillar aggregates.

3.5 Conclusion
We applied herein the pulsed HDX-MS approach described in Chapter 2 to investigate further the effect of different ligands on Aβ42 aggregation. We utilized the on-line pepsin digestion
Figure 3.4 Pulsed HDX results for three peptic peptides from Aβ_{42} interacting with AMF. Aggregation was done at 37 °C, with agitation at 150 rpm, in presence of AMF, either in the absence (upper panel) or presence of Cu^{2+} (lower panel). Peptic peptides 1-19 are represented in A and D, 20-35 in B and E, and 36-42 in C and F, respectively. All lines are fits using an F-W model in MathCAD.
protocol that was introduced in Chapter 2 to examine the binding effects on a peptide level. Importantly, we obtained Aβ42 aggregation kinetic parameters for peptides corresponding to different regions of Aβ42. In particular, BFCs L1 and L2 bind to Aβ42 and stabilize small Aβ42 oligomers in the absence of Cu²⁺, suggested by a shortening of the lag phases in the aggregation process. The kinetic curves indicate that BFCs accelerate the Aβ42 structural reorganization process to form large aggregates. When both BFC and Cu²⁺ are present, the kinetic curves show a significant lag phase for L1, but no lag phase for L2, suggesting that these two ligands interact differently with Aβ42 and Cu²⁺. In addition, in the presence of the natural product AMF, Aβ42 aggregation kinetic curves show no significant lag phase, and are identical in the absence and presence of Cu²⁺, suggesting that AMF binds Aβ42 significantly differently compared to the two BFCs. Most importantly, it seems that AMF eliminated the “Cu effect” of stabilizing the small oligomers and promotes the formation of large Aβ42 aggregates.

In conclusion, we envision this approach to be general for investigating various protein aggregation processes. Moreover, this method could be used to screen lead compounds as potential drug candidates for AD. To do this, we need to compare the outcome of the aggregation with biological effects of these putative drug candidates in animal testing. In addition, by applying electron capture dissociation (ECD) or electron transfer dissociation (ETD), we can obtain residue level information to guide further drug development. Furthermore, alternative footprinting methods, for example, fast photochemical oxidation of proteins (FPOP) may provide complementary information. Our hope is that others can apply this approach to investigate protein aggregation processes involved in other neurodegenerative diseases.

3.6 Acknowledgements
I acknowledge Dr. Anuj Sharma for synthesizing L1 and L2, and Don Rempel for modeling.
3.7 References


Chapter 4: Differential Hydrogen Deuterium Exchange Mass Spectrometry Studies of Protein Arginine Deiminase 2 (PAD2)

This work is part of a manuscript submitted to ACS Chem Biol.

4.1 Abstract
We applied differential hydrogen deuterium exchange mass spectrometry (HDX-MS) to understand the calcium dependence of the protein arginine deiminases or peptidylarginine deiminase (PADs), a family of enzymes whose activity is dys-regulated in inflammatory disease and cancer. Calcium binding triggers conformational changes, both within the active site cleft and at distal regions, and these increase PAD activity by at least 10,000 fold. Although PADs play an important role in regulating human gene transcription via the citrullination of histone arginines, their activation in cells is not understood because the calcium concentration required for activity is >10 times higher than physiologically relevant concentration. In this chapter, we describe the use of HDX-MS in a study of protein-ligand interaction, especially in calcium binding proteins. Wild type PAD2 and a PAD2 C647A mutant were analyzed by HDX-MS to monitor the calcium binding-induced protein conformational changes and to show that HDX-MS is sensitive to the dynamics of calcium binding. Multiple regions in PAD2 are located that have conformational changes upon calcium binding. The results of HDX-MS are consistent with the crystal structures and provide dynamic information for PAD2-calcium binding in solution.

4.2 Introduction
Mammals encode five PAD isozymes, PADs 1-4 and 6, which convert arginine into citrulline (Arg→Cit) (Figure 4.1). Given that arginines are important for protein-protein and protein-DNA/RNA interactions, the scope and importance of citrullination to human physiology is likely to be broad, especially when one considers that this post-translational modification (PTM) converts a positively charged guanidinium group into a neutral urea. Consistent with this statement is the fact that PAD activity influences multiple cellular processes, including apoptosis, gene regulation, and cell growth1-3.
Figure 4.1 Function of PAD.

arginine (positively charged)  →  citrulline (neutral)
The PAD family is made up of four active enzymes (PAD1-4) and catalytically inactive PAD6. They share 50-55% sequence identity\(^1\). The physiological roles of the individual isozymes and how they relate to disease remains poorly understood at the molecular level. The most well established connection to disease is overexpression of PADs 2 and 4 in the joints of Rheumatoid arthritis (RA) patients; overexpression induces formation of antibodies towards citrullinated proteins and localized activation of the immune system\(^4\). Although PADs 2 and 4 are most often associated with disease, PADs 1 and 3 are also involved in regulation of epidermal functions\(^5\), respectively, suggests that all active isozymes in PAD family represent potential therapeutic targets.

Based on previous reported crystal structure of calcium-free wild type PAD4\(^6\), this enzyme adopts an “elongated shape that N-terminal domain forms two immunoglobulin-like subdomains, and the C terminal domain forms an \(\alpha/\beta\) propeller structure consisting of five \(\beta\beta\alpha\beta\) modules arranged circularly in a pseudo-five-fold symmetric” structure\(^7\). Approximately 67% of sequence homology on the C terminus of all PADs suggests that the structures of these C-terminal domains are similar. Five calcium binding regions were identified in the holo form of PAD4. All five calcium binding regions do not have typical EF-hand motif (helix-loop-helix motifs) found in common calcium-binding proteins like Calmodulin. Among five calcium binding regions, Ca1 and 2 are at the bottom of the active site cleft, whereas Ca3-5 are positioned near the molecular surface of sub-domain 2 in the N terminal domain. Based on their positions, Ca1 and 2 are crucial for substrate binding and enzyme activity. By comparing calcium-free and bound PAD4, we can conclude that calcium binding induces conformational changes that generate the active site cleft in the C terminal domain where the target substrate, arginine, is converted to citrulline. In PAD4, a nucleophilic cysteine, C645, hydrolyzes the guanidinium of an arginine residue to
form citrulline. Interestingly, Ca3-5 binding regions are well-conserved among all PAD family members. The protein structures surrounding Ca3-5 shifted from disorder to order when calcium ions are bound, inducing significant conformational changes. Thus, the structure of the holo form of PAD is stabilized. On the contrary, Ca3-5 does not affect PAD enzymatic activity.

Unlike PAD4, there is no high resolution structure published for PAD2 that is also present in the synovial tissues of RA patients⁴. Beside RA, highly expressed PAD2 in luminal breast cancers have been reported⁸. As potential drug target, PAD2 have been studied by comparing its distribution, catalytic activities, calcium dependency, optimal pH range of activities and substrate specify with PAD4⁹,¹⁰. Recent studies proposed that PAD2 uses a new substrate-assisted mechanism of catalysis¹¹. Biological importance of PAD2 has promoted the deep structural studies of PAD2.

In this collaborative project, we planned to use X-ray crystallography, mutagenesis and HDX-MS to determine the location, order of calcium binding and the dissociation constants for the individual calcium-binding regions. These studies may identify the key calcium binding regions that are required for the high calcium dependence of the enzyme. The Thompson lab has developed a cell line for expressing pure PAD2 protein and crystallization method for holo form of PAD2. This crystallization-based approach can provide high resolution structure model for holo form of PAD2. However, it lacks information about PAD2 in solution phase, which can be obtained through HDX-MS.

As I presented in the introduction chapter (Chapter 1), HDX-MS is a new powerful biophysical approach in study of protein-ligand interactions. Proteins in their native state (solution phase) were labeled by deuterium in buffer solution. The protein conformational information as
represented by deuterium uptake is recorded by mass spectrometry in an LC-MS experiment. Our group has reported several HDX-MS studies of calcium-binding proteins including Calmodulin\(^{12, 13}\) and Troponin \(C^{14}\). In this chapter, we started with HDX-MS investigation of PAD2-Ca\(^{2+}\) interactions. Both apo and holo forms of PAD2 were analyzed by HDX at seven exchange time points at two temperatures, 4 °C and 25 °C, respectively. HDX kinetic curves for each identified peptic peptides of PAD2 were plotted in both the apo and holo forms. Calcium binding induced conformational changes (either direct or allosteric interactions) were identified for both wild type and mutant PAD2. This HDX-MS approach provides a solid basis for HDX-MS based calcium binding affinity studies, our PLIMSTEX approach, as will be described in the following chapter.

### 4.3 Methods and Materials

#### 4.3.1 Chemicals and Proteins
All chemicals and reagents were purchased from Sigma Aldrich (St. Louis, MO) unless stated otherwise. Calcium free PAD2 was provided by the Thompson lab.

#### 4.3.2 HDX-MS
We examined the effects of calcium-binding to PAD2 at the peptide level by using HDX. The protein sample (14 μM PAD2 in 20 mM Tris pH 7.6, 0.5 M NaCl, 0.5 mM TCEP, 10% glycerol) was pre-incubated with either 0.5 mM EDTA or 2.5 mM CaCl\(_2\) for 60 min at 4 °C. All HDX experiments were carried out at either 4 °C or 25 °C in D\(_2\)O (20 mM Tris in D\(_2\)O, 0.5 M NaCl, 0.5 mM TCEP, pD 7.6) where 4 μL protein solution was mixed with 16 μL D\(_2\)O buffer. HDX was analyzed at eight time points (0, 0.17, 0.5, 1, 2, 15, 60, 240 min). HDX was quenched by adding 30 μL of 3 M Urea, 1% TFA at 4 °C. A total of 50 μL of quenched protein sample was injected for proteolysis with on-line desalting and digestion (immobilized pepsin) and analysis.
with a fast gradient HPLC (5 min). The HPLC was coupled to an LTQ-Orbitrap XL mass spectrometer (Thermo Scientific, San Jose, CA). MS/MS experiments were also performed with the LTQ Orbitrap as preliminary for the HDX to identify the peptides produced by pepsin proteolysis. Product-ion spectra were acquired in a data-dependent mode, and the six most abundant ions were selected for product-ion analysis.

4.3.3 Peptide Identification and Data Processing
The raw data files were converted to *.mgf files (MassMatrix Mass Spectrometric Data File Conversion Tools, Version 3.9) and then submitted to MassMatrix (Version 2.4.2) for peptide identification\textsuperscript{15,16}. Peptides included the set used for PLIMSTEX to be described in the following chapter had a MassMatrix score of 10 or greater, when using peptide and MS/MS tolerance as 10 ppm and 0.8 Da, respectively. The MS/MS MassMatrix search was also performed against a decoy (reverse) sequence, and ambiguous identifications were ruled out. The product-ion (MS/MS) spectra of all peptide ions from the MassMatrix search were manually inspected, and only those verifiables were followed in the HDX experiments.

For the HDX experiments, the centroid masses of isotopic envelopes were calculated with HDX WorkBench\textsuperscript{17}: deuterium level (%) = ((m(P) – m(N))/(m(F) – m(N))) x 100%, where m(P), m(N), and m(F) are centroid values of partially deuterated peptides, nondeuterated peptides, and fully deuterated peptides, respectively. To accommodate the situation where a fully deuterated control is not available, m(F) was determined as m(F) = m(N) + (n-p-2)/z, where n is the number of amino acids in the peptide, p is the number of prolines, and z represents charge. Prolines, with no amide hydrogen, were not considered. The value “2” is subtracted in the equation because the first two amino acids do not retain deuterium upon back exchange. No correction was made for
back exchange because all values are relative and reflect the same back exchange, thus giving reliable trends.

In addition, we compared the HDX results at different temperatures. It is well known that HDX rate is affected by temperature (Equation 4.1)\(^{18}\). Thus, to compare HDX at two temperatures, the HDX exchange time scales of the experiments must be adjusted; for example, by increasing time at the lower temperature.

\[
\frac{k_{ch1}}{k_{ch2}} = \exp\left(\frac{E_a}{RT_1} - \frac{E_a}{RT_2}\right) = \exp\left(\frac{E_a - E_a}{RT_1 - RT_2}\right)
\]

Equation 4.1

The calculated reaction rate at 25 °C is 9.3 times faster than that at 4 °C. We modified the timescale of HDX data at 25 °C. All time points are adjusted to 9.3 times longer than the original scale. For example, the extents of HDX at the seven 25 °C time points of 0.17, 0.5, 1.0, 2.0, 15, 60, 240 min were converted to same extents but at 1.55, 4.65, 9.3, 18.6, 139.5, 558, 2232 min.

4.4 Results and Discussion

We identified ~280 peptic peptides from the outcome of the LC-MS/MS experiments. These peptides cover 96% sequence of PAD2 (Figure 4.2). Thus, almost all PAD2 regions can be monitored in HDX studies. More importantly, there are many overlaps of peptic peptides from proposed calcium binding regions.

4.4.1 HDX Studies of Apo and Holo Wild-Type (WT) PAD2 at 4°C

In the first set of experiments, both apo and holo forms of PAD2 were incubated in D\textsubscript{2}O buffer at 4 °C. The kinetic curves for both apo and holo forms were plotted for seven time points (Figure 4.3). Data from apo form of PAD2 at 4 °C are highlighted on the dynamic maps of PAD2.
Figure 4.2 HDX peptide coverage map for WT PAD2 (96%).
476-491 (+3)
479-490 (+2)
479-490 (+3)
479-491 (+3)
479-491 (+2)
491-497 (+1)
491-514 (+4)
498-512 (+3)
498-512 (+4)
498-512 (+5)
498-514 (+3)
498-514 (+4)
498-514 (+5)
501-512 (+2)
501-512 (+3)
501-512 (+4)
501-514 (+3)
Figure 4.3 HDX kinetic curves for WT PAD2 (4 °C) of both apo (solid line) and holo (dash line) states. Numbers in parentheses with “+” sign are the charge states of the peptides.
structure obtained from unpublished data from the Thompson lab (Figure 4.4). The conformational changes on smaller regions of the protein can be mapped from the HDX data of multiple overlapping peptides. PAD2 regions with different dynamic features can be identified on the map. Several regions (highlighted in blue in Figure 4.4) are still protected from HDX even for 240 min, indicating the rigidity of the structure and the strong H bonds in that region. The least dynamic regions of PAD2 are β sheets (β4, β17 and β18, represented by several peptides in regions 40-51, 266-273 and 290-300) of the N-terminal domain (residues 1-295), other β sheets (β22, β23 and β34, represented by several peptides in regions 355-369 and 565-573) and an α12 helix (represented by peptides in region 538-555) of the C-terminal domain (residues 296-665). All the numbers for the secondary structure are based on PAD4. Those regions exhibit less than 10% deuterium uptake after 240 min incubation in D2O buffer. On the contrary, all calcium binding regions undergo more HDX in the apo state (highlighted in warm color in Figure 4.4). Calcium-binding motifs, containing many acidic residues, exhibit higher dynamic features in the apo state relative to the holo. In general, calcium binding proteins take advantage of these flexible regions to interact rapidly with free calcium ions in solution and form stable conformations through non-covalent interactions.

For PAD2, five of the six calcium binding regions overlapped (Ca1-5) in the sequence (Table 4.1). Residues 156-171 are shared by Ca3-5 binding regions, and 352-355 by Ca 1 and 2 binding regions. Only five regions, 349-351 and 406-449 for Ca1, 370-384 for Ca2, 152-155 for Ca3, 385-391 for Ca4, and PAD2 121-131 for Ca6, are not shared by multiple calcium binding regions (Table 4.1). We compared the HDX kinetic curves of peptides from the above regions for both apo and holo WT PAD2 to examine calcium binding-induced conformational changes.
Figure 4.4 Dynamics map for HDX of PAD2 (4 °C). The deuterium uptake levels for the apo state of WT PAD2 are mapped with color for each exchange point. Color codes are shown on the right edge.
Table 4.1 Calcium binding regions in PAD2. “Y” represents that this region is involved in the corresponding Ca binding region.

<table>
<thead>
<tr>
<th>regions</th>
<th>Ca1</th>
<th>Ca2</th>
<th>Ca3</th>
<th>Ca4</th>
<th>Ca5</th>
<th>Ca6</th>
</tr>
</thead>
<tbody>
<tr>
<td>121-131</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>152-155</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>156-171</td>
<td>Y</td>
<td></td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>172-180</td>
<td></td>
<td>Y</td>
<td></td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>349-351</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>352-355</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>370-384</td>
<td></td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>385-391</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>406-449</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For the Ca1 binding region, there are no significant differences between the HDX kinetic curves for the apo and holo states at PAD2 406-449 (Figure 4.3), which is the region involved only in Ca1 binding and is not overlapping with other Ca binding regions. HDX data from shared region between Ca1 and 2 binding regions, PAD2 349-355, show similar trends. Thus, no calcium binding-induced conformational change is identified for Ca1 binding region from the HDX data at 4°C.

Peptic peptides covering Ca 2-5 binding regions are heavily overlapped. Overlapped peptides from region 156-171 (for Ca3-5 binding) show a significant decrease in deuterium uptake when Ca\(^{2+}\) is added, indicating that these regions undergo large conformational changes upon calcium binding. We observed slight HDX differences in PAD2 250-257, a region that is involved only in Ca5 binding and is not overlapping with other Ca binding regions, between apo and holo states. Based on sequence identity and similarity of PAD4 and PAD2 (Figure 4.5), the active site, as well as regions surrounding calcium binding regions Ca2-5, are strongly conserved. Such high conserved sequences suggest that PAD2 and PAD4 share similar binding capability for these four calcium binding regions. Our HDX data confirmed calcium binding on these four regions in PAD2, which is consistent with an unpublished X-ray crystallography data from the Thompson lab. The published X-ray crystallography data for PAD4 point to these four regions as Ca binding regions as well. PAD4 showed significant conformational changes particularly in Ca3-5 binding regions going from disordered in the apo state to become ordered in the holo state. Consistent with this are the significant differences in deuterium uptake in our PAD2 HDX experiment, indicating that similar structural changes from disorder to order in Ca3-5 binding regions occur for the solution protein.
Figure 4.5 Sequence alignment of PAD2 and PAD4. Symbols below represents degree of identity or similarity of the residues between the two sequences. For example, “*” stands for identity in the sequence, “:” for better similarity than “.” based on the algorithm from ref. PAD4 sequence lies above and PAD2 lies below. Different Ca binding sites are color coded as shown at bottom.
For the Ca6 binding region, only three peptic peptides from PAD2 121-151 cover this region. There are significant HDX differences observed between the apo and holo PAD2 for peptide 121-131, indicating that this region becomes protected upon calcium binding. Peptide 121-151 shows less but still significant differences that are reduced owing to an averaging effect from other residues (132-151) that are not involved in Ca binding. A previous X-ray structure shows that PAD4 only binds five calciums and lacks a 6th Ca binding region. The X-ray structure of PAD2 from the Thompson lab, however, also confirmed a 6th Ca coordination contacts that fit well within standard calcium bond lengths (2.3-2.5 Å).

We also observed deuterium uptake difference between apo and holo WT PAD2 at several regions that are not calcium-binding (Figure 4.3). Interestingly, we observe increased deuterium uptake for peptides 5-18 and 273-287 upon calcium binding. These regions may involve in PAD2 dimerization. PAD2 dimer dominates in the apo form (unpublished native spray mass spectrometry data from Gross lab) but may subject to change in the holo form. Multiple regions, PAD2 78-117, 203-226, 336-348 and 631-657, are protected (show less HDX) upon calcium binding, which may relate to PAD2 active cleft formation. The PAD4 study shows that there are no structural differences between PAD4-Ca and PAD4-Ca-substrate complex, indicating that the active cleft is already formed upon Ca binding.

In summary, the differential HDX experiments detected significant changes in Ca2-6 binding regions between apo and holo form of WT PAD2. On the contrary, there are no calcium-induced conformational changes identified at the Ca1 binding region (Table 4.1 and Figure 4.3). There might be two reasons for such observation: (1) this region does not bind calcium in WT PAD2 at 2.5 mM Ca$^{2+}$ concentration; (2) calcium binding at this region is weak and still as dynamic as it is in the apo state. To examine the second possibility, we need to perform HDX at a higher
temperature to obtain higher extent of hydrogen deuterium exchange in the apo state, thus making the differences between apo and holo states more significant and available to measure.

### 4.4.2 HDX Studies of Apo and Holo WT PAD2 at 25°C

We also performed HDX at a higher temperature. We increased the HDX incubation temperature to 25°C to identify weak calcium binding interactions that we may have missed at 4 °C. These data are presented also as kinetic curves for both apo and holo PAD2 (Figure 4.6). A dynamic map was generated by using HDX data from apo PAD2 (Figure 4.7). As expected, we observe faster deuterium exchange at 25 °C than 4 °C, as seen by comparing the dynamic maps under these two conditions (Figure 4.7 vs. Figure 4.4). Most of PAD2 regions have more than 30% deuterium uptake after 240 min incubation at 25 °C, larger than what we observed at 4 °C. The α helix and β sheets regions on either N- or C-terminal regions that are found to exhibit low level deuterium uptake in 4 °C experiments (i.e., peptides in regions 40-51, 266-273, 290-300, 355-369, 543-555, and 565-573) take less than 30% deuterium uptake in 25 °C experiments. We conclude that these regions are consistently the least dynamic region among PAD2 at both temperatures.

Kinetic curves of the peptides covering the calcium-binding regions were compared between apo and holo PAD2 at the two temperatures. For the Ca1 binding region, there are no significant changes upon Ca binding (Figure 4.6) at either temperature. For PAD4, the Ca1 binding region is at the bottom of an active-site cleft where the substrate arginine binds. The citrullination of the substrate arginine is initiated by a nucleophilic cysteine (C645 in PAD4 and C647 in PAD2) hydrolyzing the guanidinium group of arginine. This reaction is induced by a conformational change on residue C645 in PAD4, C645 is far away from active site cleft in calcium free PAD4 but it moves ~5 Å and be close to the cleft upon Ca1 binding from inactive to active state. The
Figure 4.6 HDX kinetic curves for WT PAD2 (25 °C) of both apo (solid line) and holo (dash line) states. Numbers in parentheses with “+” sign are the charge states of the peptides.
Figure 4.7 Dynamics map for HDX of PAD2 (25 °C). The deuterium uptake levels for the apo state of WT PAD2 are mapped with color for each exchange point. Color codes are shown on the right edge.
highly conserved sequence of Ca1 binding region between PAD2 and 4 suggests C647 of PAD2 undergoes similar conformational change upon Ca1 binding. In the parallel X-ray based studies from Thompson lab, the residue C647 of PAD2 is inactivate and far away (12 Å apart) from the active site cleft when PAD2 is incubated with 10 mM calcium. No conformational change in HDX is in consistence with the inactive state PAD2 observed in X-ray. The absence of calcium on Ca1 binding site of holo PAD2 may suggest the different activation mechanism between PAD2 and PAD4.

For the other five calcium-binding regions (Ca 2-6), significant decreases in deuterium uptake are observed upon calcium binding at 25 °C, and these results are consistent with HDX data at 4 °C. There are several other regions that are not involved in calcium binding but still undergo conformational changes in the holo states according to the HDX experiment at 4 °C. For example, regions 7-18 and 274-286, exhibit increasing deuterium uptake in the holo state at 25 °C, and they may be involved in forming a PAD2 dimer interface as discussed previously. We also observed decreased deuterium uptake at other regions at 25 °C. Regions 78-87, 113-117, 336-348 and 632-657 undergo decreased HDX as was observed at 4 °C.

4.4.3 Comparing HDX Studies (4 °C vs. 25 °C) of Apo WT PAD2
We adjusted the HDX data at 25 °C experiment and plotted them with the HDX data at 4 °C (Figure 4.8). In general, we do not observe significant changes upon increasing the temperature for apo PAD2. Almost all regions show the same trends in the kinetic curves. There are no significant changes or breaks between curves taken at the two temperatures. If PAD2 has conformational changes at 25 °C, the kinetic curves should show significant differences. That little change occurs indicates that PAD2 remains as a highly conserved structure at higher
Figure 4.8 HDX kinetic curves for WT PAD2 apo state comparison between two temperatures. 4 °C was shown in black dots and 25 °C in red. Numbers in parentheses with “+” sign are the charge states of the peptides.
temperature (4 °C vs. 25 °C). In addition, these data speak to the good precision and high consistency of our experimental data.

4.4.4 HDX Studies of Mutant PAD2
The enzyme activity of PAD2, arginine deimination, requires several residues in the binding pocket to interact with the substrate. Based on proposed mechanism of PAD2 catalysis, two aspartic acid residues (Asp351 and 473) formed hydrogen bonds with Arginine substrate while the active site thiolate of C647 (equal to C645 in PAD4) attacks the guanidinium carbon of arginine to initiate the Arg→Cit conversion. In previous structural studies of PAD4-substrate complex, the key residue, C645, for Arg→Cit conversion is replaced by Alanine. This PAD4 C645A mutant is catalytically inactive and form stable calcium bound PAD4-substrate complexes. Similarly, PAD2 C647A mutant results in a loss of PAD2 enzyme activity. Evidence from unpublished X-ray crystallography data from the Thompson lab demonstrated that PAD2 C647A mutant still binds to the substrate, but no conversion from arginine to citrulline occurs. The PAD2 C647A mutant likely has a different higher order structure in the active site cleft than does WT PAD2. Calcium binding-induced conformational changes of PAD2 C647A remain unknown. The HDX data from PAD2 C647A mutant could provide such conformational differences. Thus, we extended our HDX studies to PAD2 C647A mutant and compared the calcium binding-induced conformational changes between the mutant and WT PAD2.

Sequence coverage of PAD2 mutant is similar to WT PAD2. Multiple overlapped peptic peptides cover all six proposed calcium binding regions (Figure 4.9). To provide a view of the dynamics of the protein, we show a map with color coding for the extent of HDX for the apo state of the mutant (Figure 4.10). Unlike for WT PAD2, there is a significant decrease in deuterium uptake
Figure 4.9 HDX kinetic curves for PAD2 C647A mutant of both apo (solid line) and holo (dash line) states at 4 °C. Numbers in parentheses with “+” sign are the charge states of the peptides.
Figure 4.10 Dynamics map for HDX of C647A PAD2 (4 °C). The deuterium uptake levels for the apo state of PAD2 C647A are mapped with color for each exchange point. Color codes are shown on the right edge.
for Ca1 binding regions (PAD2 340-354 and 406-411) of PAD2 C647A mutant as indicated by the peptic peptides covering these regions. These results indicate that the Ca1 binding region undergoes conformational changes upon calcium binding. As suggested by unpublished X-ray crystallography studies by the Thompson lab, the Ca1 binding region in WT PAD2 may resist change in conformation by interactions with C647. For C647A, the lack of C647 clearly releases such inhibition and makes the Ca1 region available for calcium binding.

For the Ca 2-5 binding regions of PAD2 C647A mutant, similar decreases in deuterium uptake are observed upon calcium binding as for WT. Thus, the HDX experiments on PAD2 C647A mutant provide evidence that the five calcium-binding regions, Ca1-5, are occupied by calcium in the holo form. These data are consistent with a highly dynamic apo state, particularly for the Ca2-5 regions, that become constrained upon calcium-binding, consistent with the unpublished X-ray crystal structure from the Thompson lab. For Ca6, there is no significant difference in deuterium uptake between the apo and holo forms of the mutant, strongly suggesting there is little or no calcium binding at Ca6 binding region of the mutant.

There are several regions that are not involved in calcium binding but that show higher deuterium uptake in the holo form than in the apo form of the PAD2 C647A mutant. As discussed for the WT PAD2, those regions may be involved as an interface in PAD2 dimer reorganization. In addition, some regions, 227-249, 310-329 and 431-478, show significant decrease in deuterium uptake upon calcium binding in the mutant but not the WT (Figure 4.9). Those conformational changes may be involved in deactivation of PAD2 function by C647A mutation.
4.5 Conclusion
We conducted differential, solution HDX experiments to study PAD2 upon Ca binding. Several regions show dramatic decreases in deuterium uptake upon addition of Ca$^{2+}$. Four calcium binding sites, Ca2-5, has similar conformational changes upon calcium binding, which is consistent with previous studies of PAD4 and unpublished X-ray studies on PAD2. Calcium binding on those regions induces the order on protein conformation and stabilized the following substrate binding and citrullination. Solid state structure suggests Ca1 binding, but this is not seen in solution by HDX at 2.5 mM Ca$^{2+}$ concentration. This suggests different activation mechanism between PAD families and requires further investigation. HDX data represents not only conformational changes on calcium binding regions, but also the potential dimer interface of PAD2. Information from our extended HDX studies of PAD2 at elevated temperature and on mutant PAD2 has similar trends on most regions. The only difference between WT and mutant PAD2 was on the Ca1 and 6 binding sites. It indicates conformational differences induced by the mutation.

PAD2 is a complicated enzyme system that contains up to six calcium binding regions. Detailed interpretation of the calcium binding in each region requires more information including Ca’s binding affinity and binding order. The unpublished x-ray data provide only a crude estimation of the calcium dissociation constants for the individual region, and based on comparison to the concentration of calcium required for half maximal activity, these values are shifted substantially from that occurring in solution, a deviation that may be due to crystal artifacts. In the following chapter, we will use PLIMSTEX methodology to confirm the order of calcium binding and estimate the dissociation constants for the individual binding in solution.
4.6 Acknowledgements
I acknowledge Dr. Daniel Slade for PAD2 synthesis and purification.
4.7 References


Chapter 5: PLIMSTEX studies of calcium binding of protein arginine deiminase 2 (PAD2)
5.1 Abstract
The Protein Arginine Deiminase (PAD) family plays important roles in inflammatory diseases and cancer. One of the family members, PAD2, is involved in the onset and progression of neurodegenerative human disorders and multiple sclerosis. Although PAD2 is an attractive therapeutic target, its calcium binding-induced activation mechanism is poorly understood. In the previous chapter, we described calcium binding-induced conformational changes that were determined by differential HDX-MS. A total of six calcium binding sites were identified in WT PAD2 and PAD2 C647A mutant. In this chapter, we focus on WT PAD2 and extended our HDX-MS studies by applying a Protein–Ligand Interactions by Mass Spectrometry, Titration, and H/D EXchange (PLIMSTEX) strategy coupled with online-pepsin digestion to study binding affinity of each Ca$^{2+}$ binding site. The calcium titration experiment was designed based on the differential HDX-MS kinetic results described in the previous chapter. In the PLIMSTEX study, PAD2 protein was incubated with different concentration of calcium ions in a titration format. The deuterium uptake measured by MS is plotted as a titration curve to afford information of binding affinity. Such information further improves our understanding about the calcium-induced activation of PAD2.

5.2 Introduction
PADs are enzymes that catalyze a post-translational modification called arginine deimination or citrullination, which alters a positive charge to neutral on Arginines. These proteins are often up-regulated in inflammatory diseases and cancer$^{1,2}$. There are five members in the family, PAD1-4, and 6, and PAD4 is the only that is well characterized. PAD2 is involved in citrullination of histones H3 and H4. This post-translational modification causes activation or repression of genes that are regulated by estrogen receptor (ER) and p53$^{3,5}$. Although PAD2 is an attractive
therapeutic target, the mechanism of enzyme activation is poorly understood. We do know that PAD2 needs to bind several Ca$^{2+}$ ions to carry out the enzymatic reaction. Understanding the details of calcium binding for each PAD2 calcium binding region will further improve our knowledge about calcium binding and the subsequent activation of PAD2.

As discussed in the introduction, information about protein conformational changes can be obtained by comparing two states of a protein, apo and holo states, by using HDX-MS. We reported an investigation of the PAD2-calcium interaction by differential HDX-MS in the previous chapter (Chapter 4), and we used this approach to identify the calcium binding-induced conformational changes and to infer binding sites. Such traditional HDX is a powerful footprinting method, but it provides little information about binding affinity at each of the calcium binding sites of PAD2. Previous x-ray crystal studies on PAD4 raise the question about the functions of the five calcium binding sites in PAD4$^6$. Similar to PAD4, there are five calcium binding sites, as identified by our HDX-MS experiments, on WT PAD2. The Ca1 binding region was not identified by differential HDX-MS at a 2.5 mM Ca$^{2+}$ concentration. An important extension to this work is to obtain detailed information like calcium binding affinity in these regions.

There are several HDX-MS approaches to determine protein-ligand binding affinity$^7,^8$. One development by our group$^8$ is PLIMSTEX. A detailed introduction of PLIMSTEX is covered in the introduction (Chapter 1). Thus, we will only give a brief description here. In a typical PLIMSTEX experiment, one compares deuterium uptake level of two states of the protein, not as a function of HDX incubation time, but of the ratio between total ligand and total protein concentration$^9,^{11}$. By fitting the ligand titration curve, the binding affinity can be elucidated.

When accompanying by pepsin digestion, PLIMSTEX curves can generate binding information.
for peptide or even residue specific level\(^9\). Herein, we conduct PLIMSTEX studies on peptide level for PAD2-Ca\(^{2+}\) binding system. Based on our HDX experiments, we conducted the PLIMSTEX titration at 4 °C and at 30 sec of exchange. We titrated apo PAD2 at 41 different concentrations of calcium to afford a deuterium uptake on the peptide level as was monitored by LC-MS coupled with online pepsin digestion. Information from this further provides an estimate of the binding in PAD2-Ca\(^{2+}\) system.

**5.3 Materials and Methods**

**5.3.1 Expression of WT-PAD2; Peptide Level HDX; Peptide Identification and HDX Data Processing**

All of this information is covered in the previous chapter (Chapter 4).

**5.3.2 PLIMSTEX**

PLIMSTEX experiments were done similarly as the differential solution HDX experiments except that a 3 μM PAD2 after incubation with different concentrations of Ca\(^{2+}\) or ethylenediaminetetraacetic acid (EDTA), to obtain higher binding affinity information and avoid the “sharp break” curve discussed in the introduction (Chapter 1). In total, the PLIMSTEX experiments were interrogated at two concentrations of EDTA ([EDTA]:[PAD2] = 36 or 198) and 41 concentrations of Ca\(^{2+}\) ([Ca\(^{2+}\)]:[PAD2] = 0 to 3968) to obtain a full binding profile. We chose an incubation temperature at 4 °C based on the HDX-MS experiment discussed in the previous chapter. In the PLIMSTEX experiments, the protein sample was incubated with Ca\(^{2+}\) or EDTA at 4 °C for exactly 1 h, followed by D\(_2\)O incubation at the same temperature for 30 s. A standard quench, desalting and online-digestion procedures were performed following a procedure described in Chapter 4. The resulting peptic peptides were analyzed by LC-MS on an Orbitrap XL LC/mass spectrometer (Thermo, San Jose, CA).
5.3.3 PLIMSTEX Curve Fitting
All of this information is covered in a previous chapter (Chapter 1). In brief, a nonlinear least squares (NLLS) regression is performed by procedures implemented with MathCAD 14 (PTC Inc., Needham, MA). The PLIMSTEX data are fitted using a 6:1 ([ligand]:[protein]) binding system. The best fit is obtained by searches, changing all the variable parameters (i.e., $\beta$, $D_0$, and $\Delta D_i$) to minimize the error between the fitted curve and the experimental data by iterating through many trials.

5.4 Results and Discussion
Our goal is to determine the binding affinities of the PAD2-$\text{Ca}^{2+}$ binding system by PLIMSTEX. In all of these experiments, the protein was labeled in in solution with Tris buffer. The conformational changes were revealed by following the deuterium uptake as a function of the concentration ratio of the calcium ions and WT PAD2 under the same conditions. We conducted a total of 129 experiments (including biological triplicates) at 41 different calcium concentrations and two different EDTA concentrations to ensure that the PLIMSTEX results can reliably report the conformational changes induced by any perturbations (i.e., ligand concentration changes).

In a typical PLIMSTEX experiment, one first performs a standard HDX-MS experiment, to select a D$_2$O incubation time that permits a maximum deuterium uptake difference between the apo and holo states. In this case, the D$_2$O incubation time is 30 s based on our previous differential HDX-MS results (Chapter 4). Consequently, all PLIMSTEX experiments were done at an exactly 30 s of D$_2$O incubation time, with only the EDTA or Ca$^{2+}$ concentrations as variables. Adding EDTA titration into our experiment is to assure an apo state of PAD2.

Although we designed to keep protein (PAD2) concentration constant and change the ligand (Ca$^{2+}$) concentration, selection of the protein concentration requires careful consideration. The
protein concentration needs to be low enough to follow binding at high affinity sites and avoid “sharp break” curves (as discussed in Chapter 1), but sufficiently high to obtain decent MS signals for peptides representing the ligand-binding regions. Fortunately, the ligand-binding regions for Ca binding protein have many acidic residues, making the peptides hydrophilic and more readily detected by MS than hydrophobic peptides. Based on these two criteria, we conducted a series of pilot experiments and selected 3 μM PAD2 for all PLIMSTEX experiments. As usual, we plotted deuterium uptake level as a function of [Ca²⁺]/[protein] ratio to obtain PLIMSTEX curves.

To assure our PLIMSTEX experiments were reproducible and no systematic error were introduced, HDX PAD2-Ca²⁺ titration results from “control region of PAD2” were analyzed. We plotted deuterium uptake level from three peptides as a function of [Ca²⁺]/[protein] ratio (e.g., 18-25, 61-69, and 266-272, Figure 5.1), which covered “control region” that are remote from the Ca²⁺ binding sites. These three peptides also range from protected area (showing no deuterium uptake) to flexible regions (showing extensive deuterium uptake). We did not observe any significant deuterium uptake difference with increasing the [Ca²⁺]/[protein] ratio for any of these three peptides, suggesting there is no significant systematic error. More importantly, these results suggest that the changes observed in our PLIMSTEX experiments can reflect effectively and accurately any conformational changes induced by Ca²⁺.

For Ca²⁺ binding, we chose a total of four peptides that covered Ca2, 3/5, 4, and 6 binding sites (370-391, 168-183, 378-391, and 121-151, respectively) to investigate the binding by PLIMSTEX. These choices reflect that Ca3 and Ca5 bind in the same region (based on the unpublished X-ray results from the Thompson lab and HDX results from Chapter 4). We did not
Figure 5.1 PLIMSTEX data for regions that do not show deuterium uptake decrease.
cover Ca1 because we did not observe any significant deuterium uptake changes in that binding region at 2.5 mM Ca\(^{2+}\), as discussed in Chapter 4.

Unlike the control experiments, we observed significant decreases in deuterium uptake for these peptides upon increasing the \([\text{Ca}^{2+}]/[\text{protein}]\) ratio (Figure 5.2), which agrees with the HDX kinetics data in Chapter 4 that show deuterium uptake differences between apo and holo states in almost all Ca\(^{2+}\) binding regions. Surprisingly, all PLIMSTEX curves corresponding to different regions show similar trends, indicating of a high level of cooperativity or regulatory interactions.

These results agree with the unpublished X-ray structure from the Thompson lab, which shows that the Ca3, 4 and 5 (called a “tripartite calcium switch”, TCS) bind cooperatively, whereas the TCS can regulate binding of Ca2. It may also affect the low-affinity binding at Ca1. The TCS apparently plays a key role in PAD2 enzymatic activity by releasing the active site for the substrate arginine to interact. Thus, even if Ca1 and Ca2 binding is not cooperative with the TCS, it is still understandable why all binding curves look similar.

Unfortunately, we cannot obtain the binding order of these calcium ions because the PLIMSTEX curves are similar in shape and location. To assign the binding affinities to each individual Ca, we start with the unpublished X-ray crystallography data. We divide all the curves into three deuterium uptake decrease stages. The curves all start with a sharp turn at \([\text{Ca}^{2+}]/[\text{protein}]\) ratio less than one, follow by two immediate decreases in deuterium uptake level with the second one less significant than the first. After these three stages there is a plateau, suggesting that Ca binding is complete for PAD2. We then fit three of the four binding curves (including 370-391 for Ca2, 168-183 for Ca3/5 and 378-391 for Ca4) simultaneously and obtain binding affinities following procedures described in Chapter 1 (Figure 5.2). The reason for leaving out peptide 121-151 representing Ca6 is that the deuterium uptake decrease is less than 5%, which causes
Figure 5.2 PLIMSTEX data for five Ca\(^{2+}\) binding regions. Dashed lines are curve fitting for obtaining the binding affinities, based on the 1-3-1-1 binding pattern.
difficulties for identifying the best fitting parameters. We set the total number of bound Ca\(^{2+}\) as six during the fitting process, and we enabled a “cooperative binding” function in the fitting program, where we can specify which sites bind Ca\(^{2+}\) cooperatively (Table 5.1). In this function, we specify how many calcium ions bind to the protein cooperatively, and how many bind independently. In addition, we order the binding system by binding affinity (i.e., higher binding affinity occurs “before” lower). For example, 1-3-1-1 binding pattern indicates to the fitting program that there are three Ca\(^{2+}\) bind fully cooperatively with second highest affinity, whereas the other sites bind Ca\(^{2+}\) independently.

We tested all possibilities for six Ca\(^{2+}\) binding by fitting the three curves using different binding patterns (Table 5.2). We applied several criteria to pre-mark some of the binding patterns “untrue”. For example, we observed a sharp break when [Ca\(^{2+}\)]/[protein] ratio is less than 1, which means the first Ca\(^{2+}\) binding should not be cooperative with any other Ca\(^{2+}\) binding, and it does have high binding affinity (less than 3 \(\mu\)M). Thus, the first number in the binding pattern is “1”, which means all the possible binding patterns start with a number larger than one (i.e., two to six) should not be considered. In addition, the third turn of the curve is not as sharp as the second, so we need less than Ca\(^{2+}\) for the fitting of the last stage of the curve. Unfortunately, current PLIMSTEX data analysis and fitting process cannot deal with partially cooperative binding behavior (i.e., the binding of one Ca\(^{2+}\) partially affects or regulates the binding behavior of at another Ca\(^{2+}\) binding site). Such fitting and simulation requires larger computational capacity than we currently have.
Table 5.1 Examples of the Ca binding cooperativity

<table>
<thead>
<tr>
<th>Ca binding order</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; Ca</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; Ca</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt; Ca</th>
<th>4&lt;sup&gt;th&lt;/sup&gt; Ca</th>
<th>5&lt;sup&gt;th&lt;/sup&gt; Ca</th>
<th>6&lt;sup&gt;th&lt;/sup&gt; Ca</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A: All calcium ions bind PAD2 independently.
B: Only the first two calcium ions bind PAD2 cooperatively.
C: Only the 2<sup>nd</sup> and 3<sup>rd</sup> calcium ions bind PAD2 cooperatively.
D: The 2<sup>nd</sup> and 3<sup>rd</sup> calcium ions bind PAD2 cooperatively; as well as 4<sup>th</sup> and 5<sup>th</sup> calcium ions. However, these two groups of calcium ions (i.e., 2<sup>nd</sup> and 3<sup>rd</sup> vs. 4<sup>th</sup> and 5<sup>th</sup>) do not have cooperativity.
<table>
<thead>
<tr>
<th>1(^{st}) Ca</th>
<th>2(^{nd}) Ca</th>
<th>3(^{rd}) Ca</th>
<th>4(^{th}) Ca</th>
<th>5(^{th}) Ca</th>
<th>6(^{th}) Ca</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
By comparing the shape of the fitted curves, we determined that the best fit for a binding pattern is 1-3-1-1. We plotted the fits together with the data in Figure 5.2. In addition, we obtained binding affinities for each individual site in the 1-3-1-1 binding profile.

Parallel studies are needed to generate complementary information; for example, if we knew the binding order, we may be able to resolve more fully the complicated calcium binding system. The Thompson lab has investigated a calcium titration experiment by using X-ray crystallography (unpublished) to estimate the affinities. To define the calcium activation pathway and determine the binding order of six calcium binding sites, they generated an eight-point crystallographic dose response curve. PAD2 crystals were soaked with increasing concentrations of calcium (0-10 mM), including the concentration required for half-maximal activity ($K_{0.5} = 180 \mu$M) as determined from an enzymatic activity assay. The peak electron density for each calcium ion was quantified as a reporter in the titration and normalized against the density of a conserved water molecule. Using the binding order from the crystallography data, we assign that Ca6 has a $K_d$ of less than 3 $\mu$M whereas Ca3, 4 and 5 sites have affinities approximately 420 $\mu$M. Ca2 has an affinity of 1 mM whereas Ca1 has an affinity around 8 mM. These results are in approximate agreement with the crystallography data (Table 5.3) except on Ca1 binding region, even though the crystallography data refer to a solid-state structure.

We assign the lowest binding affinity (8 mM) for Ca1 because it agrees with the differential HDX results. At 2.5 mM Ca$^{2+}$ concentration, only 24% of PAD2 bound to Ca$^{2+}$ at Ca1 binding region, which may not be sufficient to be observed by differential HDX. Unfortunately, peptides that are involved in Ca1 binding region do not provide enough MS signal. Thus, we are not able to use these peptides for PLIMSTEX analysis. It is noted that according to the unpublished X-ray
Table 5.3 Comparison of the data between PLIMSTEX and X-ray crystallography (unpublished)

<table>
<thead>
<tr>
<th>Ca binding regions</th>
<th>Binding affinity measured by PLIMSTEX</th>
<th>Binding affinity measured by X-ray crystallography</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ca6</td>
<td>&lt; 3 μM</td>
<td>&lt; 1 μM</td>
</tr>
<tr>
<td>Ca3, 4, and 5</td>
<td>420 μM</td>
<td>280 μM</td>
</tr>
<tr>
<td>Ca2</td>
<td>1 mM</td>
<td>N/A</td>
</tr>
<tr>
<td>Ca1</td>
<td>8 mM</td>
<td>630 μM</td>
</tr>
</tbody>
</table>
crystallography data from the Thompson lab, Ca1 binding behavior is affected by the TCS. Thus, it is reasonable to postulate that the PLIMSTEX curve in Ca1 binding region, if can be monitored, will have similar shape as the others. In addition, it is very suspicious of the biological relevant function of Ca1 binding site if it has such a low binding affinity.

Furthermore, we did analysis of two different ratio of [EDTA]/[protein], 36 and 198, to assure a that the apo state of PAD2 was completely free of Ca$^{2+}$. We observed no significant changes in any peptide when EDTA was added. The deuterium uptake levels for Ca 2, 3/5, 4, and 6 were approximately 47%, 42%, 38%, and 26%, respectively, for all three conditions: no EDTA, [EDTA]/[protein] ratio = 36 and = 198. These outcomes suggest that either the native state PAD2 (the original sample state) is pure apo state and does not bind to Ca$^{2+}$ ions or that the first binding of Ca$^{2+}$ to PAD2 is sufficiently strong that it is unaffected by EDTA. If such a high-affinity binding of Ca$^{2+}$ to the protein occurred, we cannot determine the binding affinity with present mass spectrometry-based technology even if we use more EDTA to remove all the calcium ions from the protein.

5.5 Conclusion
In this study, we applied HDX-based mass spectrometry biophysics approach to a protein calcium-binding system. This study is challenging because six calciums bind to PAD2 and their binding behaviors are highly cooperative. We modified the PLIMSTEX fitting program was respond to the complexities of PAD2 and then applied it to investigate the Ca$^{2+}$ binding affinities for different sites on the protein. Although the results are not completely convincing, it is worth noting that this would be difficult by any other equilibrium tool except for X-ray crystallography.
Unlike other calcium binding proteins we studied in our group\textsuperscript{11-14}, PAD2 does not have a typical EF-hand motif for each binding site. The cooperative binding to calcium ions of PAD2 has been identified by previous studies including differential HDX-MS (Chapter 4). All these factors in the PAD2 system pose difficulties for our PLIMSTEX experiment. The PLIMSTEX curves for each binding site indeed show similar trends and are consistent with the cooperative binding scheme we propose. Combining the data from the X-ray crystallography-based binding order studies done in the Thompson lab, we successfully obtained binding affinities for each binding sites, and confirmed that Ca6 has the highest binding affinity whereas Ca1 has the lowest.

We envision the peptide level PLIMSTEX to be useful for other protein ligand binding systems. It certainly provides more detailed information than global-level PLIMSTEX. In addition, a bootstrap strategy can be used in data analysis to obtain precision of binding affinities. Further investigations including electron capture or electron transfer dissociation can also be carried out to give even residue-level information, giving this approach unique capability.

5.6 Acknowledgements
I acknowledge Dr. Daniel Slade for PAD2 synthesis and purification, and Don Rempel for modeling.
5.7 References

Chapter 6: Structural Analysis of Diheme Cytochrome c by Hydrogen Deuterium Exchange Mass Spectrometry and Homology Modeling

This work is part of the following publication:

6.1 Abstract
Chapter 4 and 5 describe development and application of HDX-MS approaches on protein or protein-ligand complex based on known structures. A lack of X-ray or NMR structures of proteins, however, can inhibit their further study and characterization, motivating the development of new ways of analyzing structural information without crystal structures. The combination of hydrogen deuterium exchange mass spectrometry (HDX-MS) data in conjunction with homology modeling (HM) can provide improved structure and mechanistic predictions. Here a unique diheme cytochrome c (DHCC) protein from *Heliobacterium mosdesticaldum* is studied with both homology modeling and HDX to bring some definition of the structure of the protein and its role. Specifically, HDX data were used to guide the homology modeling to yield a more functionally relevant structural model of DHCC.

6.2 Introduction
Cytochromes c are essential metalloproteins in the electron-transfer chain of most living organisms, including all photosynthetic taxa. Metalloproteins move electrons around the cell to produce energetic compounds that drive cellular metabolism. Many of the photosynthetic electron-transfer proteins of interest belong to the cytochrome c family. Understanding their structure and function will contribute not only to understanding photosynthesis, its evolution, and its role in early Earth history but also to harnessing solar energy.

Cytochrome c’s are electron-transport hemoproteins that are covalently bound through two thioether bonds between the vinyl groups of a heme macrocycle and the sulfhydryls of a CXXCH heme-binding motif. Two heme propionate groups are exposed and can participate in hydrogen bonding. The main function is to mediate single-electron transfer reactions between protein electron donors and acceptors via reversible oxidation/reduction of Fe$^{2+}$ and Fe$^{3+}$. The
axial coordination of the heme iron and its effect on other parts of the larger protein environment play key roles in determining redox, electron transfer, and other properties of cytochrome c\textsuperscript{2,3}.

Prokaryotic diheme cytochrome c’s, as representatives of electron-transfer proteins with two iron centers, have been studied mainly with various spectroscopic methods\textsuperscript{4-6}. For example, cytochrome c\textsubscript{4} from \textit{Pseudomonas stutzeri}, which shares the same origin and has a similar amino-acid sequence as DHCC\textsuperscript{7}, possesses a strong hydrogen bond involving the heme propionate groups, and the protein facilitates intramolecular, inter-heme electron transfer\textsuperscript{8}. Both hemes contain hexa-coordinated iron involving histidine and methionine axial ligands. The hydrogen bond in \textit{P. stutzeri} closes one edge of the heme, allowing it to act as the donor and the other as the acceptor.

DHCC from \textit{Heliobacterium modesticaldum}, an early-evolving, gram-positive phototrophic anaerobic prokaryote, plays a role as the terminal electron acceptor in the high-potential electron-transfer chain of the cytochrome \textit{bc} complex. This complex participates in the photosynthetic electron-transport cycle by oxidizing the quinone pool, sending electrons to the reaction center and pumping protons to establish an electrochemical gradient that drives ATP production. DHCC replaces the canonical monoheme cytochrome c\textsubscript{1} in the cytochrome \textit{bc}_{1} complex, or cytochrome \textit{f} in the cytochrome \textit{b}_{6f} complex\textsuperscript{9,10}. DHCC has two \textit{c}-type hemes that are predicted to reside in similar \textit{c}-type cytochrome folds. The heme 1 domain is closer to the N-terminal helix, whereas the heme 2 domain begins approximately halfway through the sequence. The unusual diheme architecture is an interesting contrast with those of the more common monoheme cytochromes. In general, an electron transfer of one metal center induces conformational changes in the site of the other, and that usually enhances the rates of subsequent steps\textsuperscript{11}. Studying the structure, function, and redox properties of this protein is crucial to the
understanding of the function of the entire heliobacterial cytochrome bc complex. Furthermore, an understanding of the *Heliobacterium* electron transport chain (ETC), which is simpler and smaller than later evolved organisms, can provide clues to how life and photosynthesis evolved on Earth and possibly on other planets. Similarly, the simple nature of the ETC makes it an attractive target for biomimicry in solar energy production.\textsuperscript{12-15}

Redox-dependent conformations of cytochrome c’s have been studied extensively for decades by using various biochemical and biophysical approaches\textsuperscript{16-23}. Although some comparative X-ray crystallographic studies of tuna-heart cytochromes c show little or no difference between the backbone structures\textsuperscript{24}, almost all solution-based studies show a clear conformational change between the two redox states\textsuperscript{25-27}. Specifically in the case of horse heart cytochrome c, the radius of the oxidized state is significantly larger than that of the reduced one\textsuperscript{25}. Despite the success of these approaches, there is a need for other methods. For example, many biophysical techniques usually require a large amount of sample\textsuperscript{28-30}, which is not always available, and their structural resolution is relatively low\textsuperscript{28}. In contrast, mass-spectrometry (MS) approaches have high sensitivity and moderate structural resolution. For studying protein conformations, hydrogen deuterium exchange (HDX) coupled with MS is an effective and now commonly employed approach\textsuperscript{31-35}. The exchange rates for HDX are good readouts for hydrogen bonding and solvent accessibility at global, peptide and even amino acid levels\textsuperscript{36}. It is also possible to use HDX data to adjudicate protein subunit docking and ligand-binding structural models\textsuperscript{37-39}.

In HDX studies, one often compares the exchange rate of the same protein regions for two different states that are produced by some protein perturbation (e.g., ligand binding, change in redox status, formation of a complex). Although each amino acid has different intrinsic deuterium exchange rate constants, this causes no problem because the experimental exchange
rates between two states of the protein are compared. Previously, HDX coupled with infrared spectroscopy was used to study the conformations of reduced and oxidized horse heart cytochrome \(c^{40}\). Viala and co-workers\(^{41}\) reported an HDX-MS study of the same system, indicating of a more open structure in the oxidized state. Thus far, however, there are no HDX-MS studies of the DHCC system. To test our proposed mechanism of the DHCC subunit of the heliobacterial cytochrome \(bc\) complex, which involves a more closed conformation upon reduction\(^{42}\), we designed an HDX study of both the reduced and oxidized states of the heliobacterial DHCC in an effort to achieve peptide-level resolution of structural changes.

Homology Modeling (HM) is another common approach to protein structure and function especially when data from other structural approaches (e.g., X-ray crystallography and NMR) are not available. Although HM statistics for assigning model quality do not necessarily reflect the in vivo structure of the protein, the approach should be more accurate when coupled with experimental approaches. HDX-MS with HM was used previously to investigate docking and binding studies in combination with Electron Microscopy\(^{43-46}\). These semi-quantitative approaches rely on heavy computational techniques and assess quaternary structure of the protein. Here, we are using HDX-MS results to direct HM to obtain a low resolution understanding of protein tertiary structure and to predict a reaction mechanism for the complex.

Our method relies on readily available applets for modeling and HDX for making qualitative structural distinction between a buried or exposed region to adjudicate the models, quite unlike the previous studies that use HDX to study the docking of ligands or subunits. Success will not only provide insight to DHCC but also to establish the utility of the combined approach for other protein systems.
6.3 Materials and Methods

6.3.1 Protein Expression and Purification
The diheme cytochrome c subunit from *Heliobacterium modesticaldum* bc complex was expressed in *Escherichia coli* with the cytochrome c insertion plasmid and purified as previously described, yielding only the holo-protein\(^2\). The N-terminal transmembrane helix region was truncated to express only the fully soluble portion of the protein. Post-processing did not eliminate the hemes, both being covalently bound to the protein.

6.3.2 HDX of Reduced and Oxidized DHCC
Differential, solution HDX experiments were performed at 4 °C. DHCC in 20 mM Tris-HCl buffer, pH 7.5 solution was incubated with either sodium dithionite (reductant), or potassium ferricyanide (oxidant) at a final concentration of 25 μM for both the protein and the reductant/oxidant. The samples were equilibrated for 1 h prior to initiating HDX. Continuous labeling was initiated by incubating 2 μL of the mixture with 18 μL of D\(_2\)O for seven predetermined times (0.17, 0.5, 1, 2, 15, 60, and 240 min). The exchange reaction was quenched by mixing the D\(_2\)O solution with 30 μL of 3 M urea, 1% trifluoroacetic acid (TFA) at 0 °C. The quenched solution was then flash-frozen in liquid nitrogen and stored at -80 °C for less than 48 h prior to analysis.

The quenched sample was then digested by passing it through a custom-packed pepsin column (2 mm x 2 cm) at 200 μL/min; the peptides were captured on a C\(_8\) trap column (2 mm x 1 cm, Agilent Inc., Santa Clara, CA, USA) and desalted with a 3 min flow. Peptides were then separated by using a C\(_{18}\) column (2.1 mm x 5 cm, 1.9 μm Hypersil Gold, Thermo Fisher Scientific, Waltham, MA, USA) with a 5 min linear gradient of 4% - 40% CH\(_3\)CN in 0.1% formic acid. Protein digestion and peptide separation were carried out in an ice-water bath to
minimize back exchange. All analyses were with a hybrid LTQ Orbitrap (Thermo Fisher Scientific, San Jose, CA, USA) at an ESI capillary temperature of 225 °C. Each experiment was performed in duplicate.

6.3.3 Peptide Identification and HDX Data Processing
All of this information is covered in the previous chapter (Chapter 4).

6.3.4 Homology Modeling and Model Refinement
Homology modeling of DHCC was carried out with three different modeling platforms. The first was with Modeller\textsuperscript{47} v9.10 and v9.12 using alignment algorithms ClustalW2 and ClustalΩ, respectively, with both pdb files 1ETP\textsuperscript{8} and 3MK7c\textsuperscript{48} generating 100 models from each run. The best models from each run were selected based on the molpdf, GA341 and DOPE scores. The best model from each run was then tested by taking the experimental HDX data for each peptide as a measure of the secondary and tertiary structure of that portion of the protein. The amount of exchange in each section was compared to the secondary and tertiary structure predicted by the model. For regions that disagreed, the template to DHCC alignment was manually adjusted and another round of models with the new alignment was generated until the best match was obtained with the HDX data.

The second approach used models that were generated using the Phyre2\textsuperscript{49} and I-TASSER\textsuperscript{50} on-line modeling suites. The models generated from these two suites were compared to each other because they do not allow for manual adjustment of the alignments or parameters or further analysis options. Disorder prediction was run on the best model in Phyre2. The top resulting model was compared to the HDX data and to the Modeller generated models.
The third approach was built on previous work$^{10}$ where a complete model of the cytochrome $bc$ complex was built using the Phyre2 server to analyze the whole complex structure and predict mechanism for electron transfer within the complex.

6.3.5 Solvent Accessible Surface Area (SASA) Assessment of Homology Models
SASA was calculated for all four homology models using the program Get Area$^{51}$. The SASA for the backbone only was chosen as the value to be used in our assessment because the only HDX that was counted was of the backbone hydrogens not the whole residue. The backbone SASA was calculated for each residue and is reported in square Angstroms. To compare to the HDX data, the SASA was averaged for all residues in each of the MS peptic peptides. The percent of deuterium uptake was reported calculated by taking the average of the initial and final point.

6.4 Results and Discussion
Prior to conducting HDX, we tested whether charge-state distributions of the intact protein could provide coarse structural data to distinguish the reduced and oxidized states of DHCC. Experiments revealed that the conditions of native ESI are sufficiently oxidative to preclude obtaining mass spectra for the pure reduced species. Expression of the protein with the covalently bound heme precluded any comparisons with the apo state. Thus, we turned to HDX footprinting to obtain structural data for the DHCC protein.

6.4.1 HDX Shows the Oxidized Form of DHCC is More Flexible
An on-line pepsin digestion of expressed and purified DHCC afforded hundreds of peptides, of which 69 detected with high signal/noise ratio and validated MS/MS information were chosen
Figure 6.1 HDX peptide coverage map for DHCC showing 95% coverage. The same peptides with different charge states are also shown.
for the subsequent peptide mapping. The coverage is 95% (shown in Figure 6.1), allowing us to assess nearly fully the protein in both of the redox states. As mentioned before, no correction for back exchange of the protein was made because all data were used in two-state comparisons of the redox-state samples under identical conditions. All samples are subjected to back exchange after quench and processing, and this converted any side chain deuteriums (e.g., on -NH$_2$) back to hydrogens. Thus, it is the back-bone deuteriums that remain on the polypeptide for the LC/MS analysis. Changes in protein dynamics, as reflected by HDX, reflect changes to the protein induced by reduction. The resulting perturbations alter the conformation of the protein backbone and its H bonding as reflected by changes in HDX. Therefore, all fluctuations in HDX rates reflect only reduction-induced changes. For example, we expect that less constrained regions should exhibit relatively fast HDX, whereas structured and buried regions exchange slowly.

We plotted the HDX kinetic curves for each of the peptides from both the reduced (Figure 6.2, solid lines) and oxidized (Figures 6.2, dashed lines) states of the DHCC. When peptides of different charge cover the same region, we found that the results were nearly identical, indicating that the HDX results have good precision. More importantly, we conducted each experiment in duplicate and show the error bars (average deviations) in each plot. To show more clearly the precision, we enlarged one of the typical HDX plots to show the precision of the experiments (Figure 6.3). In summary, all the HDX experiments afforded a 0-3% relative average deviation for duplicate determinations.

The extents of HDX for most regions of the protein are different for the two states (oxidized and reduced), indicating that the structural conformation of DHCC is redox-state-dependent. For those regions that show differences, the oxidized state always shows greater deuterium uptake compared to the reduced state, indicating that the oxidized state is more flexible (less structurally
**Figure 6.2** Kinetic curves of all the peptides used for HDX mapping for reduced (solid line) and oxidized (dash line) states of the DHCC. The peptides involved in heme c binding pockets are highlighted in red frames. Numbers in parentheses with “+” sign are the charge states of the peptides.
Figure 6.3 Selected HDX kinetic curves for demonstration of the precision for each experiment. The curves are typical through the sequence.
constrained). This is consistent with other horse-heart-cytochrome c HDX studies that show that the oxidized state is more dynamic and flexible than the reduced. The largest differences in HDX kinetics occur near the heme-binding CXXCH motif (i.e., for peptides 1-29 and 109-127) (Figure 6.2). These differences indicate that conformational opening in the heme-binding region accompanies oxidation. This result suggests that redox-dependent conformational changes begin at the metal center. Comparing the changes in HDX for the oxidized and reduced states at the two heme binding pockets, we find similar outcomes. Specifically, the oxidized protein undergoes between 30 and 50% HDX at the two sites whereas the reduced protein undergoes between 20 and 40% exchange (Figure 6.2). Our previous reported potentiometric titration of DHCC shows a single observable midpoint potential, which is consistent with the conclusion that the two heme binding pockets have similar chemical environments.

6.4.2 HDX Reflects Secondary Structure
Although four regions display no changes upon reduction (e.g., 45-49, 130-138, 143-149, and 169-172, Figures 6.2), they can be viewed as controls because they indicate that regions insensitive to oxidation do not show differences in HDX. Importantly, these regions also show no detectable HDX as a function of time, indicating that these regions are heavily H-bonded and/or buried in the protein. We assign these four regions to be involved in tertiary structural helix bundles (Category I). We take this result to indicate that the core structural components of DHCC do not change dramatically upon reduction, whereas flexible regions are more likely to be affected and participate in the conformational changes that accompany redox changes. In contrast, region 173-208 undergoes extensive exchange, even at short time (Figure 6.2). We categorize regions of this type to be part of loops (Category II).
Regions that constitute a third category also show increasing HDX with time but at a more modest extent (none showed 90% or higher deuterium uptake). This lower HDX occurs in part because the peptides are long and likely cover regions containing both H-bonded helices and flexible loops, thus reporting an “average” of these structural elements. For example, the peptides for region 78-108 show deuterium uptake ranging from 20-50% for the reduced state and 30-60% for the oxidized state (Figure 6.2). The helix region undergoes less HDX, whereas the peptide bonds in the flexible loop exchange more readily. Similarly, regions 78-91 and 92-108, also comprise helix and loop structures and, thus, show an intermediate level of deuterium uptake (~ 60%). This intermediate exchange also pertains to region 109-127, and 150-168.

We also identified a fourth category where HDX at the shortest times in near zero but increases slowly with time (Figures 6.2). We assigned these regions, including 38-44, 50-57, and 128-142, to be likely helices but not part of buried helix bundles.

6.4.3 Homology Model Building and Assessment
A series of homology models were generated with the software Modeller and PHYRE2 (Figure 6.4 and Table 6.1). To begin, we chose template 1ETP because it is a bacterial cytochrome c₄ protein, and the evolutionary origin of DHCC may be the cytochrome c₄ protein family. Likewise, 1ETP undergoes electron transfer from a donor to an acceptor protein. We selected 3MK7c as the second template on the basis of bioinformatic analysis. A BLAST search of the Protein Data Bank (PDB) with DHCC’s sequence returns 3MK7c as the most similar protein by alignment statistics. Functionally, 3MK7c is a soluble member of a protein complex, but it resides in the cytoplasm and accepts electrons from a cytoplasmic donor and shuttles them into the complex, the opposite direction from which DHCC should function. Both 1ETP and 3MK7c have hemes with two different redox potentials. Although the generated homology
Figure 6.4 Homology models A to D. Regions that are highlighted in red do not agree with HDX results.
Table 6.1 Homology modeling data.

<table>
<thead>
<tr>
<th>Model</th>
<th>Template</th>
<th>Alignment Algorithm</th>
<th>Percent Similarity</th>
<th>Percent Identity</th>
<th>Modeling Software</th>
<th>Molpdf score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model A</td>
<td>1ETP</td>
<td>ClustalW2</td>
<td>49% (100/205)</td>
<td>15% (32/205)</td>
<td>Modeller v9.10</td>
<td>0.24 (0.49-highest)</td>
</tr>
<tr>
<td>Model B</td>
<td>1ETP</td>
<td>Clustal Omega</td>
<td>45% (92/205)</td>
<td>16% (33/205)</td>
<td>Modeller v9.10, v9.12</td>
<td>0.12 (0.28-highest)</td>
</tr>
<tr>
<td>Model C</td>
<td>3MK7c</td>
<td>Clustal W2</td>
<td>57% (116/205)</td>
<td>28% (68/205)</td>
<td>Modeller v9.10</td>
<td>0.46 (highest)</td>
</tr>
<tr>
<td>Model D</td>
<td>3MK7c</td>
<td>Phyre2</td>
<td>57% (116/205)</td>
<td>28% (58/205)</td>
<td>I-TASSER, PHYRE2</td>
<td>3.2Å Resolution N/A</td>
</tr>
</tbody>
</table>
models do not incorporate the cofactor heme by nature of the program, the templates we chose do include the cofactors within the protein. Even though we did not add the hemes in the HM modelling, the protein structure should still be representative. Functional considerations of the template are a way to assess model quality.

Homology Modeling (HM) statistics assess model quality in a more quantitative way. The percent similarity and percent identity examine the alignment of the template and model sequences. Amino-acid residues that align are examined to see if they are identical or in the same category. The percent identity increases from Model A to Model B as a result of manual alignment changes implemented on the basis of the HDX data. Further increases in both percent similarity and percent identity occur for models C and D, which use 3MK7 as a template (Table 6.1). The higher these values, the more similar a model is to the template, and the more faithful is the homology model. The 28% identity achieved for Models C and D is just under that preferred for HM, but it is sufficient for further consideration. We also view it as acceptable because there are so few diheme cytochromes with solved structures in the PDB. After running HM in Modeller, score reports are generated for each model. The molpdf score can be used to compare models across runs. Model C has the highest molpdf score indicating it is the better homology model. Model B may be a better model functionally as discussed below, but it has a lower molpdf score because manual alignment changes were made (Table 6.1). Homology modeling data indicate that Models C and D are better models; therefore, 3MK7c is the better template. This result is only based on alignment and threading algorithms, however, and does not speak to function or biochemical properties of the models and templates. Thus, biochemical and/or biophysical data are needed to assess further these models.
6.4.4 HDX Analysis of DHCC with Homology Models

HDX data generated by MS can provide biophysical data that, when mapped onto the four homology models, should be able to test the correctness of secondary and tertiary structures generated by the models (Figure 6.4 and Table 6.2). For Model A, most of the regions that show discrepancies between the HDX results and the HM outcomes are the helical regions (Figure 6.4A and Table 6.2). For example, peptides in Category I show almost no deuterium uptake in either of the states (Figures 6.2), suggesting they participate in forming tertiary structure, most likely a helix bundle. In Model A, however, only region 143-149 has the helix-bundle motif. Peptides 45-49 and 130-138 are modelled to be a loop or a singular helix structure. After manually adjusting the alignment based on this observation, Model B and C show better fits for 45-49, 143-149, and 169-172, as these regions are predicted to be helices in helix bundles. Furthermore, Model A HM did not identify the second heme binding motif and the alignment and subsequent modeling are poor in this region, 109-127.

To resolve regions for which there is a lack of agreement, we generated Model B (Figure 6.4B and Table 6.2) by manually adjusting the alignment based on the results of Model A as evaluated by HDX. Model B showed improved agreement between HDX and the HM structure. The helix bundle and 2nd heme regions predicted by Model B are consistent with the experimental results, but achieving these improvements sacrificed other regions that had agreed with Model A. For example, 33-37 shows as a loop, but it takes up only ~50% of possible deuterium. On the contrary, a mostly loop region, 173-208, does show ~70% HDX. In addition, residues 173-205 show the poorest agreement with HDX results compared to other models, as the sequence does not align with the template. 1ETP is a smaller protein, and in adjusting for the heme regions, residues at the end of DHCC were left without template residues. As a result, this region is
### Table 6.2 Comparison of HDX data and Homology models.

<table>
<thead>
<tr>
<th>Region</th>
<th>Category</th>
<th>Model A</th>
<th>Model B</th>
<th>Model C</th>
<th>Model D</th>
<th>HDX agreement on the models&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-29</td>
<td>III</td>
<td>h/l</td>
<td>h/l</td>
<td>h/l</td>
<td>h/l</td>
<td>**          **          **          **</td>
</tr>
<tr>
<td>33-37</td>
<td>III</td>
<td>h/l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>**          *</td>
</tr>
<tr>
<td>38-44</td>
<td>IV</td>
<td>h bundle</td>
<td>h/l</td>
<td>h/l</td>
<td>h/l</td>
<td>*</td>
</tr>
<tr>
<td>45-49</td>
<td>I</td>
<td>h</td>
<td>h bundle</td>
<td>h/l</td>
<td>h bundle</td>
<td>**          **</td>
</tr>
<tr>
<td>50-57</td>
<td>IV</td>
<td>l</td>
<td>l</td>
<td>h bundle</td>
<td>h/l</td>
<td>*</td>
</tr>
<tr>
<td>58-65</td>
<td>II</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>**          **          **          **</td>
</tr>
<tr>
<td>78-91</td>
<td>III</td>
<td>h/l</td>
<td>h/l</td>
<td>h/l</td>
<td>h/l</td>
<td>**          **          **          **</td>
</tr>
<tr>
<td>92-108</td>
<td>III</td>
<td>h/l</td>
<td>h/l</td>
<td>h/l</td>
<td>h/l</td>
<td>**          **          **          **</td>
</tr>
<tr>
<td>109-127</td>
<td>III</td>
<td>l</td>
<td>l</td>
<td>h/l</td>
<td>h/l</td>
<td>**          **</td>
</tr>
<tr>
<td>128-142</td>
<td>IV</td>
<td>h/l</td>
<td>l</td>
<td>l</td>
<td>l</td>
<td>*</td>
</tr>
<tr>
<td>143-149</td>
<td>I</td>
<td>h bundle</td>
<td>h bundle</td>
<td>h</td>
<td>h bundle</td>
<td>**          **          *</td>
</tr>
<tr>
<td>150-168</td>
<td>III</td>
<td>h/l</td>
<td>h/l</td>
<td>h/l</td>
<td>h/l</td>
<td>**          **          **          **</td>
</tr>
<tr>
<td>169-172</td>
<td>I</td>
<td>h</td>
<td>h bundle</td>
<td>h bundle</td>
<td>h bundle</td>
<td>**          **          **          **</td>
</tr>
<tr>
<td>173-208</td>
<td>II</td>
<td>h/l</td>
<td>l</td>
<td>h/l</td>
<td>h/l</td>
<td>**          **          **          **</td>
</tr>
</tbody>
</table>

**a.** Regions that determined by HDX. Most of the regions are covered by multiple peptides, and they give similar results.

**b.** Different regions are signed in four categories based on their HDX behavior.

**c.** Secondary/tertiary structure elements in each model. “h” stands for helix, “l” stands for loop, “h/l” stands for mixture of helix and loop, and “h bundle” stands for helix bundle.

**d.** Measures of agreement degrees between HDX and modeling results. “**” stands for agreement, “*” stands for agreement with restrictions (information from order prediction by Phyre2), and blank cells represent lack of agreement.
unstructured according to Model B. Even though this region shows the highest HDX for the whole protein (~70%), it still should have some secondary structure. Despite these areas of disagreement, Model B is better than Model A because the active regions of the protein are consistent with the outcomes of both HDX and HM. Overall, the transition from Model A to Model B demonstrates that HDX data can inform the manual refinement and improve homology models.

Model C (Figure 6.4C and Table 6.2) was developed in the same manner as Models A and B, but used 3MK7c as a template. Model C gives better agreement for the second heme-binding region, 109-127, than do Models A and B (Table 6.2). It also decreased the number of regions of disagreement. The model points to a highly open structure, however, and it does not form the expected two cytochrome c-type folds as seen in the other models. Thus, Model C, although affording better agreement in the second heme region, is not a good model overall for DHCC.

Model D (Figure 6.4D and Table 6.2) was generated from the online suites I-TASSER and Phyre2, and nearly identical results were obtained from running both I-TASSER and Phyre2. Both packages identified 3MK7c as the best template, gave highly similar alignments and consequently identical models. Model D is presented as that obtained by using Phyre2. It is consistent with the HDX results in all regions of the subunit except 38-44, 50-57, and 130-138 (Figure 6.4D and Table 6.2). These regions are loops in Model D; however, they undergo lower HDX, more typical of a rigid region. The Phyre2 software includes a disorder prediction, identifying secondary and tertiary structure (Figure 6.5), and a disorder prediction analysis was performed on Model D. More order is predicted for all three regions of disagreement, suggesting that, although they are not alpha helices, there is more order in the region than can be explained by a loop. The order in these regions is likely caused by the interaction of the chains across the
Figure 6.5 Model D with disorder prediction data. Warmer color and larger number indicates more disordered region. Color codes are shown on the right edge.
domains, as these regions are in the middle where the two domains interact. Thus, the regions of disagreement are actually areas of partial agreement. The region containing residues 130-138 is the only region that is inconsistent with all four models and across the three software packages, suggesting that the region is different than that seen in the templates. Taking all the evidence together, we conclude that Model D best agrees with the HDX data.

6.4.5 SASA Calculation Shows Model D Correlates Best with the HDX Data
A correlation of SASA at the peptide level (calculated as an average of the SASA for the individual amino acids comprising the region) and the average extents of HDX for the corresponding peptide regions confirmed our assignment of model D as the best homology model of DHCC (Figure 6.6). This model gave the best correlation coefficient in comparison to correlations with Models A-C. In fact, the SASA analysis only disagreed in only one area with Model D. Model C has the most open structure, thus with larger SASA values. Overall, the SASA analysis is a more quantitative assessment to illustrate that Model D is the best model presented thus far.

6.4.6 HDX Data Mapping Onto Model D
To view the two redox states at the protein level, we mapped the HDX differences between the two redox states of DHCC taken as an average of the differences for each time point onto homology Model D (Figure 6.7). The most prominent difference occurs for the short helix in the heme 2 domain (colored in blue in Figure 6.7). In agreement with the kinetic traces, most helical and loop regions around the heme sites have larger differences, indicating that structural changes accompany the redox chemistry. Adding credibility to this conclusion is that the cores of the two helix bundles show no significant changes.
Figure 6.6 Correlation of HDX and SASA data for all models A (A), B (B), C (C) and D (D). The Pearson correlation coefficients $R$'s are shown in red.
Figure 6.7 The deuterium uptake differences between the oxidized and reduced states of DHCC are mapped onto Model D. Color codes show the differences (reduced state subtracted by the oxidized state).
a. Reduced state HDX kinetics

0.17 min 0.5 min 1 min 2 min 15 min 60 min 240 min

No Data
<0.1
<0.2
<0.3
<0.4
<0.5
<0.6
≥0.6

Reduced state HDX kinetics

2 min 15 min 60 min 240 min

No Data
<0.1
<0.2
<0.3
<0.4
<0.5
<0.6
≥0.6
Figure 6.8 The deuterium uptake levels for all peptides mapped onto Model D with color for each exchange point. The figures for oxidized state are in (a) and reduced in (b). Color codes are shown on the right edge.
We also mapped the extent of HDX for all the peptide regions onto our best model, Model D, with a color scheme indicating the amount of exchange at each time point (Figure 6.8). Instead of an overall picture of kinetics information as in Figure 6.7, these maps show the dynamics of the protein at each time point. By comparing the same state at different times, we see that most of the regions of the oxidized state are dynamic, acquiring more and more deuterium as seen by the increasingly warmer colors.

General features of the protein’s secondary structure are seen by the different HDX extents. Loop regions achieve either a higher level of HDX than helices or a similar level at an earlier time. Those loops undergoing less HDX are more ordered and constrained by interactions within the protein. The core regions within the helix bundle are always seen as cold colors, indicating they are so inflexible that HDX is small even after 4 h of incubation. The data reveal important trends about the impact of changing redox states as well as the secondary and tertiary structural and solvent accessibility changes of DHCC.

6.4.7 Proposed Function of DHCC
HDX provides kinetics and dynamics information on the peptide-level of DHCC in both reduced and oxidized states, whereas HM provides a potential protein structure even in the absence of a crystal structure. HM can only yield models that are as good as the information entered, and they cannot provide the high-resolution structural information of a crystal structure. The approach described here takes HM statistics and complements them with HDX results to assess model agreement. We used HDX data to refine Model A and correct areas of disagreement to generate Model B. Likewise, we used HDX to determine if Model B were more informative, and we found that it is in better agreement with HDX than Model A, demonstrating the utility of HDX data for testing and refining a homology model. Comparing all four models, we found that
Model D had the highest HM statistics and the best agreement with HDX, including a good correlation with SASA. We subsequently selected it as the best model for the protein structure.

Functionally, the choice of templates will bias the model predicted. The DHCCs 1ETP and 3MK7c are the best templates for structural alignments and are both soluble proteins that contain two $c$-type hemes. Neither protein has been chemically oxidized or reduced; thus, the template structure is probably a mix of both forms. 1ETP, from *Pseudomonas stutzeri*, is a periplasmic diheme cytochrome $c_4$. It does not belong to a complex but instead shuttles electrons between complexes in an electron-transport chain. The two hemes are located in different protein environments, yielding two distinct redox potentials and allowing for inter-heme electron transfer. The protein 3MK7c, from *P. stutzeri*, is a subunit of the cbb$_3$ cytochrome oxidase. Like DHCC, 3MK7c is a soluble subunit with an N-terminal anchor. The cytochrome cbb$_3$ complex, however, moves electrons in the opposite direction of the cytochrome bc complex. As for 1ETP, 3MK7c has two distinct redox potentials, but they differ to a greater extent than do those of 1ETP. Inter-heme electron transfer is still observed. Based on their biological roles, 3MK7c is closer functionally to DHCC. This agrees with the observation that Model D accommodates better the HM parameters and the HDX data.

Numerous solution-based spectroscopic and computational studies have shown that there is a conformational change between the reduced and oxidized states of horse heart cytochrome $c^{40,41}$. These results are in apparent conflict with high-resolution crystallographic data that show the crystal structures of the two states are similar$^{52}$. One explanation is that the solid-state X-ray structure does not reflect the solution structure, which is likely to be dynamic. It was proposed that the oxidized state undergoes lower-frequency and larger-amplitude motions than the reduced state$^{53}$. Our HDX results for DHCC suggest that there are significant structural differences.
between the two different redox states. It is also possible that both X-ray structures represent the reduced form of the protein, produced by X-ray irradiation. Just as the HDX kinetics suggest a redox-dependent conformational change for DHCC, they also indicate areas that are likely more solvent-accessible (less H-bonded). The heme 2 domain particularly exhibits faster exchange. It follows that the heme 2 domain interacts with the soluble electron acceptor cytochrome $c_{553}$. Similarly, in vivo, the heme 1 domain would have an N-terminal helix anchoring the protein to the $b$ and IV subunits of the $bc$ complex. Heme 1 would also then interact with the Rieske subunit. These subunit-subunit interactions would lower the HDX rates in the heme 1 region, as is observed.

For monoheme cytochromes, electron transfer is a two-step process. After the electron donor binds with and reduces the cytochrome, the cytochrome binds the electron acceptor and passes the electron to it. DHCC, as other diheme cytochrome $c$’s, adds one more step, the intramolecular inter-heme electron transfer. In DHCCs from other species, the two hemes usually possess two distinct macroscopic redox potentials (e.g., 240 mV and 330 mV for cytochrome $c_4$ from $P. stutzeri$). The values are consistent with the negative and positive electrostatic charges of the two domains$^{11}$. The direction of the electrostatic field within the protein gives rise to the distinct heme redox potentials and directs the flow of the electron from the lower to higher potential heme. The DHCC from $Heliobacterium$ cytochrome $bc$ complex, as studied here, however, has only one redox mid-potential$^{42}$, indicating similar chemical environments around the two heme-binding pockets. The HDX kinetics data are consistent with this observation. The significant conformational changes upon reduction, as revealed by HDX, may work to facilitate an inter-heme electron transfer despite the observation of a single mid-point potential in vitro. Our results point to formation of a closed structure upon reduction of the protein, providing a
closer distance between the two hemes. In *P. stutzeri*, the inter-heme electron transfer occurs upon hydrogen-bond formation between the heme propionate groups and a surrounding hydrogen-bond network established by a conformational change to the reduced state, producing the closer heme-heme distance and an altered heme environment. Such alterations change the redox potential and facilitate inter-heme electron transfer. It is likely that only one redox potential is observed *in vitro* because the heme environments are sensitive to the truncated N-terminal helix and subunit interactions that are missing in the expressed DHCC. Thus, our prediction is consistent with the *in vitro* data.

On the basis of the evidence discussed above, we can propose a potential complex mechanism for the *H. modelsicaldum bc* complex (Figure 6.9). As for other *bc* or *b* complexes, menaquinol enters at the Qo site. Two electrons and two protons are removed, one at a time in a bifurcated electron transfer. The first electron is transferred quickly to the Rieske iron-sulfur cluster, reducing the Rieske cluster and causing it to move towards the periplasm, preventing rapid back transfer of the electron to Qo. The Rieske cluster is then in position to transfer an electron to heme 1 of DHCC. Docking of the Rieske cluster is probably the first step in inducing the redox-based conformational changes of DHCC. Based on the mechanism of interheme electron transfer within other diheme cytochrome c’s, reducing heme 1 changes the conformation and causes formation of a hydrogen-bonding network that permits rapid inter-heme electron transfer. Heme 2 then donates the electron to soluble cytochrome *c*553, which departs, allowing the Rieske to return to Qo. The second electron would then travel down the b heme pathway, reducing menaquinone at the Qi site. Another menaquinol docks and the process repeats, completing one turnover of the *bc* complex and translocating a net of two protons to the periplasm.
Figure 6.9 Proposed function of diheme cytochrome c. (A) and (B) Assumed dimer of the *H. modesticaldum* cytochrome *bc* complex. Blue- cytochrome *b* subunit, Purple- Fe-S subunit, Green- cytochrome *f* subunit. A) Imagined left half of dimer shown as homology modelled subunits of the *H. modesticaldum* cytochrome *bc* complex. B) Right side of the dimer shown as the solved crystal structure of the cytochrome *bc*1 complex from *R. sphaeroides* used as the template for 7A. Heme cofactors highlighted in red. (C) Proposed *H. modesticaldum* cytochrome *bc* complex bifurcated electron transfer steps and mechanism. Q- quinones, Qo- quinone oxidation site, Qi- quinone reduction site, R- Rieske Fe-S, b1, H- type hemes, H1,2- DHCC c-type hemes, c553- electron acceptor soluble cytochrome c553.
Such a mechanism is efficient for many species with a monoheme but not with a diheme cytochrome. Dihemes often result from gene duplication that could be random\(^6\). However, multi-heme systems in nature serve to transport electrons rapidly or to store electrons\(^5\). In the mechanism proposed here, the diheme is fulfilling both of these functions. The interheme electron transfer has been observed to be very rapid compared to transfer to a donor and can be used to prevent back reaction with the Rieske cluster. Additionally, if cytochrome \(c_{553}\) is scarce or turning over at a slower rate than the \(bc\) complex, the diheme can store an electron and wait for cytochrome \(c_{553}\) without significantly hindering the turnover rate of the \(bc\) complex. Further study, particularly crystallographic information and inhibitor studies, will clarify these mechanistic proposals.

6.5 Conclusion

HDX data, when used in conjunction with HM and verified by SASA, improves protein structure models and allows mechanistic predictions. For DHCC, HDX kinetic analysis reveals that the oxidized state is more dynamic and open than the reduced one. This is seen in loop regions that generally undergo more extensive HDX at shorter times. The bundle core regions do not show significant differences between the redox states, suggesting that they are stable and undergo little conformational change during redox cycling. Redox-dependent conformational change facilitates the formation of potential hydrogen bonds, permitting inter-heme electron transfer. The dynamics data reveal faster uptake on the heme 2 domain, suggesting that it is the site of electron donation to cytochrome \(c_{553}\). This evidence is used to predict the mechanism for the \(bc\) complex of \(H. modesticaldum\).

The study provides insight on the relation of conformational change to the binding of the electron donor Rieske subunit and soluble electron accepter, cytochrome \(c_{553}\), and on the
function of the DHCC subunit within the bc complex of *H. modesticaldum* photosynthetic electron-transfer chain. In addition, we believe that the combined approach of modeling and protein footprinting will be useful in other problems in structural biology where protein crystal or NMR structures are lacking. Information at the amino-acid level, such as that can be provided by HDX with electron-capture dissociation or electron-transfer dissociation or by OH radical footprinting (e.g., FPOP), will improve the ability to distinguish models.

### 6.6 Acknowledgements

I acknowledge Dr. Bruce Pascal at The Scripps Research Institute for helpful discussions and support in our utilization of HDX WorkBench, and Dr. Manolo D. Plasencia for help on using calculation of SASA data. I also acknowledge Dr. Hai Yue for DHCC synthesis and purification, and Erica Majumdar for Homology modeling and mechanism proposal.
6.7 References


Chapter 7: Conclusions and Future Directions
Mass spectrometry (MS)-based biophysical approaches are new “tools” for protein structure characterization owing to their sensitivity and fast turnaround rate. One of the emerging MS-based biophysical approaches is protein footprinting. One can obtain information such as protein high-order structure or dynamics through protein footprinting by counting the number or extent of the covalent labels. Such information can be readout on peptide level through bottom-up strategy where the target proteins undergo enzymatic digestion process. One of the MS-based protein footprinting methods is hydrogen deuterium exchange (HDX).

In this thesis, we focused on developments and applications of HDX. Basic principles of HDX-MS and description of HDX-MS-based strategies on characterization of protein-ligand binding are introduced in Chapter 1. Chapters 2 and 3 focused on development of a pulsed HDX-MS platform on characterization of protein aggregation, and its extension on investigation of drug candidate for Alzheimer’s Disease. Chapters 4 and 5 investigated on a protein and six Ca$^{2+}$ ions binding system, focusing on both mapping of binding regions and measuring binding affinities. In the last Chapter, we used a combined approach of HDX-MS and homology modeling (HM) to determine protein structure when high-resolution X-ray or NMR data are not available.

HDX-MS is a powerful technique and is widely used in characterization of protein structures. We can monitor structural behavior of every single residue except proline through HDX-MS. By combining with pepsin digestion and dissociation approaches such as electron capture dissociation (ECD) and electron transfer dissociation (ETD), residue-level information can be elucidated. In addition, automatic procedures provide better reproducibility and less user effort compared to the traditional HDX-MS platform. Furthermore, several available software packages ease the users from tedious data analysis process.
HDX-MS still requires efforts, however, in its basic understandings. HDX is a reversible labeling technique. One must perform the experiment fast between the quenching and LC-MS analysis steps, usually within 3-10 min for desalting and peptide trapping. Thus, this disadvantage limits application of HDX-MS in analyzing complex samples, which typically requires long-time separation and digestion procedures before MS. Other MS-based footprinting approaches, such as fast photochemical oxidation of proteins (FPOP) and carboxyl group labeling (GEE), are irreversible labeling techniques that allow the labeled proteins to be separated and digested using different techniques. Thus, HDX, FPOP, and GEE labeling can be complementary and provide complete view of the system better than each of them separately.

In addition, HDX report solvent accessibility and hydrogen bonding because the backbone hydrogen atoms are sensitive to both conditions. Thus, it is difficult to extract reliably the solvent accessibility information only. On the contrary, FPOP and GEE reaction is only sensitive to the movement of the side chain groups, thus, reporting only the solvent accessibility information. Again, one needs to combine these three techniques together for a complete structure analysis.
Curriculum Vitae

Ying Zhang
CB1134, One Brookings Dr., St. Louis, MO, 63130
314-805-3381
ying.zhang@wustl.edu

SUMMARY

• Ph.D. trained in Prof. Michael L. Gross lab. Research assistant with 5+ years’ experience in mass spectrometry (MS)-based protein structural characterization.

• Diverse background in MS-based protein structural characterization approaches, including hydrogen deuterium exchange (HDX-MS) and hydroxyl radical footprinting coupled with bottom-up analysis.

ACHIEVEMENTS

• Able to handle multiple projects simultaneously, including collaboration with Genentech Inc. (Protein Analytical Chemistry) on mass spectrometry based structural characterization of protein therapeutics.

• Developed an HDX-MS based platform in studies of protein aggregation.

EDUCATION

Ph.D. in Chemistry
Washington University in St. Louis, St. Louis, MO
Advisor: Profs. Michael L. Gross and Liviu M. Mirica
2015
Spring

B.S. in Chemistry
Nanjing University, Nanjing, Jiangsu, China
2009

PUBLICATIONS


MANUSCRIPT


