


2-1-2020

Covalent Labeling-Mass Spectrometry for Characterizing Protein-Ligand Complexes

TIANYING LIU

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

 Part of the [Analytical Chemistry Commons](#), [Biochemistry Commons](#), and the [Other Biochemistry, Biophysics, and Structural Biology Commons](#)

Recommended Citation

LIU, TIANYING, "Covalent Labeling-Mass Spectrometry for Characterizing Protein-Ligand Complexes" (2020). *Doctoral Dissertations*. 1847.
<https://doi.org/10.7275/fdz4-4x24> https://scholarworks.umass.edu/dissertations_2/1847

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

**COVALENT LABELING AND MASS SPECTROMETRY FOR
CHARACTERIZING PROTEIN-LIGAND COMPLEXES**

A Dissertation Presented

by

TIANYING LIU

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

DOCTOR OF PHILOSOPHY

February 2020

Department of Chemistry

© Copyright by Tianying Liu 2020

All Rights Reserved

**COVALENT LABELING AND MASS SPECTROMETRY FOR
CHARACTERIZING PROTEIN-LIGAND COMPLEXES**

A Dissertation Presented

by

TIANYING LIU

Approved as to style and content by:

Richard W. Vachet, Chair

Igor A. Kaltashov, Member

Jeanne A. Hardy, Member

Stephen J. Eyles, Outside Member

Ricardo B. Metz, Department Head
Chemistry

ABSTRACT

**COVALENT LABELING AND MASS SPECTROMETRY FOR
CHARACTERIZING PROTEIN-LIGAND COMPLEXES**

FEBRUARY 2020

TIANYING LIU

TIANYING LIU, B.S., JILIN UNIVERSITY
Ph.D. UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Richard W. Vachet

This dissertation focuses on applying covalent labeling (CL) and mass spectrometry (MS) for characterizing protein-ligand complexes. Understanding protein-ligand interactions has both fundamental and applied significance. Covalent labeling is a protein surface modification technique that selectively modifies solvent-exposed amino acid side chains of proteins. A covalent bond is formed between the functional groups of labeling reagent and protein's side chain. One of the key factors that affects CL reactivity is a side chain's solvent accessibility. Ligand binding protects residues on the protein surface from being labeled, and residues involved in ligand binding can be indicated via decreases in labeling extents.

The main goal of this study is to develop strategies that apply CL-MS to characterize protein-ligand complexes. Diethyl pyrocarbonate (DEPC) is the labeling reagent we focused on. First, we developed a strategy that can identify ligand binding site as well as determine the ligand binding affinity to the protein. We characterized the complexes between β -2 microglobulin (β 2m) and three amyloid inhibiting molecules under Cu(II)-induced amyloid forming conditions. The rest of the dissertation focused on comparing the information from two complementary MS-based methods, hydrogen deuterium exchange (HDX)-MS and CL-MS. Using three model protein-ligand systems, we demonstrate that the two labeling techniques can provide synergistic structural information about protein-ligand binding when reagents like DEPC are used for CL because of the differences in the intrinsic reaction rates of DEPC-based CL and HDX.

This dissertation highlights the power of CL-MS for characterizing protein-ligand complexes. The understanding of how three amyloid inhibiting molecules bind to Cu(II)- β 2M could facilitate future library screening for new drug candidates. Our work also

indicates CL-MS is capable of characterizing protein-ligand complexes that are difficult to study by other methods such as X-ray crystallography and nuclear magnetic resonance spectroscopy.

TABLE OF CONTENTS

	Page
ABSTRACT.....	iv
LIST OF TABLES.....	x
LIST OF FIGURES	xi
CHAPTER	
1. INTRODUCTION	1
1.1 Mass Spectrometry for Characterizing Protein-Ligand Complexes	1
1.2 Covalent Labeling-Mass Spectrometry for Characterizing Protein-Ligand Complexes	1
1.2.1 Covalent Labeling-Mass Spectrometry.....	1
1.2.2 Diethyl pyrocarbonate.....	3
1.2.3 Other Amino Acid Specific Labeling Reagents.....	7
1.2.4 The Basic Idea of CL-MS for Characterizing Protein-Ligand Complexes.....	8
1.2.5 Experiment Design and Workflow	10
1.3 Synergistic Information from HDX-MS and CL-MS for Characterizing Protein-Ligand Complexes	12
1.4 β -2-microglobulin and its amyloid inhibiting molecules.....	14
1.4.1 β -2-microglobulin and its Cu(II) induced amyloid forming pathway.....	14
1.4.2 Amyloid Inhibiting Molecules for Cu(II)- β 2m.....	15
1.5 Summary.....	17
1.6 Reference	19
2. USING CL-MS TO IDENTIFY PROTEIN BINDING SITS OF AMYLOID INHIBITING MOLECULES.....	23
2.1 Introduction.....	23
2.2 Experimental and Methods	25
2.2.1 Materials	25
2.2.2 ESI-MS Titration	26
2.2.2.1 Sample preparation	26
2.2.2.2 ESI-MS Titration Experiments	26
2.2.3 Size Exclusion Chromatography.....	28
2.2.4 Circular Dichroism Spectroscopy	28
2.2.5 Fluorescence Spectroscopy	29
2.2.6 Covalent Labeling.....	30
2.2.6.1 Sample Preparation.....	30
2.2.6.2 DEPC Labeling.....	30
2.2.6.3 BD Labeling.....	31
2.2.6.4 EDC/GEE Labeling	31

2.2.7 Proteolytic Digestion	31
2.2.8 Liquid Chromatography-Mass Spectrometry (LC-MS).....	32
2.2.9 Determination of Modification Percentages	33
2.2.10 Ion Mobility Mass Spectrometry	34
2.2.11 Protein-Ligand Docking.....	35
2.3 Results and Discussion	36
2.3.1 K_d Values for $\beta 2m$ with Selected Molecules.....	36
2.3.2 Size Exclusion Chromatography.....	37
2.3.3 Rifamycin SV Binding.....	39
2.3.4 Doxycycline Binding.....	46
2.3.5 Suramin Binding	52
2.4 Conclusion	56
2.5 Reference	58
3. PROTEIN-LIGAND AFFINITY DETERMINATIONS USING COVALENT LABELING-MASS SPECTROMETRY	60
3.1 Introduction.....	60
3.2 Materials and Methods.....	63
3.2.1 Materials	63
3.2.2 Sample Preparation	63
3.2.3 HNSB Labeling.....	64
3.2.4 DEPC Labeling	64
3.2.5 Proteolytic Digestion	65
3.2.6 Liquid Chromatography–Mass Spectrometry (LC/MS).....	65
3.2.7 Determination of Covalent Labeling Modification Percentages	67
3.2.8 Calculation of K_d	68
3.2.9 Solvent Accessible Surface Area (SASA) Calculations	69
3.3 Results and Discussion	70
3.3.1 Effect of Labeling Extent on the Apparent K_d Value	70
3.3.2 Lysozyme-NAG ₃ Binding.....	72
3.3.3 K_d of Maltose to Maltose Binding Protein.....	76
3.3.4 K_d of EGCG to $\beta 2m$	81
3.4 Conclusion	86
3.5 Reference	88
4. SYNERGISTIC STRUCTURAL INFORMATION FROM COVALENT LABELING AND HYDROGEN-DEUTERIUM EXCHANGE MASS SPECTROMETRY FOR PROTEIN-LIGAND INTERACTIONS.....	90
4.1 Introduction.....	90
4.2 Experimental and Methods	93
4.2.1 Materials	93
4.2.2 Sample Preparation	93
4.2.3 HDX Experiments.....	94
4.2.4 CL-MS Experiments	96
4.3 Results and Discussion	98
4.3.1 Myoglobin.....	98
4.3.2 Bovine Carbonic Anhydrase II	106

4.3.3 Maltose Binding Protein	113
4.4 Conclusion	125
4.5 Reference	126
5. SUMMARY AND FUTURE DIRECTIONS	128
5.1 Summary	128
5.2 Limitation of the Method	131
5.3 Future Directions	133
5.3.1 CL-MS and Protein-Ligand Docking for Screening New β 2m Amyloid Inhibiting Molecules	133
5.3.2 Distinguish Conformational Change from Ligand Binding Site in a CL-MS for Characterizing Protein-Ligand Complexes Experiment	134
5.4 Reference	135
BIBLIOGRAPHY	136

LIST OF TABLES

Table	Page
Table 2.1 Suramin docking score.....	56
Table 3.1 SASA value of CL modified residues of MBP.....	78
Table 4.1 Peptides from myoglobin with statistically significant differences in the relative deuterium uptake.....	103
Table 4.2 Peptides from MBP with statistically significant differences in the relative deuterium uptake.....	115
Table 4.3 SASA values of residues on MBP that are modified by DEPC or HNSB.....	124

LIST OF FIGURES

Figure	Page
Figure 1.1 Modification reaction of DEPC labeling.....	5
Figure 1.2 Amino acid specific labeling reagents.....	8
Figure 1.3 Idea of applying CL-MS for characterizing protein-ligand complexes.....	9
Figure 1.4 Experiment workflow for CL-MS to characterizing protein-ligand complexes. ...	10
Figure 1.5 Basic idea of applying HDX-MS for characterizing protein-ligand complexes. ...	12
Figure 1.6 Summary of Cu(II) induced β 2m amyloid formation process.....	16
Figure 1.7 Structure of dimer and tetramer in the proposed Cu(II) induced β 2m amyloid formation process.....	17
Figure 1.8 Amyloid inhibiting molecules for Cu(II)- β 2m and example TEM figure of amorphous aggregates comparing to β 2m fibril.	17
Figure 2.1 Extracted ion chromatogram for relative labeling ratio calculation.	34
Figure 2.2 Example ESI-MS spectrum of suramin-Cu(II)- β 2m.	36
Figure 2.3 ESI-MS titration curve indicating the binding affinity of suramin to Cu(II)- β 2m.....	37
Figure 2.4 Example SEC chromatograph for β 2m incubated with Rifamycin SV.....	38
Figure 2.5 Example SEC chromatograph for β 2m incubated with doxycycline and suramin.....	39
Figure 2.6 CL-MS results of rifamycin SV with Cu(II)- β 2m.....	40
Figure 2.7 BD labeling assists with rifamycin SV binding site identification.....	42
Figure 2.8 EDC/GEE labeling assists with rifamycin SV binding site identification.	42
Figure 2.9 Structure similarity of Cu(II)- β 2m with doxycycline, rifamycin SV, or suramin to Cu(II)- β 2m monomer.....	45
Figure 2.10 Effect of small molecules on thermal stability of β 2m.....	46

Figure 2.11 CL-MS results of doxycycline with Cu(II)- β 2m.....	48
Figure 2.12 EDC/GEE labeling assists with doxycycline binding site identification.	51
Figure 2.13 BD labeling assists with doxycycline binding site identification.....	51
Figure 2.14 Labeling experiment result and docking result presenting on the protein surface structure: doxycycline.....	51
Figure 2.15 CL-MS results of suramin with Cu(II)- β 2m.....	53
Figure 2.16 EDC/GEE labeling assists with suramin binding site identification.	55
Figure 2.17 BD labeling assists with suramin binding site identification.	55
Figure 2.18 Top Glide docking predicted suramin binding pose compared to experiment result.....	56
Figure 3.1 Relationship between apparent K_d and real K_d explained by a hypothetical model.....	72
Figure 3.2 Identify Trp62 of lysozyme modified by HNSB.....	73
Figure 3.3 HNSB covalent labeling of lysozyme in the presence of (NAG) ₃ using a 1.5 mM concentration of HNSB and 10 s reaction time	74
Figure 3.4 HNSB covalent labeling of lysozyme in the presence of (NAG) ₃ at different extents of labeling.....	75
Figure 3.5 CL for determining significantly changed residues of MBP.....	76
Figure 3.6 CL for determining K_d of MBP and maltose.....	79
Figure 3.7 Example residues that do not have clear trend in labeling ratio with increasing concentration of maltose.....	81
Figure 3.8 EGCG binding to Cu(II)- β 2m results in decreases and increases in DEPC labeling at several residues.	83
Figure 3.9 Fluorescence spectroscopy determining Cu(II) binding to β 2m.	84
Figure 3.10 Selected residues in Cu(II)- β 2m for determining K_d of EGCG to the protein.....	85
Figure 3.11 Peak heights from SEC chromatography to determine K_d of EGCG to β 2m.....	86

Figure 4.1 Relative fractional deuterium uptake for myoglobin shown on the holo myoglobin structure.....	99
Figure 4.2 Deuterium uptake plots for myoglobin.....	102
Figure 4.3 DEPC labeling percentages for myoglobin	104
Figure 4.4 Residues of myoglobin identified to undergo significant decreases in DEPC labeling upon heme binding.....	104
Figure 4.5 A comparison of the differential CL and HDX results for myoglobin mapped on its crystal structure.	105
Figure 4.6 Overlay of BCA and HCA-brinzolamide structures.	107
Figure 4.7 Relative fractional deuterium uptake for BCA shown on its structure.	107
Figure 4.8 Deuterium uptake plots for BCA.....	112
Figure 4.9 DEPC labeling percentages for BCA.	112
Figure 4.10 DEPC labeling results of BCA on crystal structure.	113
Figure 4.11 Structure of ligand free and maltose bound state for the maltose binding protein.	114
Figure 4.12 Relative fractional deuterium uptake for MBP shown on MBP-maltose structure.	115
Figure 4.13 Deuterium uptake plots for MBP.	120
Figure 4.14 Regions identified to undergo significantly decreased HD exchange upon maltose binding of MBP.	121
Figure 4.15 Summary of CL result for MBP upon maltose binding.	122
Figure 4.16 A comparison of the differential CL and HDX results for the maltose binding protein mapped on its crystal structure.	124

CHAPTER 1

INTRODUCTION

1.1 Mass Spectrometry for Characterizing Protein-Ligand Complexes

Protein-ligand interactions, a type of molecular recognition process, are fundamental in all living organisms. Understanding the mechanisms of protein-ligand interactions is an integral step to understanding biology at molecular level [1]. For drug discovery, design, and development purpose, intensive characterizations of the protein-ligand complexes of interest is also essential for the pharmaceutical industry [2]. With all these purposes, various kinds of analytical methods have been developed to study different aspects of the protein-ligand interactions, such as ligand binding affinity of the protein, ligand binding site of the protein, and whether ligand binding will induce structural changes to the protein. In this work, we aimed to achieve the residue-level structural information of the protein-ligand complexes, and mass spectrometry (MS) based methods were chosen as methods of choice to study.

Mass spectrometry (MS) is an analytical technique that measures mass-to-charge ratios (m/z) of gas-phase ions. In one way, analyte in solution phase can be ionized by an ion source to generate gas-phase chemical species, and a mass analyzer sorts the ions into a spectrum based on their m/z values [3]. However, as its name indicated, MS reports only “mass”, as indicated by m/z of the analyte, it does not readily gives us the “picture”, i.e. structural analysis of the protein-ligand complex. Thus, some other techniques are required to work together with MS that the structural information can be encoded and delivered. These MS-based structural methods do not require the crystal of the protein-ligand complex be available, nor are they limited by the size of the protein. These

features make them possible to serve as an alternative when other well-established methods such as X-ray crystallography or nuclear magnetic resonance (NMR) are not applicable [4, 5].

1.2 Covalent Labeling-Mass Spectrometry for Characterizing Protein-Ligand Complexes

1.2.1 Covalent Labeling-Mass Spectrometry

The MS-based method that we focused on in this thesis is covalent labeling-mass spectrometry (CL-MS). Covalent labeling (CL) is a protein surface modification technique that selectively modifies solvent-exposed amino acid side chain of the protein, a covalent bond will be formed between the functional groups of labeling reagent and protein's side chain which can encode the structural information around the certain side chain, e.g. topology, solvent accessibility, and microenvironment. The label will cause a mass shift compared to the unmodified side chain such that a modification could be identified by MS, which complete the method as CL-MS [6, 7].

There have been many CL techniques being developed and they can be divided into two categories: (a) amino acid specific labeling techniques (e.g. Lys-specific labeling, Trp-specific labeling, carboxylic acid labeling) [6] and (b) non-specific labeling techniques (e.g. hydroxyl radical footprinting [8] and DEPC labeling [7]). In the work presented in this thesis, the most applied CL technique is DEPC labeling, although some other kinds of amino acid specific labeling reagents were also applied.

1.2.2 Diethyl pyrocarbonate

Diethyl pyrocarbonate (DEPC) is a commercially available labeling reagent that can react with a range of nucleophilic residues: Cys, His, Lys, Thr, Tyr, Ser, and N-termini via nucleophilic substitution reactions (Figure 1.1). Comparing to some other non-specific labeling techniques, DEPC is capable of modifying few types of amino acids. For example, fast photochemical oxidation of proteins (FPOP) can modify at least 14 kinds of common amino acid residues [8, 9], and some more recently developed labeling reagent such as $\cdot\text{CF}_3$ can modify 18 kinds of residues [10]. However, being able to modify up to 6 kinds of residues allows DEPC to cover about up to 25% sequence in a typical protein, which is rather informative. Also, the corresponding data interpretation of DEPC is considerably easier. DEPC labeling will only generate a single type of product, which corresponds to a mass shift of +72.021 Da. This greatly simplifies the identification of labeled sites. Manual data interpretation is rather easy as shown in the work for Chapter 2 and 3. Although a customized data processing software for DEPC is also available [11, 12], which has been applied in the work for Chapter 4. While for some non-specific labeling reactions, especially hydroxyl radical based, the reaction could lead to multiple kinds of products [8, 9]. This makes the data interpretation of these reactions impossible to be derived manually in reasonable time [13, 14].

DEPC labeling reaction is also very easy to perform. DEPC will directly label the protein when being added to the solution that no sophisticated equipment (e.g. laser or flow cell) is required at all. While to generate hydroxyl radicals [15], carbenes [16, 17], or $\cdot\text{CF}_3$ [10], a laser is mandatory in most cases. A laser source is usually expensive, and lasers can potentially be dangerous. Besides this, majority of these reactions also require

both precise and continuous reaction time control that a set of specialized experimental setup (usually a flow cell) must be accompanied. All these restrictions cause radical based reactions to only be performed by a well experienced experimenter. For DEPC labeling, the requirement for performer's experience or equipment is minimum.

Besides the differences described above, the most distinctive difference between DEPC labeling and other popular non-specific labeling reactions is the intrinsic reaction rate. The intrinsic reaction rate for DEPC is about tens of seconds [18], while hydroxyl radicals [15, 19, 20], carbenes [16, 17], or $\cdot\text{CF}_3$ [10] all have much faster reaction rate range from nanoseconds to microseconds, which is comparable to or faster than protein folding rates. This difference makes DEPC (and CL reactions which have similar reaction rate) more likely to adopt a different strategy from those extremely fast labeling reactions when being applied to characterize protein-ligand complexes. For example, Gross and co-workers have developed a set of strategies applying FPOP for characterizing protein-ligand complexes and protein's higher order structures (HOS) [21]. When being applied to determine the ligand binding affinity to the protein, due to hydroxyl radical's extremely fast reaction speed, the strategy was developed based on monitoring protein's dynamic changes induced by ligand binding [22, 23]. While for DEPC, the reaction is too slow to monitor protein's dynamic change, also the competition between labeling reaction and ligand's $k_{\text{on}}/k_{\text{off}}$ must be considered, that a different strategy monitoring the fraction of ligand free protein in the system was developed (Chapter 3). Another example is, when being used as a complementary method of hydrogen-deuterium exchange (HDX) MS for characterizing protein-protein interactions, FPOP is considered able to provide more information of subtle differences in conformation or dynamics, when the exchange

at the local region is too rapid to the HDX time scale [24]. While for DEPC, again, we propose DEPC is too slow to indicate rapid changes in dynamics of the protein, that the result of DEPC labeling will be more focusing on the overall decrease in solvent accessibility of ligand binding region (Chapter 4). The reader should keep in mind that this thesis will only focus on CL reactions whose intrinsic reaction rate are about tens of seconds.

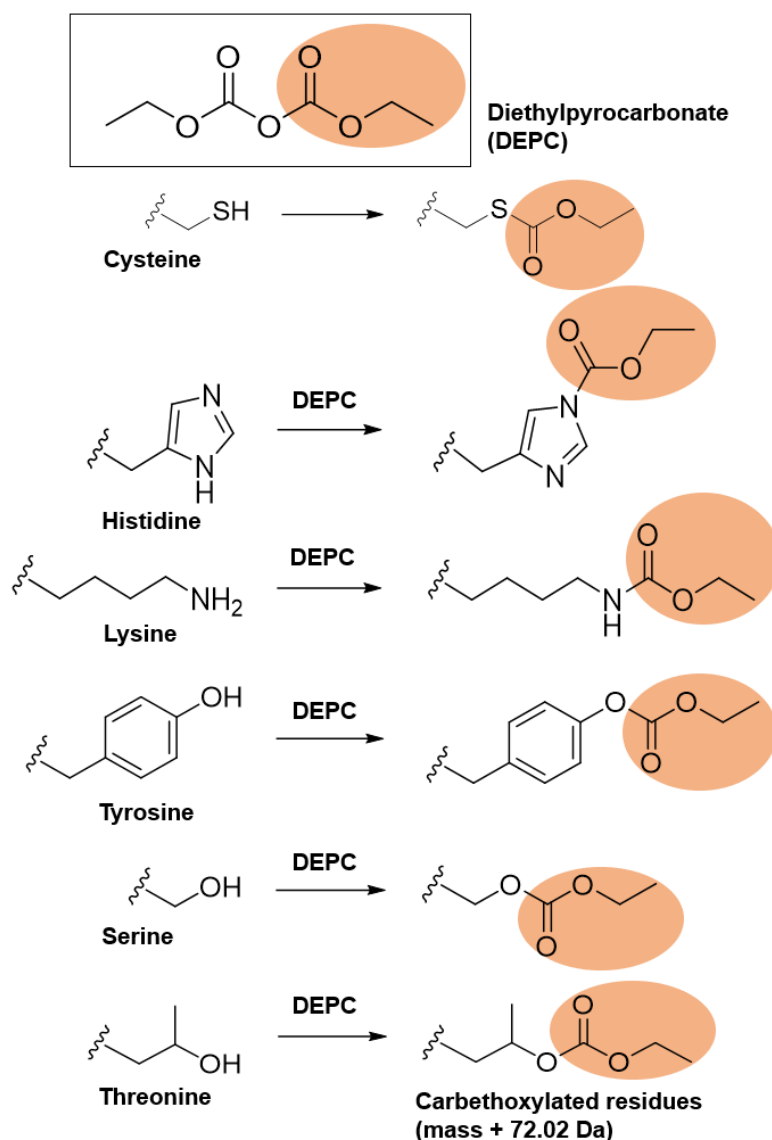


Figure 1.1 Modification reaction of DEPC labeling.

Our lab has been developing the strategies using DEPC as a CL reagent over a decade and the technique has been improved along with the studies. The first work using DEPC for protein surface mapping was reported in 2008 [25]. In this work, the basic strategy of applying this CL reagent was developed, with the understanding of second-order reaction between DEPC and proteins, an estimation of intrinsic reaction rates, as well as the discovery that multiple kinds of residues can be modified. Following this first work, to maximize the resolution of DEPC as a labeling reagent for CL-MS, Zhou et al. optimized the “bottom-up” CL-MS strategy (Section 1.2.5) for DEPC labeled proteins. In the first set of studies, the lability of modified residues through the protein denaturing and digesting process was investigated. Besides the observation that the labels are considerably labile, label scrambling to Cys could happen if free Cys residues are not alkylated [26]. Also, minimizing the time between the labeling reaction and MS measurements could avoid label loss, preserve the information encoded by CL, and increase the acquired protein structural resolution [27]. In another set of studies, tandem MS methods that are suitable for the DEPC labeled peptides were discussed. This work suggests scrambling during tandem MS fragmentation can be minimized if using electron transfer dissociation instead of collision-induced dissociation if used [28]. In the latest studies, Limpikirati and coworkers explored the effect from microenvironment to the DEPC labeling and found microenvironment in the intact protein tuned weakly nucleophilic side chains (Ser, Thr, and Tyr) so that they can be effectively modified, while the ones from unstructured peptides only have limited or no reactivity [29]. Also, the study of correlating DEPC labeling results with SASA at each modified residue is still

going on, which is expected to be utilized for model protein complex structures in the near future.

For application and related method developments, DEPC labeling was first applied to resolve the Cu(II) involved β -2-microglobulin (β 2m) amyloid forming pathway (see Section 1.4). This part of work includes development of a set of strategies using CL-MS to assess the structure of Cu(II)- β 2m monomer [30] and investigate the structure of β 2m dimer [31] and tetramer [32] formed in the Cu(II) involved amyloid forming pathway. In more recent studies, DEPC is further applied for characterizing therapeutic proteins. Different from previous works that also characterize protein complexes, therapeutic proteins are especially challenging due to their size and complexity. Borotto et al. and Limpikirati et al. demonstrated that DEPC labeling can identify specific protein regions that mediate aggregation and regions that undergo subtle conformational changes upon mishandling of these proteins [11, 12]. Importantly, DEPC labeling provides excellent structural resolution (up to 30% of the surface residues) that allows probing subtle conformation changes that are not detectable by common biophysical techniques [12].

1.2.3 Other Amino Acid Specific Labeling Reagents

In this study, other kinds of amino acid specific labeling reagents were also applied: 1) 2,3-butanedione (BD) that selectively modifies arginine [33], 2) 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide and glycine ethyl ester (EDC-GEE) labeling pair that selectively modifies aspartic acid and glutamic acid [34], 3) Dimethyl(2-hydroxy-5-nitrobenzyl)sulfonium bromide (HNSB) that selectively modifies tryptophan [35].

(Shown in Figure 1.2)

These amino acid specific labeling reactions have intrinsic reaction rates in the same range as DEPC, so they can be applied with the same strategy as DEPC for characterizing protein-ligand complexes. In some cases, an amino acid specific labeling reaction serve as the main labeling methods when ligand binding information is not available by DEPC. Otherwise, these techniques were used to improve the overall confidence of the CL-MS result by combining the information with that of DEPC.

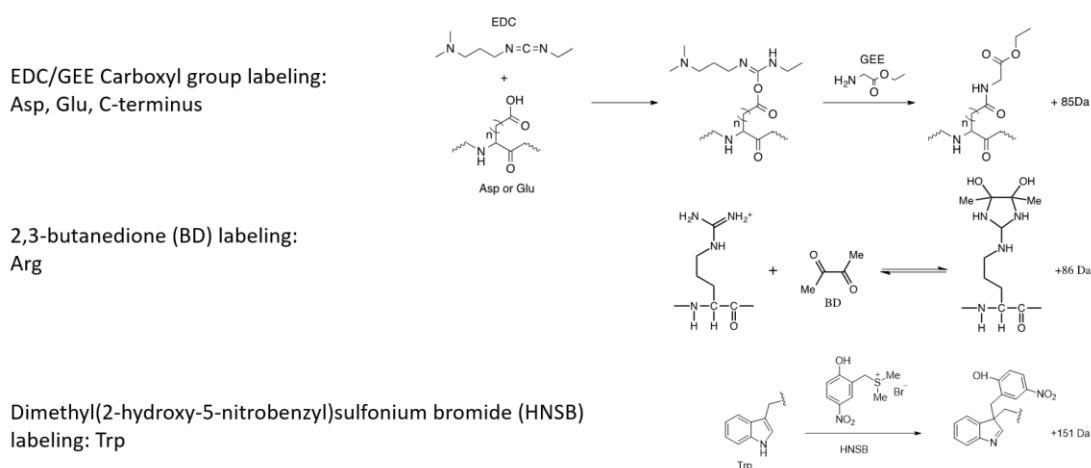


Figure 1.2 Amino acid specific labeling reagents.

1.2.4 The Basic Idea of CL-MS for Characterizing Protein-Ligand Complexes

Protein-ligand interactions are primarily mediated by non-covalent interactions, e.g. hydrogen bonds or van der Waals interactions between ligand's functional group and amino acid side chains. Besides these interactions, ligand binding can also protect the involved region of the protein's surface. As a result, protein-ligand interactions can lead to a decrease in solvent accessibility of the side chains involved in ligand binding. Because side chain's solvent accessibility is one of the key factors that affect CL reactivity [7]. In its most straightforward implementation, ligand binding protects residues on the protein surface from being labeled. This idea is the main concept that we

can apply CL-MS to characterize protein-ligand complexes and will be discussed and demonstrated through the entire thesis. When using CL-MS to identify the ligand binding site, the residues involved in ligand binding are indicated via decreases in labeling extents (summarized by Figure 1.3). When using the labeling ratio of specific residues in the protein to indicate the fraction of ligand-free protein in the system, ligand binding affinity can be determined. As prior to the point that the binding site is fully occupied by the increasing concentration of the ligand, the lower the ligand-free fraction is, the smaller the labeling extent is. Additionally, a significant change of residue's solvent accessibility might also be caused by major conformational changes of the protein upon ligand binding in some cases. CL-MS is also capable to report such conformational changes by changes in labeling extents.

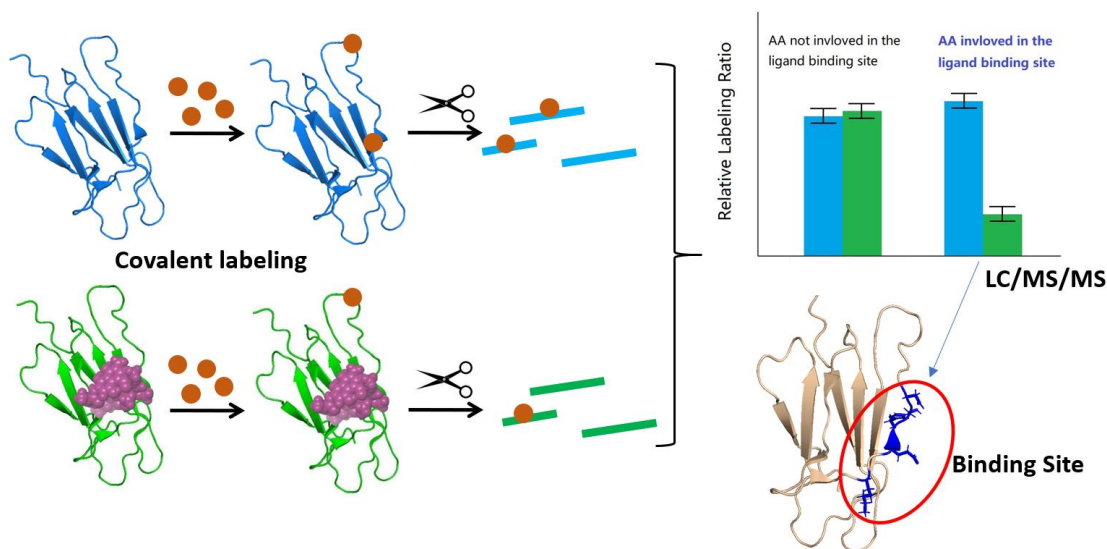


Figure 1.3 Idea of applying CL-MS for characterizing protein-ligand complexes.

1.2.5 Experiment Design and Workflow

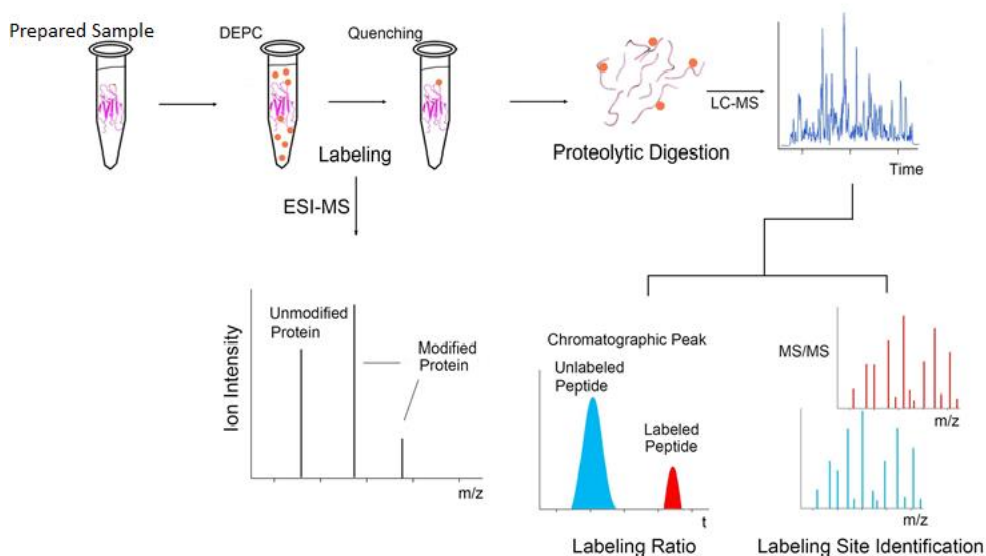


Figure 1.4 Experiment workflow for CL-MS to characterizing protein-ligand complexes.

During CL experiments, the solvent exposed residues can be modified, thus providing indirect structural information. While the buried residues should be unmodified. As illustrated in Figure 1.3, ligand binding information obtained from a CL experiment comes from comparing the labeling ratio of the protein reacted under at least two conditions: usually, it is with and without ligand conditions. The resulting differential reactivity is used to deduce ligand binding information. Besides following the basic principle of CL experimental design that the labeling reagent must be used in a way that will not perturb protein's HOS to avoid any change to the protein during labeling, the amount of ligand being applied is also critical when designing a set of experiments to characterize protein-ligand complexes. Ideally, to ensure optimal sensitivity for detecting changes in labeling, maximum protein occupancy is desired. Usually, this can be easily achieved when the dissociation constant or related physiological study has been

performed that a valid range of ligand concentration is known. However, for the less characterized protein or complex, optimizing ligand concentrations and allowing proper time for complex formation should be implied. The caveat is that protein aggregation might occur in some circumstances, given improper conditions being used. When this is the case, the false-positive ligand binding site will be reported, which could be burial resulting from protein-protein interactions, but not protein-ligand binding. Such a kind of problem was frequently encountered in experiments involving amyloid forming protein β -2-microglobulin, as described in Chapter 2 and 3.

After the labeling reaction being performed and quenched, the labeled protein sample can be directly analyzed by MS or LC-MS to get information at a global level. Otherwise, a bottom-up proteomic method can be applied. In the bottom-up approach, the labeled protein is proteolytically digested prior to LC-MS/MS analysis (Figure 1.4). Besides CL-MS experiments require complete or nearly complete protein sequence coverage to obtain the desired structural information, the residue-level spatial resolution is also desired. Comparing to top-down sequencing [36-38], bottom-up sequencing [36, 39, 40] is easier to achieve high spatial resolution as MS/MS is more effective on smaller peptides. Usually, labeled sites are pinpointed down to 1 or 2 amino acids, meaning ligand binding site can be identified by this technique at residue-level resolution. Moreover, modification extents can usually be determined as low as 0.1% or even lower, allowing CL modified residues to be detected at good sensitivity. Although bottom-up sequencing of covalently labeled proteins still suffers from the usual challenges associated with this method, such as biases associated with proteolytic cleavages and peptides going undetected during LC/MS experiments.

Another important note is, unless being corrected by additional strategy (e.g. using internal standard calibration), MS acquired labeling extent is only semi-quantitative and not absolute quantitation. This is mainly due to the fact that modified and unmodified species have different ionization efficiencies, and they are eluted at the different retention times at which the mobile phase composition are also different. When applying CL-MS for characterizing protein-ligand complex is often a comparison of labeling extent between two conditions (e.g. with and without ligand), absolute labeling extent is not necessarily required.

1.3 Synergistic Information from HDX-MS and CL-MS for Characterizing Protein-Ligand Complexes

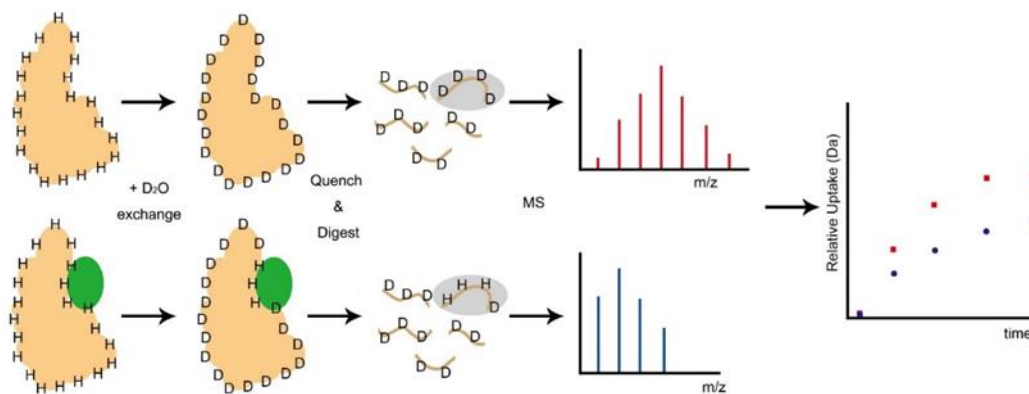


Figure 1.5 Basic idea of applying HDX-MS for characterizing protein-ligand complexes.

Hydrogen deuterium exchange (HDX) MS is a well-established method to study protein's HOS used often in the pharmaceutical industry [41-43]. HDX relies on changes in mass resulting from amide hydrogen exchange with deuterium solvent. In a commonly applied time-course HDX experiment, the reaction can be easily started by dilute protein into high percent D₂O. After a certain exchanging time, the reaction will be quenched by lowering both pH and temperature at which the exchanging rate of amide hydrogen is

minimal. Then, the deuterium-labeled protein will be digested by proteolysis and analyzed by LC-MS. The acquired information is the deuterium uptake value of a peptide/residue at different deuterium exchange time period [44, 45]. These values can be further processed and used to characterize protein HOS, protein folding [46-48], and protein aggregation kinetics [49].

Deuterium exchange rates vary based on a number of factors including protein secondary structure, solvent accessibility, pH, and temperature. For an unstructured backbone amide hydrogen under neutral pH, the intrinsic exchange rate is on the order of milliseconds [50]. However, in a folded protein this exchange rate can vary significantly from minutes to days as a result of the sequence structure of the protein, the accessibility to solvent and the dynamics of the local region [51]. Upon ligand binding, usually, it is expected that local region will have both decreased accessibility to solvent and dynamics [52, 53]. As a result of these expected decreases, HDX can be applied to characterize the protein-ligand complex [54-56] (see Figure 1.5).

However, because HDX responds to changes in both solvent accessibility and structural fluctuations, it sometimes provides ambiguous information with regard to ligand binding site. Information from CL-MS might be applied to help distinguish what acquired by HDX. The difference in intrinsic reaction rates between slow covalent labeling techniques like DEPC labeling and HDX are 2-3 orders of magnitude. Such a slow reaction rate of DEPC allows it to be transparent to cases which solvent exposed states of the region are too short-lived to be labeled. While HDX is sensitive to these changes because the H to D exchange reaction is orders of magnitude faster. For ligand binding regions, decreases in solvent accessibility will be identified by both methods.

Thus, when used together, HDX and CL have the potential to provide clearer information about protein-ligand binding sites and binding-induced stabilization.

1.4 β -2-microglobulin and its amyloid inhibiting molecules

1.4.1 β -2-microglobulin and its Cu(II) induced amyloid forming pathway

β -2-microglobulin (β 2m) is a component of the major histocompatibility complex (MHC) class I molecules which presents on all nucleated cells (excludes red blood cells) [57]. It is a 99-residue protein composed of seven strands arranged in an anti-parallel sandwich motif held together by a lone disulfide bond [58]. Despite its common existence in human physiology, for patients on long-term hemodialysis the protein can aggregate into amyloid fibers, deposit in joint spaces and eventually lead to deconstruction of the joints. This disease is known as dialysis-related amyloidosis (DRA) [59], a kind of amyloid diseases. Like many other kinds of amyloid diseases, namely Alzheimer's and Parkinson's [60-62], β 2m amyloid formation process *in vivo* is not fully understood yet. *In vitro*, several conditions can convert the protein from soluble to insoluble amyloids: low pH condition, co-incubation with a truncated version of β 2m, trifluoroethanol (TFE), thermal denaturation, sonicating the protein with sodium dodecyl sulfate [63-66], and incubation with catalytic amounts of Cu(II) [30, 67-72]. The strategy to induce amyloid formation that our group applies most is the Cu(II) binding induced strategy, which is summarized in Figure 1.6 [30]. Cu(II) concentrations in dialysis patients are higher than normal [67], suggesting a molecular role for Cu(II). In addition, the binding of Cu(II) is specific at the N-terminus, Ile1-Gln2, His31, and Asp59 of the protein [68, 72] with a binding constant of approximately 2 μ M. Cu(II) binding will induce some side chain

arrangement of the protein monomer, which facilitates the oligomer forming process [71, 73, 74]. Although a detailed structure of Cu(II)- β 2m is not available, the β 2m dimer [31] and tetramer [32] structures in this pathway have also been characterized in our group's previous study using CL-MS (Figure 1.7). Similar to the idea of using CL-MS for characterizing protein-ligand complexes as explained in Section 1.2.4, characterizing the structure of protein-protein complexes also relies on decreased labeling ratio caused by decreased SASA of certain residues. Decreases in SASA of these residues are caused by the protein-protein interactions, i.e. another protein. For example, comparing the labeling ratio of a residue from the β 2m monomer and the dimer, a residue within the dimer interface is expected to have a significantly decreased labeling ratio, while a residue not involved in the interface should remain unchanged.

1.4.2 Amyloid Inhibiting Molecules for Cu(II)- β 2m

Currently, there is no treatment for DRA [75, 76]. Among the efforts to develop future drugs targeting DRA, several amyloid formation inhibiting molecules have been discovered *in vitro* [77-83]. Some of them are currently studied under Cu(II) involved amyloid forming conditions by our lab: doxycycline, rifamycin SV [77] (Chapter 2) and epigallocatechin gallate [78] (Chapter 3) (Figure 1.8). Briefly, these three amyloid inhibiting molecules work in a way that redirects the amyloid forming process to form dissolvable amorphous aggregates [77, 78] (Figure 1.8). Although it is still arguable if the β 2m amorphous aggregates are noncytotoxic, some basic understanding of the complex formed by Cu(II)- β 2m and these amyloid inhibiting molecules would still facilitate drug design in the future.

In the work presented in this thesis, we investigated the binding site of the inhibitor on the Cu(II)- β 2m monomer. Some previous work by our group indicates such protein-ligand complex exists when inhibitor is involved in the Cu(II) induced amyloid formation process [77, 78]. As a follow-up step, we identified the ligand binding site on the protein and shed light on the mechanism of these inhibitors. The key challenge to experimentally identify the binding site on β 2m, or any amyloid-forming protein, was how rapidly the monomeric protein could convert to oligomers and aggregates. This feature of amyloid forming protein makes conventional protein structural analysis techniques, such as X-ray crystallography and NMR, unsuitable. Because CL reaction can be completed within minutes, with all the structural information encoded, this time scale appears to be suitable for providing protein–ligand binding information for systems that are difficult to study by other methods.

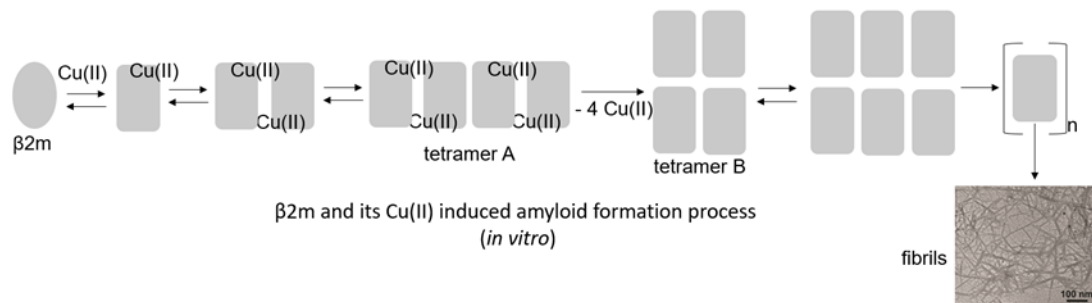


Figure 1.6 Summary of Cu(II) induced β 2m amyloid formation process.
 TEM figure from [77].

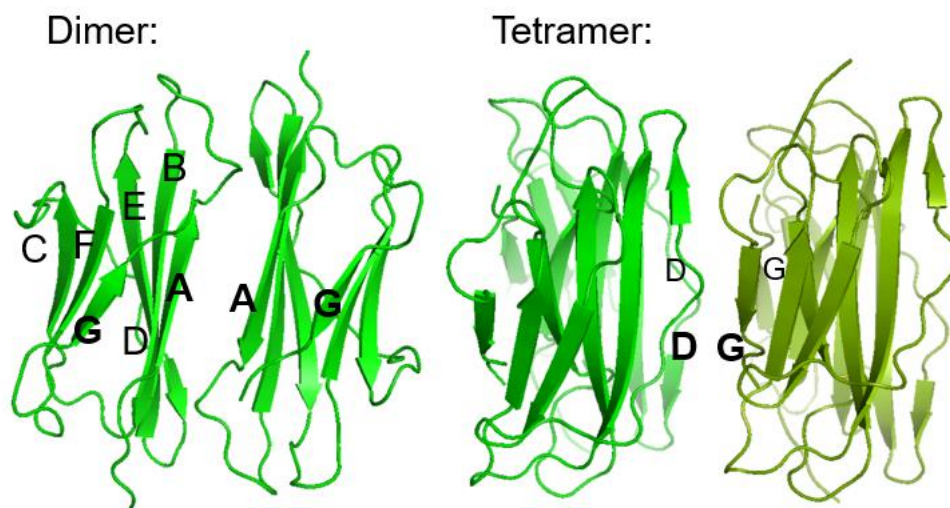


Figure 1.7 Structure of dimer and tetramer in the proposed Cu(II) induced β 2m amyloid formation process.

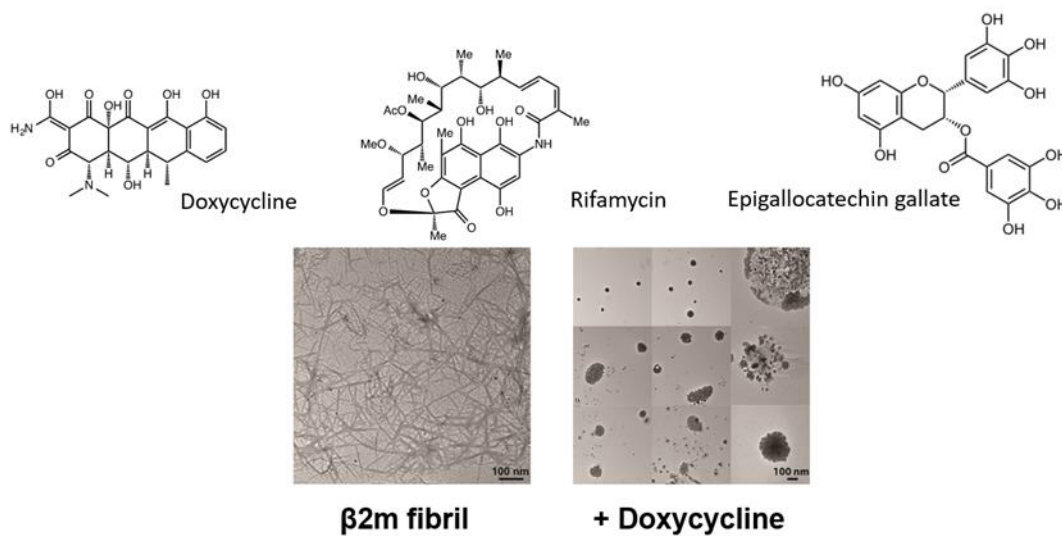


Figure 1.8 Amyloid inhibiting molecules for Cu(II)- β 2m and example TEM figure of amorphous aggregates comparing to β 2m fibril.
TEM figure from [77].

1.5 Summary

Characterizing protein-ligand complexes facilitate both scientific significance and practical applications in drug discovery. In this thesis, we developed a set of methods that rely on CL-MS to characterize protein-ligand complexes. Ligand binding sites on

proteins, ligand binding affinity of the protein, and whether ligand binding will induce structural changes to the protein can all be determined by the CL-MS based methods. In addition, we also discussed and compared the information of two MS-based methods, CL-MS and HDX-MS, to characterize the same protein ligand complexes.

In Chapter 2, we applied CL-MS to identify the protein binding sites of amyloid inhibiting molecules. We show that CL appears to be suitable for providing protein ligand binding information for systems that are difficult to study by other methods.

In Chapter 3, the development of a strategy that uses CL-MS in a ligand-based titration to determine the dissociation constant of ligand to protein is described.

In Chapter 4, we investigated three model systems and demonstrated the synergistic structural information available from CL and HDX-MS for characterizing protein-ligand interactions.

Finally, conclusions and future directions of this work are included in Chapter 5.

1.6 Reference

- [1] Baron, R.; McCammon. *Annu. Rev. Phys. Chem.* **2013**, *64*, 151-175.
- [2] Babine, R. E.; Bender, S. L. *Chemical reviews.* **1997**, *97*, 1359-1472.
- [3] Maher, S.; Jjunju, F. P. M.; Taylor, S. *Rev. Mod. Phys.* **2015**, *87*, 113-135.
- [4] Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; Liu, S.-Q. *Int. J. Mol. Sci.* **2016**, *17*, 144.
- [5] Riccardi Sirtori, F.; Altomare, A.; Carini, M.; Aldini, G.; Regazzoni, L. *Methods.* **2018**, *144*, 152-174.
- [6] Mendoza, V. L.; Vachet, R. W. *Mass Spectrom. Rev.* **2009**, *28*, 785-815.
- [7] Limpikirati, P.; Liu, T.; Vachet, R. W. *Methods* **2018**, *144*, 79-93.
- [8] Xu, G.; Chance, M. R. *Chem. Rev.* **2007**, *107*, 3514-3543.
- [9] Wang, L.; Chance, M. R. *Anal. Chem.* **2011**, *83*, 7234-7241.
- [10] Cheng, M.; Zhang, B.; Cui, W.; Gross, M. L. *Angew. Chem. Int. Ed.* **2017**, *56*, 14007-14010.
- [11] Borotto, N. B.; Zhou, Y.; Hollingsworth, S. R.; Hale, J. E.; Graban, E. M.; Vaughan, R. C.; Vachet, R. W. *Anal. Chem.* **2015**, *87*, 10627-10634.
- [12] Limpikirati, P.; Hale, J. E.; Hazelbaker, M.; Huang, Y.; Jia, Z.; Yazdani, M.; Graban, E.M.; Vaughan, R.C.; Vachet, R.W.; *mAbs* **2019**, *11*, 463-476.
- [13] Kaur, P.; Kiselar, J. G.; Chance, M. R. *Anal. Chem.* **2009**, *81*, 8141-8149.
- [14] Rinas, A.; Espino, J. A.; Jones, L. M. *Anal. Bioanal. Chem.* **2016**, *408*, 3021-3031.
- [15] Gau, B. C.; Sharp, J. S.; Rempel, D. L.; Gross, M. L. *Anal. Chem.* **2009**, *81*, 6563-6571.
- [16] Jumper, C. C.; Bomgarden, R.; Rogers, J.; Etienne, C.; Schriemer, D. C. *Anal. Chem.* **2012**, *84*, 4411-4418.
- [17] Jumper, C. C.; Schriemer, D. C. *Anal. Chem.* **2011**, *83*, 2913-2920.
- [18] Mendoza, V. L.; Vachet, R. W. *Anal. Chem.* **2008**, *80*, 2895-2904.
- [19] Hambly, D. M.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 2057-2063.
- [20] Watson, C.; Janik, I.; Zhuang, T.; Charvátová, O.; Woods, R. J.; Sharp, J. S. **2009**, *81*, 2496-2505.
- [21] Zhang, B.; Cheng, M.; Rempel, D.; Gross, M. L. *Methods* **2018**, *144*, 94-103.
- [22] Liu, X. R.; Zhang, M. M.; Rempel, D. L.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 213-217.
- [23] Liu, X. R.; Rempel, D. L.; Gross, M. L. *Anal. Chem.* **2019**, *30*, 213-217.
- [24] Li, K. S.; Chen, G.; Mo, J.; Huang, R. Y.-C.; Deyanova, E. G.; Beno, B. R.; O'Neil, S. R.; Tymiak, A. A.; Gross, M. L. *Anal. Chem.* **2017**, *89*, 7742-7749.
- [25] Mendoza, V. L.; Vachet, R. W. *Anal. Chem.* **2008**, *80*, 2895-2904.
- [26] Zhou, Y.; Vachet, R. W. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 899-907.
- [27] Zhou, Y.; Vachet, R. W. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 708-717.
- [28] Borotto, N. B.; Degraan-Weber, N.; Zhou, Y.; Vachet, R. W. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1739-1746.
- [29] Limpikirati, P.; Pan, X.; Vachet, R. W. *Anal. Chem.* **2019**, *91*, 8516-8523.
- [30] Srikanth, R.; Mendoza, V. L.; Bridgewater, J. D.; Zhang, G.; Vachet, R. W. *Biochemistry.* 2009, *48*, 9871-9881.

- [31]Mendoza, V. L.; Antwi, K.; Barón-Rodríguez, M. A.; Blanco, C.; Vachet, R. W. *Biochemistry*. **2010**, *49*, 1522-1532.
- [32]Mendoza, V. L.; Barón-Rodríguez, M. A.; Blanco, C.; Vachet, R. W. *Biochemistry*. **2011**, *50*, 6711-6722.
- [33]Fliss, H.; Viswanatha, T. *Can. J. Biochem.* **1979**, *57*, 1267-1272.
- [34]Kaur, P.; Tomechko, S. E.; Kiselar, J.; Shi, W.; Deperalta, G.; Weckler, A. *T. mAbs* **2015**, *7*, 540-552.
- [35]Xu, Y.; Strickland, E. C.; Fitzgerald, M. C. *Anal. Chem.* **2014**, *86*, 7041-7048.
- [36]Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. *J. Am. Chem. Soc.* **1999**, *121*, 806-812.
- [37]Siuti, N.; Kelleher, N. L. *Nat. Methods*. **2007**, *4*, 817-821.
- [38]McLafferty, F. W.; Breuker, K.; Jin, M.; Han, X.; Infusini, G.; Jiang, H.; Kong, X.; Begley, *FEBS J.* **2007**, *274*, 6256-6268.
- [39]Meng, F.; Forbes, A. J.; Miller, L. M.; Kelleher, *Mass Spectrom. Rev.* **2005**, *24*, 126-134.
- [40]Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C., & Yates III, J. R. *Chem. Rev.* **2013**, *113*, 2343-2394.
- [41]Rogstad, S.; Faustino, A.; Ruth, A.; Keire, D.; Boyne, M.; Park, J. *Am. Soc. Mass Spectrom.* **2017**, *28*, 786-794.
- [42]Beck, A.; Diemer, H.; Ayoub, D.; Debaene, F.; Wagner-Rousset, E.; Carapito, C.; Van Dorsselaer, A.; *TrAC Trends Anal. Chem.* **2013**, *48*, 81-95.
- [43]Kaltashov, I. A.; Bobst, C. E.; Abzalimov, R. R.; Wang, G.; Baykal, B.; Wang, *Biotechnol. Adv.* **2012**, *30*, 210-222.
- [44]Konermann, L.; Pan, J.; Liu, Y. *Chem. Soc. Rev.* **2011**, *40*, 1224-1234.
- [45]Masson, G. R.; Burke, J. E.; Ahn, N. G.; Anand, G. S.; Borchers, C.; Brier, S.; Bou-Assaf, G. M.; Engen, J. R.; Englander, S. W.; Faber, J.; et al. *Nat. Methods* **2019**, *16*, 595-602.
- [46]Tsutsui, Y.; Cruz, R. D.; Wintrode, *Natl. Acad. Sci.* **2012**, *109*, 4467-4472.
- [47]Hu, W.; Walters, B. T.; Kan, Z.-Y.; Mayne, L.; Rosen, L. E.; Marqusee, S.; Englander, S. W. *Stepwise. Proc. Natl. Acad. Sci.* **2013**, *110*, 7684-7689.
- [48]Liyanaage, R.; Devarapalli, N.; Pyland, D. B.; Puckett, L. M.; Phan, N. H.; Starch, J. A.; Okimoto, M. R.; Gidden, J.; Stites, W. E.; Lay, J. O. *Mass Spectrom.* **2012**, *330-332*, 63-70.
- [49]Zhang, Z.; Vachet, R. W. *Anal. Chem.* **2015**, *87*, 11777-11783.
- [50]Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins Struct. Funct. Bioinforma.* **1993**, *17*, 75-86.
- [51]Englander, S. W. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1481-1489.
- [52]Powell, K. D.; Ghaemmaghani, S.; Wang, M. Z.; Ma, L.; Oas, T. G.; Fitzgerald, M. C. *J. Am. Chem. Soc.* **2002**, *124*, 10256-10257.
- [53]Zhu, M. M.; Rempel, D. L.; Du, Z.; Gross, M. L. *J. Am. Chem. Soc.* **2003**, *125*, 5252-5253.
- [54]Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; West, G. M.; Griffin, P. R. *Expert Rev. Proteomics* **2011**, *8*, 43-59.
- [55]Jacob, R. E.; Engen, J. R. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1003-1010.
- [56]Percy, A. J.; Rey, M.; Burns, K. M.; Schriemer, D. C. *Anal. Chim. Acta.* **2012**, *721*, 7-21.

- [57] Güssow, D.; Rein, R.; Ginjaar, I.; Hochstenbach, F.; Seemann, G.; Kottman, A.; Ploegh, H. L. *J. Immunol.* **1987**, *139*, 3132-3138.
- [58] Katou, H.; Kanno, T.; Hoshino, M.; Hagihara, Y.; Tanaka, H.; Kawai, T.; Hasegawa, K.; Naiki, H.; Goto, Y. *Protein Sci.* **2002**, *11*, 2218-2229.
- [59] Floege, J.; Ketteler, M. *Kidney Int.* **2001**, *59*, S164-S171.
- [60] Hazenberg, B. P. C. *Rheum. Dis. Clin.* **2013**, *39*, 323-345.
- [61] Doig, A. J. *Curr. Opin. Struct. Biol.* **2015**, *30*, 50-56.
- [62] Cheng, B.; Gong, H.; Xiao, H.; Petersen, R. B.; Zheng, L.; Huang, K. *Biochim. Biophys. Acta BBA - Gen. Subj.* **2013**, *1830*, 4860-4871.
- [63] McParland, V. J.; Kad, N. M.; Kalverda, A. P.; Brown, A.; Kirwin-Jones, P.; Hunter, M. G.; Sunde, M.; Radford, S. E. *Biochemistry.* **2000**, *39*, 8735-8746.
- [64] Sasahara, K.; Yagi, H.; Naiki, H.; Goto, Y. *J. Mol. Biol.* **2007**, *372*, 981-991.
- [65] Eichner, T.; Radford, S. E. *FEBS J.* **2011**, *278*, 3868-3883.
- [66] Ookoshi, T.; Hasegawa, K.; Ohhashi, Y.; Kimura, H.; Takahashi, N.; Yoshida, H.; Miyazaki, R.; Goto, Y.; Naiki, H. *Nephrol. Dial. Transplant.* **2008**, *23*, 3247-3255.
- [67] Morgan, C. J.; Gelfand, M.; Atreya, C.; Miranker, A. D. *J. Mol. Biol.* **2001**, *309*, 339-345.
- [68] Eakin, C. M.; Knight, J. D.; Morgan, C. J.; Gelfand, M. A.; Miranker, A. D. *Biochemistry.* **2002**, *41*, 10646-10656.
- [69] Eakin, C. M.; Attenello, F. J.; Morgan, C. J.; Miranker, A. D. *Biochemistry.* **2014**, *43*, 7808-7815.
- [70] Antwi, K.; Mahar, M.; Srikanth, R.; Olbris, M. R.; Tyson, J. F.; Vachet, R. W. *Protein Science.* **2014**, *17*, 748-759.
- [71] Dong, J.; Joseph, C. A.; Borotto, N. B.; Gill, V. L.; Maroney, M. J.; Vachet, R. W. *Biochemistry.* **2014**, *53*, 1263-1274.
- [72] Lim, J.; Vachet, R. W. *Anal. Chem.* **2004**, *76*, 3498-3504.
- [73] Borotto, N. B.; Zhang, Z.; Dong, J.; Burant, B.; Vachet, R. W. *Biochemistry.* **2017**, *56*, 1095-1104.
- [74] Torbeev, V. Y.; Hilvert, D. *Proc. Natl. Acad. Sci.* **2013**, *110*, 20051-20056.
- [75] Yamamoto, S.; Kazama, J. J.; Maruyama, H.; Nishi, S.; Narita, I.; Gejyo, F. *Nephrol.* **2008**, *70*, 496-502.
- [76] Gejyo, F.; Kawaguchi, Y.; Hara, S.; Nakazawa, R.; Azuma, N.; Ogawa, H.; Koda, Y.; Suzuki, M.; Kaneda, H.; Kishimoto, H.; et al. *Artif. Organs* **2004**, *28*, 371-380.
- [77] Marcinko, T. M.; Dong, J.; LeBlanc, R.; Daborowski, K. V.; Vachet, R. W. *J. Biol. Chem.* **2017**, *292*, 10630-10638.
- [78] Marcinko, T. M.; Drews, T.; Vachet, R. W. Manuscript in preparation.
- [79] Quaglia, M.; Carazzone, C.; Sabella, S.; Colombo, R.; Giorgetti, S.; Bellotti, V.; Lorenzi, E. D. *Electrophoresis.* **2005**, *26*, 4055-4063.
- [80] Regazzoni, L.; Colombo, R.; Bertoletti, L.; Vistoli, G.; Aldini, G.; Serra, M.; Carini, M.; Facino, R. M.; Giorgetti, S.; Stoppini, M.; Caccialanza, G. *Anal. Chim. Acta* **2011**, *685*, 153-161.
- [81] Woods, L. A.; Platt, G. W.; Hellewell, A. L.; Hewitt, E. W.; Homans, S. W.; Ashcroft, A. E.; Radford, S. E. *Nat. Chem. Biol.* **2011**, *7*, 730.
- [82] Giorgetti, S.; Raimondi, S.; Pagano, K.; Relini, A.; Bucciantini, M.; Corazza, A.; Fogolari, F.; Codutti, L.; Salmona, M.; Mangione, P.; et al. *J. Biol. Chem.* **2011**, *286*, 2121-2131.

[83]Carazzone, C.; Colombo, R.; Quaglia, M.; Mangione, P.; Raimondi, S. Giorgetti, S.; Caccialanza, G.; Bellotti, V.; De Lorenzi, E. *Electrophoresis*. **2008**, *29*, 1502-1510.

CHAPTER 2

USING CL-MS TO IDENTIFY PROTEIN BINDING SITES OF AMYLOID INHIBITING MOLECULES

The work described in this chapter has been published as:

Liu, T., Marcinko, T. M., Kiefer, P. A., & Vachet, R. W. (2017). Using covalent labeling and mass spectrometry to study protein binding sites of amyloid inhibiting molecules. Analytical Chemistry, 89(21), 11583-11591.

M.T.M. conducted circular dichroism spectroscopy and fluorescence spectroscopy experiments; K.P.A. conducted most ESI-MS titration experiments; L.T. conducted all the rest experiments.

2.1 Introduction

Amyloid diseases, such as Alzheimer's and Parkinson's, are characterized by the accumulation of insoluble aggregated proteins in cells, tissues, and organs [1]. Dialysis-related amyloidosis (DRA) [2], which occurs in patients undergoing long-term hemodialysis due to renal dysfunction, involves amyloid deposits of the protein β -2-microglobulin (β 2m) in the musculoskeletal system [3-5]. Currently, there is no treatment for DRA [6, 7], but recent in vitro studies have identified several molecules that can redirect β 2m amyloid formation, suggesting that these molecules might act as prototypes for future drug design [8-10]. To facilitate drug design efforts, it would be valuable to identify where these molecules bind on β 2m, so that targeted screening of compound libraries could be conducted to find even more potent molecules. A key challenge to experimentally identifying the binding site on β 2m, or any amyloid-forming protein, is how rapidly the monomeric protein is converted to oligomers and aggregates, which makes traditional protein structural analysis techniques, such as X-ray crystallography and NMR, unsuitable. To address this challenge, we are exploring methods based on

covalent labeling and mass spectrometry (MS) to rapidly map binding sites before aggregation occurs.

Covalent labeling is a protein surface modification technique that relies on selective [11] (e.g. succinimides) or non-selective [12-18] (e.g. hydroxyl radicals) labeling reagents to covalently modify solvent-exposed amino acid side chains that can then be identified by MS and tandem MS, often in conjunction with bottom-up sequencing. Covalent labeling approaches can be particularly valuable for finding protein-protein or protein-ligand binding sites, as they probe changes in side chain solvent accessibility. Our group has found that diethylpyrocarbonate (DEPC) is a valuable pseudo-selective reagent for studying protein/protein interactions [19-25]. DEPC has some advantages over other non-selective reagents, such as hydroxyl radicals, in that it requires no special equipment (e.g. laser or synchrotron), and it results in only a single reaction product, thereby simplifying MS analyses and improving detection sensitivity. DEPC provides good structural coverage as it can react with up to 30% of the residues in the average protein, providing an effective resolution around 8-10 Å [23]. While this level of structural detail can often help define protein-protein interaction sites, this level of resolution may not be sufficient for identifying small molecule binding sites. Consequently, we are exploring the combination of information obtained from DEPC labeling with information from other labeling reagents. In this study, we show that other labeling reagents, namely 2,3-butanedione (BD), which labels Arg residues [26], and the 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC)-glycine ethyl ester (GEE) pair, which labels Asp and Glu residues [27], can be used with DEPC to better pinpoint protein-small molecule binding sites. To demonstrate the effectiveness of this combined

labeling approach, we determine the binding sites of three small molecules known to bind to β 2m—doxycycline [8], rifamycin SV [9], and suramin [28]. The former two molecules are known to inhibit Cu(II)-induced β 2m amyloid formation [29, 30] by diverting the reaction toward amorphous aggregates, while suramin has no effect on the amyloid formation reaction [10]. The identified binding sites are consistent with computational modeling and previous biochemical studies, providing validation of the obtained labeling results. We predict that our combined labeling approach should be applicable to other protein-ligand systems that are difficult to study by more traditional methods.

2.2 Experimental and Methods

2.2.1 Materials

Human full-length β 2m was obtained from Lee Biosolutions (Maryland Heights, MO). Diethylpyrocarbonate (DEPC), doxycycline hyclate, glycine ethyl ester hydrochloride (GEE), imidazole, iodoacetamide, MOPS, MOPS sodium salt, rifamycin SV sodium salt, suramin sodium salt, 2,3-butanedione (BD), tris(2-carboxyethyl)phosphine (TCEP), urea, N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC), and L-arginine were obtained from Sigma-Aldrich (St. Louis, MO). Immobilized trypsin and chymotrypsin and triethylamine acetate (pH 8.0, 1 M) were obtained from Princeton Separations (Adelphia, NJ). Acetonitrile, ammonium acetate, CuSO₄, formic acid, potassium acetate, and HPLC grade water were all purchased from Fisher Scientific (Fair Lawn, NJ). Centricon molecular weight cutoff (MWCO) filters were obtained from Millipore (Burlington, MA).

2.2.2 ESI-MS Titration

2.2.2.1 Sample preparation

Each sample contained 2.5 μM β2m , 0 or 20 μM Cu(II)SO_4 , and 50 mM ammonium acetate. The concentrations of the small molecules were varied over an appropriate range to cover the K_d value. For the doxycycline- β2m samples, 20 μM of suramin was added and served as a reference ligand since β2m -doxycycline complexes were not readily measured in the mass spectrometer. The samples were incubated for up to 3 h at 37°C before analyzing the samples by ESI-MS.

2.2.2.2 ESI-MS Titration Experiments

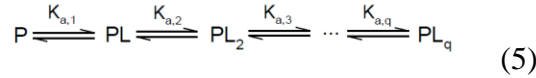
Except for the cases of doxycycline-Cu(II)- β2m , all samples were measured by a Bruker AmaZon quadrupole ion trap mass spectrometer by direct infusion at a flow rate 50 $\mu\text{L/h}$ using a modified Bruker nanoflow sprayer. For the titration experiments, the MS-signal was optimized for the β2m charge state 7⁺ and 6⁺ and the associated non-covalent complexes. The acquisition time for each sample was at least 3 min. To identify the appropriate small molecule concentrations for the covalent labeling studies, the dissociation constant (K_d) values for doxycycline, suramin and rifamycin SV bound to β2m were determined using previously described ESI-MS titration methods [31]. The values for suramin and rifamycin SV were determined directly. K_d of suramin was determined based on the equations below:

$$K_d = \frac{1}{K_a} \quad (1) \quad K_a = \frac{[PL]_{eq}}{[P]_{eq} \cdot [L]_{eq}} \quad (2)$$

$$\frac{[PL]_{eq}}{[P]_{eq}} = \frac{\sum_n I(PL^{n+})}{\sum_n I(P^{n+})} = R \quad (3)$$

$$\frac{R}{R+1} = \frac{1+K_a[P]_0+K_a[L]_0-\sqrt{(1+K_a[P]_0-K_a[L]_0)^2+4K_a[P]_0}}{2K_a[P]_0} \quad (4)$$

The following equations were used for rifamycin SV to account for the binding of two rifamycin SV molecules. The determined K_a values were then used to determine the K_d values, by assuming that K_a and K_d are inversely related.



$$K_{a,q} = \frac{R_q}{R_{q-1} \left([L]_0 - \frac{(R_1+2R_2+\dots+qR_q)[P]_0}{1+R_1+R_2+\dots+R_q} \right)} \quad (6)$$

While for doxycycline a reference ligand method was applied [32] because the doxycycline- β 2m complex is not preserved in the gas phase during the electrospray process. K_a of doxycycline is calculated by the following equations, based on previous work and the K_d value was determined by assuming that K_a and K_d are inversely related

$$K_a = \frac{R}{[L]_0 - \frac{R}{1+R+R_{ref}}[P]_0} \quad (7)$$

$$R = \frac{R_{ref}}{R_{exp}} - 1 \quad (8)$$

R_{exp} is known from experiment result:

$$\frac{[PL_{ref}]_{eq}}{[P]_{eq}} = \frac{\sum n I(PL_{ref}^{n+})}{\sum n I(P_{ref}^{n+})} = R_{exp} \quad (9)$$

R_{ref} can be calculated with known K_a of reference ligand (suramin) from the following equation:

$$R_{ref} = K_{a,ref} \left([L_{ref}]_0 - [P]_0 \frac{R_{exp}}{1+R_{exp}} \right) \quad (10)$$

The K_d of doxycycline to Cu(II)- β 2m were measured by the same Waters Synapt G2-Si quadrupole time-of-flight (QTOF) described in the main text. The electrospray

capillary voltage was set to 1.0 kV, while the source temperature was 30°C. The source offset was 10 and sampling cone was 150. Each sample contained 10 μM $\beta 2\text{m}$, 20 μM Cu(II)SO_4 , 100 mM ammonium acetate, and 80 μM suramin. Concentration of doxycycline varies from 80 to 600 μM .

2.2.3 Size Exclusion Chromatography

For covalent labeling and size-exclusion chromatography (SEC) measurements, an 81 μL solution of 75 μM $\beta 2\text{m}$ was prepared in 18.5 mM MOPS (pH 7.4), 0.37 M urea and 0.11 M potassium acetate, with 150 μM Cu(II) (CuSO_4). Stock solutions of 0.1 M doxycycline, 0.05 M rifamycin SV, and 0.025 M suramin were freshly prepared in distilled and deionized water. Samples containing $\beta 2\text{m}$ and doxycycline or suramin were incubated at 37 °C for 1 h before subjecting $\beta 2\text{m}$ to the measurements of interest. Rifamycin SV was incubated with $\beta 2\text{m}$ at 37 °C for 1 min before conducting the measurements of interest.

Protein samples were separated by a SuperSW2000 SEC column (Tosoh Bioscience, LLC; Tokyo, Japan) with an HP Agilent (Wilmington, DE) 1100 HPLC system. The mobile phase contained 150 mM ammonium acetate (pH 6.8), and a flow rate 0.35 mL/min was used. Before injection, the SEC column was equilibrated by flowing the mobile phase for 30 min. 20 μL of a prepared sample or calibration standard was injected, and eluting proteins were detected by UV absorbance at 214 nm. A solution containing 5 μM bovine serum albumin (MW = 66000), 5 μM ovalbumin (MW = 45000), 5 μM carbonic anhydrase (MW = 29040), and 5 μM $\beta 2\text{m}$ (MW= 11731) was used for molecular weight calibration.

2.2.4 Circular Dichroism Spectroscopy

Circular dichroism was performed on a JASCO J-1500 (Easton, MD) to measure the impact of the small molecules on the secondary structure of β 2m. Data were collected by scanning from 250 to 195 nm at 25 °C at a data pitch of 0.5 nm with a scan rate of 20 nm/min. Each sample contains 50 μ M β 2m and 100 μ M selected small molecule, with 25 mM MOPS and 150 mM potassium acetate at pH 7.4 (the urea was omitted to generate higher quality CD spectra). Three scans were averaged per sample. Baseline subtraction and spectral smoothing was applied in JASCO Spectra Analysis. Data were then plotted in Origin. To estimate secondary structure content, JASCO's CD Multivariate SSE was used. A protein data library was constructed using proteins with known structures and CD spectra provided by JASCO, and principal component analysis was used to calculate percent secondary structure values.

2.2.5 Fluorescence Spectroscopy

Each sample contains 2 μ M β 2m and 100 μ M selected small molecule, with 25 mM MOPS, 500 mM urea and 150 mM potassium acetate at pH 7.4. Prior to measurement, the samples were equilibrated at ambient room temperature (21 °C) for 10 minutes.

To measure the impact of the small molecules on the tertiary structure of β 2m, a PTI Quantamaster 300 was used (Edison, NJ). Intrinsic fluorescence was measured via excitation separately at both 283 and 295 nm. Emission was monitored from 305 to 400 nm. Data were collected at 1 nm intervals with an integration time of 1 second. Five scans were averaged per sample. Peaks were normalized from 0 to 1 to better compare λ_{\max} due to fluorescence quenching from inner filter effect.

For thermal stability experiments, data were collected on a JASCO 1500 CD Spectrophotometer equipped with a scanning emission monochromator and emission detector for fluorescence detection. Intrinsic tryptophan emission of β 2m was measured every 1°C from 25 to 80°C via excitation at 295 nm while monitoring emission from 305-400 nm. Prior to individual scans, samples were equilibrated at the new temperature for 1 minute. λ_{max} of each spectrum was determined by fitting the emission spectra to fourth order polynomial functions in Origin (Northampton, MA). For the control, the sample consisted of 3 μ M β 2m, 25 mM MOPS, 500 mM urea, and 150 mM potassium acetate at pH 7.4. A second control containing 6 μ M CuSO_4 was also measured. For the samples with suramin, doxycycline, or rifamycin SV added, the samples contained 3 μ M β 2m, 6 μ M CuSO_4 , 100 μ M of the small molecule. Prior to all replicates, samples were incubated for 15 minutes at room temperature (21 °C). Samples were measured in triplicate and errors bars represent standard error of the mean.

2.2.6 Covalent Labeling

2.2.6.1 Sample Preparation

Sample preparation is the same as described in 2.2.3.

2.2.6.2 DEPC Labeling

Stock solutions of DEPC were freshly prepared in acetonitrile. The reaction was initiated by adding a 0.8 μ L aliquot of the DEPC stock solution so that there was a DEPC : protein molar excess of 4:1. In all cases, the final volume of acetonitrile was less than 1%. The reaction was allowed to proceed at 37 °C for 1 min, before being quenched by the addition of imidazole at a final concentration of 10 mM. These modification

conditions lead to 0.8 to 1.2 labels per protein on average, which is necessary to maintain the structure integrity of the protein during the modification reaction. These conditions have been established by numerous previous studies of DEPC labeling of β 2m by our group [19, 21-23].

2.2.6.3 BD Labeling

Stock solutions of BD were freshly prepared in water. The reaction was initiated by adding a 3.3 μ L aliquot of the BD stock solution so that there was a BD : protein molar excess of 1250 : 1. The reaction was allowed to proceed at 37 °C for 1 min, before being quenched by the addition of an arginine solution to a final concentration of about 140 mM. These modification conditions lead to about 0.9 - 1.1 labels per protein on average, which ensures the structural integrity of the protein during the labeling reaction.

2.2.6.4 EDC/GEE Labeling

Stock solutions of EDC and GEE were freshly made in 25 mM MOPS (pH 7.4) at a concentration of 50 mM and 2.0 M respectively. The reaction was initiated by adding an 18 μ L aliquot of the EDC stock solution and an 18 μ L aliquot of the GEE stock solution simultaneously; the EDC : GEE : protein molar excess was 150:6000:1. The reaction was allowed to proceed at room temperature for 30 min, before being quenched by the addition of formic acid to a final concentration of 0.1% (v/v). These modification conditions lead to about 0.4 labels per protein on average which ensures the structural integrity of the protein during the labeling reaction.

2.2.7 Proteolytic Digestion

Before digestion of the proteins for LC/MS/MS analysis, the labeled protein samples were diluted with HPLC grade water to a volume of 400 μ L and then concentrated using a 10,000 MWCO filter to a final volume of 20 μ L. The samples were then brought to pH 8.0 by adding 70 μ L of 100 mM triethylamine acetate. The disulfide bond was reduced at 37 °C for 10 min using TCEP at a protein : TCEP molar ratio of 1:80, and the resulting reduced cysteines were alkylated in the dark for 30 min using iodoacetamide at a protein : iodoacetamide molar ratio of 1:80. Next, the samples were incubated with 10% (v/v) acetonitrile at 50 °C for 45 min to denature the protein. The protein sample was then applied to the selected enzyme at an enzyme : substrate ratio of 1:10. Digestion of DEPC and BD labeled proteins were performed with immobilized chymotrypsin. EDC/GEE labeled proteins were divided into tubes with equal volumes before either immobilized chymotrypsin or immobilized trypsin was added. After 2 h of digestion at 37 °C, the enzyme was separated from the mixture by centrifugation, and the precipitate was discarded. In the case of the EDC/GEE labeled proteins, the peptide mixture generated by the two enzymes was then mixed at equal volume and analyzed at the same time.

2.2.8 Liquid Chromatography-Mass Spectrometry (LC-MS)

All protein digests were measured immediately after proteolytic digestion by LC-MS. Mass analysis was carried out on a Bruker AmaZon (Billerica, MA) quadrupole ion trap mass spectrometer equipped with an electrospray ionization source. Typically, the electrospray needle voltage was kept at ~4 kV, and the capillary temperature was set at 220 °C. Either collision-induced dissociation (CID) or electron transfer dissociation (ETD) was applied to identify the position of the labels.

HPLC separation was conducted by a Thermo Scientific (Waltham, MA) Ultimate 3000 HPLC with a Thermo Acclaim PepMap RSLC C18 column (300 μm x 15 cm, 2 μm particle size). Peptide mixtures from the proteolytic digests were eluted using a gradient of acetonitrile containing 20% water and 0.1% formic acid that increased from 5% to 45% for 45 min at a flow rate of 4 $\mu\text{l}/\text{min}$.

2.2.9 Determination of Modification Percentages

The modification percentage of each amino acid was determined in the same manner as described previously [21-25] Briefly, the ion abundance of each peptide of interest (i.e. labeled or unlabeled) was determined from extracted ion chromatograms (e.g. Figure 2.1). For peptides that contain more than one modified amino acid, it was possible to separate the isomeric peptides (e.g. Figure 2.1), thereby allowing measurements of individual amino acid modification levels. The % modification was calculated based on eq 11. The modified ion intensities ($I_{unmodified}$) used in equation 1 are the sum of the total ion intensities of all the peptides that contain the modified amino acid of interest. Averaged ion intensity across the selected peak areas of individual peaks were used to determine ion intensity. The percent modification is a relative rather than an absolute value because the modified and unmodified peptides have different ionization efficiencies.

$$\% \text{ modification} = \frac{I_{modified}}{(I_{modified} + I_{unmodified})} \times 100 \quad (11)$$

Reported percent modifications are averages from three separate experiments. A two-sample unpaired student's t-test at a 95% confidence interval was used to determine if changes in modification levels with and without small molecules were significantly different.

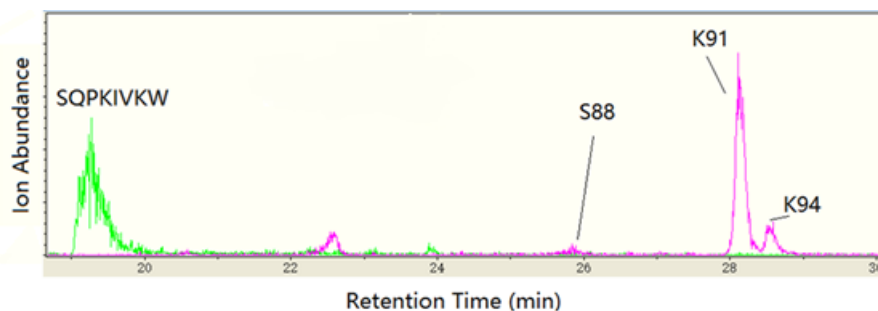


Figure 2.1 Extracted ion chromatogram for relative labeling ratio calculation. Extracted ion chromatogram of m/z 329.2 (green), which is the +3 ion of the unmodified peptide $^{88}\text{SQPKIVKW}^{95}$; extracted ion chromatogram of m/z 529.3 (pink), which is the +3 ion of the single DEPC unmodified peptide $^{88}\text{SQPKIVKW}^{95}$.

2.2.10 Ion Mobility Mass Spectrometry

A Waters Synapt G2-Si quadrupole time-of-flight (QTOF) that was equipped with a nanospray source was used to collect native mass spectral data. Both the mass spectral and ion mobility data were obtained at the University of Massachusetts Mass Spectrometry Center. Electrospray capillaries were prepared in-house by sputter coating gold using a Cressington 108 onto pulled borosilicate thin-wall capillaries (Harvard Apparatus, cat#30-0035, Holliston, MA) in a manner similar to established protocols [33]. The electrospray capillary voltage was set to 1.0 kV, while the source temperature was 30°C. The m/z scale was calibrated from 500-8,000 using perfluoroheptanoic acid (PFHA). The source offset and sampling cone were set to 20 V. Samples were incubated for 6 days at 37°C in solution conditions similar to the labelling experiments. Immediately prior to analysis, the samples were removed from the incubation chamber and desalted into 100 mM ammonium acetate using a HiTrap column (GE Healthcare, Chicago, Illinois). The desalting process yielded a final sample concentration of approximately 10 μM . Data for each sample was acquired for 10 minutes and averaged over that duration. Collisional cross section (CCS) values were estimated from a linear

calibration curve of native-like protein ions using a method described elsewhere [34] The R2 for this curve was 0.98. Data analysis was carried out using Waters MassLynx and Driftscope. The extracted arrival time distributions (ATDs) were plotted using Origin.

2.2.11 Protein-Ligand Docking

Protein-ligand docking was carried out using Glide (Version 2016-1, Schrodinger, LLC, New York). The β 2m monomer NMR structure (PDB ID 1JNJ, first one of the 20 structures) was used as a starting point and was processed and optimized by Protein Preparation Wizard, setting PROPKA at pH=7.0. The 3D SDF structures of doxycycline and rifamycin SV, 2D structure of suramin (3D structure not available) were downloaded from PubChem (pubchem.ncbi.nlm.nih.gov). All the small molecule structures were then prepared with LigPrep with forcefield OPLS3 applied. Chiralities of 3D structure were maintained and possible states at the target pH 7.4 ± 0.5 were generated. A grid was generated with centroid of specified coordinates (centered on β 2m) that enabled the entire β 2m monomer structure to be covered by adjusting the size of ligand diameter midpoint box, and no other constraint was applied. The docking was performed using Ligand Docking with Extra Precision (XP) mode. The number of poses to be reported was not limited, but post-docking minimization was applied with 20 poses per ligand included with strain correction terms. After the above steps, all the reported results were ranked by Docking Scores, and poses with Docking Score above zero were excluded.

2.3 Results and Discussion

2.3.1 K_d Values for β 2m with Selected Molecules

We first determined the K_d values for each of the three molecules when binding to monomeric β 2m. Measuring the K_d values was important for confirming that the molecules bind to β 2m at the relevant concentrations used in the subsequent covalent labeling experiments. K_d values were measured via ESI-MS titration using an approach that is similar to previous work [31, 32]. Figure 2.2 shows example mass spectra, Figure 2.3 shows an example of the titration data obtained for suramin. The determined K_d values for doxycycline and suramin are $200 \pm 20 \mu\text{M}$ and $40 \pm 2 \mu\text{M}$, respectively, whereas for rifamycin SV, two K_d values are measured: $K_{d1} = 250 \pm 25 \mu\text{M}$ and $K_{d2} = 80 \pm 20 \mu\text{M}$. These results confirm that the molecules bind to the β 2m monomer, even though their affinities for β 2m are relatively weak.

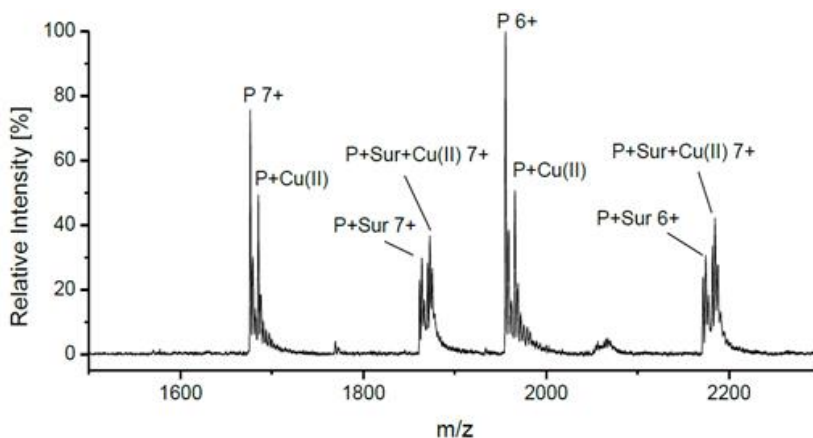


Figure 2.2 Example ESI-MS spectrum of suramin-Cu(II)- β 2m.

Sample included β 2m ($2.5 \mu\text{M}$) with suramin ($12.5 \mu\text{M}$) and $20 \mu\text{M}$ Cu(II), in ammonium acetate buffer (50 mM).

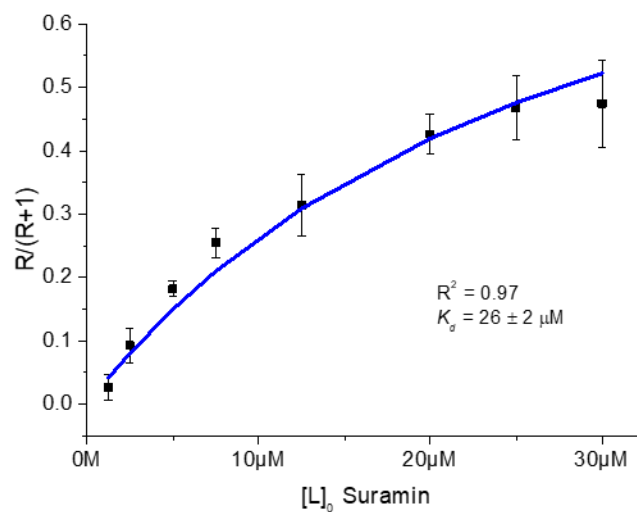


Figure 2.3 ESI-MS titration curve indicating the binding affinity of suramin to Cu(II)-β2m.

2.3.2 Size Exclusion Chromatography.

The measured K_d values allow us to identify concentrations of each small molecule needed to maximize the β2m-molecule complex concentration so that the covalent labeling experiments could be as sensitive as possible. Because we wanted to measure binding sites under conditions in which β2m is known to form amyloid and because we previously found that the inhibitors (i.e. rifamycin SV and doxycycline) can influence the formation of off-pathway pre-amyloid oligomers [10], we needed to optimize the small molecule concentrations used in the covalent labeling experiments to ensure that Cu(II)-β2m was still monomeric in the presence of the molecules of interest. Thus, we identified the highest concentration of each small molecule that could be used without observing any protein aggregation during the 1 min covalent labeling reaction. SEC was used to measure the extent of β2m oligomer formation at different concentrations of doxycycline, rifamycin SV, and suramin. From the SEC results (see example chromatograms in Figures 2.4 and 2.5), we found that up to 1.8 mM of

doxycycline and 0.22 mM of suramin could be added without significant β 2m aggregation being observed within 1 h. In both cases, these concentrations correspond to over 85% of the protein being bound to the given molecule. In contrast, only 0.25 mM of rifamycin SV could be added without generating a significant amount of β 2m oligomers in 1 min. Under these conditions, about 33% of the Cu(II)- β 2m complexes are not bound to rifamycin SV.

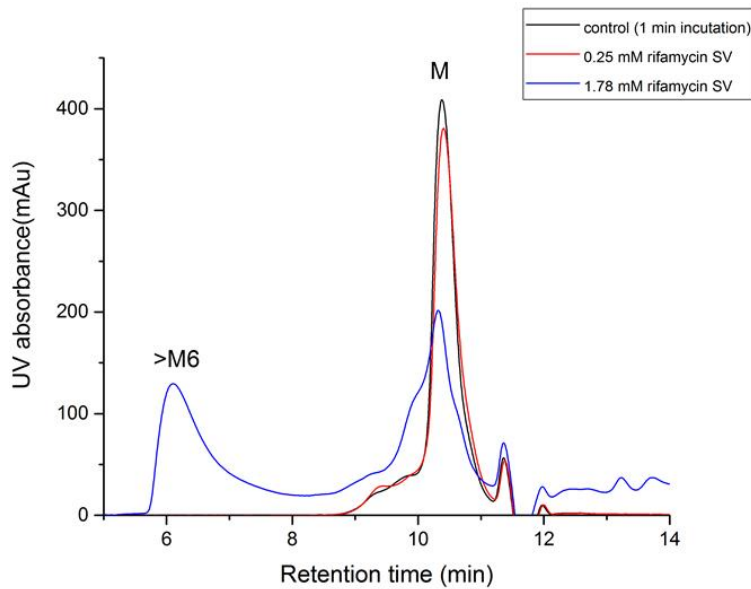


Figure 2.4 Example SEC chromatograph for β 2m incubated with Rifamycin SV. High concentration of rifamycin SV (1.78 mM) will generate β 2m oligomers (>M6), while 0.25 mM rifamycin SV will not.

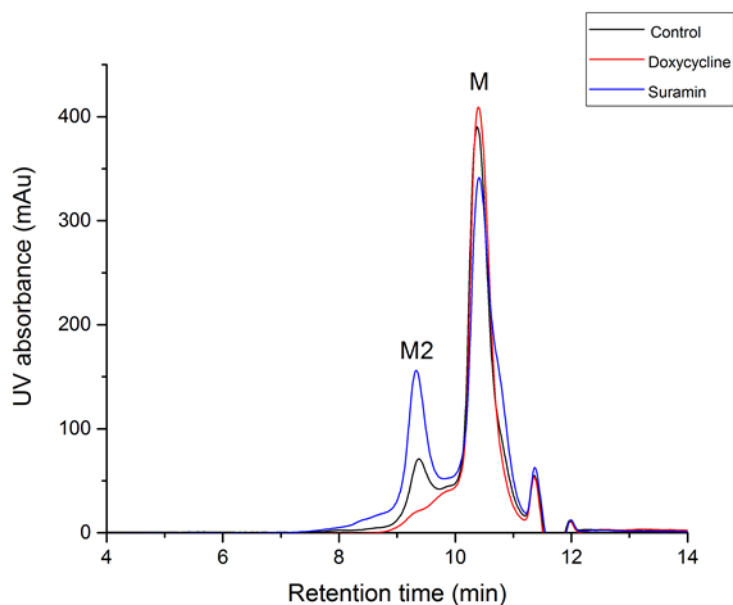


Figure 2.5 Example SEC chromatograph for β 2m incubated with doxycycline and suramin.

High concentration of doxycycline (1.8 mM) or suramin (0.22 mM) will only generate minor amount of oligomer under selected experimental labeling conditions.

2.3.3 Rifamycin SV Binding

The β 2m-Cu(II) complex was covalently labeled in the presence and absence of rifamycin SV as a way to identify the binding site of this molecule. DEPC was first used because it can modify a wide range of amino acids, including His, Lys, Ser, Thr, Tyr, and the N-terminus. Results from the DEPC labeling experiments are shown Figure 2.6.1. Upon binding rifamycin SV, Lys6, Lys91 and Lys94 are found to undergo significant decreases in labeling, while the N-terminus undergoes a significant increase in labeling. It should be noted that the extents of the labeling changes are small in many cases, but the reproducibility of the experiments is usually sufficient to confidently distinguish even small changes. The reason for the small changes in labeling extent is likely attributed to the fact that about 33% of Cu(II)- β 2m complex is not bound to rifamycin SV for the conditions under which we performed the covalent labeling experiments.

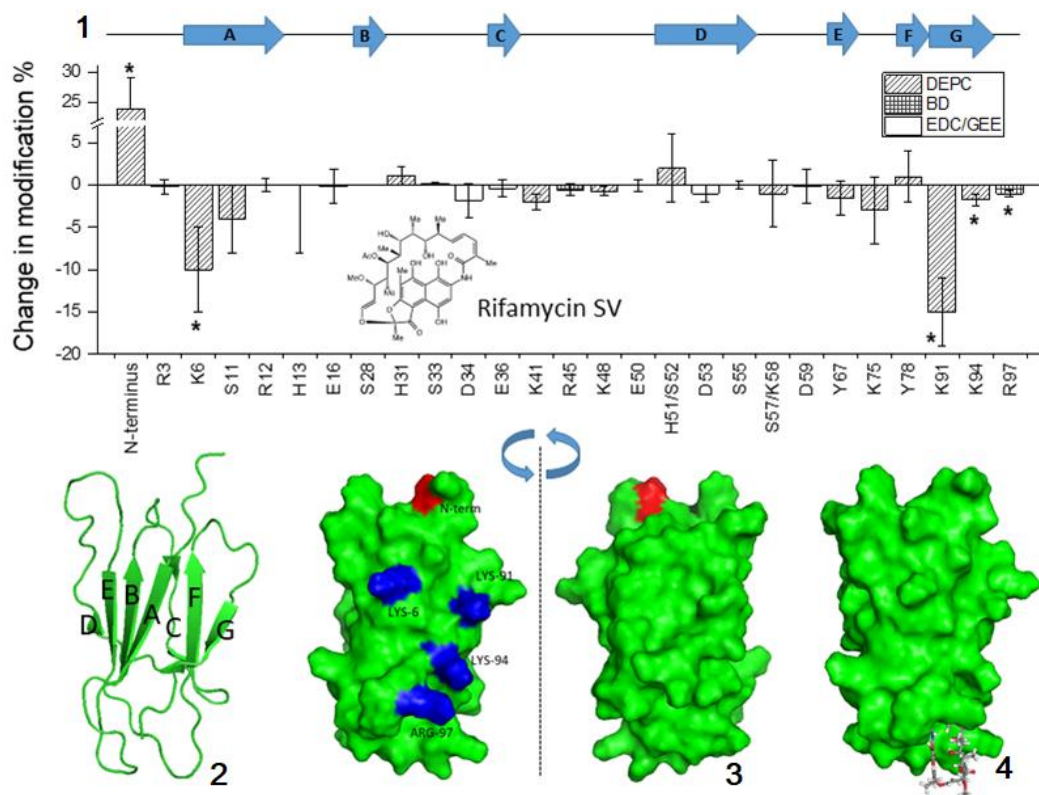


Figure 2.6 CL-MS results of rifamycin SV with Cu(II)-β2m

(1) Changes in covalent labeling modification percentages with rifamycin SV bound to the Cu(II)-protein complex. Error bars represent a single standard deviation from three replicate measurement. Asterisks above the bars represent the residues that undergo a significant change in modification level at a 95% confidence interval as determined by a two-sample unpaired Student's t-test. The arrows at the top of the graph indicate the locations and directions of the seven β strands in β2m. (2) Ribbon structure of β2m, showing the seven β strands labeled A through G. (3) β2m surface structure illustrating the sites of significant changes in covalent labeling induced by rifamycin SV. Sites that increase in labeling are colored red, while sites that decrease in labeling are colored blue. (4) Protein-ligand docking result.

To assess the potential rifamycin SV binding site, the covalent labeling results were mapped on the NMR structure (PDB ID 1JNJ) of the Cu(II)-free protein. This structure was selected as the model because other work had demonstrated that Cu(II) binding does not dramatically change the native β2m structure [19, 35]. (Note: no Cu(II)-β2m monomer structures are available.) Upon considering the protein model structure, it

is clear that Lys6, Lys91 and Lys 94 are in relatively close proximity to each other, with about 17 Å between Lys6 and Lys91 and 12 Å between Lys91 and Lys94. Given the size of rifamycin SV, which is approximately 14 x 7 Å, the flexibility of the lysine side chains, and the fact that Cu(II)-binding causes only minor structural changes to β 2m [19, 35], the DEPC labeling results narrow down the binding site to an area near the A and G β strands. The reason for the increased labeling at the N-terminus is not clear at this stage, but it could potentially reflect an allosteric change caused by rifamycin SV binding.

To obtain additional confirmation of the binding site and further narrow down the β 2m residues that interact with rifamycin SV, we employed two other types of labels. The arginine-selective reagent 2,3-butanedione (BD) was selected because Arg3 and Arg12 flank the residues that are part of the A β strand, Arg97 is near the G β strand, and Arg45 is on the opposite side of the protein (see Figure 2.7). We predicted that changes in the extent of labeling of the nearby Arg residues would further help narrow potential rifamycin SV binding sites. From the BD labeling results (Figure 2.6.1), only Arg97 is found to undergo a significant change in labeling, as its labeling percentage decreases from 6.4 ± 0.3 to 5.4 ± 0.2 in the presence of rifamycin SV. Insignificant labeling changes to Arg3, Arg12, and Arg45 along with a decrease in Arg97 labeling suggest that rifamycin SV binds predominantly along the G β strand.

We also used the reagent pair EDC/GEE to monitor changes in the solvent accessibility of Asp and Glu residues. Seven Asp and Glu residues are modified by EDC/GEE, including Glu16, Asp34, Glu36, Asp38, Glu50, Asp53, and Asp59, but none of these residues undergo a significant change in labeling, which is not surprising considering that all but one of these residues are located on the opposite face of the

protein (see Figure 2.8) from the residues identified by DEPC and BD labeling. The fact that there is no change in labeling on Glu16, which is located on the loop that is between the A and B β strands, further confirms that rifamycin SV is not binding in this region of the protein but is binding closer to the G β strand.

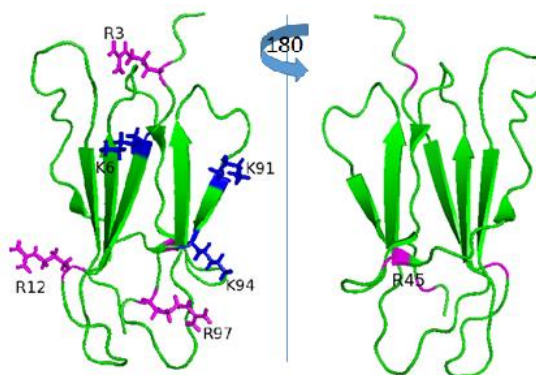


Figure 2.7 BD labeling assists with rifamycin SV binding site identification. Lys6, Lys 91 and Lys94 form a cluster that rifamycin SV could possibly near. Arg3, Arg12 and Arg97 are near this cluster, and their labeling could further narrow down the binding site.

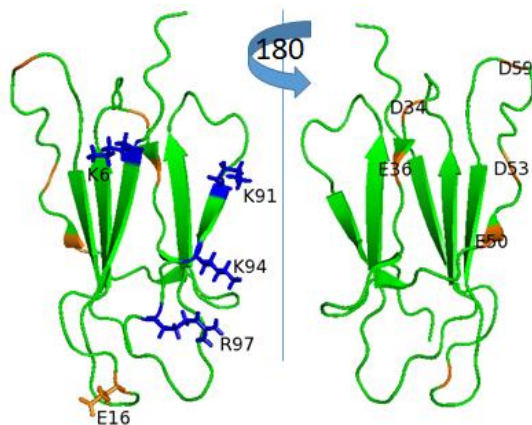


Figure 2.8 EDC/GEE labeling assists with rifamycin SV binding site identification. The cluster of significantly decreased sites include Lys6, Lys91, Lys 94 and Arg97. Asp16 is the only Asp/Glu that is close to this cluster.

Protein-ligand docking was also performed using the program Glide from the Schrödinger software suite to predict rifamycin SV's binding site on β 2m. The Cu(II)-free β 2m monomer (PDB ID 1JNJ) was used as the protein model, even though β 2m's

CD spectroscopy (Figure 2.9.1), ion mobility (Figure 2.9.3), its fluorescence λ_{\max} (Figure 2.9.5 and 2.9.6), and thermal stability (Figure 2.10) data reveal rifamycin SV binding slightly perturbs the native $\beta 2m$ structure. It should be noted that while it is known that Cu(II) binding does not have a dramatic effect on $\beta 2m$ structure [19, 35], it does have a measurable effect on $\beta 2m$ thermal stability (Figure 2.10). In comparison, rifamycin binding to the Cu(II)-bound protein stabilizes the structure to be more similar to the native. Thus, in the absence of any other $\beta 2m$ structural model, we feel that the chosen model structure is a reasonable starting point for docking experiments. Upon using this structural model for docking, we find that the predicted binding site is the pocket between Lys75 and Arg97 (Figure 2.6.4). In light of our covalent labeling, we conclude that rifamycin SV likely binds near Lys94 and Arg97 and perhaps Lys91, which is along the G β strand of the protein. A perfect match between the docking and labeling results is not expected given the fact that rifamycin SV mildly perturbs the native structure of $\beta 2m$ (Figure 2.10), but the similarity is encouraging. The decreased labeling of Lys6 may be due to its proximity to the other residues, subtle structural changes that the protein undergoes upon binding to rifamycin SV, or could perhaps be due to burial in the small fraction of protein molecules that have a second rifamycin SV bound to them.

More importantly, the binding site that is identified by covalent labeling is consistent with what is known about pre-amyloid oligomeric $\beta 2m$ structures [21, 22] and our recent work on $\beta 2m$ amyloid inhibition by rifamycin SV [10]. Our group recently showed that rifamycin SV inhibits $\beta 2m$ amyloid formation by preventing the $\beta 2m$ dimer from forming an amyloid-competent tetramer. We had previously found that the $\beta 2m$ tetramer is formed by two dimers interacting between the G β strands and the D β strands

[22]. Considering the binding site identified in the current work, rifamycin SV's proposed mode of action for preventing β 2m amyloid formation is to bind near the G β strands (i.e. Figure 2.6) and prevent two dimers from forming an amyloid-competent tetramer.

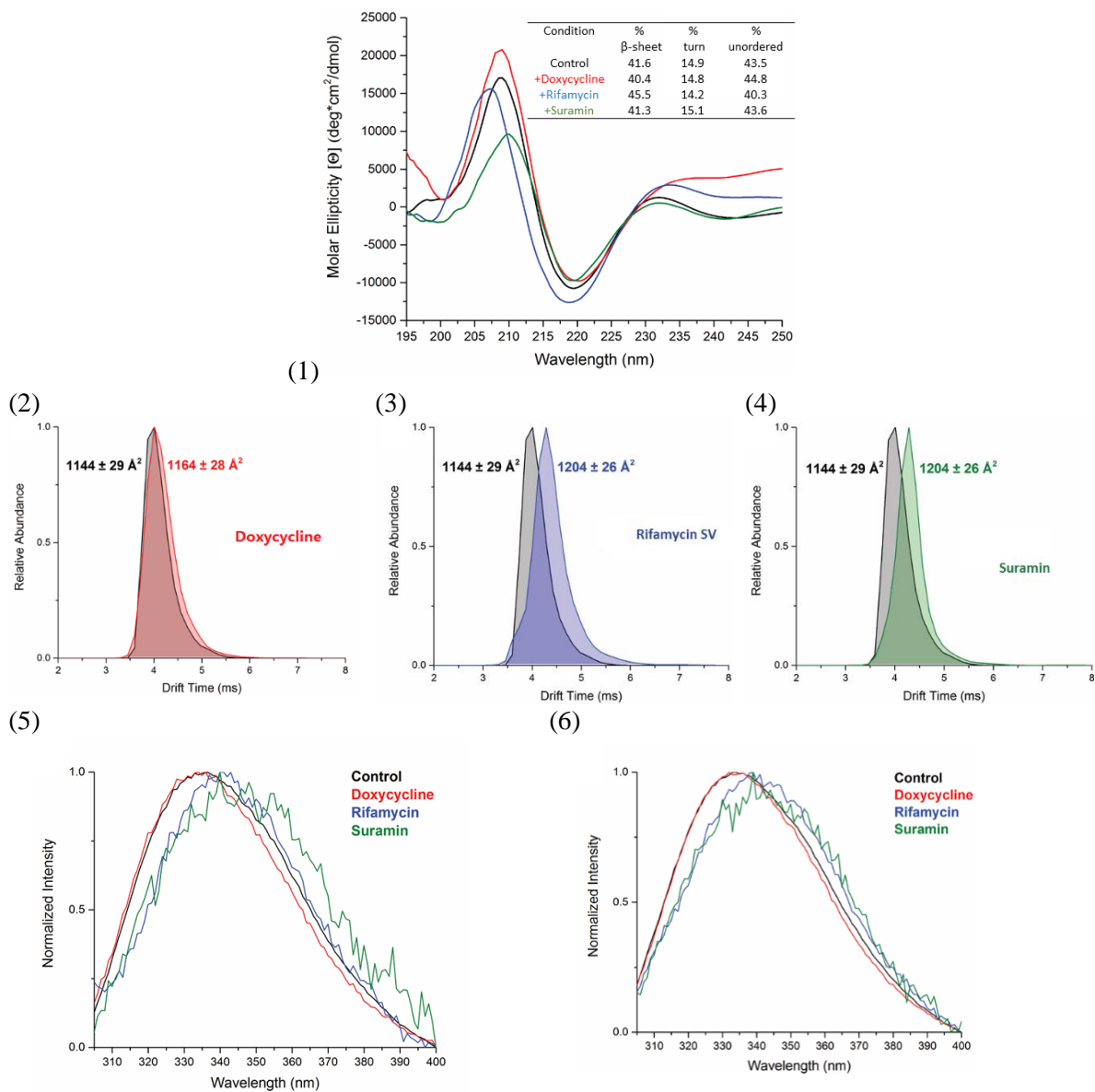


Figure 2.9 Structure similarity of Cu(II)-β2m with doxycycline, rifamycin SV, or suramin to Cu(II)-β2m monomer.

(1) Comparison of changes in secondary structure of β2m in the presence of small molecules. (2-4) comparing arrival time distribution and collisional cross section reveals expanded states in the gas phase for monomeric β2m when bound to small molecules, doxycycline (2), rifamycin (3), and suramin (4). (5-6) comparison of changes in tertiary structure via intrinsic fluorescence of β2m in the presence of small molecules. Emission spectra following excitation at 295 nm (5) and 283 nm (6).

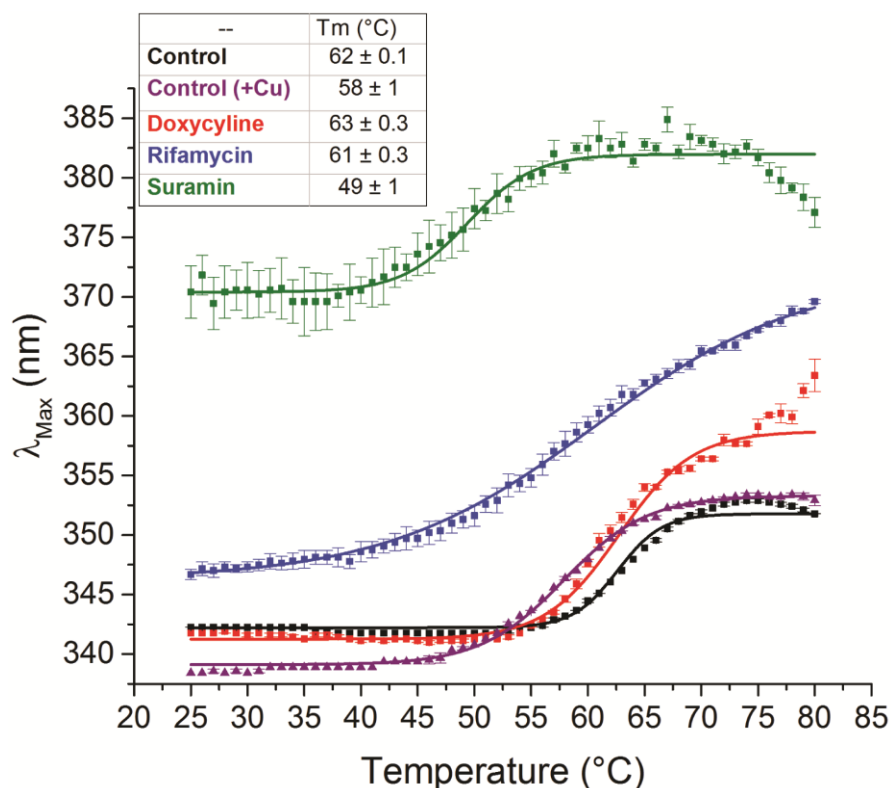


Figure 2.10 Effect of small molecules on thermal stability of β 2m.

T_m of β 2m, Cu(II)- β 2m, and Cu(II)- β 2m with rifamycin SV, doxycycline, or suramin were measured. The thermal stability of β 2m was assessed using intrinsic tryptophan fluorescence at 295 nm in triplicate. Calculated midpoint T_m values are summarized in the inset table. Error bars represent the standard error of the mean.

2.3.4 Doxycycline Binding

The binding site of doxycycline was determined in a similar manner to rifamycin SV. DEPC labeling was first applied and numerous residues undergo significant changes in labeling (Figure 2.11). More sites undergo significant changes than with rifamycin SV, and this increased number of residues is probably because more than 90% of Cu-protein complex is bound to doxycycline. Again, the Cu(II)-free β 2m NMR structure (PDB ID 1JNJ) is used to map the labeling changes for reference. The sites that undergo decreased labeling can be clustered into two primary groups: 1) Lys6, Lys91, and Lys94, and 2) Lys41, Lys48 and His51/Ser52. This clustering of residues is based on the distribution of

these sites on the protein surface and the size of one doxycycline molecule (about 13 x 7 Å). Lys75 and Ser57/Lys58 are not included in either group because they are somewhat distant from each cluster of residues. It is difficult to exclude either group based on the extent of the labeling decrease because allosteric changes or changes in the local microenvironment can alter the reactivity of residues for reasons other than decreased solvent accessibility. These possible changes are particularly true for Lys residues, as disruption of salt bridges or changes in pK_a , due to microenvironment changes, can influence their reactivity with DEPC [36].

Interestingly, the N-terminus, His13, Ser28, His31 and Ser33 all undergo significant increases in labeling. Among these five residues, the N-terminus, Ser28, His31 and Ser33 are close to or part of the known Cu(II) binding site [19, 35]. Increased labeling at these residues suggests that Cu(II) binding is affected by doxycycline binding. Indeed, as we reported in recent work, $\beta 2m$'s affinity for Cu(II) decreases by approximately 25-fold in the presence of doxycycline [10]. The increased labeling extent of His13 is more difficult to explain but may be due to an allosteric change.

EDC/GEE labeling was again applied to narrow down potential binding sites. As shown in Figure 2.12, Glu50 and Asp53 are near His51, offering a potential means of distinguishing the second group of residues identified by DEPC labeling (i.e. Lys41, Lys48 and His51/Ser52) from the first group (i.e. Lys6, Lys91, and Lys94). We find that only Glu50, of all the Asp and Glu residues, undergoes a significant decrease in labeling, which is consistent with doxycycline binding near the cluster defined by Lys41, Lys48, and His51/Ser52 but not the cluster defined by Lys6, Lys91, and Lys94. Moreover, Asp34, Glu36, and Asp38 are on the same C β strand as Lys41, but these residues do not

undergo a decrease in labeling, indicating that doxycycline is not binding to residues on the C β strand. The fact that both Glu50 and Asp53 are on the same D β strand as Lys48 and His51/Ser52, but only Glu50, not Asp53, undergoes a significant labeling decrease suggests that doxycycline only binds to the beginning of D β strand.

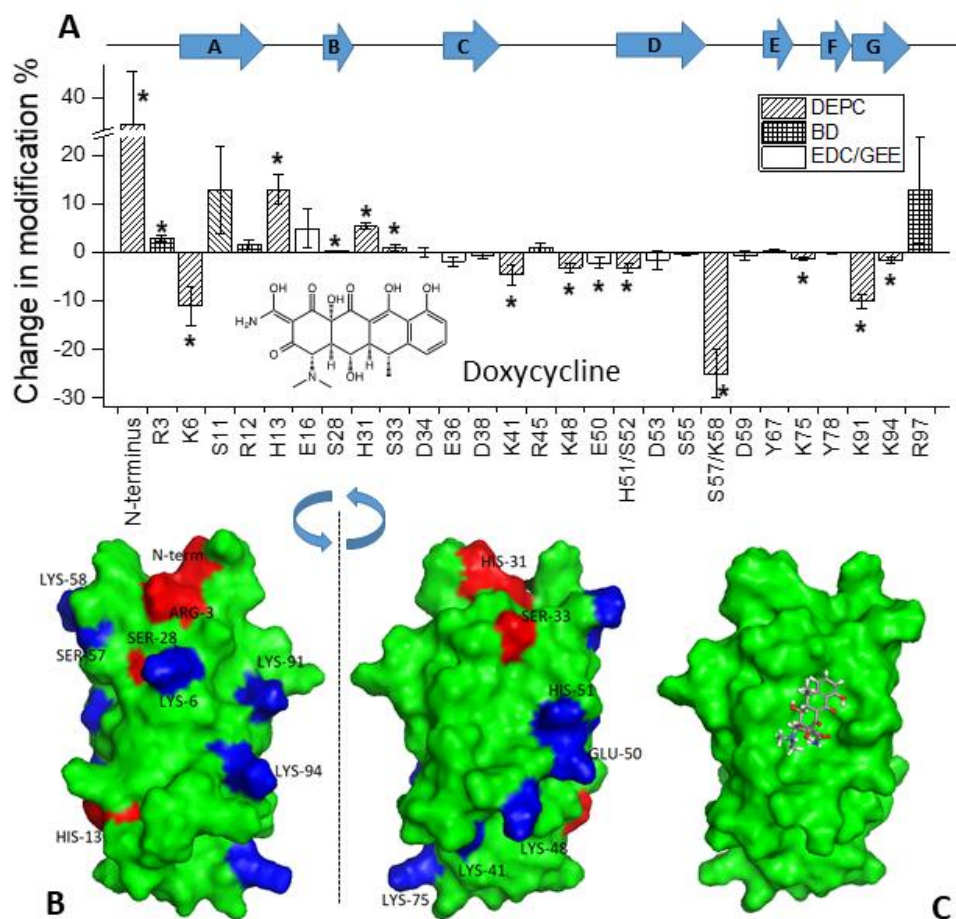


Figure 2.11 CL-MS results of doxycycline with Cu(II)- β 2m

(1) Changes in covalent labeling modification percentages with doxycycline bound to the Cu(II)-protein complex. Error bars represent a single standard deviation from three replicate measurement. Asterisks above the bars represent the residues that undergo a significant change in modification level at a 95% confidence interval as determined by a two-sample unpaired Student's t-test. The arrows at the top of the graph indicate the locations and directions of the seven β strands in β 2m. (2) β 2m surface structure illustrating the sites of significant covalent labeling changes induced by doxycycline. Sites that increase in labeling are colored red, while sites that decrease in labeling are colored blue. (3) Protein-ligand docking result.

Because no Asp or Glu residues are near the first cluster of residues identified by DEPC labeling, we further investigated the binding site by BD labeling. Arg3, Arg45, and Arg97 surround cluster 1 (Figure 2.13). Arg3 and Arg97 undergo an increase in labeling, suggesting that doxycycline is not binding near Lys6, Lys91, or Lys94. The changes in the labeling of these Lys residues may be caused by DEPC reactivity decreases caused by changes in the nearby microenvironment. Interestingly, the labeling extent of Arg45, which is near Lys41, does not change significantly. Alongside the lack of labeling changes associated with Asp34, Glu36, and Asp38, no apparent change in the labeling extent at Arg45 further suggests that doxycycline does not bind to the C β strand. Considering the labeling results as a whole, we conclude that doxycycline binds near Lys48, Glu50 and His51/Ser52. These residues all undergo significant decreases in labeling ratio and all are located on the beginning of D β strand.

We also used protein-ligand docking with Glide to predict the doxycycline binding site on β 2m. The Cu(II)-free β 2m monomer (PDB ID 1JNJ) was again used as the protein model because the CD spectroscopy (Figure 2.9.1), ion mobility (Figure 2.9.2), its fluorescence λ_{\max} (Figure 2.9.5 and 2.9.6), and thermal stability (Figure 2.10) data indicate that doxycycline binding only slightly perturbs the protein's structure. Indeed, the doxycycline appears to perturb the structure even less than rifamycin SV, suggesting that the chosen model is even more appropriate for docking. Using this structural model, docking predicts a binding site that is in the pocket between the C and D β strands (Figure 2.11.3). This docking position is very consistent with the labeling data that shows several residues on the D β strand, including Lys48, Glu50, and His51/Ser52, undergoing significant changes in labeling (Figure 2.14). Given the fact that

several residues on C β strand have no significant changes in the labeling degree and decreases in labeling are observed along the D β strand, it is reasonable to conclude that doxycycline binds near the D β strand.

Binding at the D β strand is consistent with our group's recent work that shows doxycycline can redirect β 2m amyloid formation by perturbing the β 2m dimer structure such that it cannot form an amyloid-competent tetramer [10]. As indicated above and as previously shown, the β 2m tetramer is formed by two dimers interacting between the G β strands and D β strands [22]. Binding of doxycycline to the D β strand, as indicated by our labeling results, reveals how the amyloid formation pathway can be redirected to form non-amyloid aggregates. It should be noted that Giorgetti et al. studied doxycycline/ β 2m binding by NMR and identified a different binding site that involves β strands A and B [8]. These measurements, however, were done under non-amyloid forming conditions in which the protein was present in pure water at a pH between 5.5 and 6. The difference in the binding site that we identify as compared to this previous study likely reflects the fact that the protein likely has a significantly different structure under amyloid-forming conditions at pH 7.4, which are the conditions we have used in our studies here.

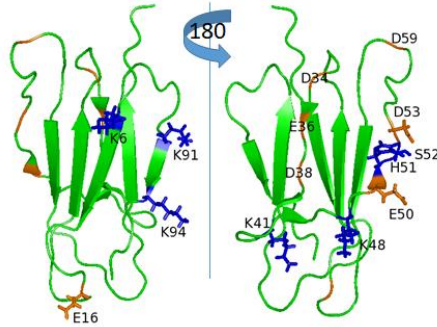


Figure 2.12 EDC/GEE labeling assists with doxycycline binding site identification. Ribbon structure of $\beta 2m$ monomer showing the distribution of glutamic and aspartic acids that can be labeled by EDC-GEE. Cluster 1 includes Lys6, Lys91 and Lys94, and Glu16 is the closest to this cluster. Cluster 2 includes Lys41, Lys48 and His51/Ser52, and Asp50 and Glu53 are near this cluster.

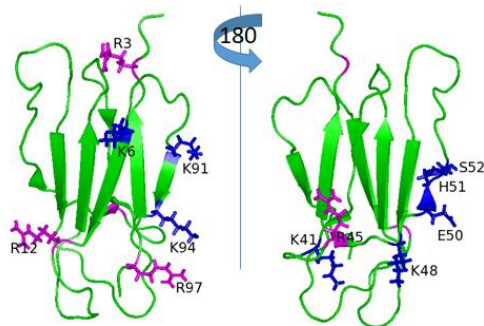


Figure 2.13 BD labeling assists with doxycycline binding site identification. Ribbon structure of $\beta 2m$ monomer showing the distribution of arginine residues that can be labeled by BD. Cluster 1 includes Lys6, Lys91 and Lys94, and Arg3, Arg12 and Arg97 are close to this cluster. Cluster 2 includes Lys41, Lys48, Asp50 and His50/Ser52, and Arg45 is next to this cluster.

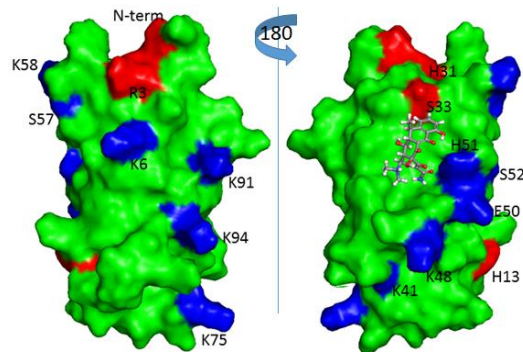


Figure 2.14 Labeling experiment result and docking result presenting on the protein surface structure: doxycycline. A residue significant change in labeling upon doxycycline binding is marked in blue (decrease) or red (increase). Docking and CL experiment results both indicate doxycycline binding site is close to D β strand.

2.3.5 Suramin Binding

Suramin binds with greater affinity to $\beta 2m$ than doxycycline and rifamycin SV, but as described previously it does not inhibit amyloid formation [10]. Thus, we hypothesized that the identity of the suramin binding site would be different from the doxycycline and rifamycin SV binding sites such that it would not interfere with oligomer formation. DEPC labeling results in four sites that decrease in labeling (Figure 2.15.1), and these sites are clustered into two groups based on proximity and suramin's size: 1) Lys41 and Lys48 and 2) the N-terminus and Ser57/Lys58. Suramin is a larger molecule than doxycycline and rifamycin SV, and it is less conformationally constrained than the other two, which suggests that it could potentially cover a larger area of $\beta 2m$'s surface.

With relatively little distinguishing information obtained from DEPC labeling alone, we also used EDC/GEE labeling to narrow further the binding site. Several Asp and Glu residues along the C and D β strands can be probed (Figure 2.16), and strikingly Asp34, Glu36, and Glu50 all undergo significant decreases in labeling (Figure 2.15.1). These three residues are positioned near Lys41 and Lys48, suggesting that suramin binds near the C and/or D β strand, rather than near the N-terminus. Curiously, Asp38 and Asp53 undergo increased labeling despite being positioned on the C and D β strands.

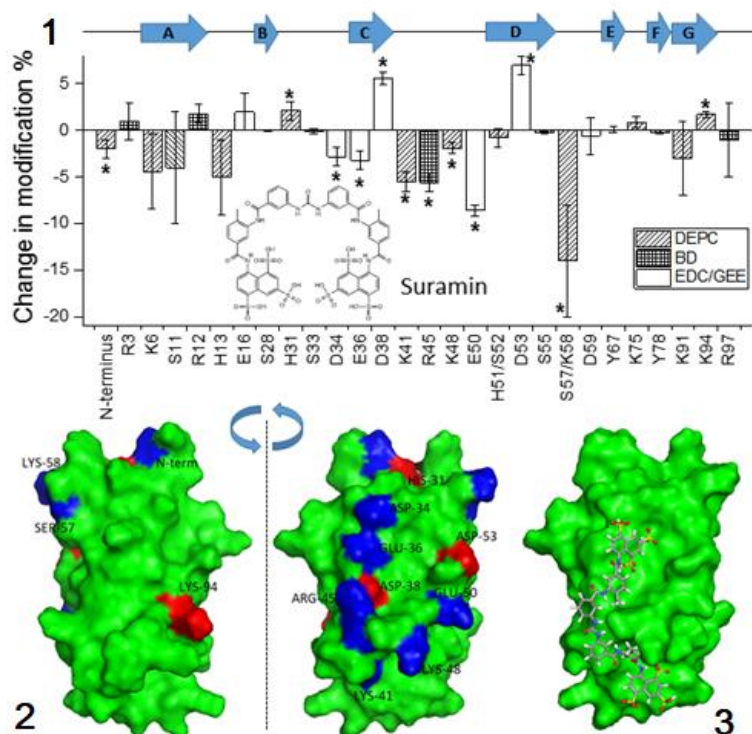


Figure 2.15 CL-MS results of suramin with Cu(II)-β2m.

(1) Changes in covalent labeling modification percentages with suramin bound to the Cu(II)-protein complex. Error bars represent a single standard deviation from three replicate measurement. Asterisks above the bars represent the residues that undergo a significant change in modification level at a 95% confidence interval as determined by a two-sample unpaired Student's t-test. The arrows at the top of the graph indicate the locations and directions of the seven β strands in β2m. (2) β2m surface structure illustrating the sites of significant covalent labeling changes induced by suramin. Sites that increase in labeling are colored red, while sites that decrease in labeling are colored blue. (3) Proposed suramin binding site based on covalent labeling data.

BD labeling was used further to identify the suramin binding site. Consistent with the EDC/GEE labeling decreases found for residues on the C β strand, Arg45, which is also on the C β strand (Figure 2.17), undergoes a significant labeling decrease (Figure 2.15.1). The three other Arg residues in the protein, including Arg3, which is near the N-terminus, do not undergo significant changes in labeling. Considering the labeling data as a whole, the suramin binding site appears to be located along the C β strand and the C-D loop, based on the decreased labeling of Asp34, Glu36, Lys41, Arg45, and Lys48 (Figure

2.15.2). The distance from Asp34 to Lys41 is approximately 23 Å, which can be easily spanned given the molecular dimensions of suramin. Moreover, suramin has negatively charged sulfonate groups that could likely bind to Lys41, Arg45, and Lys48. These sulfonate groups might also explain why Asp38, and perhaps Asp53, undergo increased labeling as they are electrostatically repelled by the sulfonate groups and made more solvent accessible. As expected from the outset, this binding site along the C β strand is in contrast to the doxycycline and rifamycin SV binding sites.

Once again, protein-ligand docking was considered for predicting the suramin binding site on β2m. Because the CD spectroscopy (Figure 2.9.1), ion mobility (Figure 2.9.3), its fluorescence λ_{max} (Figure 2.9.5 and 2.9.6), and thermal stability (Figure 2.10) measurements suggest β2m's structure is significantly perturbed upon suramin binding, we did not expect good agreement between the docking and covalent labeling results. Indeed, all 7 poses with docking scores below zero place suramin near the N-terminus or near the A or D β strands (Figure 2.18 and Table 2.1). These poses do not agree well with the labeling data. In fact, these docking results position suramin near the known β2m dimer interface [21, 22], which is inconsistent with suramin's inability to inhibit β2m amyloid formation or the progression of the oligomers that precede amyloid formation [10]. Thus, we conclude that the docking results are inaccurate, highlighting the value of experimental labeling data. This inconsistency between theory and experiment likely arises because the β2m model used for docking is probably not a good one.

Based on the labeling results, we conclude that suramin binds along the C β strand and near the C-D loop (Figure 2.15.3), which is consistent with the observation that suramin is not capable of inhibiting β2m amyloid formation [10]. From our previous

studies of β 2m oligomer structures [21, 22], the C β strand has not been implicated as part of the interface in either dimer or tetramer. Binding of suramin to this region of the protein then likely explains why it does not substantially influence β 2m amyloid formation.

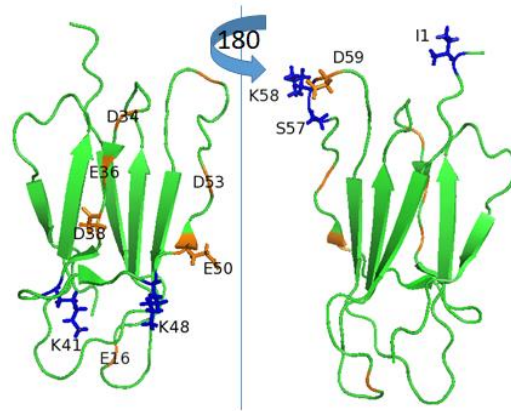


Figure 2.16 EDC/GEE labeling assists with suramin binding site identification. Ribbon structure of β 2m monomer showing the distribution of glutamic and aspartic acids that are labeled by EDC-GEE. Cluster 1 includes Lys41 and Lys48, and Asp34, Glu36, Asp38 and Glu50 are the close to this cluster. Cluster 2 includes N-terminus and Lys58/Ser57, and Asp59 is near this cluster.

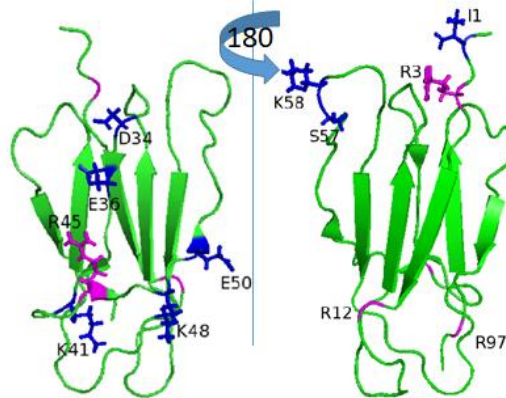


Figure 2.17 BD labeling assists with suramin binding site identification. Ribbon structure of β 2m monomer showing the distribution of arginine residues that are labeled by BD.

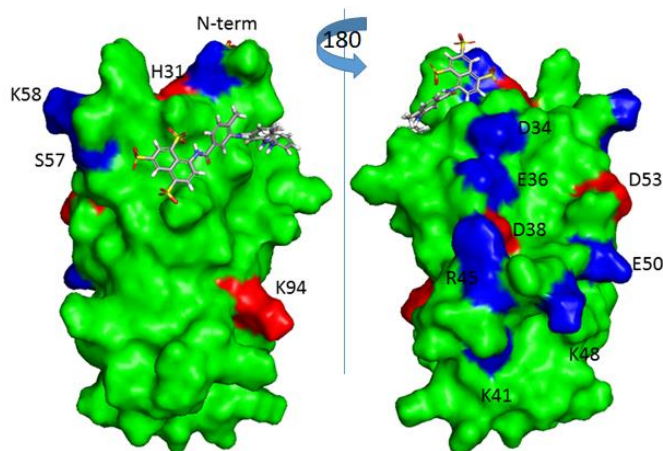


Figure 2.18 Top Glide docking predicted suramin binding pose compared to experiment result.

A residue significant change in labeling upon suramin binding is marked in blue (decrease) or red (increase). Results show inconsistent between docking result and CL proposed suramin binding site.

Table 2.1 Suramin docking score.

Pose Rank	Docking Score
1	-3.077
2	-2.685
3	-2.569
4	-2.189
5	-1.147
6	-1.011
7	-0.812

*All docking results from Glide docking, ranked by docking score from lowest to highest. The most negative docking score indicates the best docking results according to the Glide algorithm.

2.4 Conclusion

Our results indicate that covalent labeling together with MS is a useful tool for studying protein-ligand binding sites. Changes in labeling extents in clusters of adjacent amino acid residues, with and without ligand molecules present, allow ligand binding site information to be obtained. Confidence in the binding site determination is notably improved when multiple covalent labeling reagents are used together. Using DEPC, BD,

and the EDC/GEE pair as labeling reagents, we identify the binding sites of two small-molecule inhibitors of β 2m amyloid formation. The identified binding sites are consistent with the known effects of these molecules and with molecular docking results, thereby providing validation of the covalent labeling method. Rifamycin SV is found to bind to β 2m on the G β strand, while doxycycline binds on the D β strand of β 2m. These binding sites are near the known interfaces of the pre-amyloid tetramer of β 2m [22] and are consistent with previous measurements that indicate that these molecules inhibit amyloid formation by preventing the amyloid-competent tetramer from forming [10]. Moreover, the binding site of suramin, which does not inhibit β 2m amyloid formation, occurs near the C β strand and C-D loop of the protein, a region not involved in any known oligomer interface, further validating the accuracy of the covalent labeling method. The fact that three molecules bind to different sites on β 2m may seem surprising, given the small size of the protein, but its β -sandwich structure presents distinct surface regions and inter-sheet binding pockets that evidently can be exploited by the three chemically different molecules. The covalent labeling/MS method described here appears to be suitable for providing protein-ligand binding information for systems that are difficult to study by other methods. In addition, the method might be used to facilitate drug design efforts by improving the targeted virtual screening of compound libraries by narrowing the protein sites most likely to influence a protein's reactivity.

2.5 Reference

- [1] Hazenberg, B. P. *Rheum. Dis. Clin. North. Am.* **2013**, *39*, 323-345.
- [2] Floege, J.; Ketteler, M. *Kidney Int.* **2001**, *59*, S164-S171.
- [3] Drüeke, Tilman B.; Ziad A. Massy. *Semin. Dial.* **2009**, *22*, 378-380.
- [4] Heegaard, N. H. *Amyloid* **2009**, *16*, 151-173.
- [5] Gejyo, F.; Odani, S.; Yamada, T.; Honma, N.; Saito, H.; Suzuki, Y.; Nakagawa, Y.; Kobayashi, H.; Maruyama, Y.; Hirasawa, Y.; Suzuki, M. *Kidney Int.* **1986**, *30*, 385-390.
- [6] Yamamoto, S.; Kazama, J. J.; Maruyama, H.; Nishi, S.; Narita, I.; Gejyo, F. *Clin. Nephrol.* **2008**, *70*, 496-502.
- [7] Gejyo, F.; Kawaguchi, Y.; Hara, S.; Nakazawa, R.; Azuma, N.; Ogawa, H.; Koda, Y.; Suzuki, M.; Kaneda, H.; Kishimoto, H.; Oda, M. *Artif. Organs.* **2004**, *28*, 371-380.
- [8] Giorgetti, S.; Raimondi, S.; Pagano, K.; Relini, A.; Bucciantini, M.; Corazza, A.; Fogolari, F.; Codutti, L.; Salmona, M.; Mangione, P.; Colombo, L.; *J. Biol. Chem.* **2011**, *286*, 2121-2131.
- [9] Woods, L. A.; Platt, G. W.; Hellewell, A. L.; Hewitt, E. W.; Homans, S. W.; Ashcroft, A. E.; Radford, S. E. *Nat. Chem. Biol.* **2011**, *7*, 730-739.
- [10] Marcinko, T.M.; Dong, J.; LeBlanc, R.; Daborowski, K.V.; Vachet, R.W. *J. Biol. Chem.* **2017**, *292*, 10630-10638.
- [11] Mendoza, V.L.; Vachet, R.W. *Mass Spectrom. Rev.* **2009**, *28*, 785-815.
- [12] Xu, G.; Chance, M.R. *Chem. Rev.* **2007**, *107*, 3514-3543.
- [13] Wang, L.; Chance, M.R. *Anal. Chem.* **2011**, *83*, 7234-7241.
- [14] Chen, J.; Rempel, D.L.; Gau, B.C.; Gross, M.L. *J. Am. Chem. Soc.* **2012**, *134*, 18724-18731.
- [15] Zhang, Y.; Weckler, A. T.; Molina, P.; Deperalta, G.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 850-858.
- [16] Jumper, C.C.; Schriemer, D.C. *Anal. Chem.* **2011**, *83*, 2913-2920.
- [17] Jumper, C.C.; Bomgarden, R.; Rogers, J.; Etienne, C.; Schriemer, D.C. *Anal. Chem.* **2012**, *84*, 4411-4418.
- [18] Manzi, L.; Barrow, A.S.; Scott, D.; Layfield, R.; Wright, T.G.; Moses, J.E.; Oldham, N.J. *Nat. Commun.* **2016**, *7*, 13288.
- [19] Mendoza, V.L.; Vachet, R.W. *Anal. Chem.* **2008**, *80*, 2895-2904.
- [20] Srikanth, R.; Mendoza, V.L.; Bridgewater, J.D.; Zhang, G.; Vachet, R.W. *Biochemistry* **2009**, *48*, 9871-9881.
- [21] Mendoza, V. L.; Antwi, K.; Barón-Rodríguez, M. A.; Blanco, C.; Vachet, R. W. *Biochemistry* **2010**, *49*, 1522-1532.
- [22] Mendoza, V. L.; Barón-Rodríguez, M. A.; Blanco, C.; Vachet, R. W. *Biochemistry* **2011**, *50*, 6711-6722.
- [23] Zhou, Y.; Vachet, R. W. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 708-717.
- [24] Zhou, Y.; Vachet, R. W. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 899-907.
- [25] Borotto, N.B.; Zhou, Y.; Hollingsworth, S.R.; Hale, J.E.; Graban, E.M.; Vaughan, R.C.; Vachet, R.W. *Anal. Chem.* **2015**, *87*, 10627-10634.
- [26] Fliss, H.; Viswanatha, T. *Can. J. Biochem.* **1979**, *57*, 1267-1272.

- [27] Kaur, P.; Tomechko, S.E.; Kiselar, J.; Shi, W.; Deperalta, G.; Weckslar, A.T.; Gokulrangan, G.; Ling, V.; Chance, M.R. *In MAb*s **2015**, *7*, 540-552.
- [28] Regazzoni, L.; Colombo, R.; Bertolotti, L.; Vistoli, G.; Aldini, G.; Serra, M.; Carini, M.; Facino, R.M.; Giorgetti, S.; Stoppini, M.; Caccialanza, G. *Anal. Chim. Acta* **2011**, *685*, 153-161.
- [29] Eakin, C.M.; Knight, J.D.; Morgan, C.J.; Gelfand, M.A.; Miranker, A.D. *Biochemistry* **2002**, *34*, 10646-10656.
- [30] Eakin, C.M.; Attenello, F.J.; Morgan, C.J.; Miranker, A.D.; *Biochemistry* **2004**, *24*, 7808-7815.
- [31] El-Hawiet, A.; Kitova, E.N.; Klassen, J.S. *Biochemistry* **2012**, *51*, 4244-4253.
- [32] Liu, L.; Kitova, E.N.; Klassen, J.S. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 310-318.
- [33] Hernández, H; Robinson, C.V., *Nat. Protoc.* **2007**, *2*, 715-726.
- [34] Bush, M.F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C.V.; Ruotolo, B.T.; *Anal. Chem.* **2010**, *82*, 9557-9565.
- [35] Eakin, C.M.; Knight, J.D.; Morgan, C.J.; Gelfand, M.A.; Miranker, A.D. *Biochemistry* **2002**, *41*, 10646-10656.
- [36] Isom, D.G.; Castañeda, C.A.; Cannon, B.R. *PNAS* **2011**, *108*, 5260-5265.

CHAPTER 3

PROTEIN-LIGAND AFFINITY DETERMINATIONS USING COVALENT LABELING-MASS SPECTROMETRY

Marcinko, T. M. contributed to this work by conducting the fluorescence spectroscopy and size exclusion chromatography measurements for characterizing the EGCG and Cu(II)- β 2m complex described in this chapter.

3.1 Introduction

Understanding the details of protein-ligand interactions has implications for drug discovery, design, and development [1, 2]. As a part of characterizing protein-ligand complexes of interest, binding affinity determination is a critical aspect. To acquire the binding affinity of the ligand to the protein, mass spectrometry (MS) based methods have been explored as one of the options because the advantages of the technique. Inherent advantages of MS such as specificity, small sample consumption, and simplicity (e.g. no immobilization needed) are appealing aspects of this approach [3]. Native electrospray ionization mass spectrometry (ESI-MS) allows noncovalent protein-ligand complexes to be transferred intact from solution to the gas phase [4-6], which makes ESI-MS-based titration experiments possible [7, 8]. However, although the ESI process can be gentle enough to preserve the protein-ligand complexes as they are ionized, there are still possibilities that the ion mixture measured by MS does not reflect the original components in solution. For example, the protein and protein-ligand complex could have different ionization and detection efficiencies. Protein-ligand complexes might also undergo in-source dissociation either due to the ligand binding weakly to the protein,

improper instrument settings, or non-specific ligand binding to the protein or protein-ligand complex that could cause false positives [3].

As ESI-MS signals do not always reflect the analyte concentration in solution, other MS-based methods that encode protein-ligand binding information into mass shifts have been developed. For example, hydrogen deuterium exchange (HDX)-based methods have been used to measure the dissociation constants (K_d) for protein-ligand complexes. One of the first examples was the method known as SUPREX (stability of unpurified proteins from rates of H/D exchange) [9], which utilizes denaturants to monitor the relative stabilities of a protein and its ligand complex to obtain a K_d [10]. Another HDX-MS method to determine K_d values is PLIMSTEX [11, 12], which relies on labeling in the context of ligand titration. PLIMSTEX also offers information about protein-ligand stoichiometries and dynamics at the same time. When residue or peptide-level information is obtained during these experiments, the binding site can be often acquired.

For HDX-based methods, back exchange is a concern, and the need to measure multiple exchange time points can make the measurements time consuming. Covalent labeling (CL)-based methods can avoid these challenges as the covalent bond that is formed limits label loss, and multiple time-point measurements are not needed [13-16]. Recently, a method based on fast photochemical oxidation of proteins (FPOP) called LITPOMS (ligand titration, fast photochemical oxidation of proteins and mass spectrometry) [17, 18] was reported. LITPOMS uses hydroxyl radicals to modify the protein of interest, and then relies on decreased protein oxidative modification that occurs at the ligand binding site as the ligand concentration is increased. The method yields information about ligand binding affinities, binding sites, and even ligand-induced

structural changes when combined with bottom-up sequencing. As compared to HDX-based methods, CL methods like LITPOMS require little sample dilution, making it easier to study weakly binding protein-ligand complexes. Given the potential advantages of CL for determining K_d values, we were interested in assessing if other CL methods and reagents could provide reliable protein-ligand binding constants. In particular, CL methods that do not require sophisticated equipment (e.g. laser or flow cell) and are simpler to use were sought. The work described here focuses on CL methods that do not use radical-based reagents. It is known that CL-MS, including non-radical CL methods, are able to identify the ligand binding sites of proteins [13, 19-22] by identifying residues that undergo decreases in labeling due to ligand-induced decreases in solvent accessibility of the residues involved in binding. If such CL is applied in a ligand titration context, K_d values should be accessible.

A key concern with non-radical based labeling reagents is their relatively slow reaction kinetics, which are on the order of milliseconds to seconds [14, 23] as compared to radical-based CL reagents that react on the sub-millisecond timescale [24, 25]. The slower reaction timescale might mean that the ligand dissociation and re-association processes that are a normal part of the equilibrium could occur on comparable timescales, yielding incorrect binding information because the labeling distorts the original binding equilibrium. Such a concern is presumably not present with radical-based methods such as FPOP because labeling happens on a faster timescale than ligand dissociation/re-association. We predict that relatively accurate binding constants can be obtained using slower labeling reagents as long as the extent of protein labeling at the binding site is minimized enough to prevent significant label-induced changes to the equilibrium. In this

work, we use several model protein-ligand complexes to demonstrate this idea. We find that reasonably accurate K_d determinations can be made when the extent of protein labeling is kept relatively low.

3.2 Materials and Methods

3.2.1 Materials

The maltose binding protein (MBP) was obtained from MyBioSource.com (San Diego, CA). Human full-length β -2-microglobulin (β 2m) was purchased from Lee Biosolutions (Maryland Heights, MO). Lysozyme from chicken egg white and the following chemicals were obtained from MilliporeSigma (St.Louis, MO): DL-dithiothreitol (DTT), diethylpyrocarbonate (DEPC), dimethyl(2-hydroxy-5-nitrobenzyl)sulfonium bromide (HNSB), epigallocatechin gallate (EGCG), guanidine hydrochloride (GuHCL), imidazole, iodoacetamide, maltose monohydrate, MOPS, MOPS sodium salt, N,N',N'' triacetylchitotriose (NAG3), L-tryptophan, and urea. Acetonitrile, copper sulfate (CuSO_4), formic acid, potassium acetate, sodium phosphate, sodium phosphate monobasic monohydrate, HPLC grade water, and a 1 M Tris buffer (pH 8) stock solution were all purchased from Fisher Scientific (Fair Lawn, NJ). Centricon molecular weight cutoff (MWCO) filters were obtained from Millipore (Burlington, MA). Sequencing grade modified trypsin and sequencing grade chymotrypsin were obtained from Promega (Madison, WI).

3.2.2 Sample Preparation

Samples containing 30 μM lysozyme were prepared in a 20 mM MOPS (pH 7.4) buffer, and stock solutions of NAG₃, which binds lysozyme, were prepared in HPLC

grade water. Samples containing 24 μM MBP were prepared in 20 mM sodium phosphate buffer (pH 7.5), and stock solutions of maltose were prepared in the same buffer. Samples containing 30 μM β2m were prepared in 25 mM MOPS buffer (pH 7.4), 60 μM Cu(II) (CuSO_4), 500 mM urea, and 200 mM potassium acetate. EGCG stock solutions were prepared in water. Different concentrations of the small molecule ligands were mixed with the corresponding protein to conduct the titration experiments. All samples were equilibrated at room temperature (22 $^\circ\text{C}$) for 2 to 5 min after the protein was mixed with the ligand and before performing the CL reactions.

3.2.3 HNSB Labeling

HNSB stock solutions were prepared in water. The reaction was initiated by adding an aliquot of the HNSB stock solution to the buffered protein-ligand sample, was allowed to proceed at room temperature, and then was quenched by adding tryptophan to a concentration of 5 mM. The concentration of HNSB and reaction times were varied to examine the effect of different labeling extents on the measured K_d values.

3.2.4 DEPC Labeling

DEPC stock solutions were freshly prepared in acetonitrile for each experiment. The reaction was initiated by adding an aliquot of the DEPC stock solution to the buffered protein-ligand sample, with the final DEPC concentration depending on the protein and the experiment. The final volume percentage of acetonitrile was less than 1% in all experiments. The reaction was allowed to proceed at 37 $^\circ\text{C}$ for 1 min, before being quenched by the addition of imidazole at a final concentration of 10 mM. These labeling conditions are similar to our group's previous studies with this CL reagent [23, 26].

3.2.5 Proteolytic Digestion

CL-modified proteins that were subjected to proteolytic digestion and LC/MS/MS analysis were prepared by the following procedure. Before denaturing the protein, samples were diluted with water to a volume of 400 μ L and then concentrated using a 10,000 MWCO filter to a final volume of 40 μ L. Samples containing MBP were reconstituted with 0.25 M Tris (pH 8.0) and 6 M GuHCL to the final volume of 100 μ L. After incubating at 55 °C for 1 h to denature the protein, the samples were allowed to cool before being diluted again by 300 μ L of water. Then, the samples were concentrated by a 10,000 MWCO filter to a final volume of 40 μ L. The dilute-concentrate procedures were repeated one more time to reduce the concentration of GuHCL and increase the concentration of the protein. The 40 μ L samples were then reconstituted in a 0.1 M Tris (pH 8.0) buffer and digested by trypsin for 2 h at 37 °C if the protein was HNSB labeled, or by chymotrypsin digestion for 2 h at 25 °C if the protein was DEPC labeled.

The samples containing β 2m were reconstituted with 0.25 M Tris (pH 8.0), acetonitrile, and 1 M of a freshly prepared DTT stock solution to a final concentration of 0.1 M Tris, 13% (v/v) acetonitrile and 10 mM DTT. After incubating at 55 °C for 1 h, the sample was allowed to cool, and the reduced disulfide bond was alkylated by 14 mM iodoacetamide (prepared in 0.25 M Tris buffer, pH 8.0) in the dark at room temperature for 15 min. The denatured and reduced protein was then digested by chymotrypsin for 2 h at 25 °C.

3.2.6 Liquid Chromatography–Mass Spectrometry (LC/MS)

For intact protein analyses, measurements were performed on a Bruker (Billerica, MA) AmaZon quadrupole ion trap mass spectrometer equipped with an electrospray

ionization source. Typically, the electrospray needle voltage was kept at ~4 kV, and the capillary temperature was set at 220 °C. HPLC separations were performed using a Thermo Scientific (Waltham, MA) Ultimate 3000 HPLC with an OPTI-TRAP MICRO column (1 mm x 12 mm, Optimize Technologies Inc., Oregon City, OR). The proteins were eluted using a gradient of acetonitrile containing 0.1% formic acid that increased from 5 to 99% for 10 min at the flow rate of 50 μ L/min.

For peptide mixture analyses, measurements were performed on a Thermo Scientific Orbitrap Fusion mass spectrometer equipped with a nano-electrospray ionization source. The needle voltage was set at 2.1 kV, and the ion transfer tube temperature was set at 325 °C. The resolution of the Orbitrap was set at 60,000, the MS1 AGC target was set at 4×10^5 ions with a maximum injection time of 50 ms. Collision-induced dissociation (CID) with a normalized collision energy of 35% was used for tandem mass spectrometry (MS/MS). Data-dependent selection for precursor ions with ion abundances above 5,000 was used, and a dynamic exclusion of 30 s was activated after 3 spectra were acquired for any given precursor ion within 5 s. The MS/MS AGC target and maximum injection time were set to 5×10^4 ions and 100 ms, respectively. HPLC separations were conducted by a Thermo Scientific Easy-NanoLC 1000 system with a Thermo Scientific Acclaim PepMap C18 nanocolumn (15 cm x 75 μ m ID, 2 μ m, 100 Å). Peptides were eluted using a gradient of acetonitrile containing 0.1% formic acid that increased from 0 to 50% for 60 min at the flow rate of 0.3 μ L/min.

3.2.7 Determination of Covalent Labeling Modification Percentages

Residue level CL modification percentages (% labeling) were calculated by integrating the peak area of eluting peptides (i.e., modified or unmodified) using extracted ion chromatograms. The % labeling was calculated based on eq 1.

$$\% \text{ labeling} = \frac{\sum_{i=1}^n \sum_{z=1}^m A_{i,z}^{modified}}{\sum_{i=1}^n \sum_{z=1}^m A_{i,z}^{modified} + \sum_{i=1}^n \sum_{z=1}^m A_{i,z}^{unmodified}} \times 100 \quad (1)$$

$A_{i,z}$ represents the peak area of the peptide of interest. Peak areas from all of the measurable peptides (i) that contain the residue of interest and all detectable charge states (z) are included. The determined modification percentages are relative rather than absolute values because the modified and unmodified peptides have different ionization efficiencies and LC elution times.

To determine intact protein CL modification percentages, the calculation is similar but simplified. Averaged ion abundances during protein elution were used instead (eq 2), and the ion abundance of multiple charge states were summed.

$$\% \text{ labeling} = \frac{\sum_z I^{modified}}{\sum_z I^{modified} + \sum_z I^{unmodified}} \times 100 \quad (2)$$

In determining whether two labeling results were different, an unpaired student t-test was applied with a 99% confidence interval.

3.2.8 Calculation of K_d

In this work, we focus on the most frequently encountered case where only one ligand (L) binds to one protein (P), and only one kind of protein-ligand complex (PL) is formed. The dissociation constant (K_d) in such a system can be described by eq 3.

$$K_d = \frac{[P][L]}{[PL]} \quad (3)$$

where [P], [L], and [PL] are the equilibrium concentrations of P, L, and the complex PL, respectively. The total concentration of ligand ($[L]_0$) and protein ($[P]_0$) has the following concentration relationship with related fractions in the system, respectively (eq 4 and 5):

$$[L]_0 = [L] + [PL] \quad (4)$$

$$[P]_0 = [P] + [PL] \quad (5)$$

Because ligand binding decreases the extent of labeling of the residues at the binding interface, we assume there is a relationship between the covalent labeling result (represented by % labeling) and the concentration of ligand-free protein in the system ([P]). This relationship can be normalized to the extent of labeling when the protein is 100% bound to the ligand (i.e. maximum ligand) and when no ligand is present. The resulting relationship is represented by eq 6:

$$\frac{[P]}{[P]_0} = \frac{\Delta\%labeling}{\%labeling_{maximum\ ligand} - \%labeling_{no\ ligand}} \quad (6)$$

By combining equations 3-6, the following relationship that relates the relative change in labeling to K_d in terms that include the total ligand and protein concentrations (eq 7) can be derived.

relative modification fractional change

$$\begin{aligned}
 &= \frac{\Delta\%labeling}{\%labeling_{maximum\ ligand} - \%labeling_{no\ ligand}} \\
 &= (([L]_0 + [P]_0 + K_d) - (([L]_0 + [P]_0 + K_d)^2 - 4 \times [P]_0 \times [[L]_0])^{0.5}) / 2[P]_0
 \end{aligned}$$

(7)

A plot of the relative modification fractional change, based on labeling percentage measurements, as a function of total ligand concentration used during the titration experiments can be used to determine K_d . Fitting of eq 7 was performed using a customized non-linear curve fit in Origin 8 (OriginLab Corporation, Northampton, MA). The results from each experimental replicate were fit individually, then an averaged K_d value from three experiment replicates is reported.

3.2.9 Solvent Accessible Surface Area (SASA) Calculations

Solvent accessible surface area (SASA) was calculated using GETAREA [27]. A protein's structure file in PDB format was used for calculation. Radius of the water probe was set at 1.4 Å with no gradient in calculation applied. The acquired SASA percentage is the ratio of side-chain surface area to "random coil" value per residue. The "random coil" value of a residue X is the average solvent-accessible surface area of X in the tripeptide Gly-X-Gly in an ensemble of 30 random conformations. SASA (%) range from

0 to 100%. Generally, a residue is considered to be solvent exposed if the ratio value exceeds 50% and to be buried if the ratio is less than 20%.

3.3 Results and Discussion

3.3.1 Effect of Labeling Extent on the Apparent K_d Value

We predict that CL/MS based on slow labeling reactions is capable of determining the K_d values of a protein-ligand complex as long as the overall fraction of labeled protein is kept at low levels such that the perturbation to the equilibrium is limited. To illustrate this idea, consider a hypothetical protein-ligand system with a K_d value of 10 μM (K_{d1}) that upon CL of the free protein results in a significantly weaker interaction with the ligand and a new K_d value of 1000 μM (K_{d2}). The concentration relationship between each fraction in the system can be described by equations 8-11, where $[L]$ is the free ligand concentration at equilibrium, $[P]$ is the unlabeled protein concentration, $[P']$ is the labeled protein concentration, $[PL]$ is the concentration of the unlabeled protein-ligand complex, $[P'L]$ is the concentration of the labeled protein-ligand complex, $[L_0]$ is total ligand concentration, and $[P_0]$ is the initial total protein concentration.

$$[L] + [PL] + [P'L] = [L_0] \quad (8)$$

$$[P] + [P'] + [PL] + [P'L] = [P_0] \quad (9)$$

$$K_{d1} = \frac{[P][L]}{[PL]} \quad (10)$$

$$K_{d2} = \frac{[P'][L]}{[P'L]} \quad (11)$$

If one considers the percentage of the labeled protein as described in eq 12, then the apparent K_d will be defined by eq 13.

$$\% \text{ labeling} = \frac{[P'] + [P'L]}{[P_0]} \quad (12)$$

$$\text{App } K_d = \frac{[P + P'] \times [L]}{[P'L + PL]} \quad (13)$$

The relationship between the apparent K_d and the percent labeling can then be plotted for any labeling extent between 0 and 100% (Figure 3.1). Not surprisingly, when there is no labeling, the apparent K_d is equal to the real K_d (i.e. 10 μM in this hypothetical situation), whereas at 100% labeling the apparent K_d is equal to the K_d of the more weakly binding labeled protein (i.e. 1000 μM in this hypothetical situation). More interestingly is the observation that when the percent labeling is kept low (e.g. below 30%), the apparent K_d differs from the real K_d by less than a factor of 2, meaning that as long as the extent of labeling is kept low, reasonably accurate K_d values can be obtained. It should be pointed out that often the variation in measured K_d values for protein-ligand complexes differ by more than a factor of 2 when different measurement methods are compared [28]. Of course, if the percent labeling is much higher (e.g. above 70%), the apparent K_d value is 1 or more orders of magnitude higher than the real K_d value, which is typically unacceptably inaccurate.

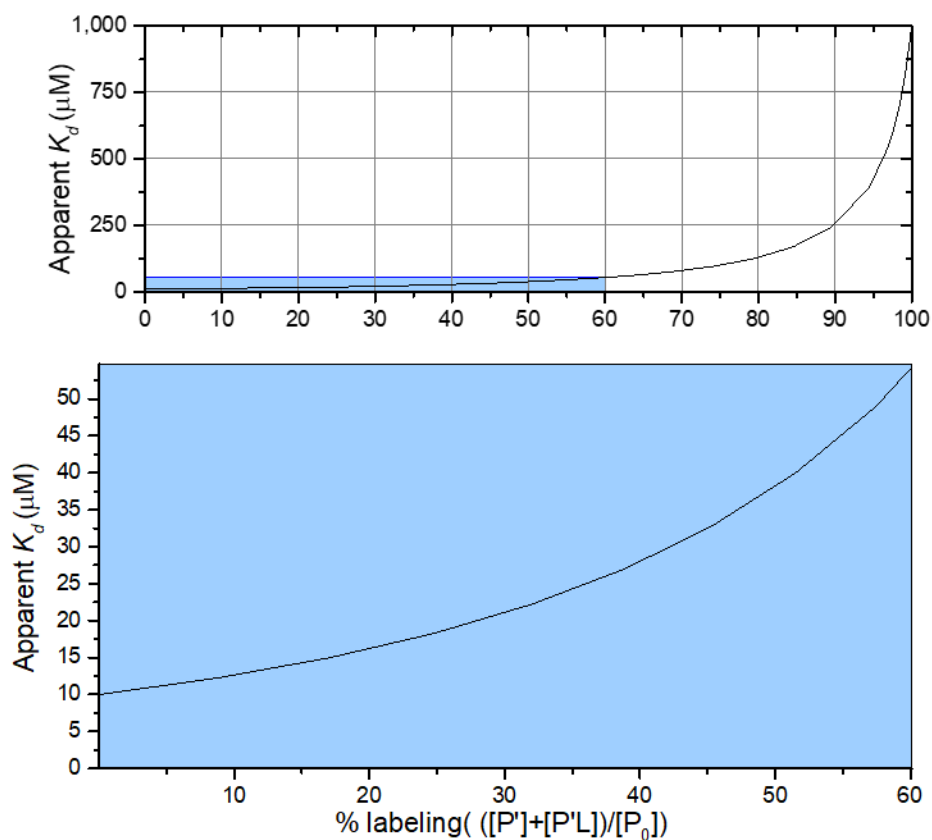


Figure 3.1 Relationship between apparent K_d and real K_d explained by a hypothetical model.

When %labeling is controlled below 50%, apparent K_d is within 5-fold range of actual K_d .

3.3.2 Lysozyme-NAG₃ Binding

To test our hypothesis, the first model system we selected was the NAG₃-lysozyme complex. Because the NAG₃ binding site includes two Trp residues, Trp62 and Trp63 [29, 30], we used HNSB, which is a Trp-specific labeling reagent. Reactions of HNSB with lysozyme with and without NAG₃ result in only Trp62 being labeled as indicated by LC/MS/MS data (Figure 3.2), as the solvent accessible surface area (SASA) of Trp63 is very low. Because only Trp62 is labeled, the labeling ratio of the intact protein could be used during the ligand titration experiments.

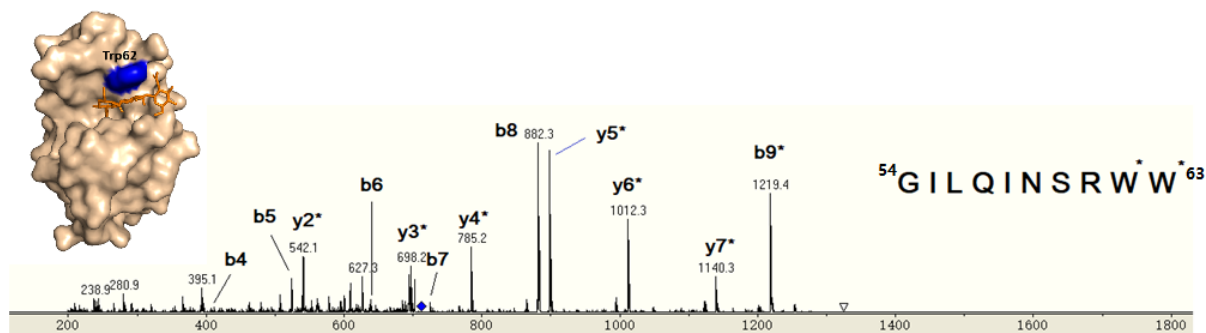


Figure 3.2 Identify Trp62 of lysozyme modified by HNSB.

Lysozyme surface structure with the (NAG)₃ binding site shown from PDB 1HEW. Trp62 is indicated in blue, and (NAG)₃ is in orange. Tandem mass spectrum of the peptide ⁵⁴GILQINSRWW⁶³ from lysozyme. The tandem mass spectrum indicates that W62 is modified by HNSB.

To determine the K_d value using CL/MS, a titration curve was generated in which the NAG₃ concentration was varied between 0 and 600 μ M at a fixed lysozyme concentration of 30 μ M. HNSB was added at a concentration of 1.5 mM and allowed to react with the protein for 10 s. This titration experiment resulted in a decrease in the percent labeling with increased total ligand concentration (Figure 3.3.1). As expected, increasing NAG₃ concentrations lead to more protein molecules in which Trp62 is protected from labeling, leading to an overall decrease in the overall percent labeling. The HNSB labeling data were then plotted and fit using eq 7 (Figure 3.3.2), and a K_d of 14 ± 2 μ M is obtained. Previous measurements of the lysozyme-NAG₃ K_d by methods such as fluorescence, UV, and ESI-MS resulted in values that range from 6 - 60 μ M [28, 31-35]. Clearly, our CL approach is able to provide a value that is consistent with prior studies.

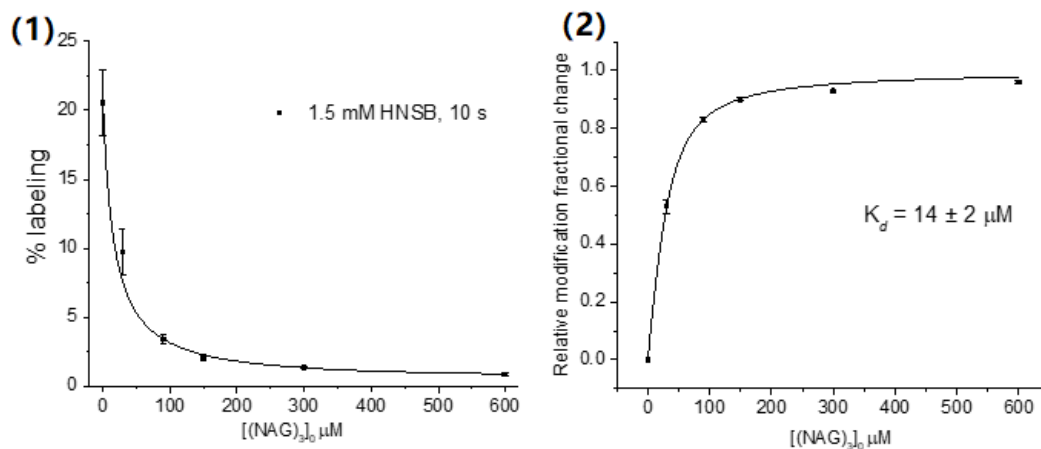
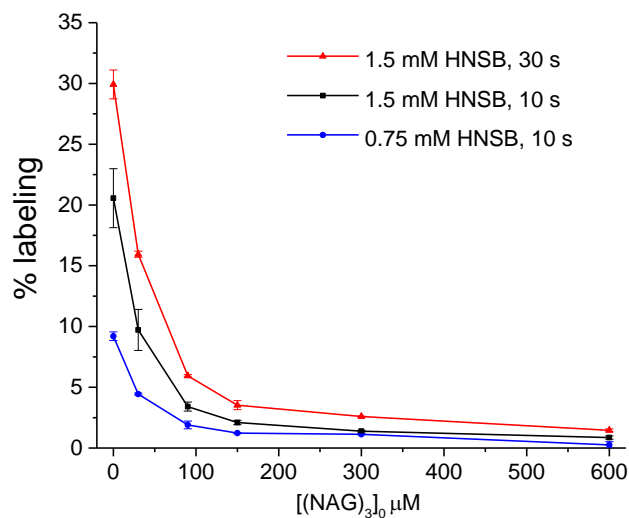


Figure 3.3 HNSB covalent labeling of lysozyme in the presence of (NAG)₃ using a 1.5 mM concentration of HNSB and 10 s reaction time.

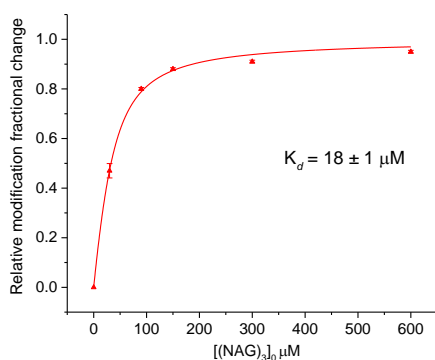
(1) HNSB covalent labeling percentage of lysozyme as a function of (NAG)₃ concentrations. (2) HNSB labeling results fit using equation 7, resulting in a lysozyme-(NAG)₃ K_d of $14 \pm 2 \mu\text{M}$.

To investigate if the CL modification extent significantly perturbs the ligand-binding equilibration (leading to variations in the apparent K_d), additional HNSB labeling conditions were explored in the context of the same titration experiment. Labeling with 1.5 mM HNSB for 30 s and 0.75 mM HNSB for 10 s resulted in more and less extensive labeling, respectively, than the conditions used to generate Figure 3.3.1. These resulting titration curves are shown in Figure 3.4.1, and the corresponding fits to eq 7 are shown in Figure 3.4.2 and 3.4.3. Comparing the K_d results from the three labeling conditions, the results are not significantly different from each other at 95% confidence interval. Moreover, there is no clear trend in the K_d value with the extent of labeling, indicating that the labeling extent does not have a strong influence on the resulting K_d value. The reason for this is likely the fact that the extent of the protein labeling at higher ligand concentrations is very low ($< 10\%$ for almost all ligand concentrations), resulting in very minimal perturbation to the protein-ligand binding equilibrium.



(1)

(2)



(3)

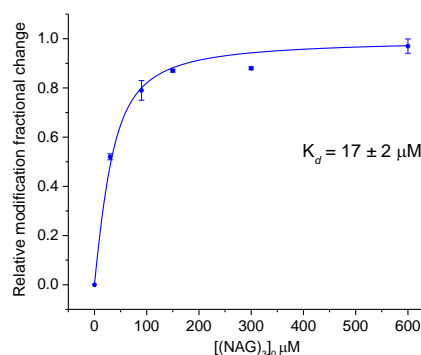


Figure 3.4 HNSB covalent labeling of lysozyme in the presence of (NAG)₃ at different extents of labeling.

(1) HNSB labeling extent for lysozyme at three different labeling conditions as a function of (NAG)₃ concentration. (2) HNSB labeling at 1.5 mM HNSB and a 30 s reaction time fit to equation 7, resulting in a lysozyme-(NAG)₃ K_d of $19 \pm 2 \mu\text{M}$. (3) HNSB labeling at 0.75 mM HNSB and a 10 s reaction time fit to equation 7, resulting in a lysozyme-(NAG)₃ K_d of $17 \pm 4 \mu\text{M}$.

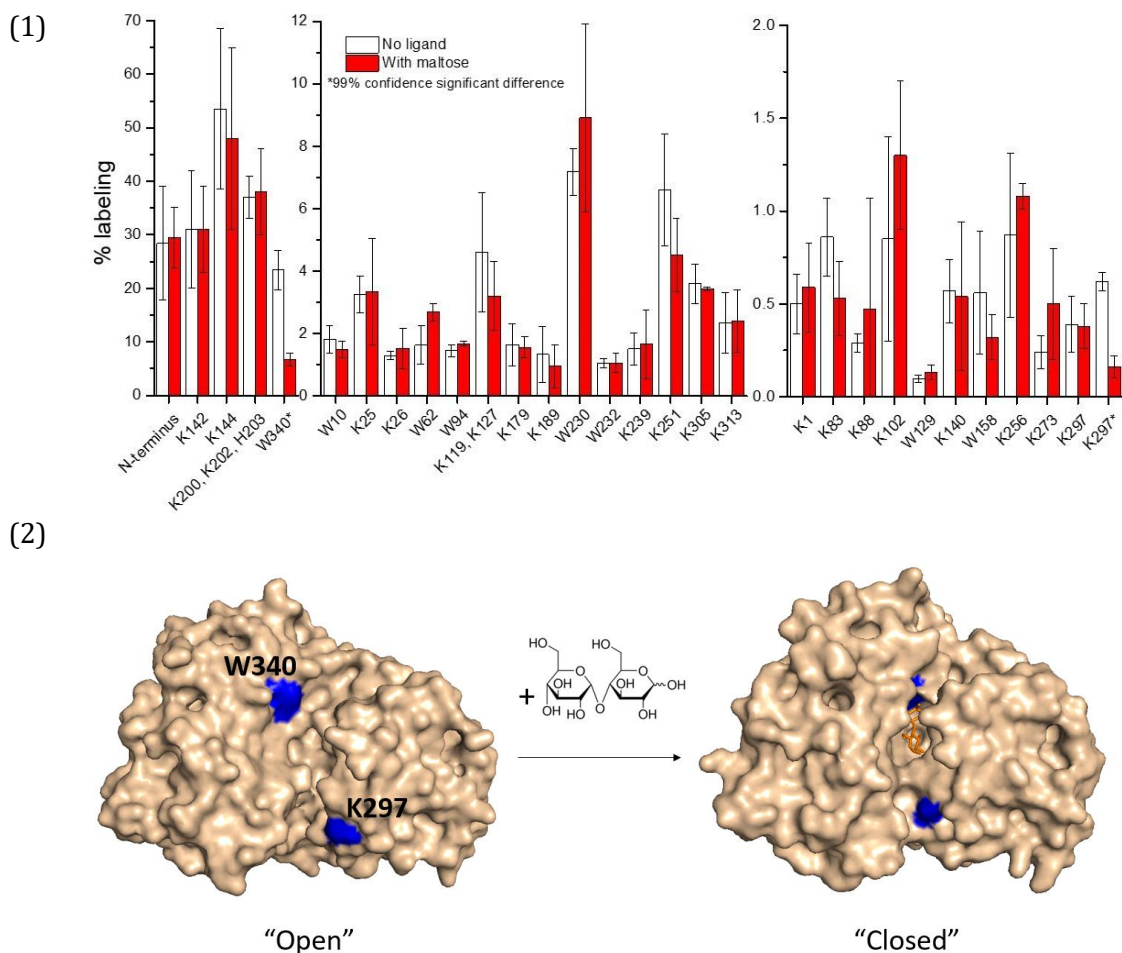


Figure 3.5 CL for determining significantly changed residues of MBP.

(1) Bar graph summarizing the labeling ratio of no ligand and with 285 μM maltose. (2) Surface structure of MBP in the ligand-free “open” state (PDB 1OMP) and the ligand-bound “closed” state (PDB 1ANF). The two residues with significantly decreased labeling ratio, Lys297 and Trp340 in blue.

3.3.3 K_d of Maltose to Maltose Binding Protein

The second model system that we investigated was the maltose binding protein (MBP) bound to maltose. This system allowed us to investigate if accurate K_d values can be acquired by using the extent of labeling at multiple residues. Both HNSB and DEPC were used to label the protein, and upon comparing MBP labeling in the presence and absence of maltose, we find that Trp340 from the HNSB labeling and Lys297 from the

DEPC labeling show statistically significant decreases upon ligand binding (Figure 3.5). These residues are expected to decrease in labeling because based on the crystal structures for the protein and its complex with maltose, their SASA values all decrease upon ligand binding [36]. Trp230's SASA also decreases upon ligand binding and is labeled extensively by HNSB, but the extent of labeling on this residue varied considerably as the maltose concentration was increased. (Table 3.1)

Separate titrations were conducted with HNSB labeling-and DEPC labeling with a fixed MBP concentration of 24 μM , while the concentration of maltose was varied from 0 to 285 μM (Figure 3.6 and 3.7). As was observed for the lysozyme-NAG₃ system, the extents of labeling at Trp340 and Lys297 decrease as the maltose concentration is increased, although because the MBP measurements are from peptide fragments after digestion and LC/MS, the errors are larger than observed for the lysozyme system (Figure 3.6.1 and 3.6.3). From the HNSB labeling data of Trp340, a K_d value of $7 \pm 4 \mu\text{M}$ is obtained (Figure 3.6.2). This value is comparable to a K_d value of 1 to 3 μM that was measured previously by fluorescence [37-40]. Interestingly, other Trp residues (e.g. Trp6, Trp10, Trp94, Trp129, Trp158, and Trp232) that are distant from the binding site or do not change in SASA upon ligand binding (Table 3.1) do not undergo any significant change in labeling extent. As examples, Trp10 and Trp158 are modified by HNSB, but because these residues are distant from the ligand binding site, their extents of labeling do not change as the maltose concentration is increased (Figure 3.7.1 and 3.7.2). Residues like Trp10 and Trp158 serve as useful controls for residues like Trp340 that undergo changes in labeling upon ligand binding.

Table 3.1 SASA value of CL modified residues of MBP.

Residue	No ligand PDB 1OMP (%)	With ligand PDB 1ANF (%)
K1	Unresolved in the crystal structure	Unresolved in the crystal structure
W10	11.9	10.6
K25	54.2	83.4
K26	67	64.7
W62	22.5	22.7
K83	96	94.1
K88	63.1	57.8
W94	3.8	3.6
K102	67.1	66.3
K119	48.3	47.1
K127	64.6	83.8
W129	0.3	1.1
K140	46.2	43.2
K142	62.5	48.5
K144	35.2	30.6
W158	0.1	1.4
K179	62.3	68.9
K189	30.5	39.9
K200	67.2	62.2
K202	68.3	79.9
H203	15.7	16.4
W230	34.3	5.2
W232	10.3	1.2
K239	87.3	82.1
K251	48.6	56.6
K256	26.5	30.8
K273	43.7	44.7
K295	69.6	67.1
K297	38	23.9
K305	61.5	47
K313	87.8	71.7
W340	25	7.1

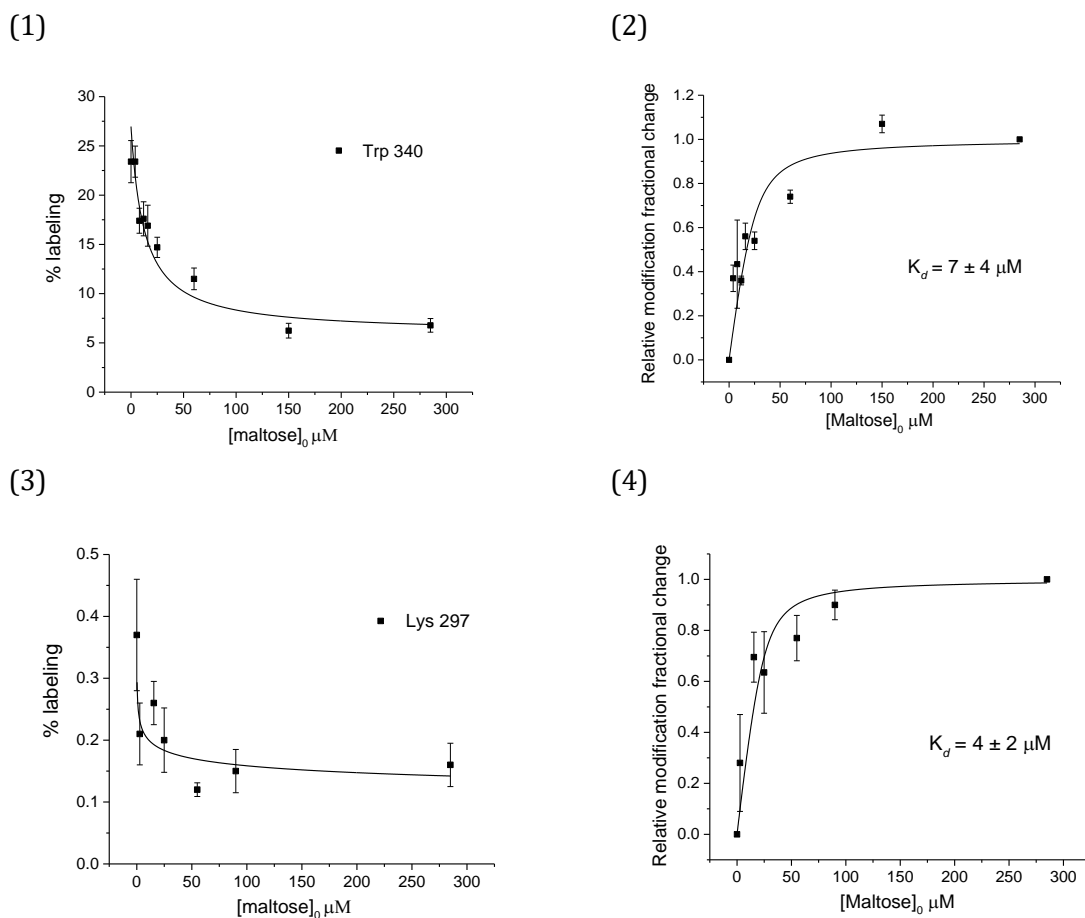


Figure 3.6 CL for determining K_d of MBP and maltose.

(1) HNSB labeling extent of Trp340 in MBP as a function of maltose concentration. (2) HNSB labeling results for Trp340 fit using equation 7, resulting in a MBP-maltose K_d of $7 \pm 4 \mu\text{M}$. (3) DEPC labeling extent of Lys297 in MBP as a function of maltose concentration. (4) DEPC labeling results for Lys297 fit using equation 7, resulting in a MBP-maltose K_d of $4 \pm 4 \mu\text{M}$.

We also explored titrations with DEPC because it is a labeling reagent that can modify up to six different types of amino acids, including Lys, His, Tyr, Ser, Thr, and Cys residues [14]. Even though DEPC can label so many residues, only Lys297 undergoes a significant decrease in labeling extent in the presence of maltose. Indeed, of the 25 residues in MBP that can be labeled by DEPC, Lys297 undergoes the greatest percent change in SASA value (from 38% to 23%) upon maltose binding. Upon plotting

the labeling data for Lys297, a K_d value of $4 \pm 2 \mu\text{M}$ is obtained (Figure 3.6.4). This value is similar to the value obtained from HNSB labeling of Trp340 and is comparable to the literature value for the MBP-maltose complex [37-40]. It is worth noting that the data for Lys297 provides a reasonable measure of the K_d even though it is not a residue directly interacting with maltose. This observation demonstrates that residues in regions that undergo a structural change upon ligand binding can also be applied for K_d determination. Several other residues, such as Lys26, Lys239, and Lys295 (Figure 3.7.3), undergo no change in labeling upon maltose binding, and thus serve as useful controls.

Like the HNSB labeling experiments of the NAG₃-lysozyme complex, the low extent of labeling observed at MBP residues that undergo changes in SASA (i.e. Trp340 and Lys297) upon ligand binding facilitates determination of a K_d value that is close to the literature value. In the case of MBP, the extents labeling of Trp340 and Lys297 are below 25% and 0.5%, respectively. Because labeling by both HNSB and DEPC is limited to about one label on average per protein molecule, which is spread across multiple modifiable residues all over the protein, the extent of modification at any given residue, including binding residues, is low (see Figure 3.5.1). Thus, the number of modified protein molecules that might perturb the protein-ligand equilibrium is expected to be low.

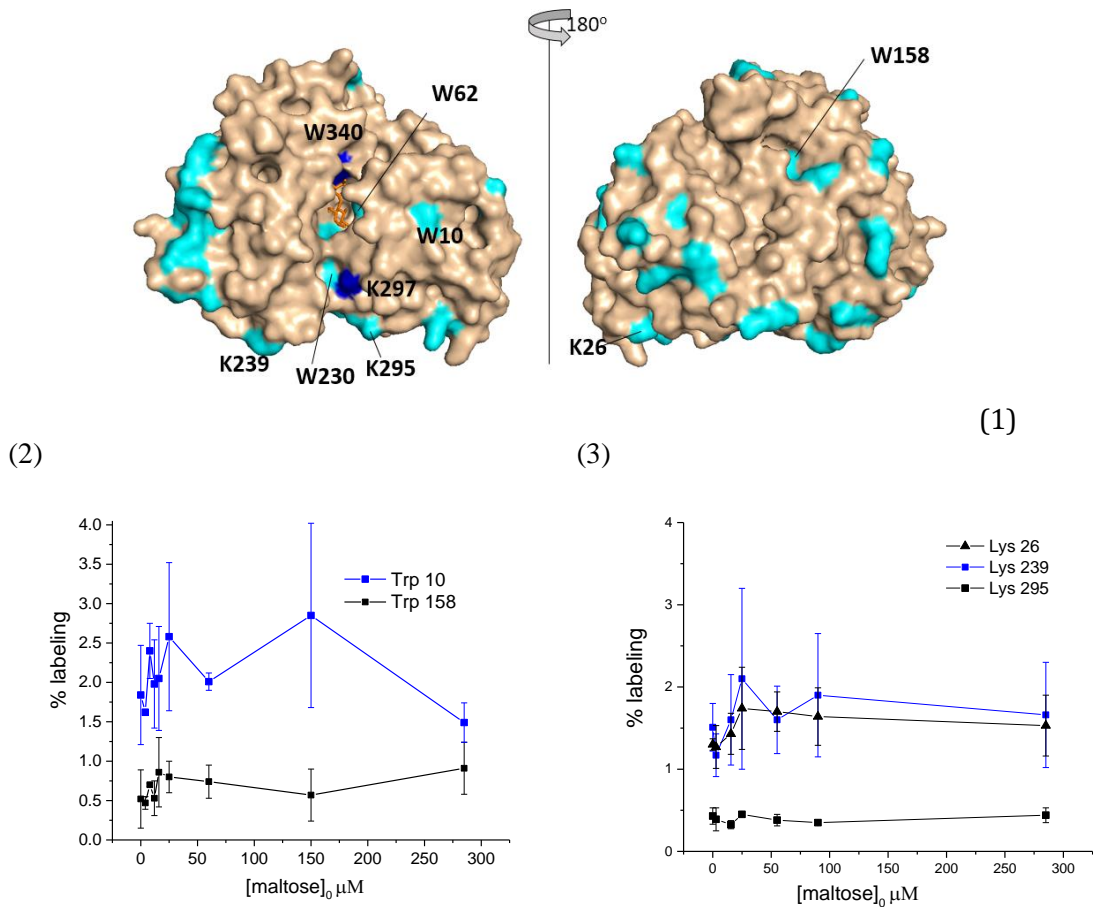


Figure 3.7 Example residues that do not have clear trend in labeling ratio with increasing concentration of maltose.

(1) Showing the location of selected residues on MBP-maltose surface structure (PDB 1ANF), residues with significantly decreased labeling ratio are in blue, no significant change in cyan. Labeling ratio of Trp10 and Trp158 (2); Lys26, Lys239, and Lys295 (3), do not have clear trend in increasing concentration of maltose.

3.3.4 K_d of EGCG to β 2m

We next applied CL/MS to determine the K_d for epigallocatechin gallate (EGCG) bound to β -2-microglobulin (β 2m). β 2m forms amyloid fibrils in patients who undergo long term dialysis due to kidney failure [41], and EGCG (Figure 3.8.1) has recently been found by our group to prevent Cu(II)-induced β 2m amyloid formation *in vitro* by

redirecting β 2m aggregation toward amorphous, re-dissolvable aggregates [42]. Thus, we decided to apply a CL-MS based method to determine the K_d of this complex.

EGCG concentrations ranging from 0 to 115 μ M and a β 2m concentration of 30 μ M were used in a titration experiment along with DEPC labeling. DEPC labeling at high EGCG concentrations indicate that several residues undergo both increased and decreased extents of labeling upon ligand binding, including the N-terminal amine, Thr4, Lys6, His13, His31, and Lys91 (Figure 3.8.2). Most of these residues, particularly the ones that decrease in labeling extent, are near the N-terminus region of the protein (Figure 3.8.3), suggesting EGCG's binding site is between one β sheet that includes residues from Lys6 to Ser11 and another that includes residues from Lys91 to Asp96.

Interestingly, some of the residues that undergo increased labeling, specifically the N-terminal amine and His31, are residues that comprise the Cu(II) binding site in the Cu(II)- β 2m complex [43, 44]. The increased reactivity of these residues with DEPC as the EGCG concentration is increased suggests that EGCG is disrupting the Cu(II) binding site to some extent. Indeed, the K_d of the Cu(II)- β 2m complex increases from 3.3 μ M to 49 μ M in the presence of EGCG (Figure 3.9), which decreases the concentration of the β 2m-Cu(II) complex. The reduced concentration of the metal-bound protein means that when EGCG is present the N-terminus and His31 are free in more protein molecules to react with DEPC.

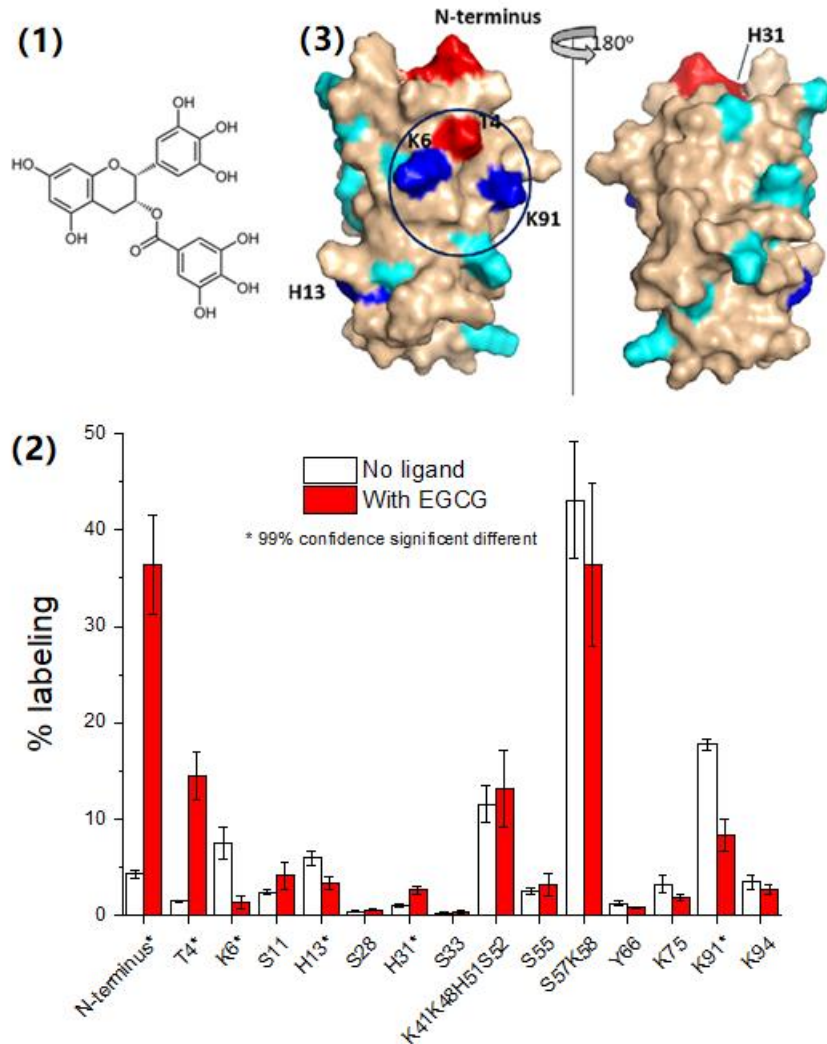


Figure 3.8 EGCG binding to Cu(II)-β2m results in decreases and increases in DEPC labeling at several residues.

(1) Structure of EGCG. (2) DEPC labeling results of Cu(II)-β2m. (3) Residues undergoing covalent labeling with DEPC mapped onto the Cu(II)-free structure of β2m (PDB 1JNJ). Note that there is no structure is available for Cu(II)-β2m. DEPC modified residues that undergo no significant change upon ligand binding are shown in cyan. Those residues that decrease or increase in labeling extent are shown in blue and red, respectively. The proposed EGCG binding site is indicated by the black circle.

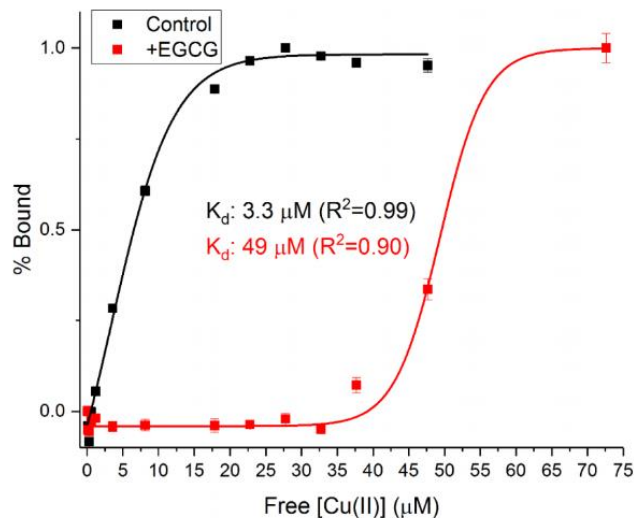


Figure 3.9 Fluorescence spectroscopy determining Cu(II) binding to β 2m. Intrinsic fluorescence of β 2m indicating Cu(II) binding affinity (as K_d) to β 2m monomer. Ligand absent in the black line, with EGCG in the red line.

DEPC labeling as a function of EGCG concentration for residues that both increase and decrease in labeling in the presence of EGCG are shown in Figure 3.10.1. The K_d acquired from the labeling results of the N-terminal amine, Lys6, and Lys91 are $12 \pm 5 \mu\text{M}$, $3.2 \pm 3 \mu\text{M}$, and $11 \pm 5 \mu\text{M}$, respectively (Figure 3.10.2, 3.10.3, and 3.10.4). These values are reasonably similar, and they indicate an EGCG- β 2m K_d value of between 3 and 12 μM . It is interesting to note that although the increase in the labeling of the N-terminal amine is likely due to the release of Cu(II) caused by ligand binding, the resulting value is consistent with the K_d value obtained from the residues that are protected from DEPC labeling. The K_d values acquired by CL-MS are also consistent with the K_d value of 6 μM obtained from size exclusion chromatography data (Figure 3.11) [42].

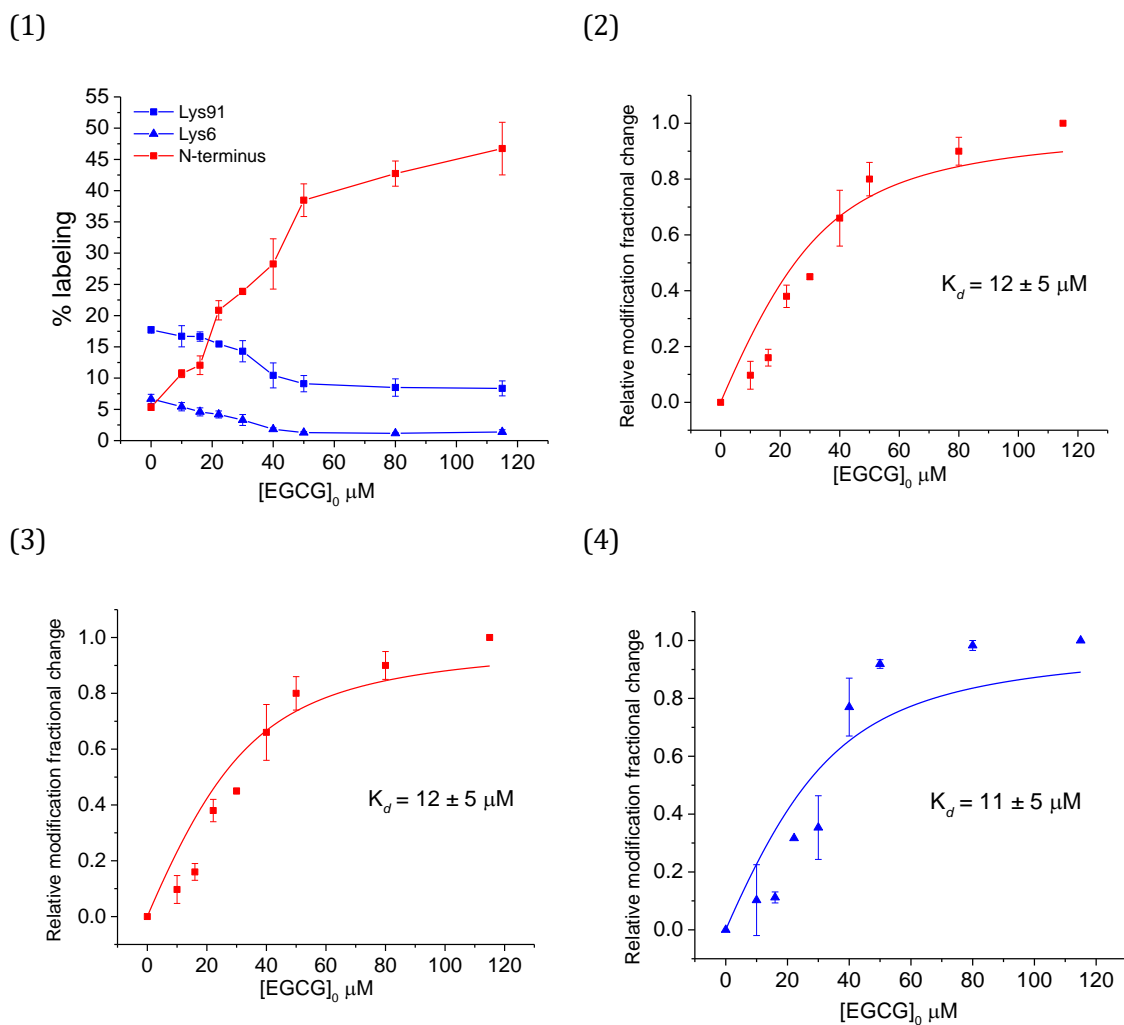


Figure 3.10 Selected residues in Cu(II)-β2m for determining K_d of EGCG to the protein.

(1) DEPC labeling extent of N-terminus, Lys6, and Lys91 in Cu(II)-β2m as a function of EGCG concentration. (2) DEPC labeling results for N-terminus fit using equation 7, resulting in a Cu(II)-β2m and EGCG K_d of $12 \pm 5 \mu\text{M}$. (3) DEPC labeling results for Lys6 fit using equation 7, resulting in a Cu(II)-β2m and EGCG K_d of $3.2 \pm 3 \mu\text{M}$. (4) DEPC labeling results for Lys91 fit using equation 7, resulting in a Cu(II)-β2m and EGCG K_d of $11 \pm 5 \mu\text{M}$.

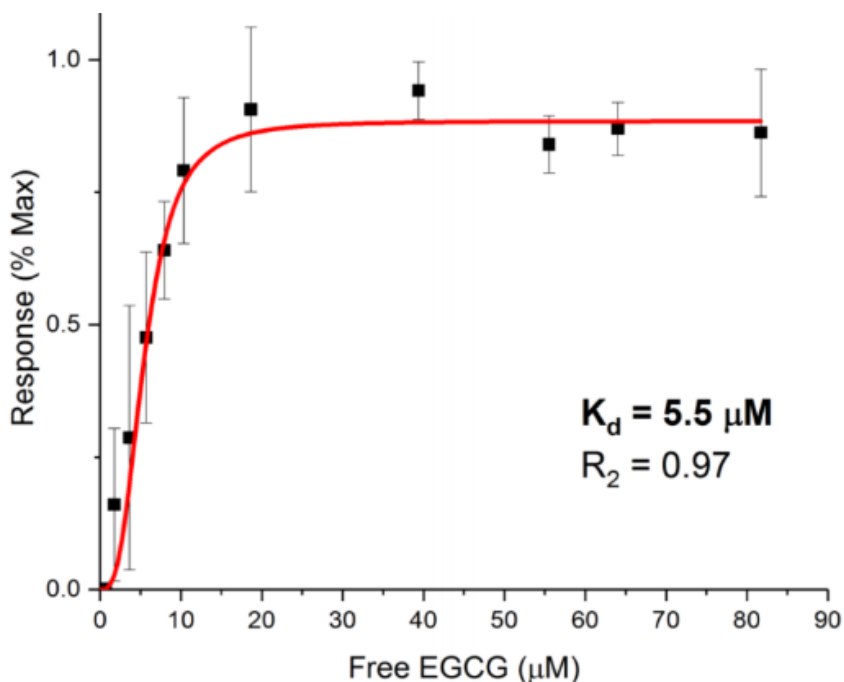


Figure 3.11 Peak heights from SEC chromatography to determine K_d of EGCG to β 2m.

SEC-HPLC data was collected on an Agilent 1100 series HPLC that was outfitted with a SuperSW2000 column purchased from Tosoh Bioscience (Tokyo, Japan). The mobile phase consisted of 150 mM ammonium acetate. The flow rate was 0.35 mL/min. The detector was set to 214 nm. “Response” is the normalized data from relative peak heights, M^*/M . Where M represent β 2m monomer. M^* represent intermediate peak that appears to elute between the dimer and monomer peaks, which can be used to represent EGCG bound Cu(II)- β 2m.

3.4 Conclusion

We have demonstrated that CL-MS can be used to obtain reasonably accurate K_d values for protein-ligand complexes. In particular, we find that reagents like DEPC and HNSB can provide this information even when labeling from these reagents might be expected to perturb the equilibrium of the protein-ligand complex. We find that accurate K_d values can be obtained when the extent of protein labeling is kept low. The experimental results from two model systems and one unknown system support this conclusion and show that our CL-MS based strategy is able to correctly determine the

protein-ligand binding affinity. An intriguing prospect of using CL-MS to determine K_d values is the ability to simultaneously identify the ligand binding site, if the residue level labeling results contain sufficiently detailed information. This CL-MS based ligand titration strategy might serve as an alternative for characterizing protein-ligand complexes that are hard to measure by some other methods, such as fluorescence spectroscopy, surface plasmon resonance spectroscopy, or nuclear magnetic resonance spectroscopy.

3.5 Reference

- [1] Babine, R. E.; Bender, S. L. *Chem. Rev.* **1997**, *97*, 1359-1472.
- [2] Chaires, J. B. *Annu. Rev. Biophys.* **2008**, *37*, 135-151.
- [3] Kitova, E. N.; El-Hawiet, A.; Schnier, P. D.; Klassen, J. S. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 431-441.
- [4] Ganem, B.; Li, Y. T.; Henion, J. D. *J. Am. Chem. Soc.* **1991**, *113*, 7818-7819.
- [5] Katta, V.; Chait, B. T. *J. Am. Chem. Soc.* **1991**, *113*, 8534-8535.
- [6] Loo, R. R. O.; Goodlett, D. R.; Smith, R. D.; Loo, J. A. *J. Am. Chem. Soc.* **1993**, *115*, 4391-4392.
- [7] Wang, W.; Kitova, E. N.; Klassen, J. S. *Anal. Chem.* **2003**, *75*, 4945-4955.
- [8] Daniel, J. M.; Friess, S. D.; Rajagopalan, S.; Wendt, S.; Zenobi, R. *Int. J. Mass Spectrom.* **2002**, *216*, 1-27.
- [9] Ghaemmaghami, S.; Fitzgerald, M. C.; Oas, T. G. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 8296-8301.
- [10] Powell, K. D.; Ghaemmaghami, S.; Wang, M. Z.; Ma, L.; Oas, T. G.; Fitzgerald, M. C. *J. Am. Chem. Soc.* **2002**, *124*, 10256-10257.
- [11] Zhu, M. M.; Rempel, D. L.; Du, Z.; Gross, M. L. *J. Am. Chem. Soc.* **2003**, *125*, 5252-5253.
- [12] Zhu, M. M.; Rempel, D. L.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 388-397.
- [13] Mendoza, V. L.; Vachet, R. W. *Mass Spectrom. Rev.* **2009**, *28*, 785-815.
- [14] Limpikirati, P.; Liu, T.; Vachet, R. W. *Methods.* **2018**, *144*, 79-93.
- [15] Xu, G.; Chance, M. R. *Chem. Rev.* **2007**, *107*, 3514-3543.
- [16] Zhang, B.; Cheng, M.; Rempel, D.; Gross, M. L. *Methods.* **2018**, *144*, 94-103.
- [17] Liu, X. R.; Zhang, M. M.; Rempel, D. L.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 213-217.
- [18] Liu, X. R.; Zhang, M. M.; Rempel, D. L.; Gross, M. L. *Anal. Chem.* **2019**, *91*, 5508-5512.
- [19] Hambly, D.; Gross, M. *Mass Spectrom.* **2007**, *259*, 124-129.
- [20] Guan, J.-Q.; Vorobiev, S.; Almo, S. C.; Chance, M. R. *Biochemistry* **2002**, *41*, 5765-5775.
- [21] Li, Z.; Moniz, H.; Wang, S.; Ramiah, A.; Zhang, F.; Moremen, K. W.; Linhardt, R. J.; Sharp, J. S. *Biol. Chem.* **2015**, *290*, 10729-10740.
- [22] Liu, T.; Marcinko, T. M.; Kiefer, P. A.; Vachet, R. W. *Anal. Chem.* **2017**, *89*, 11583-11591.
- [23] Mendoza, V. L.; Vachet, R. W. *Anal. Chem.* **2008**, *80*, 2895-2904.
- [24] Hambly, D. M.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 2057-2063.
- [25] Chen, J.; Rempel, D. L.; Gau, B. C.; Gross, M. L. *J. Am. Chem. Soc.* **2012**, *134*, 18724-18731.
- [26] Fraczekiewicz, R.; Braun, W. *J. Comp. Chem.* **1998**, *19*, 319-333.
- [27] Zhou, Y.; Vachet, R. W. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 708-717.
- [28] Jecklin, M. C.; Touboul, D.; Bovet, C.; Wortmann, A.; Zenobi, R. *Mass Spectrom.* **2008**, *19*, 332-343
- [29] Cheetham, J. C.; Artymiuk, P. J.; Phillips, D. C. *J. Mol. Biol.* **1992**, *224*, 613-628.

- [30] Ishikawa, T.; Reddy Burri, R.; O. Kamatari, Y.; Sakuraba, S.; Matubayasi, N.; Kitao, A.; Kuwata, K. *Chem. Phys.* **2013**, *15*, 3646-3654.
- [31] Imoto, T.; Johnson, L. N.; North, A. C. T.; Phillips, D. C.; Rupley, J. *Academic Press, New York.* **1972**, 665-868.
- [32] Schindler, M.; Assaf, Y.; Sharon, N.; Chipman, D. M. *Biochemistry.* **1977**, *16*, 423-431.
- [33] Maenaka, K.; Matsushima, M.; Song, H.; Sunada, F.; Watanabe, K.; Kumagai, J. *Mol. Biol.* **1995**, *247*, 281-293.
- [34] Clark, S. M.; Konermann, L. *Anal. Chem.* **2004**, *76*, 7077-7083.
- [35] Denhart, N.; Letzel, Anal. *Bioanal. Chem.* **2006**, *386*, 689-698.
- [36] Quioco, F. A.; Spurlino, J. C.; Rodseth, L. E. *Structure.* **1997**, *5*, 997-1015.
- [37] Walker, I. H.; Hsieh, P.; Riggs, P. D. *Appl. Microbiol. Biotechnol.* **2010**, *88*, 187-197.
- [38] Miller, D. M.; Olson, J. S.; Pflugrath, J. W.; Quioco, F. A. *J. Biol. Chem.* **1983**, *258*, 13665-13672.
- [39] Seo, M. H.; Park, J., Kim, E.; Hohng, S.; Kim, H. S. *Nat. Commun.* **2014**, *5*, 3724.
- [40] Telmer, P. G.; Shilton, B. H.: *J. Biol. Chem.* **2003**, *278*, 34555-34567.
- [41] Floege, J.; Ketteler, M. *Kidney Int.* **2001**, *59*, S164-S171.
- [42] Marcinko, T. M.; Drews, T; Vachet, R.W.: The Effect of Epigallocatechin-3-gallate on Cu(II)-catalyzed β -2 Microglobulin Amyloid Formation. *Manuscript in preparation.*
- [43] Eakin, C. M.; Knight, J. D.; Morgan, C. J.; Gelfand, M. A.; Miranker, A. D. *Biochemistry.* **2002**, *41*, 10646-10656.
- [44] Lim, J.; Vachet, R. W. *Anal. Chem.* **2004**, *76*, 3498-3504.

CHAPTER 4

SYNERGISTIC STRUCTURAL INFORMATION FROM COVALENT LABELING AND HYDROGEN-DEUTERIUM EXCHANGE MASS SPECTROMETRY FOR PROTEIN-LIGAND INTERACTIONS

*The work described in this chapter has been submitted for publish as:
Liu, T., Limpikirati, P., & Vachet, R. W. (2019). Synergistic Structural Information from Covalent Labeling and Hydrogen-Deuterium Exchange Mass Spectrometry for Protein-Ligand Interactions.
Analytical Chemistry, manuscript under review*

L.P. processed the data of covalent labeling using the custom software pipeline; L.T. conducted all the rest experiments.

4.1 Introduction

Protein-ligand interactions are fundamental in all living organisms. Understanding the details of protein-ligand interactions is an important step for understanding biology at the molecular level [1]. Characterizing protein-ligand binding sites as well as ligand binding induced structural changes can facilitate drug discovery, design, and development [2].

Mass spectrometry (MS) based methods have some inherent advantages for studying protein-ligand interactions over other methods that reveal a protein's higher order structure (HOS) such as X-ray crystallography and NMR [3]. These advantages include limited sample consumption, almost no protein size limitations, and the ability to obtain information in mixtures. As a result of these advantages, many MS-based methods have been explored to study protein-ligand interactions to characterize binding stoichiometries, binding constants, and binding sites [4-7].

Hydrogen deuterium exchange (HDX) MS is a well-established method to study protein's HOS that is increasingly used in the pharmaceutical industry [8-10]. HDX relies on changes in mass resulting from backbone amide hydrogen exchange with deuterium at rates that vary based on a number of factors, including protein secondary structure, solvent accessibility, pH, and temperature. For a backbone amide hydrogen in an unstructured region of the protein in a solution at neutral pH, the intrinsic exchange rate is on the order of milliseconds [11]. However, in a folded protein this exchange rate can vary significantly from minutes to days as a result of the sequence, structure of the protein, accessibility of a given region to solvent, and protein backbone dynamics [12]. Upon ligand binding to a protein, changes in accessibility to solvent and to protein structural fluctuations can occur at the ligand binding site and elsewhere in the structure that can protect the protein from exchange [13, 14]. This allows HDX-MS to be a useful technique to characterize the protein-ligand binding sites [15-17]. However, it is also common that ligand binding can affect overall protein stability and dynamics in other structural regions distant from the binding site. This effect can make it difficult to distinguish which decreases in HDX result from local protection due to ligand binding and which are caused by allosteric effects that lead to distant protection against exchange.

Covalent labeling (CL) MS is another method that has been used to characterize protein-ligand complexes [6, 18-22]. In CL, a labeling reagent is used to modify amino acid side chains of the protein by forming a covalent bond [23, 24]. This results in a mass shift that can be detected via MS. CL-MS can be used to study protein-ligand interactions because ligand binding decreases the solvent accessibility of the side chains involved in ligand binding. Decreased labeling at specific residues can be used to determine the

ligand binding site. Covalent labeling techniques can be divided into two categories: amino acid specific labeling techniques (e.g. Lys-specific labeling) [23] and non-specific labeling techniques (e.g. hydroxyl radical footprinting, carbene labeling, and DEPC labeling) [24-26]. Non-specific labeling techniques have a range of intrinsic reaction rates, ranging from ns for carbene labeling [27, 28] to $\mu\text{sec/ms}$ for hydroxyl radical labeling [29] to 10's of seconds for DEPC labeling [30]. The faster reacting CL reagents, such as hydroxyl radicals, are likely sensitive to changes in protein dynamics, whereas reagents that react more slowly, such as DEPC, are presumably transparent to protein dynamics.

Because HDX probes backbone amide hydrogens when used with MS and CL probes the solvent accessibility of side chains, these two methods are typically considered to be complementary. We hypothesize, though, that as a result of the large differences in *intrinsic* reaction rates between HDX and CL reagents like DEPC, the two techniques can provide synergistic information regarding structural changes that take place upon ligand binding. Ligand binding usually induces a decrease in local solvent accessibility at the binding site and often a decrease in protein dynamics at distant sites as proteins are often stabilized by interactions with the ligand. Because HDX responds to changes in both solvent accessibility and structural fluctuations, it sometimes provides ambiguous information with regard to ligand binding site. We predict that the slower labeling timescale for DEPC labeling should make it only sensitive to changes in solvent accessibility and insensitive to changes in protein dynamics. Thus, when used together, HDX and CL have the potential to provide clearer information about protein-ligand binding sites and changes in protein dynamics caused by ligand binding.

Here, we test this hypothesis by comparing HDX-MS and CL-MS on three model protein-ligand complexes. From these experiments we show that when used together the two methods provide both complementary and synergistic information about protein-ligand interactions.

4.2 Experimental and Methods

4.2.1 Materials

Myoglobin from equine skeletal muscle, apomyoglobin from equine skeletal muscle and carbonic anhydrase isozyme II from bovine erythrocytes (BCA) were purchased from MilliporeSigma (St.Louis, MO). Maltose binding protein (MBP) was obtained from MyBioSource.com (San Diego, CA). Brinzolamide, deuterium oxide (99.9 atom %D), diethylpyrocarbonate (DEPC), dimethyl(2-hydroxy-5-nitrobenzyl)sulfonium bromide (HNSB), guanidine hydrochloride (GuHCL), imidazole, maltose monohydrate, MOPS, MOPS sodium salt, and L-tryptophan were also purchased from MilliporeSigma. Acetonitrile, formic acid, sodium phosphate, sodium phosphate monobasic monohydrate, HPLC grade water, and a 1 M Tris buffer (pH 8) stock solution were all purchased from Fisher Scientific (Fair Lawn, NJ). Centricon molecular weight cutoff (MWCO) filters were obtained from Millipore (Burlington, MA). Sequencing grade modified trypsin and sequencing grade chymotrypsin were obtained from Promega (Madison, WI).

4.2.2 Sample Preparation

Proteins were prepared in a 20 mM MOPS buffer at pH 7.5 using the material as received, except for the maltose-binding protein (MBP). MBP comes from the vendor in a 20 mM Tris (pH 8.0) buffer. Before analysis, the buffer was exchanged with a 20 mM

phosphate buffer at pH 7.5 using the 10K MWCO filters by several cycles of concentration and reconstitution. The concentrations before beginning the CL or HDX experiments were 25 or 35 μM for all the proteins. Stock solutions (5 mM) of the brinzolamide were prepared in 1:1 (v/v) acetonitrile and water due to the limited water solubility of this compound. For experiments involving brinzolamide bound to bovine carbonic anhydrase (BCA), the molar ratio between protein and ligand was always 1:1. This ratio was sufficient to lead to more than 99% bound, even after dilution for the HDX experiments, based on the known K_d of 0.1 nM [31]. The final concentration of acetonitrile in the prepared sample was always below 1% (v/v). Maltose stock solutions were prepared in 20 mM phosphate buffer, and a maltose concentration of 480 μM was used for the maltose-MBP binding experiments. Based on the known MBP-maltose K_d of 1.5 μM [32], this molar ratio resulted in greater than 99% protein bound during the CL experiments and greater than 90% bound during the HDX experiments.

4.2.3 HDX Experiments

D_2O was prepared in the buffer appropriate for each protein, as indicated above. The pD was adjusted via a pH meter corrected by the following relationship: $\text{pD} = \text{pH reading} + 0.41$ [33, 34]. HDX experiments were conducted using the Leap HDX Automation Manager as part of the Waters nanoACQUITY UPLC system (Waters Corporation, Milford, MA, USA). The HDX procedure consisted of the following steps: To initiate the HDX, 3.8 μL of the prepared protein or protein-ligand complex sample was diluted into 52.2 μL D_2O buffer and allowed to exchange for different amounts of time at 10 $^\circ\text{C}$. At the end of each exchange period, the reaction was quenched by mixing the sample with a quenching buffer (1:1, v/v) that contained 3.6 M GuHCl and $\sim 0.8\%$

formic acid in water (pH = 2.25) at 1 °C for 1 min. After the quench step, the sample was transferred and injected into the Waters ACQUITY UPLC System. Online digestion was performed using a Waters ENZYMATE immobilized pepsin column (ID: 2.1 length: 30 mm). The proteolytic products were collected by a trap column (HSS T3 pre-column, 100 Å, 1.8 µm, 2.1 mm X 5 mm, Waters) for 4 min. Then, trapped peptides were eluted by a Waters ACQUITY C18 column (2.1 x 50 mm, 1.8 µm) at 0 °C with a linear gradient of acetonitrile (with 0.1% formic acid) that was increased from 5% to 35% over 7 min and then increased from 35% to 85% in 1 min at a flow rate of 40 µL/min. The eluent was then directed into a Waters SYNAPT G2Si mass spectrometer for analysis in MSE mode over the m/z range of 50 - 2000. The relative deuterium uptake level of each measured peptide at different exchange time points was automatically calculated using the DynamX 3.0 software (Waters). Averaged values from triplicate experiments with propagated error are reported.

Peptides were identified as having statistically significant differences in deuterium uptake between ligand-bound and ligand-free states if at least two out of three exchange time points (i.e. 10 s, 10 min, and 4 (or 24) h) showed significantly different deuterium uptake levels. The 10 s, 10 min, and 4 h (or 24 h) time points were chosen to represent short, medium, and long exchange times, respectively. Statistical differences in uptake were determined using the following two-step statistical cutoff: 1) a difference in deuterium uptake between the ligand bound and ligand free states at a given time point was larger than the global deuterium uptake significance limit as described by Hageman and Weis [35], the value of which was decided with a 99% confidence interval; and 2)

the relative deuterium uptake of the two states at a given time point were significantly different from one another according to a Welch's t-test at a 99% confidence interval.

4.2.4 CL-MS Experiments

In all CL experiments, conditions were chosen to control labeling to 1 to 1.5 labels on average per protein molecule to maintain the structural integrity of the protein during the modification reactions [23, 24]. DEPC labeling was performed following procedures and conditions previously developed by our group [30, 36]. DEPC stock solutions were prepared in acetonitrile. The reaction was initiated by adding an aliquot of the DEPC stock solution into the prepared sample. The final concentration of DEPC was 120 μ M for myoglobin, 360 μ M for BCA, 300 μ M for MBP. The final volume of acetonitrile was less than 1% (v/v) in all experiments. The reaction proceeded at 37 °C for 1 min and then was quenched by the addition of imidazole at a final concentration of 10 mM. The HNSB labeling reactions were initiated by adding an aliquot of an HNSB solution at 2.7 mM to the prepared sample. The reactions were allowed to proceed at room temperature for 3 min before being stopped by the addition of tryptophan at 5 mM. After the protein labeling reactions, the samples were diluted with water to a final volume of 400 μ L and concentrated using a 10,000 MWCO filter to a final volume of 40 μ L. Then, the samples were reconstituted in 0.25 M Tris (pH 8.0) and 6 M GuHCl at 55 °C for 1 h to denature the protein. The samples were then diluted in 300 μ L water and concentrated again by a 10,000 MWCO filter to volume of 40 μ L. This dilution-concentration step was repeated twice to decrease the GuHCl concentration. The resulting sample was diluted to 100 μ L with 0.1 M Tris (pH 8.0) buffer and digested with trypsin

for 2 h at 37 °C if the protein was HNSB labeled, or with chymotrypsin for 2 h at 25 °C if the protein was DEPC labeled.

The digested proteins were analyzed by LC/MS on a Thermo Scientific (Waltham, MA) Orbitrap Fusion mass spectrometer with a nano-electrospray ionization source. The electrospray voltage was settled at 2.1 KV and the ion transfer tube temperature was set at 325 °C. The resolution of Orbitrap was set at 60,000, the MS¹ AGC target was set at 4×10^5 ions with a maximum injection time of 50 msec. Collision-induced dissociation (CID) with a normalized collision energy of 35% was used for tandem mass spectrometry (MS/MS) experiments. Data-dependent selection of the precursor ions with ion abundances above 5,000 was applied. On-line HPLC separation of the digested protein samples was conducted using a Thermo Scientific Easy-NanoLC 1000 system with a Thermo Scientific Acclaim PepMap C18 nanocolumn (15 cm x 75 μm ID, 2 μm, 100 Å). Peptides were eluted using a gradient of acetonitrile containing 0.1% formic acid that increased from 0 to 50% for 60 min at the flow rate of 0.3 μL/min.

For peptide identification and determination of CL extents, a custom software pipeline specifically designed for protein CL-MS studies was used [37, 38]. Residue level CL modification percentages (% labeling) were determined from the chromatographic peak areas of modified and unmodified peptides and by applying eq 1.

$$\% \text{ labeling} = \frac{\sum_{i=1}^n \sum_{z=1}^m A_{i,z}^{modified}}{\sum_{i=1}^n \sum_{z=1}^m A_{i,z}^{modified} + \sum_{i=1}^n \sum_{z=1}^m A_{i,z}^{unmodified}} \times 100 \quad (1)$$

In eq 1, $A_{i,z}$ represents the peptide peak area from any given peptide (i) that contains the residue of interest, and considers all detectable charge states (z) for that peptide. The resulting modification percentage is a relative rather than an absolute value because the modified and unmodified peptides have different ionization efficiencies and

elute at different retention times. Unpaired student t-tests with a 99% confidence interval were used to determine if a given residue underwent a change in labeling when the ligand was present as compared to when the ligand was absent.

4.3 Results and Discussion

4.3.1 Myoglobin

Heme bound and apo myoglobin were selected to test our hypothesis that DEPC labeling is insensitive to protein dynamics and thus can provide synergistic information when combined with HDX-MS. Apo myoglobin is known to be more dynamic than heme-bound myoglobin [39, 40], making myoglobin an excellent model system. Comparative HDX-MS measurements of myoglobin and apo-myoglobin result in a large number of peptides that undergo a statistically significant change in deuterium uptake (Figure 4.1, 4.2, and Table 4.1). These peptides come from four regions of the protein that include residues 12-29, 30-69, 70-106, and 137-153. We only observe decreases in deuterium uptake upon heme binding and do not observe any peptides that significantly increase in deuterium uptake upon ligand binding. Results from HDX/MS indicate that heme binding significantly stabilizes the protein, especially in the region spanned by residues 30-106, which is consistent with previous studies by NMR [39, 41] and HDX-MS [40, 42]. Interestingly, exchange decreases are apparent throughout the protein upon ligand binding and are not localized to just the ligand binding site. Peptides in the heme-binding pocket, including parts of residues 30-69 and 70-106, as well as other regions not directly involved in the heme-binding site (e.g. 137-153) or even remote from the binding pocket (e.g. 12-29) also have significantly decreased deuterium uptake. With the HDX-

MS data alone, it would be difficult to identify the heme binding site due to the widespread stabilization of the protein.

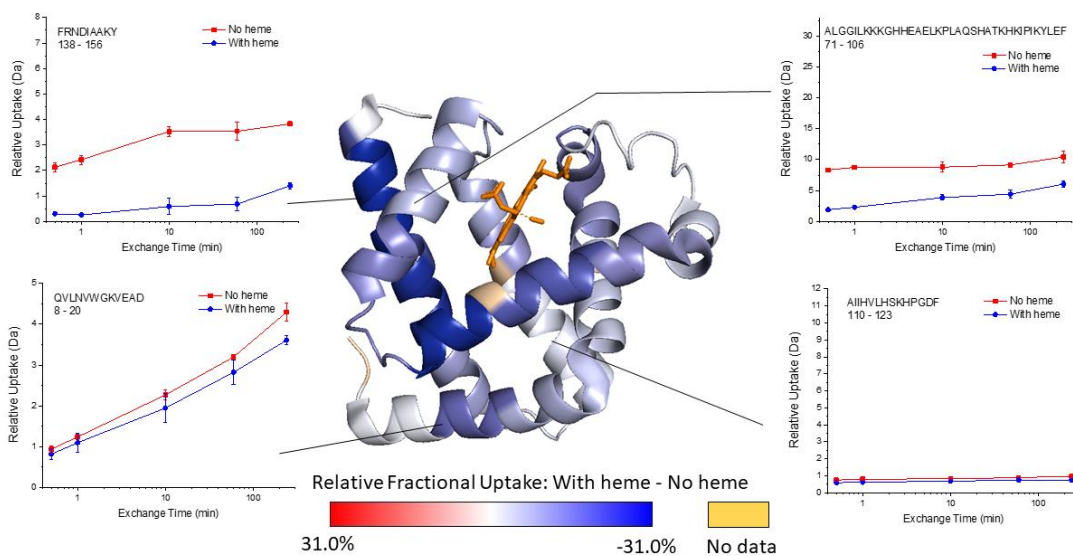
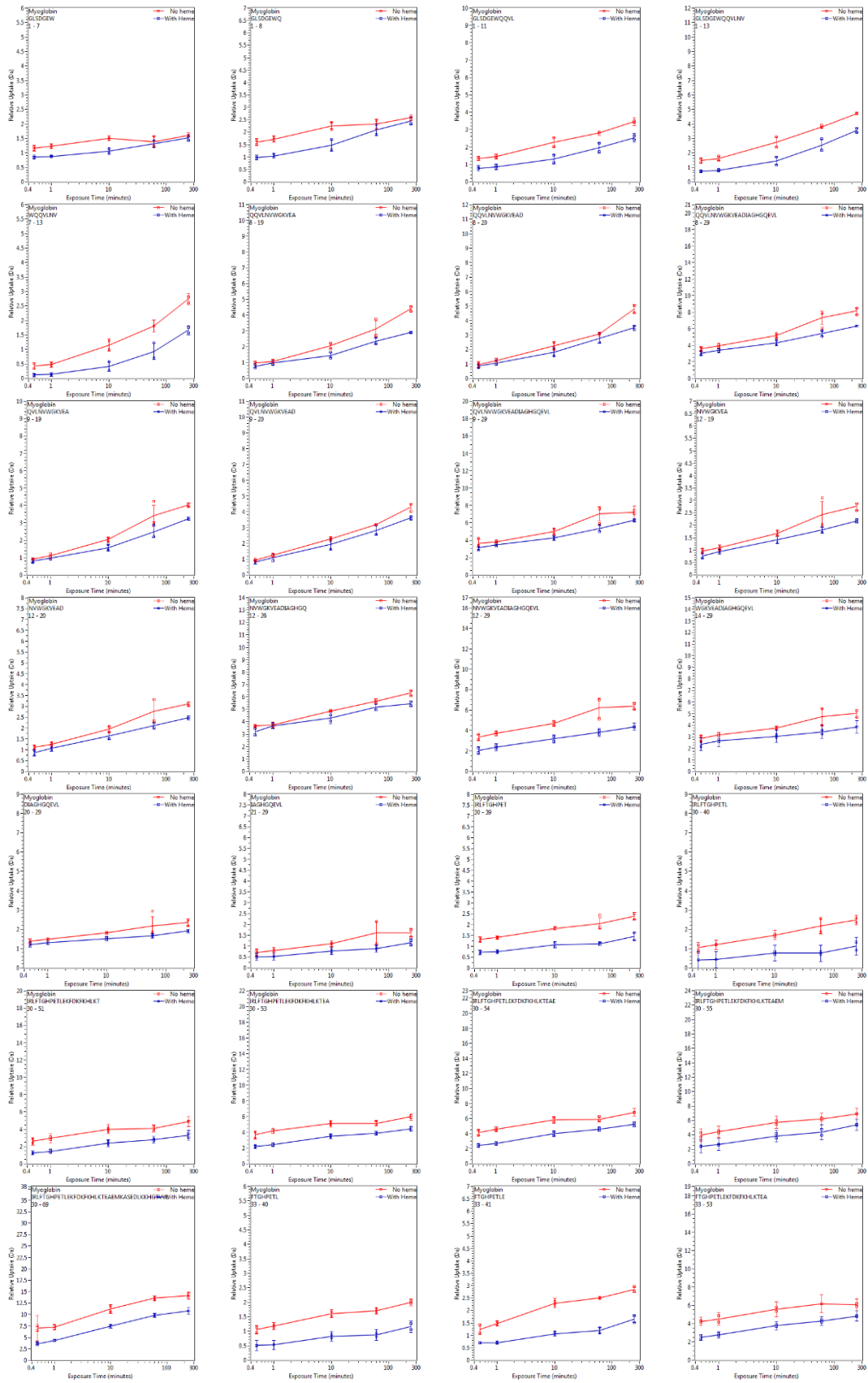
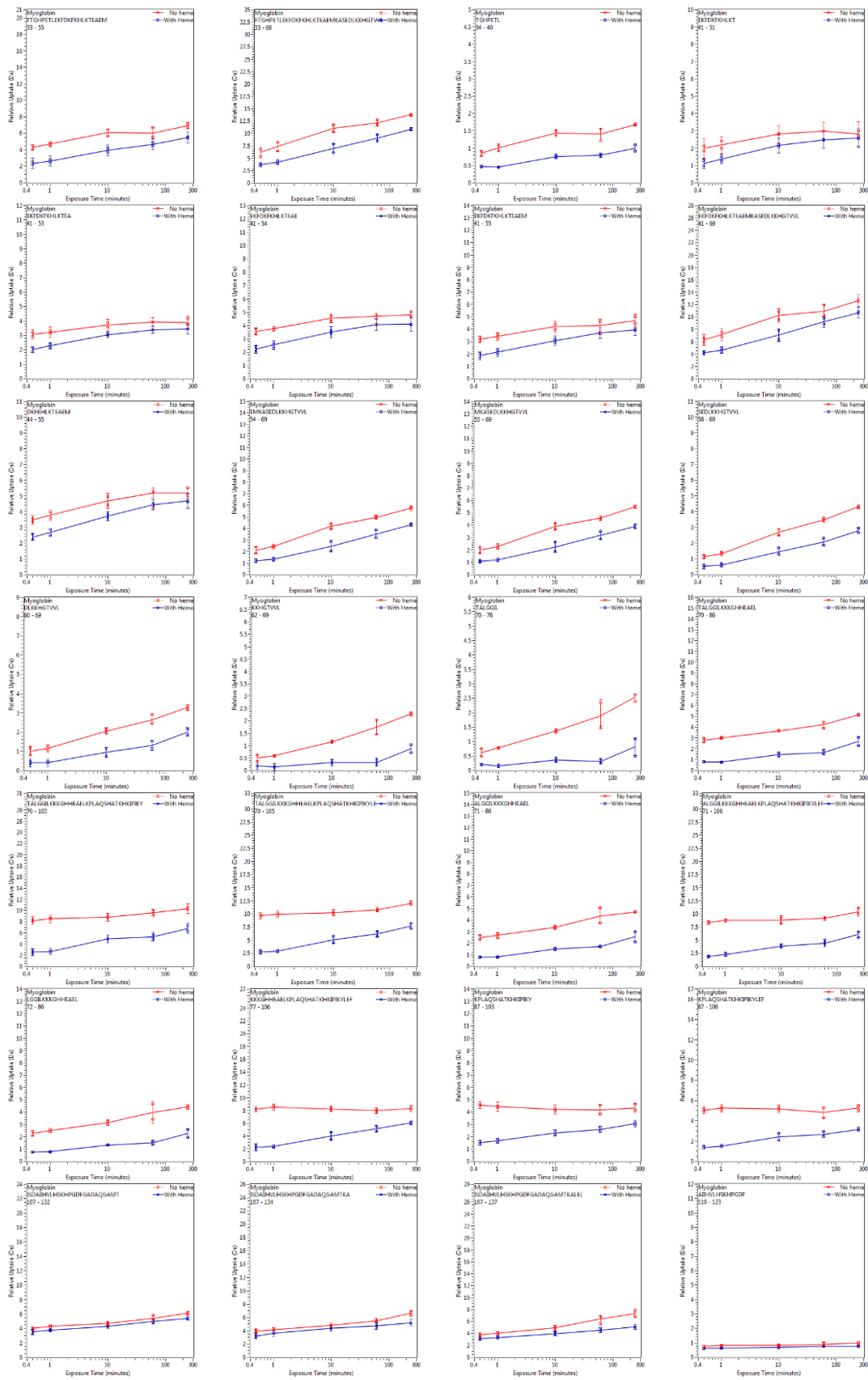


Figure 4.1 Relative fractional deuterium uptake for myoglobin shown on the holo myoglobin structure.

Representative deuterium uptake plots for myoglobin (i.e. with heme) and apo-myoglobin (i.e. without heme), and the relative fractional uptake (at the 4 h exchange time point) mapped onto the holo myoglobin structure (PDB 1DWR). The darker the blue color on the structure, the more the HDX decreases upon heme binding.





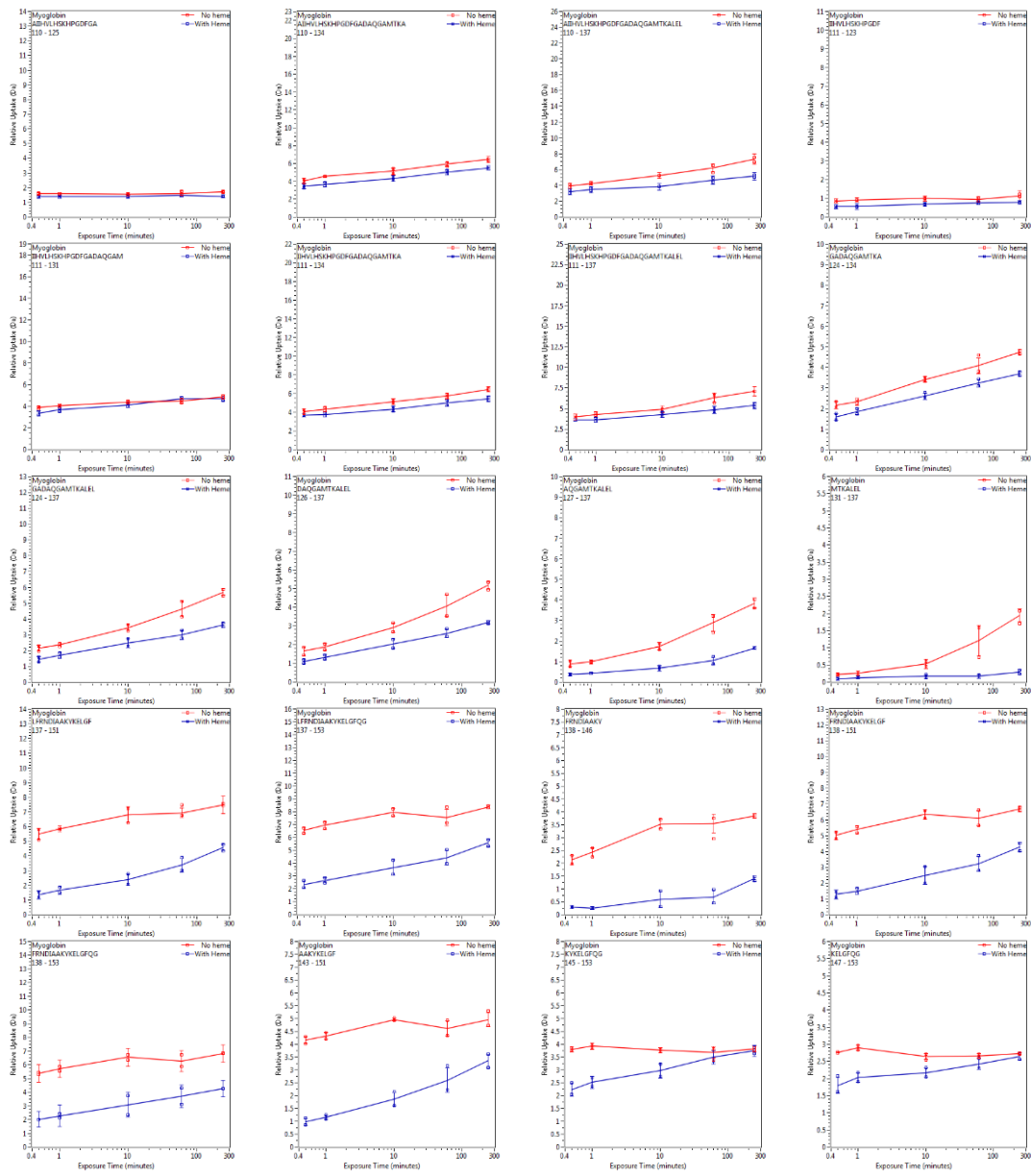


Figure 4.2 Deuterium uptake plots for myoglobin.
 Myoglobin with heme bound in blue and without heme bound in red.

Table 4.1 Peptides from myoglobin with statistically significant differences in the relative deuterium uptake.

Statistically significant differences as determined using the criteria described in experimental section. The global deuterium uptake difference significance threshold for myoglobin is 1.43 Da.

Region	Sequence position	Sequence
12 – 29	12 – 29	NVWGKVEADIAGHGQEVL
30 – 69	30 – 53	IRLFTGHPETLEKFDKFKHLKTEA
	30 – 54	IRLFTGHPETLEKFDKFKHLKTEAE
	33 – 69	FTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVVL
	54 – 69	EMKASEDLKKHGTVVL
	55 – 69	MKASEDLKKHGTVVL
70 – 106	70 – 86	TALGGILKKKGHHEAEL
	70 – 103	TALGGILKKKGHHEAELKPLAQSHATKHKIPIKY
	70 – 105	TALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLE
	71 – 86	ALGGILKKKGHHEAEL
	71 – 106	ALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEF
	72 – 86	LGGILKKKGHHEAEL
	77 – 106	KKKGHHEAELKPLAQSHATKHKIPIKYLEF
	87 – 103	KPLAQSHATKHKIPIKY
	87 – 106	KPLAQSHATKHKIPIKYLEF

Changes in CL-MS with and without the heme are not as widespread throughout the protein. Several residues have significant decreases in labeling upon heme binding, including T34, K45, K50, T51, K62, H64, S92, H93, T95, K96, H97, K98, and K102 (Figure 4.3), and these residues are primarily localized around the heme-binding site (Figure 4.4). A few of these residues (K50, T51, S58, and K62) are about 12 Å from the heme in the bound form. While these residues are not immediately next to the heme, the fact that these residues lie in an unstructured loop or in a helix that interacts with the heme causes us to surmise that heme binding induces a rearrangement of these side chains to change their solvent accessibility.

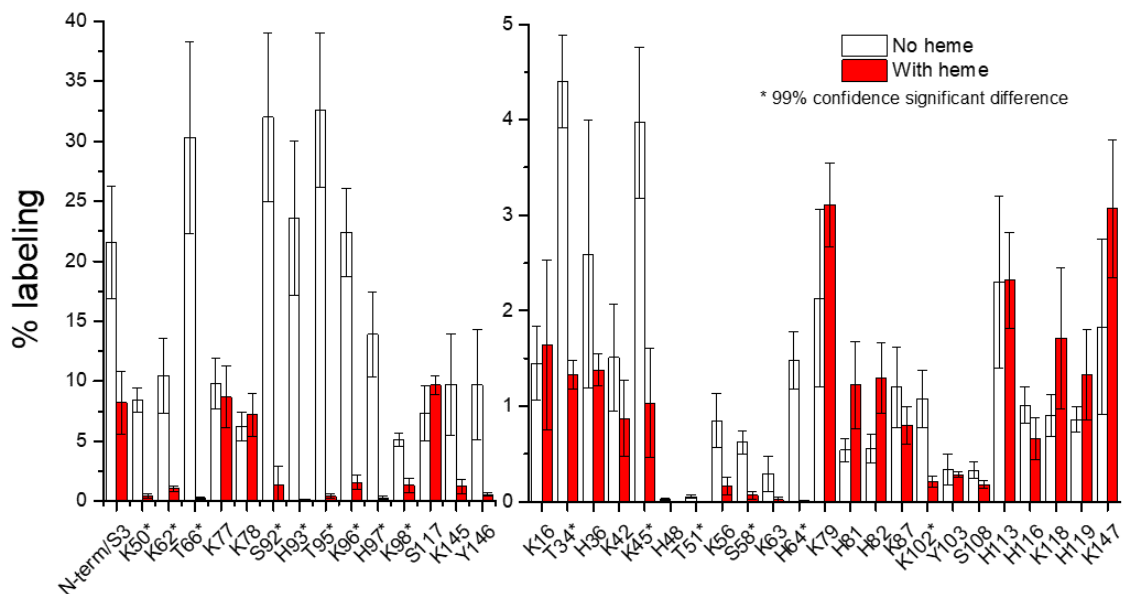


Figure 4.3 DEPC labeling percentages for myoglobin.

Residues that undergo significant decreases in DEPC labeling at a 99% confidence interval are marked with an asterisk “*”. The labeling ratio of Lys145 is $10 \pm 4\%$ and $1 \pm 1\%$ without heme and with heme, respectively, and the labeling ratio of Tyr146 is $10 \pm 5\%$ and $0.6 \pm 0.2\%$ without heme and with heme, respectively. While these two residues show a labeling decrease, they are not significant at the 99% confidence interval that was chosen for the rest of the data.

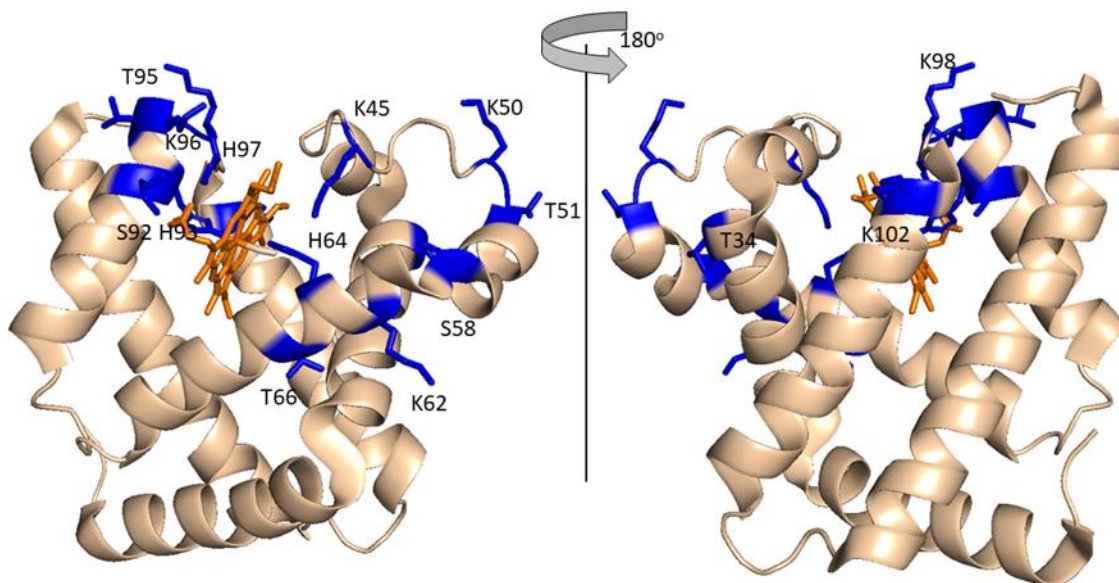


Figure 4.4 Residues of myoglobin identified to undergo significant decreases in DEPC labeling upon heme binding.

Mapping on the holo-myoglobin structure (adapted from PDB 1DWR) the residues identified to undergo significant decreases in DEPC labeling upon heme binding (blue).

More importantly, there are several protein regions that undergo significant decreases in HDX upon heme binding (e.g. 12-29, 71-86, 143-151) (Figure 4.5), but show no significant changes in CL. These regions of the protein are stabilized upon heme binding, but because they do not directly interact with the heme, they presumably do not undergo significant changes in solvent accessibility. Residues 12-29 contain a DEPC-modified residue, K16; however, this residue does not experience a significant change in labeling extent (Figure 4.3). Residues K76, K77, K78, H80, and H81, which all fall within region 71-86, are modified by DEPC, but none of them show significant decreases in CL (Figure 4.3). While residues 71-86 are part of the larger region of the protein that is significantly stabilized by heme binding [39, 40], they do not interact with the heme so their solvent exposure likely remains relatively unchanged. A similar result is obtained for residues between 143 and 151. Our results and previous HDX-MS results indicate protection in this region upon heme binding [40], but the CL changes that K145, Y146 and K147 undergo are not significant at a 99% confidence interval (Figure 4.3).

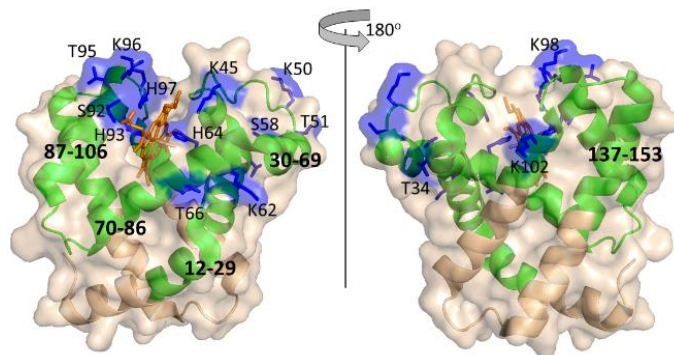


Figure 4.5 A comparison of the differential CL and HDX results for myoglobin mapped on its crystal structure.

Regions that undergo significant decreases in HDX upon heme binding are shown in green. Residues that undergo significant decreases in DEPC labeling are shown in blue. Myoglobin crystal structure: PDB 1DWR.

Overall, the myoglobin data support our hypothesis that CL by DEPC is relatively insensitive to changes in dynamics as a result of its relatively slow intrinsic reaction rate. Residues in protein regions that undergo changes in their transient folding/unfolding (i.e. dynamics) do not undergo changes in DEPC reactivity because the reagent reacts too slowly to respond to these transient changes in solvent accessibility, whereas HDX is sensitive to these changes because the H to D exchange reaction is orders of magnitude faster. Heme binding to myoglobin stabilizes the protein, causing several regions (e.g. 12-29, 71-86, and 143-151) to sample solvent exposed states less frequently (i.e. become less dynamic) and thus undergo decreased HDX. These same regions either undergo no significant change in solvent accessibility or the lifetimes of these more solvent exposed states are too short to allow labeling by DEPC. Because CL by DEPC is primarily affected by changes in solvent exposure caused by ligand (i.e. heme) binding, DEPC-based CL and HDX-MS can provide synergistic structural information. HDX-MS indicates regions that are both protected by heme binding and undergo decreased structural fluctuations, while CL by DEPC is predominantly sensitive to decreases in solvent accessibility. When used together, the techniques provide a more definitive picture of the binding site and binding-induced stabilization.

4.3.2 Bovine Carbonic Anhydrase II

The second model system we selected was bovine carbonic anhydrase II (BCA) and its inhibitor brinzolamide. Extensive studies of human carbonic anhydrase II (HCA) indicate that brinzolamide does not cause a significant structural or dynamic change to the protein [43, 44]. BCA and HCA have a high degree of sequence and structural homology (Figure 4.6), thus brinzolamide binding has a similarly small effect on BCA

[43]. This model system provides another test of the complementarity and potential synergy of DEPC-based CL and HDX-MS measurements.

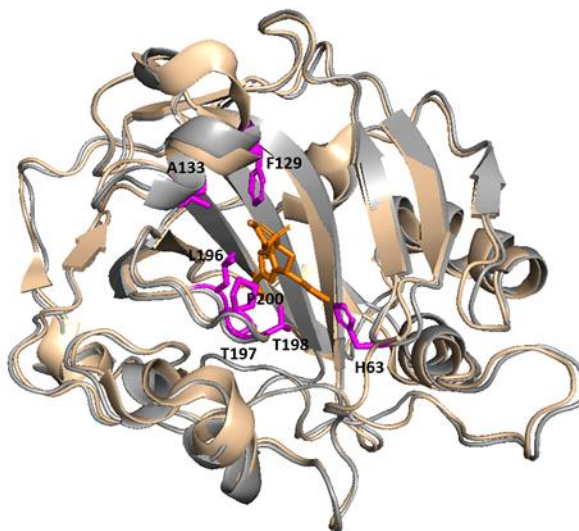


Figure 4.6 Overlay of BCA and HCA-brinzolamide structures.

Overlay of BCA (PDB 1V9E, light brown) and HCA-brinzolamide (PDB 1A42, grey) structures with residues interacting with brinzolamide (orange) shown in magenta.

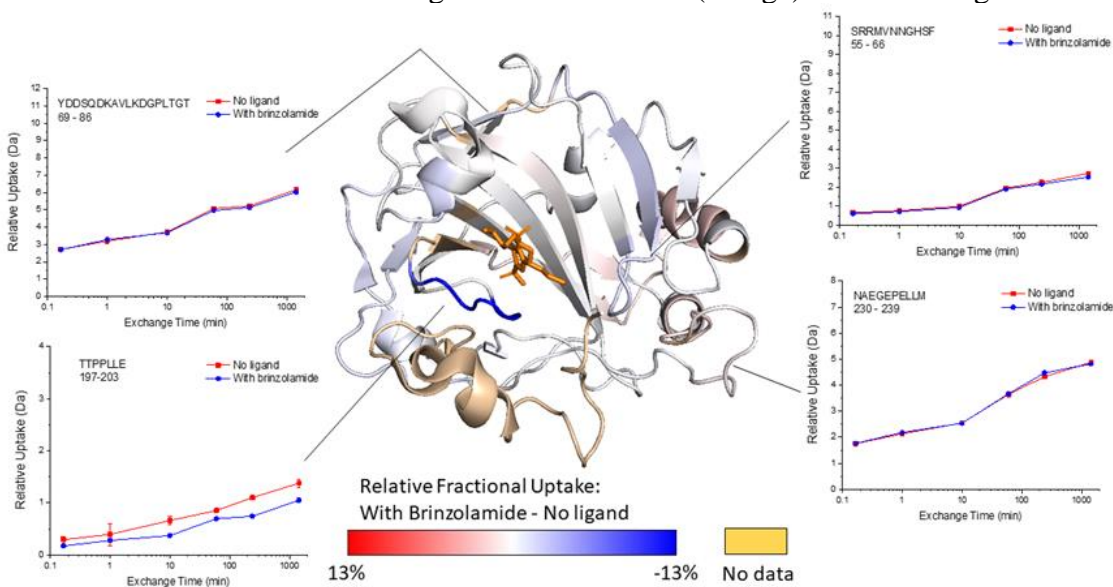


Figure 4.7 Relative fractional deuterium uptake for BCA shown on its structure.

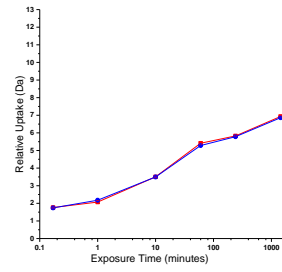
Representative deuterium uptake plots for BCA and the BCA-brinzolamide complex, and the relative fractional uptake (at the 24 h exchange time point) mapped onto the BCA crystal structure (PDB 1V9E), with the position of brinzolamide adapted from HCA-brinzolamide complex (PDB 1A42). The darker the blue color on the structure, the more the HDX decreases upon ligand binding.

The vast majority of BCA undergoes no significant change in HDX for up to 24 h when brinzolamide is bound (Figure 4.7 and 4.8), which is consistent with the known effects of ligand binding. However, there is one peptide, 197-203, that undergoes a measurable, albeit subtle change in exchange. This peptide contains Thr197, Thr198, and Pro202, which are residues known to interact with brinzolamide in the binding pocket of the protein [45]. Presumably, the decreased HDX of this peptide is due to decreased solvent exposure upon ligand binding.

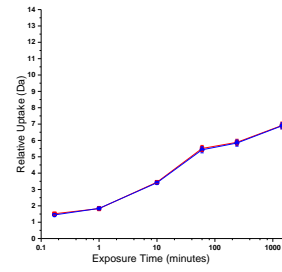
When BCA is labeled with DEPC, 19 residues are modified, but only His63, Ser64 and Lys259 undergo a significant decrease in labeling in the presence of brinzolamide, while Ser28 significantly increases in labeling upon ligand binding (Figure 4.9). Unfortunately, no labeled residues were measured between residues 197 and 203. His63 and Ser64 are on the edge of the brinzolamide binding pocket, so it is not surprising that their labeling extent decreases in the presence of the ligand (Figure 4.10.1). The decreased labeling by Lys259, which is the C-terminal residue, is more difficult to explain. Upon brinzolamide binding, this residue reorients itself but does not undergo a significant change in SASA. The increased labeling of Ser28 was unexpected because its solvent accessible surface area does not change upon ligand binding. One possible explanation is that its local microenvironment changes upon brinzolamide binding. Ser28 is about 5 Å from Thr198, which interacts with brinzolamide (Figure 4.10.2). Perhaps the presence of brinzolamide decreases the pK_a of the Ser side chain, making it more reactive with DEPC.

■ No ligand
● with ligand

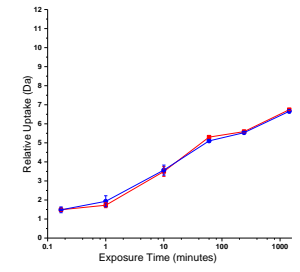
32-47
IDTKAVVQDPALKPLA



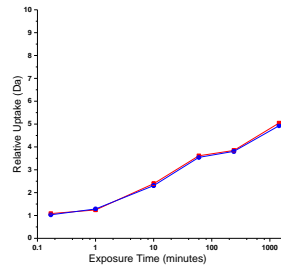
32-48
IDTKAVVQDPALKPLAL



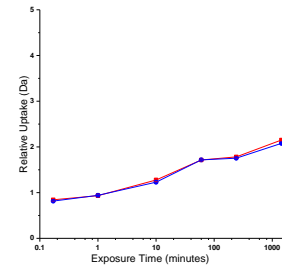
34-48
TKAVVQDPALKPLAL



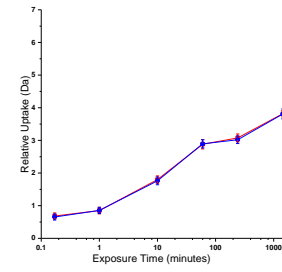
35-47
KAVVQDPALKPLA



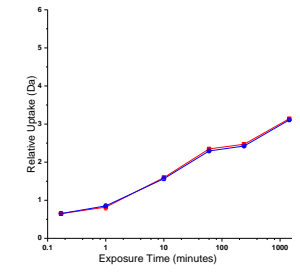
37-43
VVQDPAL



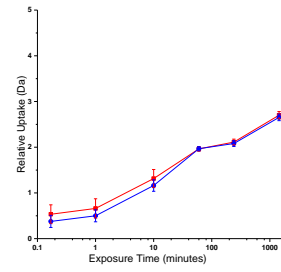
39-48
QDPALKPLAL



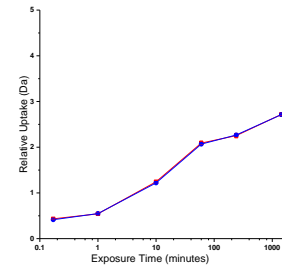
40-48
DPALKPLAL



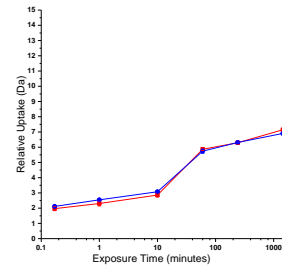
41-48
PALKPLAL



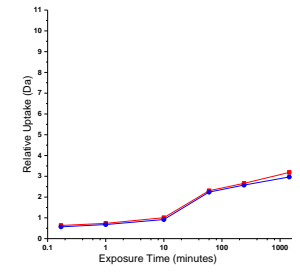
42-48
ALKPLAL



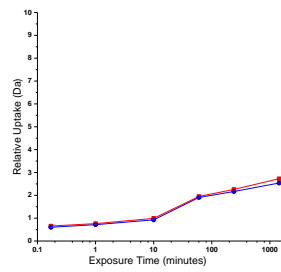
50-65
YGEATSRRMVNNGHSF



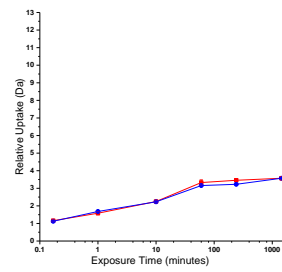
54-65
TSRRMVNNGHSF



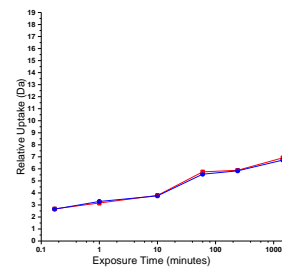
55-65
SRRMVNNGHSF



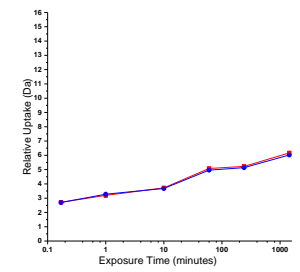
65-78
FNVEYDDSDQKAVL



66-86
NVEYDDSDQKAVLKDGPLT
GT

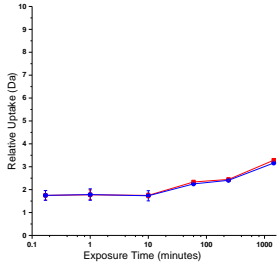


69-86
YDDSDQKAVLKDGPLTGT



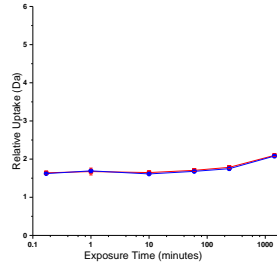
75-86

KAVLKDGPLTGT



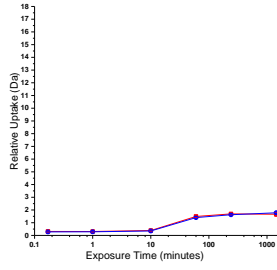
79-86

KDGPLTGT



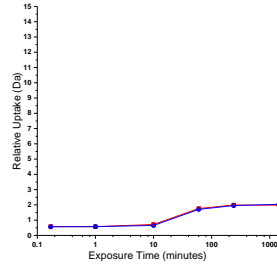
87-105

YRLVQFHFWGSSDDQGSE



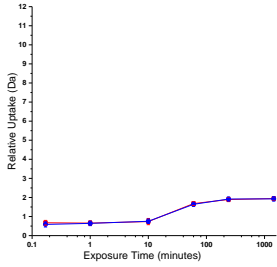
90-105

VQFHFWGSSDDQGSE



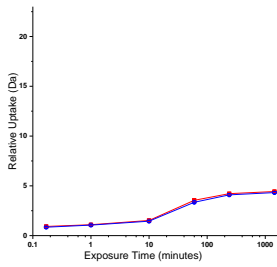
93-105

HFHWGSSDDQGSE



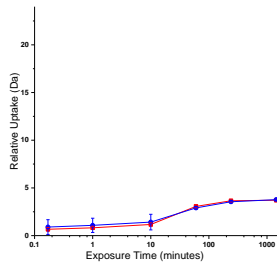
93-116

HFHWGSSDDQGSEHTVDRK
KYAAE



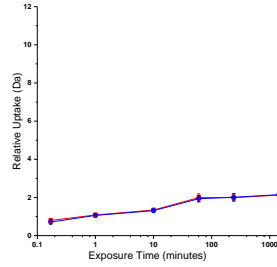
93-117

HFHWGSSDDQGSEHTVDRK
KYAAEL



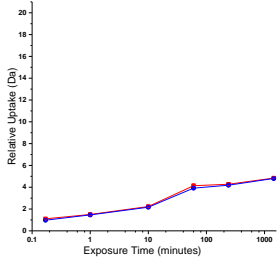
117-129

LHLVHWNTKYGDF



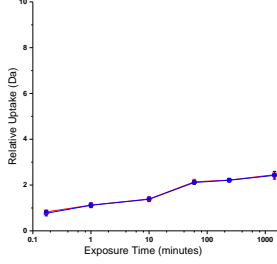
117-139

LHLVHWNTKYGDFGTAAQQ
PDGL



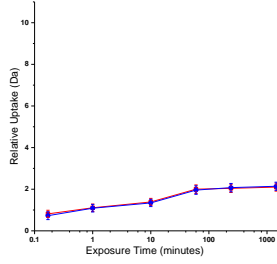
118-128

HLVHWNTKYGD



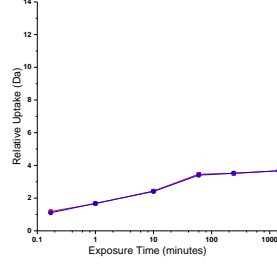
118-129

HLVHWNTKYGDF



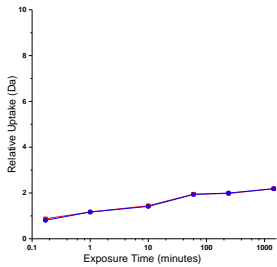
118-132

HLVHWNTKYGDFGTAA



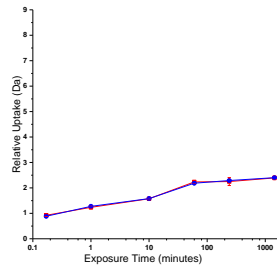
119-129

LVHWNTKYGDF



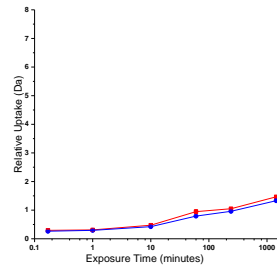
120-129

VHWNTKYGDF



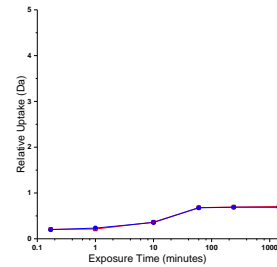
130-139

GTAAQQPDGL

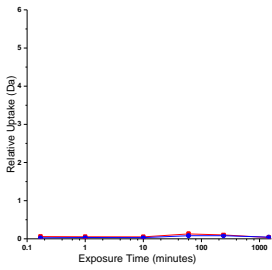


133-139

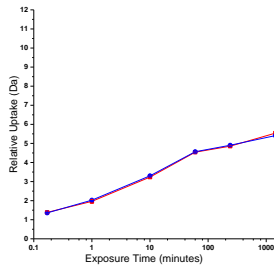
AQQPDGL



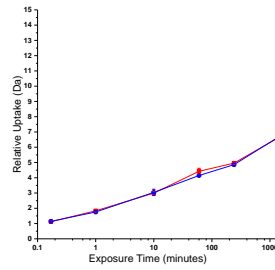
140-146
AVVGVFL



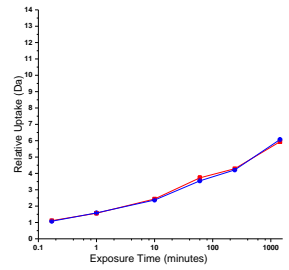
146-159
LKVGDANPALQKVL



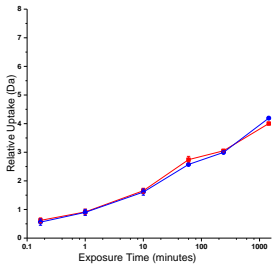
146-162
LKVGDANPALQKVLDAI



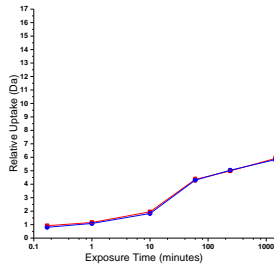
147-162
KVGDNANPALQKVLDAI



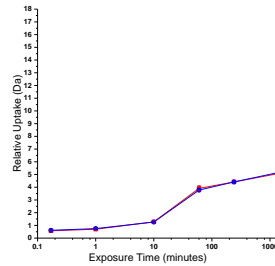
153-162
PALQKVLDAI



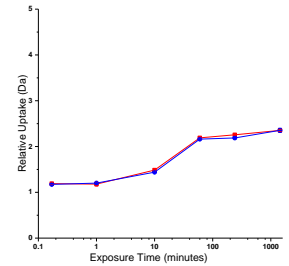
163-182
DSIKTKGKSTDFPNFDPGSL



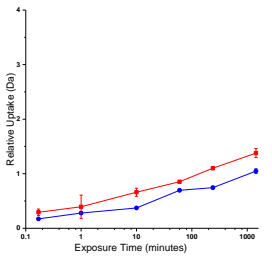
163-183
DSIKTKGKSTDFPNFDPGSLI



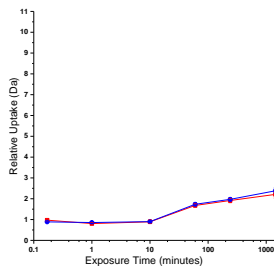
183-189
LPNVLDY



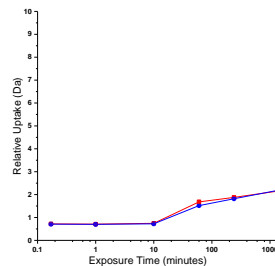
197-203
TTPPLLE



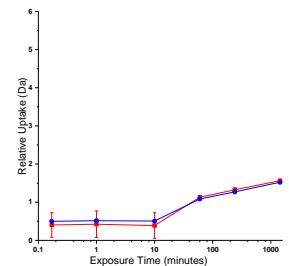
209-221
VLKEPISVSSQQM



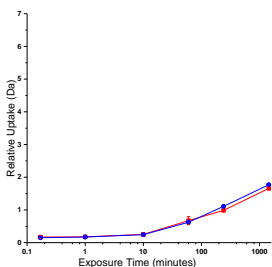
210-221
LKEPISVSSQQM



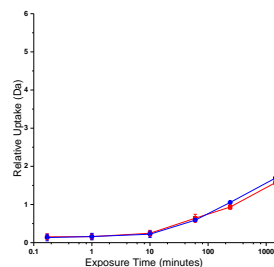
215-221
SVSSQQM



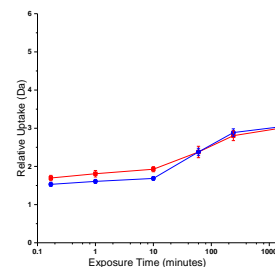
222-229
LKFRTLNF



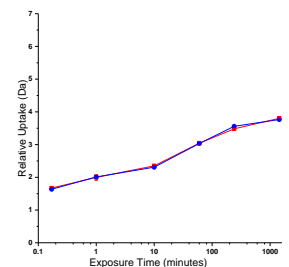
223-229
KFRTLNF



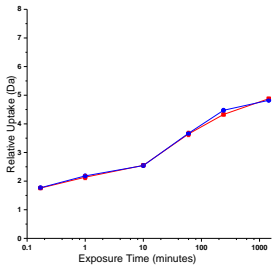
230-237
NAEGEPEL



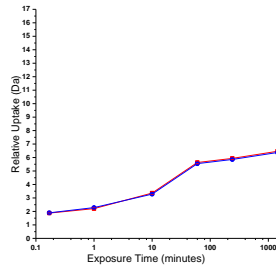
230-238
NAEGEPEL



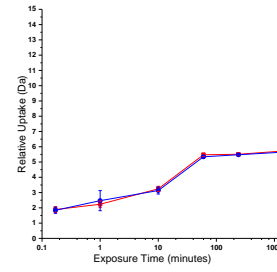
230-239
NAEGPELLM



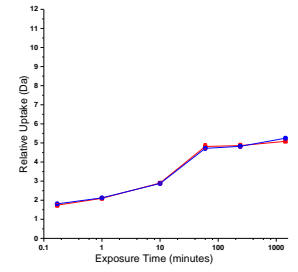
239-259
MLANWRPAQPLKNRQVRGFPK



241-259
ANWRPAQPLKNRQVRGFPK



244-259
RPAQPLKNRQVRGFPK



245-259
PAQPLKNRQVRGFPK

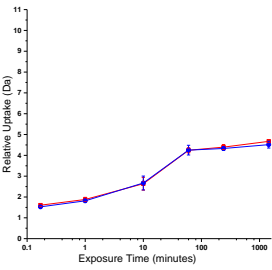


Figure 4.8 Deuterium uptake plots for BCA.

BCA with brinzolamide bound in blue and without brinzolamide bound in red.

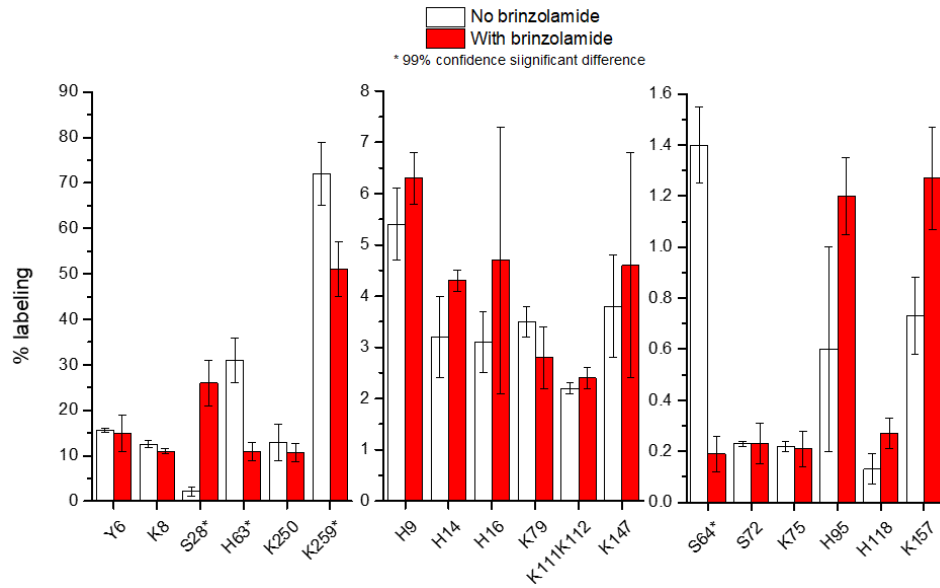


Figure 4.9 DEPC labeling percentages for BCA.

DEPC labeling results for BCA with (red) and without (white) brinzolamide bound. Residues that undergo significant decreases in DEPC labeling at a 99% confidence interval are marked with an asterisk “*”.

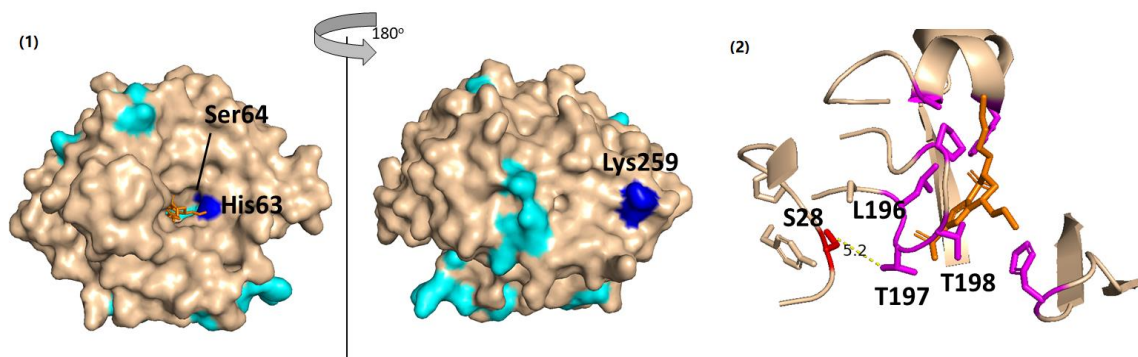


Figure 4.10 DEPC labeling results of BCA on crystal structure.

(1) Structure of bovine carbonic anhydrase with sites undergoing no change in covalent labeling (cyan) or a significant decrease in covalent labeling (blue) with binding site residues His63 and Ser64 indicated. (2) Expanded region around the brinzolamide binding site in bovine carbonic anhydrase, indicating residues (in magenta) interacting with brinzolamide (orange) and the proximity of Ser 28 (red) to the binding residues.

Applying HDX/MS and CL/MS to BCA demonstrates that the two techniques can provide complementary information. HDX reveals that the protein undergoes very little change in dynamics or structure upon ligand binding, although decreased exchange is observed in some parts of the ligand binding site. CL provides complementary insight by revealing other regions of the ligand binding site that undergo decreased solvent accessibility that are not reported by HDX/MS. Moreover, CL at Ser28 is affected indirectly by ligand binding, revealing a subtle change in structure around the binding site.

4.3.3 Maltose Binding Protein

For a third model system we selected the maltose binding protein (MBP) and its ligand maltose, which is known to undergo a conformational change upon ligand binding. Two globular domains in MBP are connected by a hinge region that includes a short helix (around residues 315-328) and a two-stranded B-sheet (around residues 167-184) [46]. These two domains adopt an “open” and “closed” state by a hinge rotation of 35°(Figure

4.11) [47-49]. According to previous NMR studies, this conformational change occurs with no significant dynamic changes in MBP induced by maltose binding [50]. Such a conformational rearrangement with minimal dynamics changes allows this protein-ligand system to reveal other aspects of the synergy provided by CL and HDX-MS measurements.

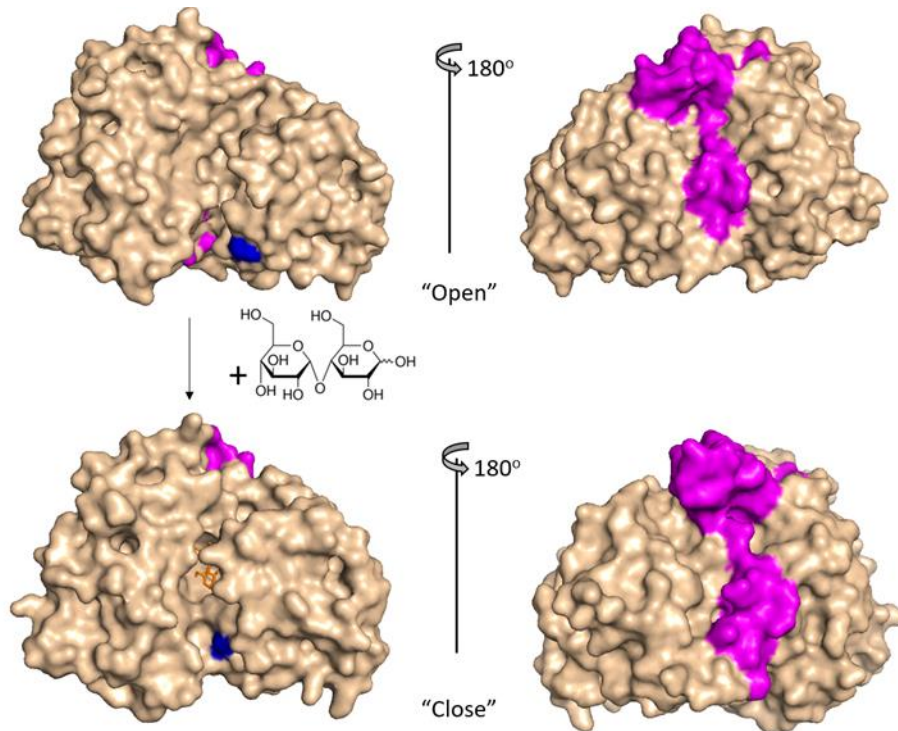


Figure 4.11 Structure of ligand free and maltose bound state for the maltose binding protein.

Ligand free “open” state (PDB 1OMP) and maltose bound “closed” state (PDB 1ANF) for the maltose binding protein. The residues that are part of the hinge region are indicated in magenta. K297 is indicated in blue.

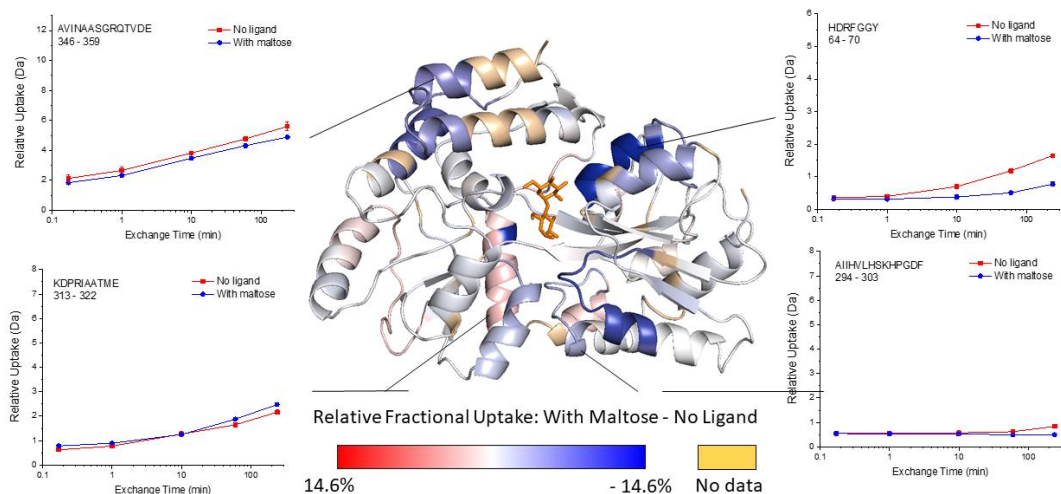


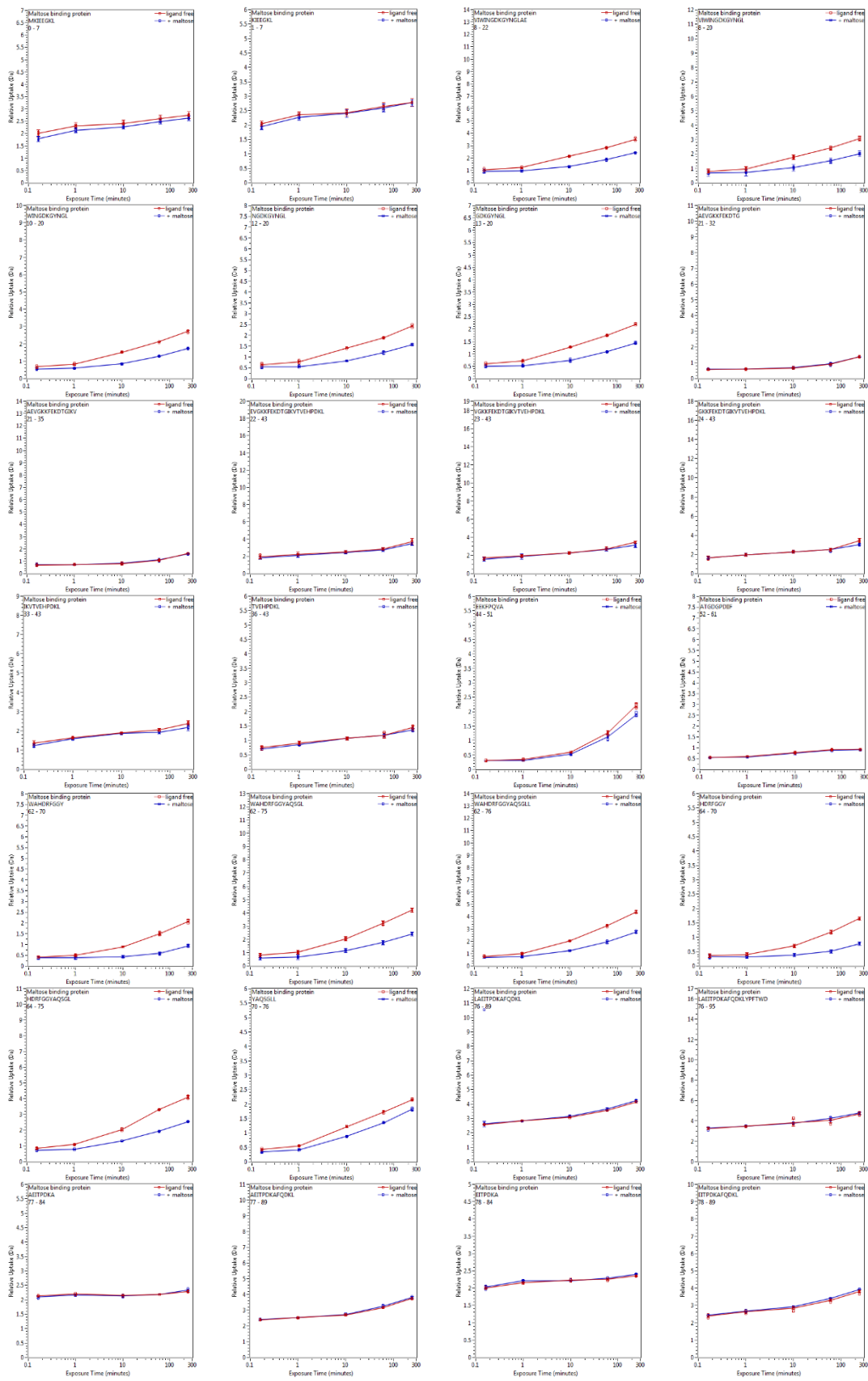
Figure 4.12 Relative fractional deuterium uptake for MBP shown on MBP-maltose structure.

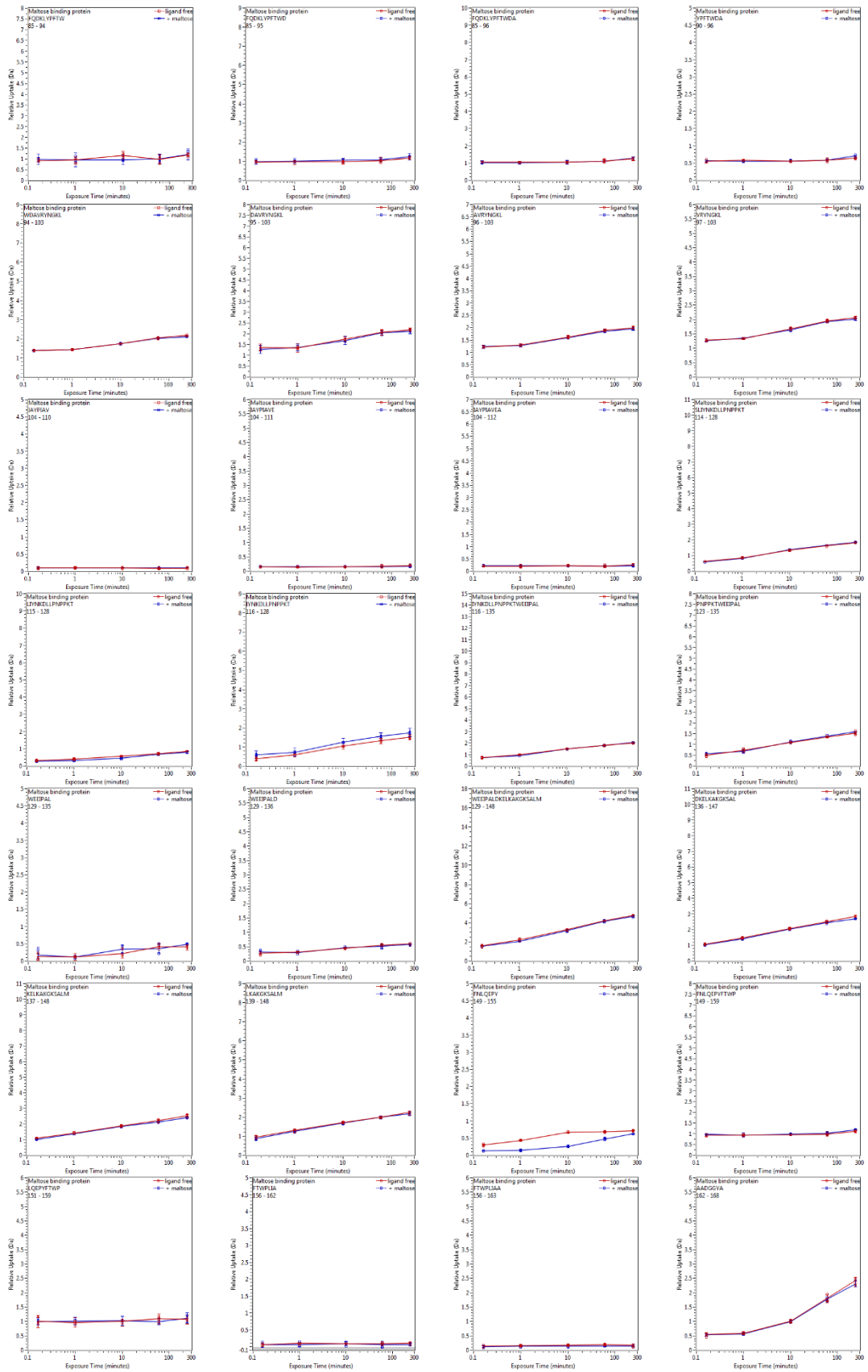
Representative deuterium uptake plots for maltose binding protein (MBP) and the MBP-maltose complex, and the relative fractional uptake (at the 4 h exchange time point) mapped onto the MBP-maltose complex crystal structure (PDB ID 1ANF). The darker the blue color on the structure, the more the HDX decreases upon ligand binding.

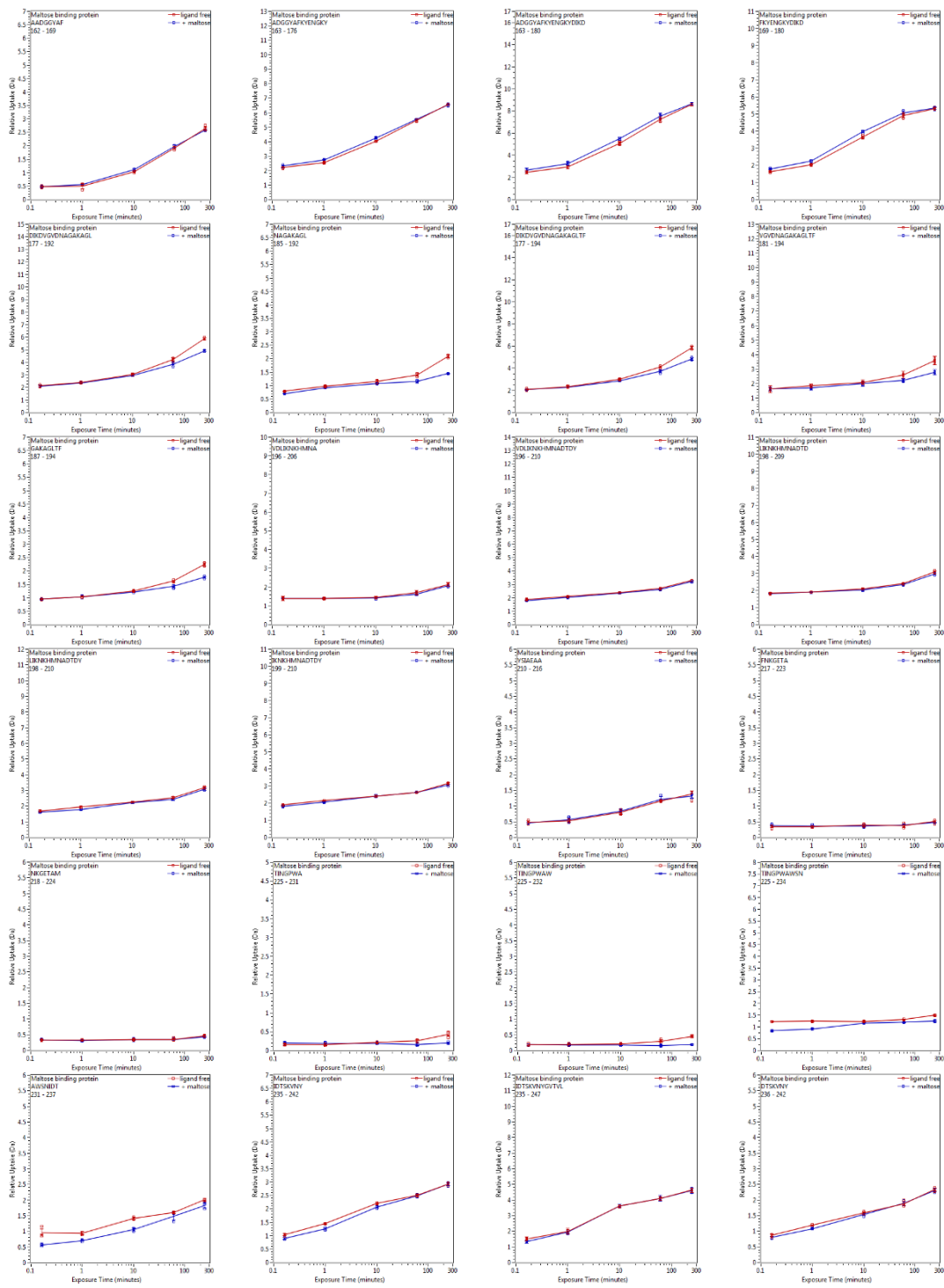
Table 4.2 Peptides from MBP with statistically significant differences in the relative deuterium uptake.

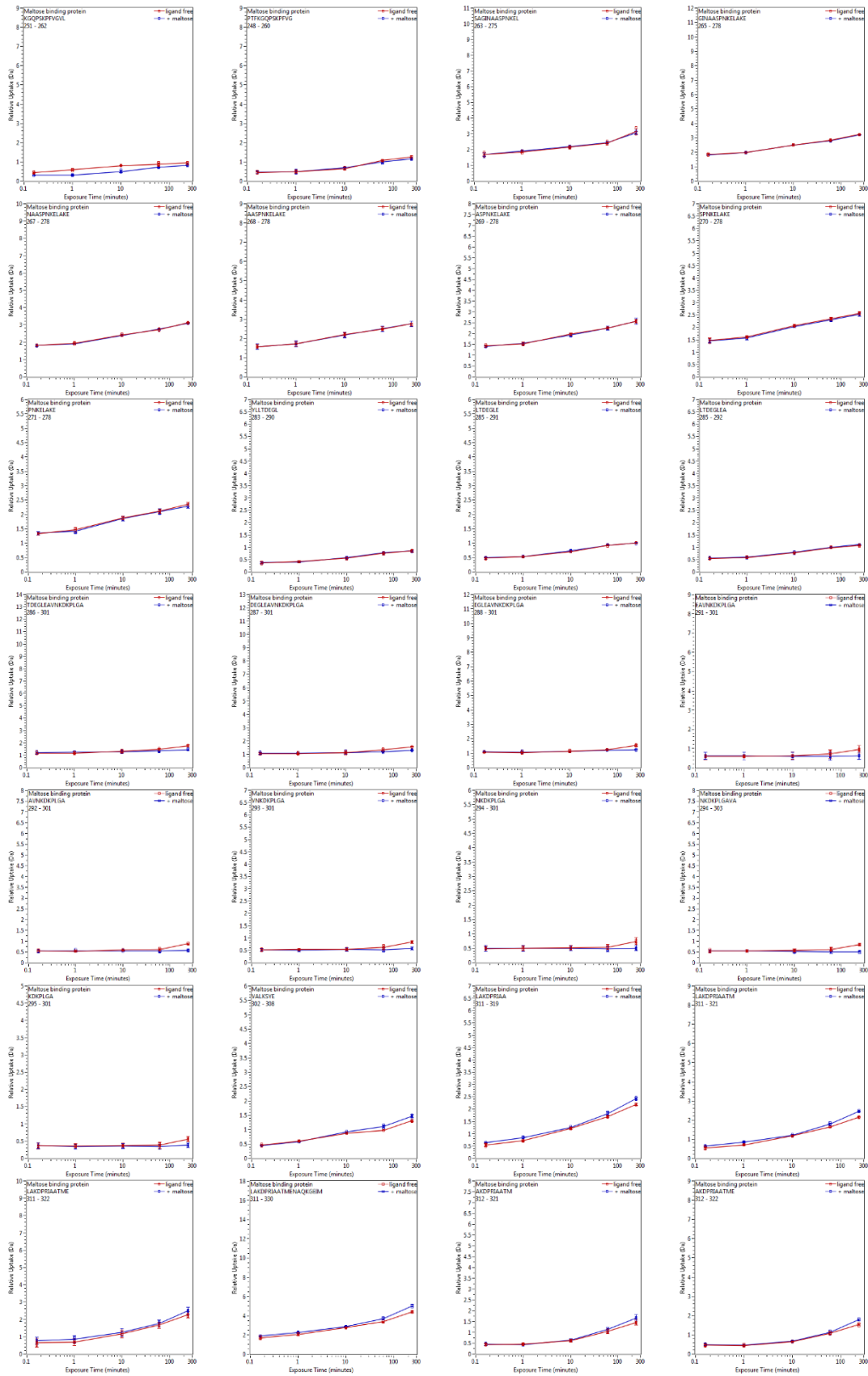
Statistically significant differences as determined using the criteria described in experimental section. The global deuterium uptake difference significance threshold for MBP is 0.34 Da.

Region	Sequence position	Sequence
8 – 22	8 – 22	VIWINGDKGYNGLAE
	10 – 20	WINGDKGYNGL
	12 – 20	NGDKGYNGL
	13 – 20	GDKGYNGL
	62 – 70	WAHDRFGGY
62 – 76	62 – 75	WAHDRFGGYAQSGL
	62 – 76	WAHDRFGGYAQSGLL
	64 – 75	HDRFGGYAQSGL
322 – 336	322 – 336	ENAQKGEIMPNIQPM
346 – 361	346 – 361	AVINAASGRQTVDEAL









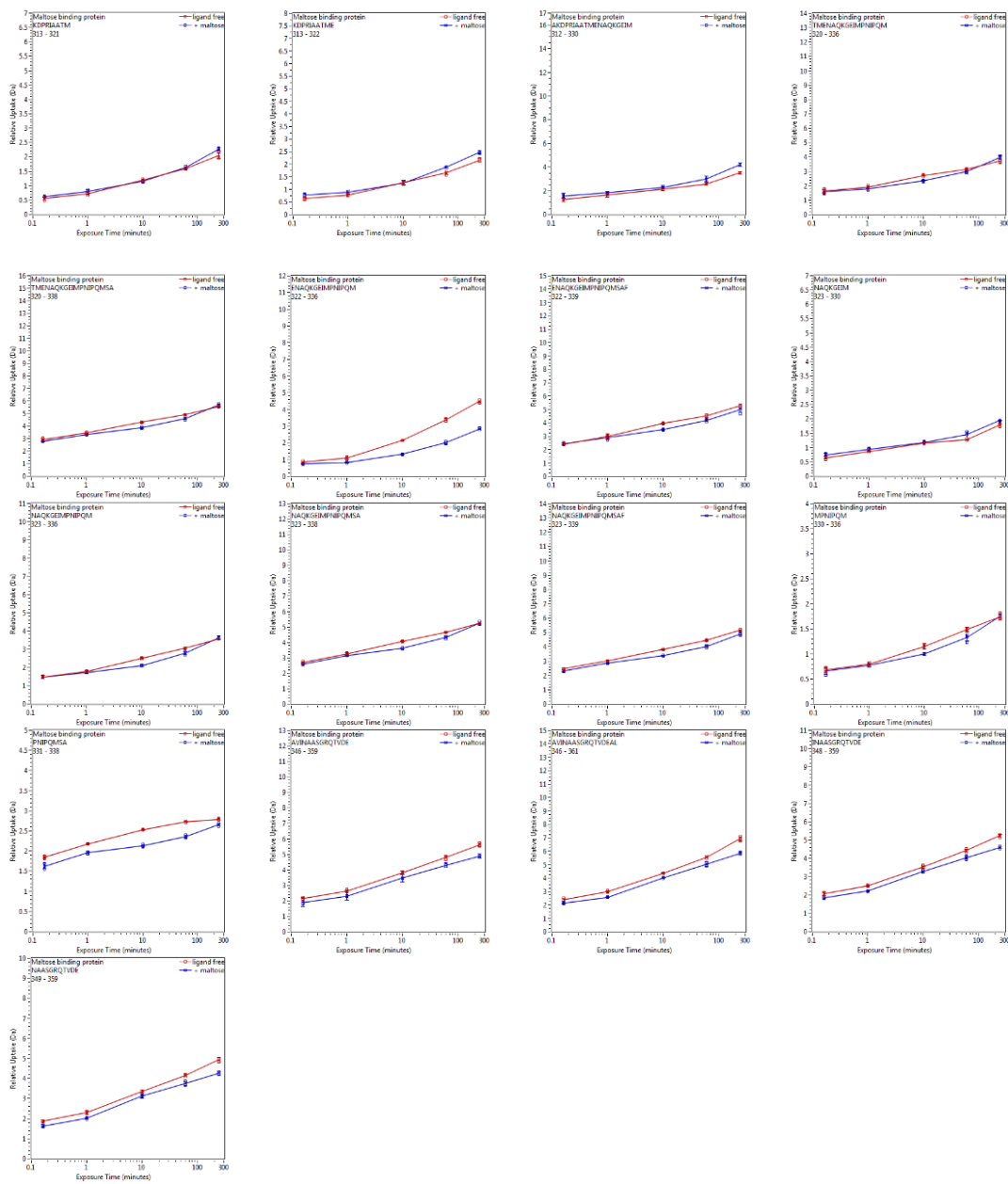


Figure 4.13 Deuterium uptake plots for MBP. MBP with maltose bound in blue and without maltose bound in red.

HDX indicates that 10 out of 125 detected peptides (Figure 4.12, 4.13 and Table 4.2) have a significant decrease in exchange, and these peptides span four regions of the protein: 8-22, 62-76, 322-336 and 346-361 (Figure 4.14). The peptides 8-22 and 62-75 contain residues that line the ligand-binding pocket and form hydrogen bonds with maltose [46, 51]. The peptide 322-336 represents part of the hinge region and also has

residues that are as close as 4 Å from the maltose binding site. The peptide 346-361 is remote from ligand binding site and also is not part of the hinge region, suggesting that this region undergoes dynamic changes that were too subtle or occur on a different timescale to have been detected in previous NMR experiments [50].

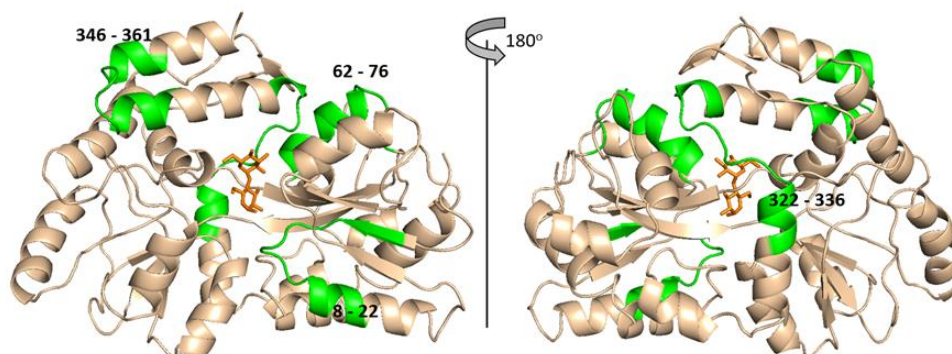


Figure 4.14 Regions identified to undergo significantly decreased HD exchange upon maltose binding of MBP.

Significantly decreased regions shown in green. The structure is from PDB 1ANF.

DEPC labeling of MBP results in 26 modified residues, yet only Lys297 shows a statistically significant decrease in labeling upon ligand binding (Figure 4.15). Unfortunately, of the seven modifiable residues (Lys15, Tyr17, His64, Tyr70, Lys326, Ser252, Thr356) in the four regions that undergo changes in HDX (i.e. 8-22, 62-76, 322-336 and 346-361), none of them are found to be labeled by DEPC. The residue that does decrease in labeling, Lys297, is partially buried when the protein undergoes the conformational change from the open to closed state. This observation is also supported by the change in solvent accessible surface area (SASA) of Lys297 from 38% in the open state to 23% in the closed state. This decrease in SASA of Lys297 is the most significant decrease among all the measured DEPC modified residues upon ligand binding and the

conformational change (Table 4.3). A few residues become relatively more exposed upon ligand binding, such as Lys25 and Lys189, but these residues are already very exposed to solvent, and we have found that residues that are already highly exposed rarely undergo significant increases in labeling [38]. In other words, CL is more sensitive to decreases in SASA.

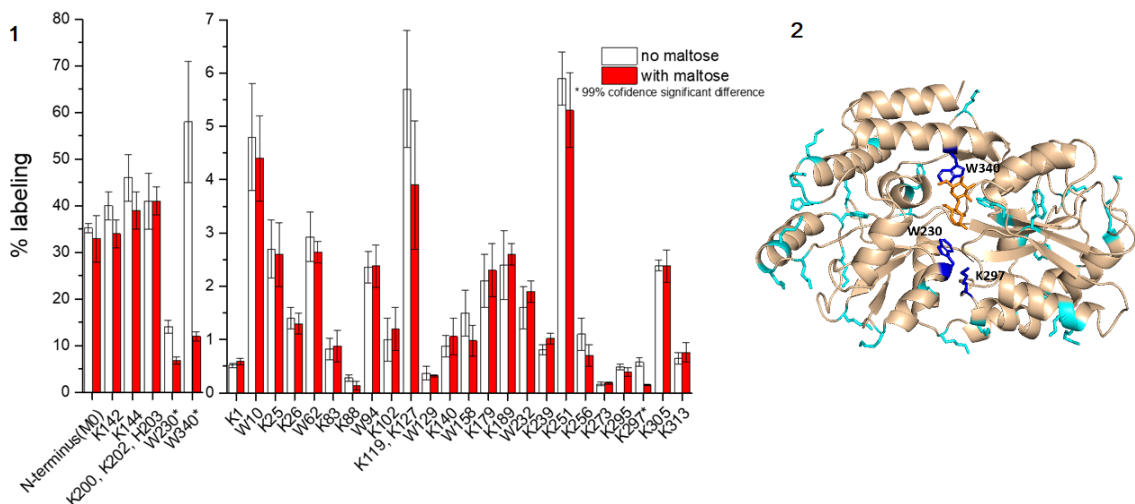


Figure 4.15 Summary of CL result for MBP upon maltose binding.

Summary of DEPC and HNSB labeling results for MBP upon maltose binding, and the location of modified residues mapped on the MBP-maltose complex structure (PDB ID 1ANF). (1) Relative covalent labeling percentages with (red) and without (white) ligand bound. (2) Residues undergoing significant decreases in labeling are indicated in blue. Residues that are covalently labeled but undergo no significant change in labeling percentage are indicated in cyan.

Because there are a limited number of residues along the maltose binding pocket that can be modified by DEPC, we applied another labeling reagent, HNSB, which specifically modifies tryptophan side chains with reaction kinetics that are similarly as slow as DEPC [23]. HNSB was chosen because there are two Trp residues (Trp230 and Trp340) near the binding site. Eight Trp residues are labeled by HNSB, and only Trp230 and Trp340 significantly decrease in labeling extent upon ligand binding. It should be noted that Trp232, which is located near the hinge region also undergoes a notable

decrease in SASA (Table 4.3) but does not significantly decrease in labeling extent. The reasons for this are not clear.

Comparing the results from HDX and CL for the MBP-maltose system further indicate the value of obtaining results from both techniques. The two methods together provide a clearer picture of the ligand binding site (Figure 4.16). CL data from both DEPC and HNSB reveal only residues that decrease in labeling due to significant decreases in SASA. Lys297, Trp230, and Trp340 all significantly decrease in SASA value upon ligand binding and the associated conformational change, but these three residues alone are somewhat insufficient to fully map the binding site. The data from HDX-MS alone is also insufficient to identify the binding site because too many non-proximate regions of the protein undergo decreases in deuterium uptake in the presence of the ligand. Considering the data together, however, allows one to more confidently map the maltose binding site.

Table 4.3 SASA values of residues on MBP that are modified by DEPC or HNSB.

Residue	No ligand PDB 1OMP	With ligand PDB 1ANF	Residue	No ligand PDB 1OMP	With ligand PDB 1ANF
K1	Not available in the structure	Not available	K179	62.3	68.9
W10	11.9	10.6	K189	30.5	39.9
K25	54.2	83.4	K200	67.2	62.2
K26	67	64.7	K202	68.3	79.9
W62	22.5	22.7	H203	15.7	16.4
K83	96	94.1	W230	34.3	5.2
K88	63.1	57.8	W232	10.3	1.2
W94	3.8	3.6	K239	87.3	82.1
K102	67.1	66.3	K251	48.6	56.6
K119	48.3	47.1	K256	26.5	30.8
K127	64.6	83.8	K273	43.7	44.7
W129	0.3	1.1	K295	69.6	67.1
K140	46.2	43.2	K297	38	23.9
K142	62.5	48.5	K305	61.5	47
K144	35.2	30.6	K313	87.8	71.7
W158	0.1	1.4	W340	25	7.1

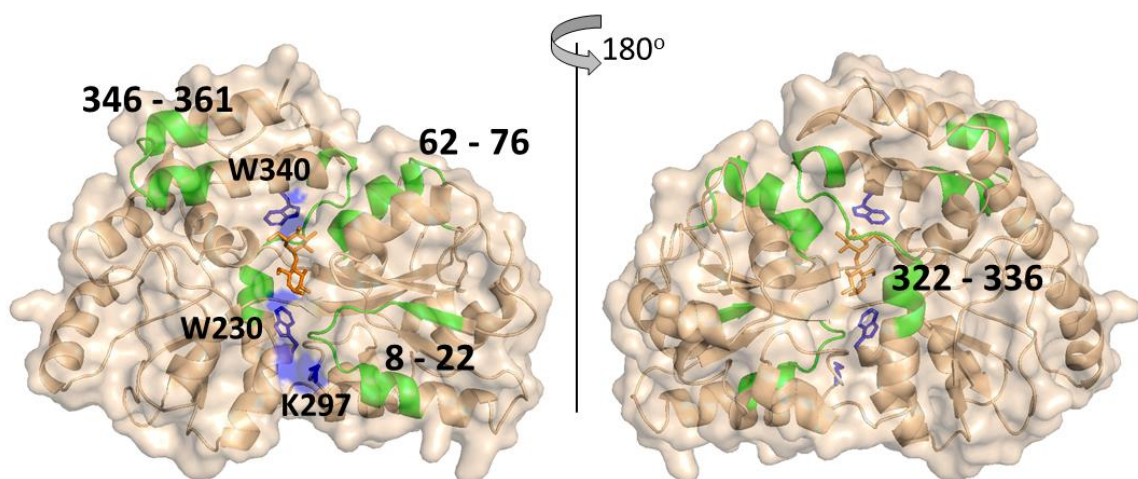


Figure 4.16 A comparison of the differential CL and HDX results for the maltose binding protein mapped on its crystal structure.

Regions that undergo significant decreases in HDX upon maltose binding are shown in green. Residues that undergo significant decreases in DEPC labeling are shown in blue. MBP crystal structure PDB 1ANF.

4.4 Conclusion

Using three model protein-ligand systems, we demonstrate that HDX-MS and CL-MS can provide complementary and synergistic information about protein-ligand interactions. For brinzolamide binding to BCA and maltose binding to MBP, the two techniques together provide separate information that more clearly indicates the ligand binding site. Changes in side chain solvent accessibility upon ligand binding cause significant decreases in CL that complement the decreases in HDX that also occur at the ligand binding site. When used together on the same protein-ligand system, the different timescales of the two labeling techniques can also provide information that is not accessible to either technique alone. This synergy is most clearly evident in heme binding to myoglobin. Heme binding to myoglobin causes decreases in HDX at the ligand binding site and distant from the ligand binding site, which would make it difficult to definitively identify the heme binding site with HDX alone. In contrast, CL is only influenced by changes in side-chain solvent accessibility at the heme binding site and is insensitive to changes in protein structural fluctuations at sites distant from the heme binding site. When used together, HDX/MS and CL/MS provide more comprehensive information, revealing more clearly the binding site and ligand induced changes in protein structural fluctuations caused by the stabilizing effect of heme binding. With this better understanding of the complementarity and potential synergy of the two MS-based labeling techniques, we predict that these two methods will find widespread usage together for more deeply understanding protein-ligand systems that are important in areas such as drug discovery.

4.5 Reference

- [1] Baron, R.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **2013**, *64*, 151-175.
- [2] Babine, R. E., & Bender, S. L. *Chem. Rev.* **1997**, *97*, 1359-1472.
- [3] Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; Liu, S.-Q. *Int. J. Mol. Sci.* **2016**, *17*, 144.
- [4] Riccardi Sirtori, F.; Altomare, A.; Carini, M.; Aldini, G.; Regazzoni, L. *Methods* **2018**, *144*, 152-174.
- [5] Kitova, E. N.; El-Hawiet, A.; Schnier, P. D.; Klassen, J. S. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 431-441.
- [6] Guan, J.-Q.; Vorobiev, S.; Almo, S. C.; Chance, M. R. *Biochemistry* **2002**, *41*, 5765-5775.
- [7] Liu, X. R.; Zhang, M. M.; Rempel, D. L.; Gross, M. L. *Anal. Chem.* **2019**, *91*, 5508-5512.
- [8] Rogstad, S.; Faustino, A.; Ruth, A.; Keire, D.; Boyne, M.; Park, J. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 786-794.
- [9] Beck, A.; Diemer, H.; Ayoub, D.; Debaene, F.; Wagner-Rousset, E.; Carapito, C.; Van Dorsselaer, A.; Sanglier-Cianfèrari, S. *TrAC Trends Anal. Chem.* **2013**, *48*, 81-95.
- [10] Kaltashov, I. A.; Bobst, C. E.; Abzalimov, R. R.; Wang, G.; Baykal, B.; Wang, S. *Biotechnol. Adv.* **2012**, *30*, 210-222.
- [11] Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins Struct. Funct. Bioinform.* **1993**, *17*, 75-86.
- [12] Englander, S. W. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1481-1489.
- [13] Powell, K. D.; Ghaemmaghani, S.; Wang, M. Z.; Ma, L.; Oas, T. G.; Fitzgerald, M. C. *J. Am. Chem. Soc.* **2002**, *124*, 10256-10257.
- [14] Zhu, M. M.; Rempel, D. L.; Du, Z.; Gross, M. L. *J. Am. Chem. Soc.* **2003**, *125*, 5252-5253.
- [15] Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; West, G. M.; Griffin, P. R. *Expert Rev. Proteomics* **2011**, *8*, 43-59.
- [16] Iacob, R. E.; Engen, J. R. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1003-1010.
- [17] Percy, A. J.; Rey, M.; Burns, K. M.; Schriemer, D. C. *Anal. Chim. Acta.* **2012**, *721*, 7-21.
- [18] Liu, T., Marcinko, T. M., Kiefer, P. A., & Vachet, R. W. *Anal. Chem.* **2017**, *89*, 11583-11591.
- [19] Li, Z.; Moniz, H.; Wang, S.; Ramiah, A.; Zhang, F.; Moremen, K. W.; Linhardt, R. J.; Sharp, J. S. *J. Biol. Chem.* **2015**, *290*, 10729-10740.
- [20] Rashidzadeh, H.; Khrapunov, S.; Chance, M. R.; Brenowitz, M. *Biochemistry* **2003**, *42*, 3655-3665.
- [21] Hambly, D.; Gross, M. *Int. J. Mass Spectrom.* **2007**, *259*, 124-129.
- [22] Zhang, H.; Gau, B. C.; Jones, L. M.; Vidavsky, I.; Gross, M. L. *Anal. Chem.* **2011**, *83*, 311-318.
- [23] Mendoza, V. L.; Vachet, R. W. *Mass Spectrom. Rev.* **2009**, *28*, 785-815.
- [24] Limpikirati, P.; Liu, T.; Vachet, R. W. *Methods* **2018**, *144*, 79-93.
- [25] Xu, G.; Chance, M. R. *Chem. Rev.* **2007**, *107*, 3514-3543.
- [26] Zhang, B.; Cheng, M.; Rempel, D.; Gross, M. L. *Methods* **2018**, *144*, 94-103.

- [27] Jumper, C. C.; Schriemer, D. C. *Anal. Chem.* **2011**, *83*, 2913-2920.
- [28] Jumper, C. C.; Bomgarden, R.; Rogers, J.; Etienne, C.; Schriemer, D. C. *Anal. Chem.* **2012**, *84*, 4411-4418.
- [29] Gau, B. C.; Sharp, J. S.; Rempel, D. L.; Gross, M. L. *Anal. Chem.* **2009**, *81*, 6563-6571.
- [30] Mendoza, V. L.; Vachet, R. W. *Anal. Chem.* **2008**, *80*, 2895-2904.
- [31] Boriack-Sjodin, P. A.; Zeitlin, S.; Christianson, D. W.; Chen, H.-H.; Crenshaw, L.; Gross, S.; Dantanarayana, A.; Delgado, P.; May, J. A.; Dean, T. *Protein Sci.* **1998**, *7*, 2483-2489.
- [32] Walker, I. H.; Hsieh, P.; Riggs, P. D. *Appl. Microbiol. Biotechnol.* **2010**, *88*, 187-197.
- [33] Glasoe, P. K.; Long, F. A. *J. Phys. Chem.* **1960**, *64*, 188-190.
- [34] Krężel, A.; Bal, W. A. *J. Inorg. Biochem.* **2004**, *98*, 161-166.
- [35] Hageman, T. S.; Weis, D. D. *Anal. Chem.* **2019**, *91*, 8008-8016.
- [36] Zhou, Y.; Vachet, R. W. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 708-717.
- [37] Borotto, N. B.; Zhou, Y.; Hollingsworth, S. R.; Hale, J. E.; Graban, E. M.; Vaughan, R. C.; Vachet, R. W. *Anal. Chem.* **2015**, *87*, 10627-10634.
- [38] Limpikirati, P.; Hale, J. E.; Hazelbaker, M.; Huang, Y., Jia, Z., Yazdani, M.; Vachet, R. W. *mAbs.* **2019**, *11*, 463-476.
- [39] Eliezer, D.; Wright, P. E. *J. Mol. Biol.* **1996**, *263*, 531-538.
- [40] Pan, J.; Han, J.; Borchers, C. H.; Konermann, L. *J. Am. Chem. Soc.* **2009**, *131*, 12801-12808.
- [41] Eliezer, D.; Yao, J.; Dyson, H. J.; Wright, P. E. *Nat. Struct. Mol. Biol.* **1998**, *5*, 148.
- [42] Johnson, R. S.; Walsh, K. A. *Protein Sci.* **1994**, *3*, 2411-2418.
- [43] Krishnamurthy, V. M.; Kaufman, G. K.; Urbach, A. R.; Gitlin, I.; Gudiksen, K. L.; Weibel, D. B.; Whitesides, G. M. *Chem. Rev.* **2008**, *108*, 946-1051.
- [44] Jain, A.; Whitesides, G. M.; Alexander, R. S.; Christianson, D. W. *J. Med. Chem.* **1994**, *37*, 2100-2105.
- [45] Stams, T.; Chen, Y.; Christianson, D. W.; Boriack-Sjodin, P. A.; Hurt, J. D.; Laipis, P.; Silverman, D. N.; Liao, J.; May, J. A.; Dean, T. *Protein Sci.* **1998**, *7*, 556-563.
- [46] Spurlino, J. C.; Lu, G. Y.; Quioco, F. A. *J. Biol. Chem.* **1991**, *266*, 5202-5219.
- [47] Sharff, A. J.; Rodseth, L. E.; Spurlino, J. C.; Quioco, F. A. *Biochemistry* **1992**, *31*, 10657-10663.
- [48] Tang, C.; Schwieters, C. D.; Clore, G. M. *Nature* **2007**, *449*, 1078-1082.
- [49] Bucher, D.; Grant, B. J.; McCammon, J. A. *Biochemistry* **2011**, *50*, 10530-10539.
- [50] Evenäs, J.; Tugarinov, V.; Skrynnikov, N. R.; Goto, N. K.; Muhandiram, R.; Kay, L. E. *J. Mol. Biol.* **2001**, *309*, 961-974.
- [51] Quioco, F. A.; Spurlino, J. C.; Rodseth, L. E. *Structure* **1997**, *5*, 997-1015.

CHAPTER 5

SUMMARY AND FUTURE DIRECTIONS

5.1 Summary

The work presented in this dissertation has focused on three main goals. First, we applied CL-MS for identifying ligand binding site on the protein. We studied two amyloid inhibiting molecules, doxycycline and rifamycin SV, identified their binding site on the Cu(II)- β 2m monomer. Besides using DEPC, a labeling reagent which can modify 6 kinds of amino acid side chains, BD labeling, which selectively modifies arginines, and the EDC/GEE labeling pair, which selectively modifies aspartic and glutamic acids, were also applied. Combining information from multiple kinds of labeling reagents increased the confidence in identifying ligand binding site. The determined doxycycline binding site on β 2m is on the D β strand and rifamycin SV is between G β strand and D β strand. This result is distinct from that of suramin, which binds to Cu(II)- β 2m, but is not an amyloid-inhibiting molecule. Suramin binds along the C β strand and near the C-D loop. More importantly, the conclusion is consistent with our group's previous finding that doxycycline and rifamycin work by preventing β 2m dimer to form tetramers. The work also suggests CL-MS could characterize protein-ligand complexes which aggregate fast and not capable to be studied by other methods such as X-ray crystallography.

In the second goal, we inspected if slower CL reactions such as DEPC is capable to correctly determine the K_d of the protein-ligand complexes. It is usually considered that due to the slow intrinsic reaction rate, non-radical CL reactions will distort the equilibrium of protein-ligand towards dissociation side, thus they cannot correctly determine the K_d . However, experiment results of two model systems: (NAG)₃-lysozyme

and maltose-maltose binding protein suggested as long as a proper strategy being applied, these CL methods are capable to correctly determine the K_d . The K_d we measured were not significantly off from what being acquired by other methods such as fluorescence and ESI-MS. We also discussed the key condition that the experiment design should follow, which is the fraction of protein that get labeled at the ligand binding site in the system must be limited at a low range (e.g. below 30%). If large fraction (e.g. more than 50%) of protein are labeled at the ligand binding site, acquired K_d would be significantly off from the actual value. We also applied the developed strategy to a system with an unknown K_d : EGCG and Cu(II)- β 2M. Besides successfully determined K_d values using residue level labeling results from two different residues, we also identified EGCG binding site of Cu(II)- β 2M monomer.

The third goal of this dissertation work was to compare the information from two complementary MS based methods, HDX-MS and CL-MS, when they are each applied to characterize the same protein-ligand complex. The intrinsic reaction rate of the two methods are 2-3 orders of magnitude difference. We expect CL, which rate is at 10's of second level, will be insensitive to ligand binding induced dynamic change to the protein. While HDX is known able to measure the dynamic change as well as protection brought by ligand binding, CL might help distinguish remote dynamic changes from protection due to decreased solvent accessibility. To examine our idea, we inspected three well-characterized model systems.

In the first model system, heme binding can significantly stabilize apomyoglobin. We observed near 70% of the myoglobin sequence decrease in deuterium uptake by HDX, while CL pinpointed 13 residues surrounding heme binding pocket. CL helped

distinguish dynamic change region from ligand binding site. In the second model system, bovine carbonic anhydrase II is a rigid protein that does not undergo dynamic or conformational change upon brinzolamide binding. While CL can easily identify ligand binding site with solvent exposure decrease of His63 caused by ligand binding. HDX only indicated limited decreases in exchange even for up to 24 h of exchange, likely due to the rigidity of the protein. The third model we inspected was maltose and maltose binding protein. Maltose binding protein will undergo significant conformational change upon maltose binding. Again, experiment result indicated CL is sensitive to changes in side chain's solvent accessibility. Overall, our experiment results lead us to the conclusion that due to its slower intrinsic reaction rate, some CL methods are insensitive to protein's dynamic changes. Thus, CL can help distinguish remote changes in protein dynamics from protections caused by ligand binding. Synergistic structural information available if CL-MS and HDX-MS are used to characterize protein-ligand complexes.

This dissertation highlights the power of CL-MS to characterize protein-ligand complexes. Our work has demonstrated that CL-MS can identify ligand binding site of the protein, determine the binding affinity of ligand to the protein, and indicate some major conformational change of the protein induced by ligand binding. If combine the information of CL-MS with HDX-MS synergistic information will allow a more comprehensive understanding of ligand binding sites and ligand induced structural changes, which can greatly assist drug discovery and related procedures. The study of several amyloid inhibiting molecules to Cu(II)- β 2M also partially explained the mechanism of these inhibitors. Determined inhibitor binding site of Cu(II)- β 2m could facilitate future library screening of new drug candidates. Our work also demonstrated

that CL-MS could characterize protein-ligand complexes that aggregate fast. This technique could be applied to study proteins that are not capable to be by other methods such as X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR).

5.2 Limitation of the Method

While we demonstrated that CL-MS can be useful for characterizing protein-ligand complexes, it does have some limitations that could be overcome with future research. First, a ligand binding site and nearby region must consist of CL modifiable residues to obtain binding site information. Through the study of several model systems as discussed in this thesis, we see that His and Lys, which are two of the most reactive residues for DEPC, are often not present within a ligand binding site. Although in a few cases a His or Lys can be found close enough to the ligand binding site and able to provide some information about ligand binding. While Ser, Tyr, or Thr can also be DEPC labeled, they seem less informative as there is no case that a ligand binding site is determined just by their labeling result. In Chapter 2, we discussed that additional information from one or more other kinds of labeling reagents that could help improve the overall confidence in the conclusion. However, suppose one is trying to determine the ligand binding site to a protein with no prior information available, starting with DEPC or residue specific labeling reagents likely will require multiple measurements to obtain the binding site information.

Also, by the nature of the strategy, information is only available from residues that are CL modified. For the rest of the sequence that is not CL covered, no information is available from the method. Depending on the situation, a cluster of such CL unmodified residues with no information could be large enough to miss out the ligand binding site or

structural changes. Or, such a cluster could decrease the overall reliability of the conclusion for the ligand binding site.

To minimize the limitations discussed above, new CL reagents which can modify more kinds of residues could help. Clearly, the more kinds of residues being modified, the higher the sequence coverage will be and less likely an area be missed out by the method. A new CL reagent is preferred to have a similar reaction rate to the CL reactions used in this thesis in order to transfer the developed strategies. One example is a series of new CL reagents being explored in our lab, that use α , β -unsaturated carbonyl compounds that can react with 13 kinds of residues.

Second, a structure of the protein is preferred to assist the data interpretation. As used many times in this thesis, a structure of the protein assists the final steps of interpreting the result from CL-MC, facilitate comparison of the spatial distribution of CL modified residues. Without the spatial distribution relationship, the result of CL will be much less meaningful, as very likely one cannot decide if CL significantly changed residues are focused as one cluster or not. Although all the proteins applied in this thesis have at least one available structure, it may not be required if effective molecular modeling could be done. The computational modeling here especially refers to the modeling strategy that use covalent labeling data to determine protein structure and interactions [1-5]. For example, Chance and co-workers have developed the ClusPro program for predicting protein complexes using the restraints from hydroxyl radical labeling to facilitate homology modeling and associated protein structural prediction for protein complexes with no NMR or X-ray crystal structure [2]. Similar work that uses DEPC labeling is being investigated in our lab.

5.3 Future Directions

5.3.1 CL-MS and Protein-Ligand Docking for Screening New β 2m Amyloid Inhibiting Molecules

In the work described in Chapter 2, we applied computational protein-ligand docking. Docking refined the structural information acquired by CL-MS and confirmed the experimentally determined ligand binding site. While the way we applied docking is one of the options, in industry, using protein-ligand docking to screen possible interactions between thousands of compounds to the target protein is a routine [6-8]. Comparing to experimental methods, docking is high throughput that can inspect thousands of compounds within a short time at low cost. However, docking cannot stand alone. Docking identified candidates still need to be verified by experiments before they are subjected to further studies [7].

In Chapters 2 and 3, we studied the binding site of doxycycline, rifamycin SV, and EGCG to Cu(II)- β 2m monomer. This information could instruct the docking process for screening new β 2m amyloid inhibiting molecules, by specifying region(s) that the candidates should bind to. For example, EGCG binding is likely among the N-terminus of the protein and EGCG-induced Cu(II) binding to the protein became weaker. New inhibitor candidates could bind in a similar manner but with a much stronger binding affinity to β 2m. Once the candidates are reported by docking, experimental methods including CL-MS can again be applied to confirm the binding site of the protein as described in Chapter 2. The final candidates could then be incubated with β 2m for further studies.

5.3.2 Distinguish Conformational Change from Ligand Binding Site in a CL-MS for Characterizing Protein-Ligand Complexes Experiment

As described in Chapters 3 and 4, our experiment results indicate CL-MS can also identify regions that undergo significant conformational changes upon ligand binding. Residues from such region will have changes in solvent accessibility, which can be reflected by residue level labeling ratio. However, as identifying ligand binding site also relies on decreases in solvent accessibility, it might be difficult to assign whether the labeling decrease is due to conformational changes or ligand binding. For example, in maltose binding protein, Lys297 significantly decreased in labeling ratio. Without a prior understanding of Lys297 get buried in during the conformational change, we might misassign the labeling decrease is due to ligand binding.

While applying HDX-MS, as described in Chapter 4, could help solve this potential problem. We hypothesize that inspecting the relative decreases of CL ratio might also help distinguish protein conformational changes from the ligand binding site. The decrease in residue's solvent accessibility induced by conformational change might not be as significant as ligand binding induced decrease. As a result, it might be reasonable to expect a residue directly involved in ligand binding has a relatively larger fractional decrease in labeling ratio, comparing to a residue from the conformational change region. At the same time, residues with increased labeling ratio could also help solve the problem. In a simple assumption, an allosteric site is more likely to have a residue become more solvent exposed during the transformation of the protein, comparing to the ligand binding site. Or, the relative fractional increase in labeling will be different enough to help distinguish the origin of two increased residues, with the increase in an allosteric site being larger. Model proteins which have multiple residues

significantly changes in labeling from both ligand binding site and conformational change region could verify this idea.

5.4 Reference

- [1] Gerega, S.K.; Downard, K.M. *Bioinformatics* **2006**, *22*, 1702-1709.
- [2] Kamal, J.K.A.; Chance, M.R. *Protein Sci.* **2008**, *17*, 79-94.
- [3] Charvátová, O.; Foley, B.L.; Bern, M.W.; Sharp, J.S.; Orlando, R.; Woods, R.J. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1692-1705.
- [4] Aprahamian, M. L.; Chea, E. E.; Jones, L. M.; Lindert, S. *Anal. Chem.* **2018**, *90*, 7721-7729.
- [5] Aprahamian, M. L.; Lindert, S. *J. Chem. Theory Comput.* **2019**, *15*, 3410-3424.
- [6] Pujadas, G.; Vaque, M.; Ardevol, A.; Blade, C.; Salvado, M. J.; Blay, M.; Fernandez-Larrea, J.; Arola, L. *Curr. Pharma. Anal.* **2008**, *4*, 1-19.
- [7] Jorgensen, W. L. *Science* **2004**, *303*, 1813-1818.
- [8] Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discov.* **2004**, *3*, 935-949.

BIBLIOGRAPHY

- Antwi, K.; Mahar, M.; Srikanth, R.; Olbris, M. R.; Tyson, J. F.; Vachet, R. W. *Protein Science*. **2014**, *17*, 748-759.
- Aprahamian, M. L.; Chea, E. E.; Jones, L. M.; Lindert, S. *Anal. Chem.* **2018**, *90*, 7721-7729.
- Aprahamian, M. L.; Lindert, S. *J. Chem. Theory Comput.* **2019**, *15*, 3410-3424.
- Babine, R. E.; Bender, S. L. *Chem. Rev.* **1997**, *97*, 1359-1472.
- Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins Struct. Funct. Bioinform.* **1993**, *17*, 75-86.
- Baron, R.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **2013**, *64*, 151-175.
- Beck, A.; Diemer, H.; Ayoub, D.; Debaene, F.; Wagner-Rousset, E.; Carapito, C.; Van Dorselaer, A.; Sanglier-Cianfèrani, S. *TrAC Trends Anal. Chem.* **2013**, *48*, 81-95.
- Borotto, N. B.; Degraan-Weber, N.; Zhou, Y.; Vachet, R. W. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1739-1746.
- Borotto, N. B.; Zhang, Z.; Dong, J.; Burant, B.; Vachet, R. W. *Biochemistry*. **2017**, *56*, 1095-1104.
- Borotto, N. B.; Zhou, Y.; Hollingsworth, S. R.; Hale, J. E.; Graban, E. M.; Vaughan, R. C.; Vachet, R. W. *Anal. Chem.* **2015**, *87*, 10627-10634.
- Bush, M.F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C.V.; Ruotolo, B.T.; *Anal. Chem.* **2010**, *82*, 9557-9565.
- Carazzone, C.; Colombo, R.; Quaglia, M.; Mangione, P.; Raimondi, S. Giorgetti, S.; Caccialanza, G.; Bellotti, V.; De Lorenzi, E. *Electrophoresis*. **2008**, *29*, 1502-1510.
- Chaires, J. B. *Annu. Rev. Biophys.* **2008**, *37*, 135-151.
- Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; West, G. M.; Griffin, P. R. *Expert Rev. Proteomics* **2011**, *8*, 43-59.
- Charvátová, O.; Foley, B.L.; Bern, M.W.; Sharp, J.S.; Orlando, R.; Woods, R.J. *J. Am. Soc. Mass Spectrom.* **2008**, *1919*, 1692-1705.
- Cheetham, J. C.; Artymiuk, P. J.; Phillips, D. C. *J. Mol. Biol.* **1992**, *224*, 613-628.
- Chen, J.; Rempel, D. L.; Gau, B. C.; Gross, M. L. *J. Am. Chem. Soc.* **2012**, *134*, 18724-18731.

- Cheng, B.; Gong, H.; Xiao, H.; Petersen, R. B.; Zheng, L.; Huang, K. *Biochim. Biophys. Acta BBA - Gen. Subj.* **2013**, *1830*, 4860-4871.
- Cheng, M.; Zhang, B.; Cui, W.; Gross, M. L. *Angew. Chem. Int. Ed.* **2017**, *56*, 14007-14010.
- Clark, S. M.; Konermann, L. *Anal. Chem.* **2004**, *76*, 7077-7083.
- Daniel, J. M.; Friess, S. D.; Rajagopalan, S.; Wendt, S.; Zenobi, R. *Int. J. Mass Spectrom.* **2002**, *216*, 1-27.
- Dennhart, N.; Letzel, Anal. *Bioanal. Chem.* **2006**, *386*, 689-698.
- Doig, A. J; *Curr. Opin. Struct. Biol.* **2015**, *30*, 50-56.
- Dong, J.; Joseph, C. A.; Borotto, N. B.; Gill, V. L.; Maroney, M. J.; Vachet, R. W. *Biochemistry* **2014**, *53*, 1263-1274.
- Drüeke, Tilman B.; Ziad A. Massy. *Semin. Dial.* **2009**, *22*, 378-380.
- Du, X.; Li, Y.; Xia, Y.-L.; Ai, S.-M.; Liang, J.; Sang, P.; Ji, X.-L.; Liu, S.-Q. *Int. J. Mol. Sci.* **2016**, *17*, 144.
- Eakin, C. M.; Attenello, F. J.; Morgan, C. J.; Miranker, A.D.; *Biochemistry* **2004**, *24*, 7808-7815.
- Eakin, C. M.; Knight, J. D.; Morgan, C. J.; Gelfand, M. A.; Miranker, A. D. *Biochemistry* **2002**, *41*, 10646-10656.
- Eichner, T.; Radford, S. E. *FEBS J.* **2011**, *278*, 3868-3883.
- El-Hawiet, A.; Kitova, E.N.; Klassen, J.S. *Biochemistry* **2012**, *51*, 4244-4253.
- Englander, S. W. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1481-1489.
- Fliss, H.; Viswanatha, T. *Can. J. Biochem.* **1979**, *57*, 1267-1272.
- Floege, J.; Ketteler, M. *Kidney Int.* **2001**, *59*, S164-S171.
- Fraczkiewicz, R.; Braun, W. *J. Comp. Chem.* **1998**, *19*, 319-333.
- Ganem, B.; Li, Y. T.; Henion, J. D. *J. Am. Chem. Soc.* **1991**, *113*, 7818-7819.
- Gau, B. C.; Sharp, J. S.; Rempel, D. L.; Gross, M. L. *Anal. Chem.* **2009**, *81*, 6563-6571.
- Gejyo, F.; Kawaguchi, Y.; Hara, S.; Nakazawa, R.; Azuma, N.; Ogawa, H.; Koda, Y.; Suzuki, M.; Kaneda, H.; Kishimoto, H.; Oda, M. *Artif. Organs.* **2004**, *28*, 371-380.
- Gejyo, F.; Odani, S.; Yamada, T.; Honma, N.; Saito, H.; Suzuki, Y.; Nakagawa, Y.; Kobayashi, H.; Maruyama, Y.; Hirasawa, Y.; Suzuki, M. *Kidney Int.* **1986**, *30*, 385-390.

- Gerega, S.K.; Downard, K.M. *Bioinformatics* **2006**, *22*, 1702-1709.
- Ghaemmaghami, S.; Fitzgerald, M. C.; Oas, T. G. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 8296-8301.
- Giorgetti, S.; Raimondi, S.; Pagano, K.; Relini, A.; Bucciantini, M.; Corazza, A.; Fogolari, F.; Codutti, L.; Salmona, M.; Mangione, P.; Colombo, L.; *J. Biol. Chem.* **2011**, *286*, 2121-2131.
- Guan, J.-Q.; Vorobiev, S.; Almo, S. C.; Chance, M. R. *Biochemistry* **2002**, *41*, 5765-5775.
- Güssow, D.; Rein, R.; Ginjaar, I.; Hochstenbach, F.; Seemann, G.; Kottman, A.; Ploegh, H. L. *J. Immunol.* **1987**, *139*, 3132-3138.
- Hambly, D. M.; Gross, M. *Int. J. Mass Spectrom.* **2007**, *259*, 124-129.
- Hambly, D. M.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 2057-2063.
- Hazenberg, B. P. C. *Rheum. Dis. Clin. North. Am.* **2013**, *39*, 323-345.
- Heegaard, N. H. *Amyloid* **2009**, *16*, 151-173.
- Hernández, H; Robinson, C.V., *Nat. Protoc.* **2007**, *2*, 715-726.
- Hu, W.; Walters, B. T.; Kan, Z.-Y.; Mayne, L.; Rosen, L. E.; Marqusee, S.; Englander, S. W. Stepwise. *Proc. Natl. Acad. Sci.* **2013**, *110*, 7684-7689.
- Iacob, R. E.; Engen, J. R. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1003-1010.
- Imoto, T.; Johnson, L. N.; North, A. C. T.; Phillips, D. C.; Rupley, J. *Academic Press, New York.* **1972**, 665-868.
- Ishikawa, T.; Reddy Burri, R.; O. Kamatari, Y.; Sakuraba, S.; Matubayasi, N.; Kitao, A.; Kuwata, K. *Chem. Phys.* **2013**, *15*, 3646-3654.
- Isom, D.G.; Castañeda, C.A.; Cannon, B.R. *PNAS* **2011**, *108*, 5260-5265.
- Jecklin, M. C.; Touboul, D.; Bovet, C.; Wortmann, A.; Zenobi, R. *Mass Spectrom.* **2008**, *19*, 332-343
- Jorgensen, W. L. *Science* **2004**, *303*, 1813-1818.
- Jumper, C. C.; Bomgarden, R.; Rogers, J.; Etienne, C.; Schriemer, D. C. *Anal. Chem.* **2012**, *84*, 4411-4418.
- Jumper, C. C.; Schriemer, D. C. *Anal. Chem.* **2011**, *83*, 2913-2920.
- Kaltashov, I. A.; Bobst, C. E.; Abzalimov, R. R.; Wang, G.; Baykal, B.; Wang, *Biotechnol. Adv.* **2012**, *30*, 210-222.

Kamal, J.K.A.; Chance, M.R. *Protein Sci.* **2008**, *17*, 79-94.

Katou, H.; Kanno, T.; Hoshino, M.; Hagihara, Y.; Tanaka, H.; Kawai, T.; Hasegawa, K.; Naiki, H.; Goto, Y. *Protein Sci.* **2002**, *11*, 2218-2229.

Katta, V.; Chait, B. T. *J. Am. Chem. Soc.* **1991**, *113*, 8534-8535.

Kaur, P.; Kiselar, J. G.; Chance, M. R. *Anal. Chem.* **2009**, *81*, 8141-8149.

Kaur, P.; Tomechko, S. E.; Kiselar, J.; Shi, W.; Deperalta, G.; Wecksler, A. T. Gokulrangan, G.; Ling, V.; Chance, M.R. *mAbs* **2015**, *7*, 540-552.

Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. *J. Am. Chem. Soc.* **1999**, *121*, 806-812.

Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discov.* **2004**, *3*, 935-949.

Kitova, E. N.; El-Hawiet, A.; Schnier, P. D.; Klassen, J. S. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 431-441.

Konermann, L.; Pan, J.; Liu, Y. *Chem. Soc. Rev.* **2011**, *40*, 1224-1234.

Li, K. S.; Chen, G.; Mo, J.; Huang, R. Y.-C.; Deyanova, E. G.; Beno, B. R.; O'Neil, S. R.; Tymiak, A. A.; Gross, M. L. *Anal. Chem.* **2017**, *89*, 7742-7749.

Li, Z.; Moniz, H.; Wang, S.; Ramiah, A.; Zhang, F.; Moremen, K. W.; Linhardt, R. J.; Sharp, J. S. *J. Biol. Chem.* **2015**, *290*, 10729-10740.

Lim, J.; Vachet, R. W. *Anal. Chem.* **2004**, *76*, 3498-3504.

Limpikirati, P.; Hale, J. E.; Hazelbaker, M.; Huang, Y.; Jia, Z.; Yazdani, M.; Graban, E.M.; Vaughan, R.C.; Vachet, R.W.; *mAbs* **2019**, *11*, 463-476.

Limpikirati, P.; Liu, T.; Vachet, R. W. *Methods* **2018**, *144*, 79-93.

Limpikirati, P.; Pan, X.; Vachet, R. W. *Anal. Chem.* **2019**, *91*, 8516-8523.

Liu, L.; Kitova, E.N.; Klassen, J.S. *J. Am. Soc. Mass Spectrom.* **2011**, *22*, 310-318.

Liu, T., Marcinko, T. M., Kiefer, P. A., & Vachet, R. W. *Anal. Chem.* **2017**, *89*, 11583-11591.

Liu, X. R.; Rempel, D. L.; Gross, M. L. *Anal. Chem.* **2019**, *30*, 213-217.

Liu, X. R.; Zhang, M. M.; Rempel, D. L.; Gross, M. L. *Anal. Chem.* **2019**, *91*, 5508-5512.

Liu, X. R.; Zhang, M. M.; Rempel, D. L.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 213-217.

- Liyanage, R.; Devarapalli, N.; Pyland, D. B.; Puckett, L. M.; Phan, N. H.; Starch, J. A.; Okimoto, M. R.; Gidden, J.; Stites, W. E.; Lay, J. O. *Mass Spectrom.* **2012**, 330-332, 63-70.
- Loo, R. R. O.; Goodlett, D. R.; Smith, R. D.; Loo, J. A. *J. Am. Chem. Soc.* **1993**, 115, 4391-4392.
- Maenaka, K.; Matsushima, M.; Song, H.; Sunada, F.; Watanabe, K.; Kumagai, J. *Mol. Biol.* **1995**, 247, 281-293.
- Maher, S.; Jjunju, F. P. M.; Taylor, S. *Rev. Mod. Phys.* **2015**, 87, 113-135.
- Manzi, L.; Barrow, A.S.; Scott, D.; Layfield, R.; Wright, T.G.; Moses, J.E.; Oldham, N.J. *Nat. Commun.* **2016**, 7, 13288.
- Marcinko, T. M.; Dong, J.; LeBlanc, R.; Daborowski, K. V.; Vachet, R. W. *J. Biol. Chem.* **2017**, 292, 10630-10638.
- Marcinko, T. M.; Drews, T.; Vachet, R.W. *The Effect of Epigallocatechin-3-gallate on Cu(II)-catalyzed β -2 Microglobulin Amyloid Formation. Manuscript in preparation.*
- Masson, G. R.; Burke, J. E.; Ahn, N. G.; Anand, G. S.; Borchers, C.; Brier, S.; Bou-Assaf, G. M.; Engen, J. R.; Englander, S. W.; Faber, J.; et al. *Nat. Methods* **2019**, 16, 595-602.
- McLafferty, F. W.; Breuker, K.; Jin, M.; Han, X.; Infusini, G.; Jiang, H.; Kong, X.; Begley, *FEBS J.* **2007**, 274, 6256-6268.
- McParland, V. J.; Kad, N. M.; Kalverda, A. P.; Brown, A.; Kirwin-Jones, P.; Hunter, M. G.; Sunde, M.; Radford, S. E. *Biochemistry.* **2000**, 39, 8735-8746.
- Mendoza, V. L.; Antwi, K.; Barón-Rodríguez, M. A.; Blanco, C.; Vachet, R. W. *Biochemistry.* **2010**, 49, 1522-1532.
- Mendoza, V. L.; Barón-Rodríguez, M. A.; Blanco, C.; Vachet, R. W. *Biochemistry.* **2011**, 50, 6711-6722.
- Mendoza, V. L.; Vachet, R. W. *Anal. Chem.* **2008**, 80, 2895-2904.
- Mendoza, V. L.; Vachet, R. W. *Mass Spectrom. Rev.* **2009**, 28, 785-815.
- Meng, F.; Forbes, A. J.; Miller, L. M.; Kelleher, *Mass Spectrom. Rev.* **2005**, 24, 126-134.
- Miller, D. M.; Olson, J. S.; Pflugrath, J. W.; Quijcho, F. A. *J. Biol. Chem.* **1983**, 258, 13665-13672.
- Morgan, C. J.; Gelfand, M.; Atreya, C.; Miranker, A. D. *J. Mol. Biol.* **2001**, 309, 339-345.
- Ookoshi, T.; Hasegawa, K.; Ohhashi, Y.; Kimura, H.; Takahashi, N.; Yoshida, H.; Miyazaki, R.; Goto, Y.; Naiki, H. *Nephrol. Dial. Transplant.* **2008**, 23, 3247-3255.

- Percy, A. J.; Rey, M.; Burns, K. M.; Schriemer, D. C. *Anal. Chim. Acta.* **2012**, *721*, 7-21.
- Powell, K. D.; Ghaemmaghami, S.; Wang, M. Z.; Ma, L.; Oas, T. G.; Fitzgerald, M. C. *J. Am. Chem. Soc.* **2002**, *124*, 10256-10257.
- Pujadas, G.; Vaque, M.; Ardevol, A.; Blade, C.; Salvado, M. J.; Blay, M.; Fernandez-Larrea, J.; Arola, L. *Curr. Pharma. Anal.* **2008**, *4*, 1-19.
- Quaglia, M.; Carazzone, C.; Sabella, S.; Colombo, R.; Giorgetti, S.; Bellotti, V.; Lorenzi, E. D. *Electrophoresis.* **2005**, *26*, 4055-4063.
- Quiocho, F. A.; Spurlino, J. C.; Rodseth, L. E. *Structure.* **1997**, *5*, 997-1015.
- Rashidzadeh, H.; Khrapunov, S.; Chance, M. R.; Brenowitz, M. *Biochemistry* **2003**, *42*, 3655-3665.
- Regazzoni, L.; Colombo, R.; Bertolotti, L.; Vistoli, G.; Aldini, G.; Serra, M.; Carini, M.; Facino, R. M.; Giorgetti, S.; Stoppini, M.; Caccialanza, G. *Anal. Chim. Acta* **2011**, *685*, 153-161.
- Riccardi Sirtori, F.; Altomare, A.; Carini, M.; Aldini, G.; Regazzoni, L. *Methods.* **2018**, *144*, 152-174.
- Rinas, A.; Espino, J. A.; Jones, L. M. *Anal. Bioanal. Chem.* **2016**, *408*, 3021-3031.
- Rogstad, S.; Faustino, A.; Ruth, A.; Keire, D.; Boyne, M.; Park, J. *Am. Soc. Mass Spectrom.* **2017**, *28*, 786-794.
- Sasahara, K.; Yagi, H.; Naiki, H.; Goto, Y. *J. Mol. Biol.* **2007**, *372*, 981-991.
- Schindler, M.; Assaf, Y.; Sharon, N.; Chipman, D. M. *Biochemistry.* **1977**, *16*, 423-431.
- Seo, M. H.; Park, J., Kim, E.; Hohng, S.; Kim, H. S. *Nat. Commun.* **2014**, *5*, 3724.
- Siuti, N.; Kelleher, N. L. *Nat. Methods.* **2007**, *4*, 817-821.
- Srikanth, R.; Mendoza, V. L.; Bridgewater, J. D.; Zhang, G.; Vachet, R. W. *Biochemistry.* **2009**, *48*, 9871-9881.
- Telmer, P. G.; Shilton, B. H.: *J. Biol. Chem.* **2003**, *278*, 34555-34567.
- Torbeev, V. Y.; Hilvert, D. *Proc. Natl. Acad. Sci.* **2013**, *110*, 20051-20056.
- Tsutsui, Y.; Cruz, R. D.; Wintrode, *Natl. Acad. Sci.* **2012**, *109*, 4467-4472.
- Walker, I. H.; Hsieh, P.; Riggs, P. D. *Appl. Microbiol. Biotechnol.* **2010**, *88*, 187-197.
- Wang, L.; Chance, M. R. *Anal. Chem.* **2011**, *83*, 7234-7241.

- Wang, W.; Kitova, E. N.; Klassen, J. S. *Anal. Chem.* **2003**, *75*, 4945-4955.
- Watson, C.; Janik, I.; Zhuang, T.; Charvátová, O.; Woods, R. J.; Sharp, J. S. **2009**, *81*, 2496-2505.
- Woods, L. A.; Platt, G. W.; Hellewell, A. L.; Hewitt, E. W.; Homans, S. W.; Ashcroft, A. E.; Radford, S. E. *Nat. Chem. Biol.* **2011**, *7*, 730.
- Xu, G.; Chance, M. R. *Chem. Rev.* **2007**, *107*, 3514-3543.
- Xu, Y.; Strickland, E. C.; Fitzgerald, M. C. *Anal. Chem.* **2014**, *86*, 7041-7048.
- Yamamoto, S.; Kazama, J. J.; Maruyama, H.; Nishi, S.; Narita, I.; Gejyo, F. *Clin. Nephrol.* **2008**, *70*, 496-502.
- Zhang, B.; Cheng, M.; Rempel, D.; Gross, M. L. *Methods* **2018**, *144*, 94-103.
- Zhang, H.; Gau, B. C.; Jones, L. M.; Vidavsky, I.; Gross, M. L. *Anal. Chem.* **2011**, *83*, 311-318.
- Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C., & Yates III, J. R. *Chem. Rev.* **2013**, *113*, 2343-2394.
- Zhang, Y.; Weckler, A. T.; Molina, P.; Deperalta, G.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 850-858.
- Zhang, Z.; Vachet, R. W. *Anal. Chem.* **2015**, *87*, 11777-11783.
- Zhou, Y.; Vachet, R. W. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 708-717.
- Zhou, Y.; Vachet, R. W. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 899-907.
- Zhu, M. M.; Rempel, D. L.; Du, Z.; Gross, M. L. *J. Am. Chem. Soc.* **2003**, *125*, 5252-5253.
- Zhu, M. M.; Rempel, D. L.; Gross, M. L. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 388-397.