Development and Applications of Chemical Labeling Protocols for Protein-Ligand Binding Analysis Using Bottom-Up Proteomics

by

Ying Xu

Department of Chemistry
Duke University

Date: ____________________

Approved:

___________________________
Michael C. Fitzgerald, Supervisor

___________________________
Jiyong Hong

___________________________
Terrence G. Oas

___________________________
Tuan Vo-Dinh

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry in the Graduate School of Duke University

2011
ABSTRACT

Development and Applications of Chemical Labeling Protocols for Protein-Ligand Binding Analysis Using Bottom-Up Proteomics

by

Ying Xu

Department of Chemistry
Duke University

Date:_______________________

Approved:

___________________________
Michael C. Fitzgerald, Supervisor

___________________________
Jiyong Hong

___________________________
Terrence G. Oas

___________________________
Tuan Vo-Dinh

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry in the Graduate School of Duke University

2011
Abstract

Proteins fold into well-defined three-dimensional structures to carry out their biological functions which involve non-covalent interactions with other cellular molecules. Knowledge about the thermodynamic properties of proteins and protein-ligand complexes is essential for answering fundamental biological questions and drug or biomarker discovery. Recently, chemical labeling strategies have been combined with mass spectrometry methods to generate thermodynamic information about protein folding and ligand binding interactions. The work in this dissertation is focused on the development and applications of two such strategies that couple chemical labeling protocols with mass spectrometry techniques including, SUPREX (stability of unpurified proteins from rates of H/D exchange) and SPROX (stability of proteins from rates of oxidation). The first part of this dissertation involves the application of SUPREX to the thermodynamic analysis of a protein folding chaperone, Hsp33, and its interaction with unfolded protein substrates. The second part of this dissertation involves the development of a new chemical labeling protocol that can be used to make protein folding and ligand binding measurements on proteomic scale.

In the first part of this work, the SUPREX technique was used to study the binding interaction between the molecular chaperone Hsp33 and four different unfolded
protein substrates including citrate synthase, lactate dehydrogenase, malate dehydrogenase, and aldolase. The results of the studies, which were performed at the intact protein level, suggest that the cooperativity of the Hsp33 folding/unfolding reaction increases upon binding with denatured protein substrates. This is consistent with the observation of significant burial of hydrophobic surface in Hsp33 when it interacts with its substrate proteins. The SUPREX derived $K_d$-values for Hsp33 complexes with four different substrates were also found to be all within a range of 3-300 nM. The interaction between Hsp33 and one of its substrates, citrate synthase (CS), was characterized at a higher structural resolution by using the SUPREX technique in combination with a protease digestion protocol. Using this protocol, the thermodynamic properties for both Hsp33 and CS were evaluated at different stages of binding, including reduced Hsp33 (inactive form), oxidized Hsp33 (active form), followed by native CS and Hsp33$_{ox}$–CS complexes before and after reduction with DTT. The results suggest that Hsp33 tends to bind unfolded proteins that still have significant amount of residual higher-order structure. Structural rearrangements of the substrate protein appear to occur upon reduction of the Hsp33-substrate complexes, which may facilitate the transfer of the substrate protein to other protein folding chaperone systems.

In the second part of this dissertation, a mass spectrometry-based covalent labeling protocol based on the amidination rate of globally protected protein amine
groups was designed and applied to the thermodynamic analysis of eight samples including: six purified proteins (ubiquitin, BCAII, RNaseA, 4OT, and lysozyme with, and without GlcNAc), a five-protein mixture comprised of ubiquitin, BCAII, RNaseA, Cytochrome C, and lysozyme, and a yeast cell lysate. The results demonstrate that in ideal cases, folding free energies of proteins and dissociation constants of protein-ligand complexes can be accurately evaluated using the protocol. Also demonstrated is the new method’s compatibility with three different mass spectrometry-based readouts including an intact protein readout using MALDI, a gel-based proteomics readout using MALDI, and an LC-MS-based proteomics readout using isobaric mass tags. The results of the cell lysate sample analysis highlight the complementarity of the labeling protocol to other chemical modification strategies that have been recently developed to make thermodynamic measurements of protein folding and evaluate protein stability on the proteomic scale.
## Contents

Abstract ......................................................................................................................................... iv

List of Tables ................................................................................................................................ xii

List of Figures ............................................................................................................................ xiii

Acknowledgements ................................................................................................................ xviii

1. Introduction ............................................................................................................................... 1
   1.1 Background ....................................................................................................................... 1
   1.2 Protein-Ligand Interactions by Mass Spectrometry .................................................... 4
   1.3 Covalent Labeling and Mass Spectrometry Based Technologies ......................... 7
      1.3.1 H/D Exchange: Continuous Amide H/D Exchange and SUPREX................. 9
      1.3.2 Protein Oxidation Reaction: Hydroxyl Radical-Mediated Footprinting and SPROX ..................................................................................................................................... 14
   1.4 Research Goals and Overview ...................................................................................... 19

2. General Protocols and Materials ........................................................................................... 23
   2.1 Materials .......................................................................................................................... 23
   2.2 General methods and Instrumentation for Mass spectrometers .................................. 24
   2.3 High-Sensitivity SUPREX protocol .............................................................................. 26
   2.4 SUPREX-digestion protocol .......................................................................................... 26
   2.5 SUPREX Data Acquisition and Analysis ..................................................................... 27
   2.6 Kd Value Determination ............................................................................................... 29
3. Thermodynamic Analysis of a Molecular Chaperone Binding to Unfolded Protein Substrates ................................................................................................................. 32

3.1 Introduction ..................................................................................................................... 32

3.2 Experimental Procedures .............................................................................................. 34

3.2.1 Materials ..................................................................................................................... 34

3.2.2 Protein Expression and Purification ....................................................................... 34

3.2.3 Chaperone Activity Measurements ........................................................................ 34

3.2.4 Sample Preparation ................................................................................................... 35

3.2.5 SUPREX Data Analysis and $K_d$ Value Determination .......................................... 36

3.3 Results .............................................................................................................................. 37

3.4 Discussion ........................................................................................................................ 40

4. Application of the SUPREX-Protease Digestion Protocol to Thermodynamic Analysis of Hsp33 Binding to its Unfolded Protein Substrate ................................................................. 49

4.1 Introduction ..................................................................................................................... 49

4.2 Experimental Section ..................................................................................................... 50

4.2.1 Sample Preparation and SUPREX-Protease Digestion Protocol .......................... 50

4.2.2. Peptide Mapping and Sequencing ......................................................................... 52

4.3. Results ............................................................................................................................. 54

4.3.1. Pepsin digestion ........................................................................................................ 54

4.3.2. Thermodynamic analyses of CS using a peptide readout ................................. 56

4.3.3. Thermodynamic analyses of Hsp33 using a peptide readout .............................. 59
4.4. Discussion ....................................................................................................................... 61

4.4.1. Prerequisites for substrate binding and releasing ................................................... 61

4.4.2. Involvement of intrinsically unfolded region in Hsp33 for substrate binding 63

4.5. Conclusions .................................................................................................................... 65

5. A Methionine Enrichment Protocol for SPROX Analysis of Protein-Ligand Binding on
   the Proteomic Scale .............................................................................................................. 75

5.1 Introduction ..................................................................................................................... 75

5.2. Experimental Section .................................................................................................... 77

5.2.1 Materials ..................................................................................................................... 77

5.2.2 Sample Preparation ................................................................................................... 77

5.2.3. Met-peptide Enhancement ...................................................................................... 79

5.2.4. LC-MS/MS-Based Proteomics ................................................................................. 81

5.2.5 Determination of protein expression level ............................................................ 82

5.3 Results and Discussion .................................................................................................. 82

6. Mass Spectrometry- and Lysine Amidination-Based Protocol for Protein Folding and
   Ligand Binding Analyses ....................................................................................................... 89

6.1 Introduction ..................................................................................................................... 89

6.2 Experimental Section ..................................................................................................... 91

6.2.1 Materials ..................................................................................................................... 91

6.2.2 SMTA Labeling Protocol .......................................................................................... 93

6.2.3 Intact Protein Readout by MALDI-TOF................................................................. 95
A.3.3 SUPREX Data Analysis and Ka Value Determination ........................................ 139

A.4 Results ........................................................................................................................... 140

A.5 Discussion..................................................................................................................... 143

References .......................................................................................................................... 155

Biography ............................................................................................................................ 170
**List of Tables**

Table 1: $C^{1/2}_{\text{SUPREX}}$ at different H/D exchange time ($t_{\text{H/D}}$) for Hsp33 and its complexes with different protein substrates. .................................................................47

Table 2: SUPREX-derived $\Delta G_f$, m-values and $K_d$ for Hsp33 and its complexes. ...............48

Table 3. Summary of the changes in thermodynamic stabilities for peptides from different stages of CS: native-state-CS (CS), CS complexed with Hsp33 (Com) and DTT-reduced CS-Hsp33 complex (DTTcom) ...................................................................................69

Table 4. Summary of the changes in thermodynamic stabilities for peptides from different stages of Hsp33: Hsp33$_{\text{red}}$ (Red), Hsp33$_{\text{ox}}$ (Ox), Hsp33-CS complex (Com) and DTT-reduced Hsp33-CS complex (DTTcom) .................................................................................. 73

Table 5: Thermodynamic parameters derived from intact protein analysis. All values derived as part of this work represent the average and standard deviation of three replicate measurements. ........................................................................................................122

Table 6: $C^{1/2}$ values determined for five peptides in the gel-based proteomic analysis of five-protein model mixture using SMTA labeling protocol. .................................................................127

Table 7: $C^{1/2}$ values determined in the LC/MS-based proteomic analysis for non-SMTA-modified-lysine peptides from an over-expressed yeast lysate sample modified by SMTA reagent. ........................................................................................................131

Table 8 $C^{1/2}_{\text{SUPREX}}$ at different H/D exchange time ($t_{\text{H/D}}$) for peptic peptides from Phb .151

Table 9: SUPREX-derived $\Delta G_f$ and m-values for Phb and its complexes. .........................153
List of Figures

Figure 1: Schematic representation of the high-sensitivity SUPREX protocol ................... 30

Figure 2: Schematic representation of the SUPREX-digestion protocol ..................... 31

Figure 3: Representative SUPREX curves obtained for Hsp33 (open circles) and its different complexes with (A) CS (filled circles), (B) MDH (filled circles), (C) Aldolase (filled circles) and (D) LDH (filled circles). Curves were obtained using the high sensitive SUPREX protocol when H/D exchange time was 15 minutes. Solid lines are the best fit of the data to a four-parameter sigmoidal equation in SigmaPlot, and the dotted arrows indicate the midpoint ($C_{1/2}^{SUPREX}$) extracted from each SUPREX curve ................. 45

Figure 4: Plots of $\Delta G_{app}$ versus $C_{1/2}^{SUPREX}$ for Hsp33 (open circles) and its different complexes with (A) CS (filled circles), (B) MDH (filled circles), (C) Aldolase (filled circles), and (D) LDH (filled circles). The resulting y-intercept and slope were taken as the $\Delta G$ and m-value, respectively. The H/D exchange time varied from 2 minutes to 45 minutes. The solid lines represent the linear least squares fitting of each data set. The correlation coefficient obtained for each data set was greater than 0.94. .................................................. 46

Figure 5: Schematic representation of Hsp33 network to prevent oxidative-induced protein aggregation. This includes: 1) Hsp33 is activated upon the oxidative conditions and binds with partially unfolded protein substrates (purple); 2) Once reducing condition restores, substrate is transferred to other chaperone proteins to be folded into its native structures and 3) Substrate-released Hsp33 returns to its inactive and reduced form .............................................................................................................................................. 66

Figure 6: Representative MALDI mass spectrum for peptic peptides obtained from A) native-CS, B) Hsp33-CS complex, C) DTT-reduced-Hsp33-CS complex, D) Hsp33 and E) mixture of CS and Hsp33. An ion signal at m/z equal to 2202 Da was zoomed in and shown as well. ............................................................................................................................. 67

Figure 7: Representative SUPREX curves for CS-peptides that can be group into three types of thermodynamic behavior when CS is at its different stages: native-CS (black lines or black filled circles), CS complex with Hsp33 (red lines or red filled circles) and DTT reduced CS complex (blue lines or blue filled circles). CS-peptides are A) 2202 Da (269-289), B) 3262 Da (38-69), C) 2673 Da (434-457), D) 1947 Da (290-309), E) 2486 Da
(310-330), F) 3129 Da (325-351), G) 3307 Da (102-131), H) 1965 Da (182-197), I) 1681 Da (141-155) and J) 6054 Da. The line represents the best fit of the data to a four-parameter sigmoid equation using SigmaPlot. The dotted line marks the $C^{1/2}$ value. H/D exchange time was 1 minutes for all the peptides with the exception of 2202 Da peptide (A, H/D = 5 minutes) and 1681 Da peptide (I, H/D = 20 minutes).

Figure 8. Thermodynamic stability changes in CS 2nd structure (A) and higher ordered structure (B) from PDB: 2CTS when peptides were from native-CS, CS-Hsp33 complex and reduced complex samples, respectively. Color labeling was based on the $\Delta C^{1/2}$ listed in Table 3. A start indicated that the standard deviation of $\Delta C^{1/2}$ at this region is relatively bigger (0.3 M) than the rest part of the protein. Gray color in B) indicated the region, from which no peptides were derived. CS helices that are most destabilized upon complex reduction are indicated by their letter codes (A, T, L, M, O and N).

Figure 9. Representative SUPREX curves for Hsp33-peptides from Hsp33<sub>red</sub> (black lines), Hsp33<sub>ox</sub> (pink lines), Hsp33-CS complex (red lines) and DTT-reduced-Hsp33-CS complex (blue lines). Hsp33-peptides are A) 2564 Da (26-47), B) 2142 Da (175-192) and C) 2112 Da (203-219). The line represents the best fit of the data to a four-parameter sigmoid equation using SigmaPlot. The dotted line marks the $C^{1/2}$ value. H/D exchange time was 1 minutes for 2142 Da peptide and 20 minutes for both 2112 Da and 2564 Da peptides.

Figure 10: Thermodynamic stability changes in Hsp33 2nd structure (A) and higher ordered structure (B) from PDB: 1VZY when peptides were from Hsp33<sub>red</sub>, Hsp33<sub>ox</sub>, Hsp33-CS complex and reduced complex samples, respectively. Color labeling was based on the $\Delta C^{1/2}$ listed in Table 4. A start indicated that the standard deviation of $\Delta C^{1/2}$ at this region is relatively bigger (0.3 M) than the rest part of the protein. Gray color in B) indicated the region, from which no peptides were derived.

Figure 11: Schematic representation of the SPROX in combination of solid-phase-based met-peptide-enrichment protocol.

Figure 12: Number of met-peptides and proteins detected by SPROX method with and without combination of met-enhancement protocol. Peptide and protein identification was used with false positive rate less than 1% for each peptide, higher than 1000 on summed iTRAQ intensity and no more than one zero intensity from each iTRAQ report ion.
Figure 13: Distribution of protein expression level from proteins detected using met-enhancement protocol (open squares), proteins detected without using met-enhancement protocol (filled squares), unique proteins that were detected only before (filled circles) and after (open circles) using the enrichment protocol and proteins (open triangles) that were commonly detected in both.

Figure 14: Schematic representation of the SMTA labeling protocol developed here.

Figure 15: Pseudo-first-order reaction curves obtained by reacting 0.4M SMTA with (A) des-Ac-a-Melanocyte stimulating hormone at different concentrations of GdmCl ranging from 1M to 6 M and (B) Ubiquitin in 6 M GdmCl. The lines represent the fitting of the data to a three-parameter single-exponential rise to max equation using SigmaPlot.

Figure 16: Representative data obtained using the SMTA-labeling protocol to analyze the thermodynamic stability of (A) ubiquitin, (B) zinc- loaded BCA II, (C) RNase A, (D) 4OT, (E) Apo-lysozyme, (F) lysozyme complex form with GlcNAc. The error bars represent +/- 1 standard deviation of the data determined from 10 replicate mass spectra. The lines represent the best fit of the data in each curve to a four-parameter sigmoid equation using SigmaPlot. The dotted lines mark the C^{1/2} value for each curve.

Figure 17: Representative SUPREX results on 4OT samples, which were reacted with the SMTA labeling reagent (0.4 M) at pH 7.4 in 20 mM phosphate buffer containing 0 and 3 M GdmCl, in (A) and (B), respectively. The line in (A) represents the best fit of the data to a four-parameter sigmoid equation using SigmaPlot. The dotted line marks the C^{1/2} value. The error bars represent +/- 1 standard deviation of the data determined from 10 replicate mass spectra.

Figure 18: SDS-PAGE gel obtained in the analysis of the five protein model mixture using the SMTA labeling protocol and the gel-based proteomics readout. Loaded in the 12 lanes are the protein samples that were labeled in the 12 denaturant containing buffers used in the SMTA labeling protocol.

Figure 19: Gel-based proteomics analysis of lysozyme (1-7) using the SMTA modification method. Representative MALDI mass spectra obtained are shown from our control experiment (A) and after SMTA modification in reaction buffers containing [GdmCl] of 0.8 and 6.1 M in (B) and (C), respectively. (D) The ΔMass_{wt,av} versus [GdmCl] plot for the typical sigmoidal curves was generated for this peptide. The error
bars represent +/- 1 standard deviation of the data determined from 10 replicate mass spectra. The dotted line marks the C\textsuperscript{1/2} value, which is listed in Table 6.

Figure 20: Representative data obtained using the SMTA labeling protocol to analyze the five protein mixture using the gel-based proteomics readout. Shown are the denaturation curves obtained for (A) the ubiquitin (25-34) peptide, (B) the RNase A(87-111) peptide, (C) the BCAII(204-236) peptide, and (D) the BCAII(204-236) peptide with an oxidized methionine. The lines represent the best fit of the data to a four-parameter sigmoid equation using SigmaPlot. The error bars represent +/- 1 standard deviation of the values determined from 10 replicate mass spectra. The dotted lines mark the C\textsuperscript{1/2} values which are summarized in Table 6.

Figure 21: Representative iTRAQ data obtained for 126 non-SMTA-modified-lysine peptides detected in yeast cell lysate. (A) A peptide yielding a denaturation curve with clearly defined pre- and post-transition baselines. (B) A peptide yielding iTRAQ data without a defined denaturation curve.

Figure 22: Distribution on normalized reporter ion intensity values from the lysine-containing peptides detected in the analysis of the yeast cell lysate sample using the LC-MS-based proteomics analysis with isobaric mass tags. The open and closed circles represented the data from the 113 and 121 tags, respectively. The dotted line marks the intensity cut-off value of 1.1 that was used to discriminate pre- and post-transition lines.

Figure 23: Pie chart with distribution of C\textsuperscript{1/2} values obtained for 112 non-SMTA-modified-lysine peptides detected in yeast cell lysate that yield well-defined denaturation curves. In cases where the transition midpoint analysis generated two possible C\textsuperscript{1/2} values, the higher value was used to in the pie chart distribution shown.

Figure 24: Crystal structure of phosphorylase b from rabbit muscle form (PDB: 3MQF)

Figure 25: Representative SUPREX curves obtained for peptic fragments from Ph\textsubscript{b}. (A) Peptide 6168 Da, (B) peptide 4505 Da and (C) peptide 4266 Da were obtained using the SUPREX-digestion protocol from Apo-Ph\textsubscript{b} (filled circles), AMP-Ph\textsubscript{b} with final concentration of AMP equal to either 0.5 mM (open circles) in SUPREX buffer and G6P-Ph\textsubscript{b} (filled triangles) with final concentration of G6P equal to 67 mM. The H/D exchange of each sample was performed with 4 minutes at room temperature. The lines represent
the best fit of the data in each SUPREX curve to a four-parameter sigmoid equation using SigmaPlot. The dotted lines mark the C^{1/2}_{\text{SUPREX}} value for each curve. .................. 150

Figure 26: Plots of $\Delta G_{\text{app}}$ vs. C^{1/2}_{\text{SUPREX}} for (A) peptide 6168, (B) peptide 4505, and (C) peptide 4266. The resulting y-intercept and slope were taken as the $\Delta G_i$ and $m$-value, respectively. These peptide fragments were obtained using the SUPREX-protease digestion protocol from Apo-Ph$b$ (filled circles), Ph$b$-AMP with a final AMP concentration of 0.5 mM (open circles), and Ph$b$-G6P with a final G6P concentration of 67 mM (filled triangles). H/D exchange times varied from 4 minutes to 225 minutes. The lines represent the results of our linear least-squares analysis of each data set. .................. 152
Acknowledgements

Looking back, I am amazed and at the same time very grateful for all I have received throughout these years. It has certainly shaped me as a person and led me to where I am now. All these years of my PhD studies are full of such gifts and it has been a wonderful and enjoyable journey.

First of all, special thanks to my advisor Dr. Michael Fitzgerald, for his invaluable advice and guidance. Investigation of my research projects was full of ups and downs, but problem-solving with Michael and sharing with him my findings are fully enjoyable. He enlightened me through his wide knowledge and deep intuitions. The enthusiasm for science that I inherited from him to a great degree leads to my success in my projects.

I would like to thank my collaborators, Dr. Ursula Jakob, Dr. Dana Reichmann and Sebastian Schmitt, not only for supplying me with my beloved protein, Hsp33, but also for their unreserved sharing of knowledge. A major part of the work presented in this dissertation could not have been accomplished without them. I would also like to thank my committee members, Dr. Jiyong Hong, Dr. Terrence G. Oas and Dr. Tuan Vo-Dinh for offering their guidance and valuable discussion in the development of this dissertation and monitoring my progress.
I would like to express my gratitude to friends and coworkers from Fitzgerald Lab. They made this experience not only enriching but also memorable: Erin Hopper, Liangjie Tang, Graham West, Patrick Dearmond, Duc Tran and Erin Strickland. I will miss them and be grateful for their friendship and companionship.

Last but not least, I would like to thank my husband Yang, who is always the first listener of my oral presentations and the first reader of my every single publication. And I would like to thank my son Andy whose smiles and hugs can always refresh me and wash away all my exhaustion and frustration. In addition, I would like to thank my parents Jiyang Xu, Lizhen Chen and my sister Yan Xu, for their unconditional love and always being there for me.
1. Introduction

1.1 Background

Proteins, as essential part of organisms, participate in almost every biological process within cells, such as cellular signaling, molecular transport, energy production, catalysis, proliferation and construction of body components. The vast majority of protein functions are achieved through protein-ligand interactions. Therefore, protein-ligand studies, including the identification of protein binding partners, the localization of protein binding sites, the determination of binding stoichiometry and the quantitation of interactions, are critical for answering fundamental biological questions and essential for drug or biomarker discovery.

Protein ligands cover a wide range of atoms and molecules that include metal atoms, vitamins, carbohydrates, drug molecules, peptides, proteins and nucleic acids. Protein-ligand binding events are usually very specific: one protein only recognizes a limited numbers of binding partners. Such specificities are mainly regulated by well defined three-dimensional structure of proteins. One such example is heat shock protein 33 (Hsp33) that is described in Chapters 4 and 5. Binding of Hsp33 with partially unfolded protein substrates requires the unfolding of its C-terminal domains. Vast majority of protein-ligand interactions are also defined by non-covalent interactions that
involve ionic bonds, hydrogen bonds and/or van der Waals forces. As a result, the association of protein binding partners to their target proteins is reversible and a dynamic equilibrium exists between protein-ligand complex (PL) and its components, protein (P) itself and ligand (L). The affinity of ligands to their protein targets is governed by the thermodynamic (i.e. equilibrium between the association and dissociation of protein-ligand complex) and kinetic (i.e. rate of the association and dissociation of the complex) properties of the protein-ligand interactions.

In the fundamental thermodynamic equation, the Gibbs binding free energy, $\Delta G_b$, is described by both enthalpy ($\Delta H$) and entropy ($\Delta S$): [4]

$$\Delta G_b = \Delta H_b - T \Delta S_b$$  \hspace{1cm} \text{Equation 1.1}

where $\Delta H_b$ is the energy cost for docking a specific ligand to protein binding sites and $\Delta S_b$ describes the dynamics of the interaction partners. T is the absolute temperature in Kevin. In any spontaneous process the free energy is minimized, $\Delta G_b$ is negative and implies that ligand association is more favorable. $\Delta G_b$ is equal to zero when dissociation and association of the protein-ligand complex reaches to equilibrium. In case of 1:1 complex, it can be expressed as:

$$P + L \leftrightarrow PL$$  \hspace{1cm} \text{Equation 1.2}
On the other hand, $\Delta G_b$ also can be expressed by the standard Gibbs binding energy, $\Delta G^\circ_b$, which is the free binding energy measured under defined standard conditions:

$$\Delta G_b = \Delta G^\circ_b + RT \ln Q$$  \hspace{1cm} \text{Equation 1.3}

In equation 1.3, $R$ is the gas constant and $Q$ is the reaction quotient that is expressed by the concentrations of the protein-ligand complex and its dissociated products:

$$Q = \frac{[PL]}{[P][L]}$$  \hspace{1cm} \text{Equation 1.4}

At equilibrium, $Q$ is equal to the association constant or binding affinity, $K_a$, thus equation 1.4 can be converted to

$$\Delta G^\circ_b = -RT \ln Q = -RT \ln K_a = RT \ln K_d$$  \hspace{1cm} \text{Equation 1.5}

Hereafter $\Delta G^\circ_b$ will be referred as the folding free energy. The binding affinity is typically used to evaluate if and/or how strongly two molecules will interact. Such relationships are the basic theoretical foundation for most affinity-based biophysical methods to detect and quantify protein-ligand interactions. A large number of techniques have been developed for the analysis of protein-ligand binding interactions. These techniques have been used to assess ligand-induced changes in different protein biophysical properties, such as protein unfolding/refolding, enthalpy, structure and protein mass to charge ratio. These techniques can be grouped into the following four categories based on the read-outs that they employ: 1) calorimetric methods like ITC
(isothermal titration calorimetry) and DSC (differential scanning calorimetry), 2) optical methods like DLS (dynamic light scattering), SPR (surface plasmon resonance) and DPI (dual polarization interferometry), 3) spectroscopic methods like CD (circular dichroism), NMR (nuclear magnetic resonance) and fluorescence and 4) spectrometric methods (mass spectrometry).

Different categories of methods described above have their advantages and disadvantages. It is necessary to take into consideration of research goals and the requirement of the protein systems, especially the amenable concentration range, protein functionalities and stability, before choosing which approach to use. [3] This dissertation is focused on the development and application of chemical labeling- and mass spectrometry-based methods described below.

1.2 Protein-Ligand Interactions by Mass Spectrometry

Two “soft” ionization methods, MALDI (matrix-assisted laser desorption/ionization) and ESI (electrospray ionization) are major MS ionization techniques for analyzing both intact and digested proteins. In MALDI, a laser beam is used to trigger the ionization. MALDI matrix consisting of organic compounds is used to protect sample molecules from being destroyed by the laser beam and facilitate their ionization. MALDI generated ions typically have few charge states than ESI (mainly
single charge) and as a result, interpretation of MALDI mass spectra are generally more straightforward than the interpretation of ESI mass spectra. Another advantage of MALDI is its fast sampling rate, ~10 seconds per sample, which makes it ideal for compound screening. However, the mass accuracy, precision and quantitative capabilities of MALDI are generally inferior to the ESI. Only a few approaches have been reported to provide quantitative measurements by using MALDI [6, 7] while quantitative analyses are typically done using ESI. In ESI, ions are generated from small and highly charged droplets in the presence of strong electric field. ESI prefers a sample free from salt and detergent. Both MALDI and ESI methods in conjunction with multiple mass analyzers can provide protein sequence information and molecular formula for small organic compounds. In addition, combination of mass spectrometry with different fractionation techniques enables us to examine protein sample in very complicated mixtures, such as cell lysate, plasma and tissues. Overall, the speed, sensitivity, quantitative capabilities and tolerance of sample complexity have made mass spectrometers particularly attractive for protein-ligand binding studies.[8]

As described above, protein-ligand interactions are dynamic processes with association and dissociation of the complex at equilibrium. The non-covalent nature of the complex renders them difficult to be directly detected in the mass spectrometry.
Fortunately, a growing number of publications indicated that such information can be obtained by using ESI-MS [9-12]. A few publications demonstrated results that this can be obtained by MALDI-MS [13, 14], which requires special precaution in sample preparation step of the analyses. Direct detection of the complex by mass spectrometry is attractive since it is label free, high throughput and low cost. However, a fundamental question still remains in debates: Does the protein-ligand complex in gas phase reflect its real assembly in the solution? Therefore, indirect methods still play a major role in protein-ligand binding studies by mass spectrometry.

One such indirect method has involved the use of chemical labeling strategies coupled with a MALDI- and/or ESI-based readout. Two types of techniques are generally used in the field: one utilizes cross-linking reagent to capture conformational states of the interacted proteins and thus probe protein-protein interactions; one relies on the modification rates of labeled amino acid residues from different positions to characterize protein in-solution-phase structures and protein unfolding/refolding properties in the absence and presence of ligands. The work in this dissertation is focused on the second technique using amide H/D exchange and protein oxidation reactions
1.3 Covalent Labeling and Mass Spectrometry Based Technologies

Previously reported experimental protocols for the second technique used chemical labeling reactions to label proteins and protein complexes as a function of time such as continuous H/D exchange and hydroxyl radical-mediated footprinting, [15-21] or as a function of chemical denaturant concentration such as SUPREX (stability of unpurified proteins from rates of H/D exchange) and SPROX (stability of proteins from rates of oxidation),[22-25] or as a function of temperature like thermal SPROX,[26] or as a function of ligand concentration such as PLIMSTEX (protein-ligand interaction by mass spectrometry, titration, and H/D exchange).[27, 28] Although they differ in data collection and analysis, all the chemical-labeling- and MS-based techniques monitor behaviors of proteins in solution phase.

Proteins and protein-ligand complexes are dynamic. They undergo conformational changes from extensively global unfolding, in which they completely lose their three-dimensional structures, to a subtle local breathing motion. At the same time, chemical labeling reagents react with exposed amino acid residues and add certain mass values to the monitored proteins. When proteins are in their transiently unfolded forms, more solvent-inaccessible reaction sites are exposed to the reagents and thus,
become labeled before protein native states are restored. Such chemical labeling mechanism can be described as:

\[
P_{\text{fold}} \xleftrightarrow[k_{\text{open}}]{k_{\text{close}}} P_{\text{unfold}} \xrightarrow[k_{\text{int}}]{k_{\text{close}}} \text{Labeled} P_{\text{unfold}} \xleftrightarrow[k_{\text{open}}]{k_{\text{close}}} \text{Labeled} P_{\text{fold}} \tag{1.5}
\]

where \( P_{\text{fold}} \) represents folded proteins and \( P_{\text{unfold}} \) represents unfolded proteins. \( k_{\text{open}} \) and \( k_{\text{close}} \) are the rate constants associated with protein unfolding and refolding reactions. \( k_{\text{int}} \) is intrinsic labeling reaction rate constant, which is usually pseudo-first order rate constant since most chemical labeling reagents’ concentrations are orders of magnitudes bigger than concentration of proteins.

The apparent labeling reaction rate (\( k_{\text{app}} \)) associated with both protein unfolding/refolding reaction and can be expressed by:

\[
k_{\text{app}} = \frac{(k_{\text{open}} \cdot k_{\text{int}})}{(k_{\text{open}} + k_{\text{int}} + k_{\text{close}})} \tag{1.6}
\]

In amide H/D exchange labeling methods, when a protein refolding rate (\( k_{\text{close}} \)) is much smaller than \( k_{\text{int}} \), a so-called EX1 exchange condition, equation 1.6 can be simplified as \( k_{\text{app}} = k_{\text{open}} \). As a result, every unfolding event generates deuterium-exchanged amide protons. Under this condition, mass spectrum collected shows two ion signals: one is from the folded and un-exchanged protein with a smaller mass value and the other is from the unfolded and exchanged protein with a bigger mass value. When a protein
refolding rate ($k_{cl}$) is much bigger than $k_{int}$, termed as EX2 exchange condition, equation 1.6 can be simplified as:

$$k_{app} = \frac{k_{op} \cdot k_{int}}{k_{cl}} = \frac{k_{int}}{K_f}$$  \hspace{1cm} \text{Equation 1.7}

where $K_f$ is the protein unfolding/refolding equilibrium constant. Under EX2 exchange conditions, only labeling reaction occurs after protein unfolding/refolding reaction and as a result, only one ion signal appears in the mass spectra. The folding free energy $\Delta G_f$ can be extracted from the derived equation 1.8

$$\Delta G_f = RT \ln K_f = RT \ln \left(\frac{k_{app}}{k_{int}}\right)$$  \hspace{1cm} \text{Equation 1.8}

Therefore, EX2 exchange condition is required to obtain the accurately determined protein folding free energy in an H/D exchange experiment and other labeling experiments. The difference in $\Delta G_f$ upon ligand binding is equal to the binding free energy for ligand-induced protein stabilization events and thus, $K_d$ values can be obtained using equation 1.3.

**1.3.1 H/D Exchange: Continuous Amide H/D Exchange and SUPREX**

Techniques based on amide H/D exchange coupled with mass spectrometry have been proven especially useful for ligand binding studies.[29-33] Such methods involve a reaction in which amide protons are replaced with solvent deuterons, resulting in one Dalton mass difference between each labeled and unlabeled site. The H/D exchange on
protons from amino acid side chain is rarely observed in the MS readout due to their either extremely fast (from side chains like Asn, Asp, Ser, Cys and Lys) or extremely slow (from side chains like His, Trp) H/D exchange rates. The intrinsic amide H/D exchange rate can be calculated either using SPHERE [34] program according to protein amino acid sequence, pH and temperature or using equation of \( k_{int} = 10^{pH - 5} \text{ min}^{-1} \) at room temperature when pH > 4. [22] At pH 2-3 and 0 °C, half-life for this exchange is over one hour. [35] Therefore, H/D exchange can be easily quenched by lowering the temperature and pH.

The most commonly used method in the area is continuous H/D exchange, a time-course experiment. In a typical time course experiment, a protein sample is allowed to undergo H/D exchange in its native form. The labeling reaction is quenched at each time point, and the protein is subsequently digested with pepsin, a protease with high activity under acidic conditions. The generated digest are then subject to MS analysis to determine the extent of H/D exchange by measuring the weight gain (\( \Delta \text{Mass} \)) at each time point until the weight gain reaches saturation. In some studies, instead of enzymatic digestion of H/D exchanged protein, protein sample can be directly fragmented inside the mass analyzer, typically in ESI-FTICR-MS. [36, 37] Ultimately, the determined \( \Delta \text{Mass} \) values are plotted against time points to obtain the relative rates of
the deuterium incorporation for each sequenced fragments. The protected amide protons have lower exchange rate than unprotected ones. At a given temperature and pH, such protection is mainly associated with hydrogen bond formation and the physical location of the amide protons in protein structure. [38] Therefore, H/D exchange rates are protein-structure-dependent and the rate on each sequenced protein fragments generated either by enzymatic digestion or FTICR-MS, provides information about protein in-solution conformation. The same experiment is performed on the ligand-bound protein, and comparisons are drawn between the uptake of deuterons for the apo and holo forms of the protein. Increased protection upon ligand binding results from either direct ligand binding on binding sites or ligand-induced conformational change, which could causes de-protection as well. Either way, those sites with changes in H/D exchange rate before and after ligand binding are involved in the protein binding events and could be important for protein fundamental studies and drug discovery.

SUPREX is another H/D exchange method that mainly uses MALDI to measure protein thermodynamic stabilities in the absence and presence of ligand. This is achieved by the use of chemical denaturant, such as guanidinium chloride (GdmCl) or urea to monitor protein global and subglobal unfolding/refolding reaction. By denaturing the protein, globally protected residues are deprotected as denaturant
concentration increases, leading to faster exchange of these amide protons and thus, producing a measurable shift in mass value, detected by MS analysis. After intact protein undergoes H/D exchange reaction under various chemical denaturant conditions, protein could be analyzed in either intact form or digested form, which is generated by incorporation of a quick digestion step, termed as SUPREX – protease digestion protocol. [39] The molecular weight increases on sample is then plotted as a function of denaturant concentration to generate typical sigmoidal curves. The transition from lower to higher Δ mass values reflects the transition of the protein from folded state to unfolded state as more amide protons become accessible for H/D exchange with increasing denaturant concentration. The slope of the transition indicates cooperativity of unfolding/folding reaction and transition midpoint indicates protein’s global (at intact protein level) or sub-global (at peptide level) thermodynamic stability. In order to obtain accurate measurements associate with protein folding, SUPREX requires these assumptions: (1) protein must exhibit reversible, two-state folding/unfolding behavior such that only the folded or unfolded states are significantly populated in the chemical denaturant-induced equilibrium of unfolding/refolding reaction; (2) H/D exchange must be under EX2 exchange conditions as discussed above and (3) unfolded protein should not contain any higher-order structure. However, for non-ideal proteins, SUPREX still
can be used to obtain accurately determined global thermodynamic information and/or protein-ligand binding affinities in combination with other approaches. For example, a pre-equilibrium protocol was reported to accurately obtain the thermodynamic stabilities for EX1 exchanged proteins.[40] For non-2-state folding protein systems SUPREX was used to quantitatively measure protein-ligand interactions with certain accuracy in the cases that the ligand interacts only with the native state of the protein and not with the partially folded intermediate state(s) that may also be populated. [23, 25, 41] In addition, combined with a surrogate ligand approach, which uses a known ligand to stabilize a protein and reduce local fluctuation, SUPREX can be used to analyze ligand binding with large (> 15 kDa), multidomain proteins that typically exhibit non-state-folding behaviors. [41] SUPREX has been used to evaluate the binding affinity of a wide range of protein-ligand complexes including proteins with small molecules, [41-44] peptides, [41, 44, 45] folded proteins [44, 46, 47] nucleic acids [44, 48] and protein folding intermediates [49] after its first report in 2000. [22]

The main difference between SUPREX and continuous H/D exchange is the use of chemical denaturant. SUPREX mainly probes global or sub-global protein unfolding/refolding reaction. Local fluctuations are not denaturant dependant under the H/D exchange condition in SUPREX. In contrast, continuous labeling H/D exchange
probes from global to local folding motions if labeling reaction is allowed to proceed for sufficient time. The second difference is the ionization techniques: MALDI is mainly used in SUPREX experiments and ESI is the major ionization method in continuous labeling H/D exchange. As a result, SPUREX is a more high-through-put method but continuous labeling H/D exchange has better sequence coverage (more than 80% compared with less than 50% in SUPREX experiments) that can probe more detailed conformational changes when protein is analyzed at the peptide level. In addition, continuous H/D exchange becomes a more automatic method with newly developed nanoACQUITY UPLC system with HDX technology from Waters Corporation (Milford, MA). However, since SUPREX experiments can be used with the combination of the enzymatic digestion, theoretically, the utility of ESI and UPLC system are able to improve the sequence coverage and automaticity in SUPREX experiment as well. In addition, SUPREX can provide quantitative measurements on protein-ligand interactions and protein folding events.

1.3.2. Protein Oxidation Reaction: Hydroxyl Radical-Mediated Footprinting and SPROX

Protein oxidation is another commonly used chemical labeling reaction (reviewed in reference [50]) Unlike the amide H/D exchange reaction that is reversible and, with time, the reversal of the reaction will make mass change more difficult to
detect, the intrinsic chemical stability of the oxidation reaction renders it compatible with commonly used peptide- and protein-fractionation techniques, leading to a maximum peptide resolution in the LC-MS experiment. In addition, the labeled sites can be directly identified using tandem MS in protein oxidation coupled with MS experiments. In contrast, deuterated sites are hard to determine in amide H/D exchange experiment due to the gas-phase scrambling of peptide protons and deuterons in collision cells to generate MS/MS spectra. [51] However, the main drawback of the protein oxidation is that oxidized proteins might have a different conformational property compared with their unoxidized species. It becomes significant when the local fluctuation events expose the buried labeling sites. The oxidation reaction can change the physical properties of these sites such as their polarity, hydrophobicity and charge state, leading to an incorrectly refolded and/or destabilized protein. Therefore, the reaction time needs to be optimized to be short enough to keep buried sites solvent-inaccessible from local fluctuation.

Hydroxyl radical-mediated footprinting experiment is very similar to continuous H/D exchange method by labeling protein native form as a function of time to assess the reaction rate at each modified site to probe in-solution protein structure with exception of the employment of protein oxidation as the labeling strategy. In contrast to targeting
protein amide protons in H/D exchange experiment, oxidation occurs at 14 amino-acid side chains with less specificity, and occasionally, cleaves protein backbone in hydroxyl radical-mediated footprinting. Its $k_{int}$ values range from $1.7 \times 10^7$ to $3.5 \times 10^{10}$ M$^{-1}$S$^{-1}$ [20] and the protein reaction time varies from microseconds to hours, depending on which method is used to generate hydroxyl radicals. The reaction needs to be limited into a short amount of time to minimize protein-backbone cleavages while maximize the oxidation of amino acid site chains.

Neither H/D exchange nor hydroxyl radical-mediated footprinting experiments to date are amenable to analysis of proteins in more complicated mixtures, such as cell lysates. The inherent lability of amide H/D exchange-based strategies renders them difficult to interface with the fractionation methods needed for the mass spectral detection and identification of proteins in complex mixtures. The non-reaction-specificity in the hydroxyl radical-mediated footprinting experiments yields multiple oxidation products from one peptide and thus complicates data interpretation on MS/MS spectra from complicated protein mixture. The extensive peptide maps that are critical to the success of continuous labeling experiments using either amide H/D exchange or footprinting are also difficult to generate on proteins in multi-component mixtures. The ability to characterize the conformational properties of proteins in multi-
component mixtures is both experimentally convenient and fundamentally important. The characterization of proteins in complex biological mixtures not only obviates the need for time-consuming protein purification steps, but it also affords the ability to study proteins in a more biologically relevant context than in their isolated form. One chemical modification-based and bottom-up proteomics-based method that has shown promise for the analysis of proteins in complex biological mixtures has been termed SPROX. [23, 24] SPROX is analogous to SUPREX except that the chemical denaturant-dependent oxidation rates of methionine side chains in a protein are used to determine the thermodynamic properties of the protein's global and/or subglobal unfolding/refolding reactions. [23] The oxidation reaction is initiated by adding H$_2$O$_2$ solution to protein samples and the $k_{\text{int}}$ reaction rate for exposed oxidation sites is around 0.04 S$^{-1}$ at pH 7.4 and room temperature. [23] The two most commonly oxidized sites are methionine residues followed by cysteine residues with relatively slower oxidation rate when they are solvent-accessible. Under the oxidation conditions, especially the optimized reaction time and H$_2$O$_2$ concentration, methionine is the mainly targeted residue observed in analysis of model protein systems [23] and a yeast cell lyaste study. [24] Oxidation reaction in SPROX experiments is more specific compared with hydroxyl radical-mediated footprinting. Like SUPREX, in order to accurately
measure the protein stability, SPROX requires: 1) a 2-state folder, 2) an EX2 condition (i.e. $k_{d} >> k_{in}$), 3) equal ionization efficiency on un-oxidized and oxidized species and 4) unchanged protein stability upon oxidation. Destabilization was observed in the studies of model protein systems.[23] However, ligand binding interactions with those destabilized protein still can be accurately measured when the partially oxidized protein have the same affinity to the ligand. [23]

SPROX was reported in a yeast cell lysate study in combination with bottom-up proteomics using tandem mass tags (TMT) strategy and MudPIT-LC-MS (multi-dimensional protein identification technology).[24] Bottom-up proteomics is a common method to identify proteins by their proteolytic digestion prior to mass spectrometry analysis. By comparing the masses of the proteolytic peptides and/or their tandem mass spectra with those predicted from a sequence database, peptides can be identified and multiple peptide identifications lead to protein identification. In this yeast cell lysate study, SPROX simultaneously assayed thermodynamic and ligand-binding properties on 886 peptides from 327 proteins. [24] A total of ten proteins appeared to be targets of the immunosuppressive drug, cyclosporine A, exhibiting a shift in thermodynamic stability upon ligand addition. [24] One of the hits is cyclophilin, a known CsA binding target. [24] More importantly, the binding affinity for each potential hit was also
provided in the assay. Globally profiling protein thermodynamic stability allows us to detect not only on-target protein stabilization, but also off-target effects leading to protein destabilization. This makes SPROX unique relative to other large-scale ligand-binding assays, such as the yeast two-hybrid assay [52-55] and affinity chromatography techniques coupled with mass spectrometry [56-59], which mostly detect on-target effects.

### 1.4 Research Goals and Overview

There were two main objectives to the work reported in this dissertation. The first was the application of a fully established protocol, SUPREX, to studies of a molecular chaperone protein binding with its partially unfolded protein substrates; the second objective was to develop SPROX and SPROX-liked techniques to facilitate thermodynamic analyses of proteins and protein-ligand complexes in complicated protein mixture.

Most proteins must fold into a well-defined three-dimensional structure in order to carry out their physiological roles within the cell. However, under different kinds of environmental stress conditions, such as elevated temperature, ultraviolet light, exposure to toxins (e.g., hypochlorous acid, arsenic, or trace metals) and infection, many proteins begin to lose their structure and function. Protein unfolding often leads to non-
specific aggregation, which is considered to be a largely irreversible process in vivo.
Molecular chaperones are a class of proteins that function in the cell to recognize and
selectively bind non-native proteins to prevent protein aggregation and facilitate their
refolding. It is generally believed that chaperones interact with unfolded protein
substrates via hydrophobic interactions (for a review, see ref. [60]). For example, peptide
binding studies using a number of different chaperones revealed that most show a
greater preference toward hydrophobic peptides than charged, hydrophilic peptides.
[61-63] Fluorescent hydrophobic probes have also been used to show that chaperones
bind unfolded protein substrates upon exposure of hydrophobic surfaces. [64-66]
Finally, analysis of several X-ray crystallographic structures of chaperones indicated
they preferentially bind to hydrophobic residues. [63, 67, 68] While there is increasing
amount of knowledge about the basic information on molecular interactions that define
chaperone-substrate interactions, much less is known about the binding affinities and
detailed chaperone-substrate interactions. A few studies published on selected
chaperone-substrate complexes have reported binding affinities in the range of 1-50 nM.
[69-71] However, it is not clear if the binding affinities of a given chaperone to different
substrate are the same and how chaperone proteins interact with their partially unfolded
protein substrates. These questions are addressed in Chapters 3 and Chapter 4
respectively in this dissertation. Evaluated in Chapter 3 is the binding affinity of the molecular chaperone Hsp33 with four different denatured protein substrates including citrate synthase (CS), lactate dehydrogenase (LDH), malate dehydrogenase (MDH), and aldolase. In Chapter 4, Hsp33 with and without CS were examined at the peptide level in the aspect of both Hsp33 and CS using a SUPREX – protease digestion protocol.

As mentioned above, SPROX in combination with bottom-up proteomics promises to provide an attractive way to study the protein-ligand interactions in complicated biological mixtures. However, the SPROX technique also has its limitations. It involves the reaction of globally protected methionine residues, which, according to the SWISS-PROT database[72], have an approximate 2.37% of natural occurrence on protein sequence. As a result, the possibility, of which a 10-amino-acid peptide contains one methionine residue detected in a typical shotgun experiment, is 23.7%. This might be the reason that only 327 protein, a small fraction of the yeast protein, were detected using SPROX methods reported in the reference [24]. Therefore, it is necessary to combine SPROX with other techniques to increase the number of protein that can be identified. A methionine enrichment protocol in combination with SPROX method is described in Chapter 5 to isolate methionine containing peptides after intact proteins are oxidized in SPROX experiment, digested and labeled with iTRAQ reagent. Isolation of
methionine-peptides reduces sample complexity and improves the efficiency in the LC-MS analysis for detection of met-peptides. In Chapter 6, a SPROX-liked technique is developed to probe additional amino acids, free amine groups from lysine residues and N-termini, to study thermodynamic stability of proteins and protein-ligand complexes.
2. General Protocols and Materials

2.1 Materials

The following reagents were purchased from Sigma-Aldrich (St. Louis, MO): deuterium oxide (99.9% atom D), sodium deuteroxide (40 wt% in D_2O, 99.5% atom D), deuterium chloride (35 wt% in D_2O, 99% atom D), porcine pepsin, sequencing-grade trypsin, insulin from bovine pancreas, trypsin inhibitor from bovine pancrease, aldolase from rabbit muscle, bradykinin, S-methyl methanethiosulfonate (MMTS), tris(2-carboxyethyl) phosphine hydrochloride (TCEP-HCl), and MALDI matrix: Sinapic acid (SA) and α-Cyano-4-hydroxycinnamic acid (HCCA). Guanidine hydrochloride (GdmCl) was purchased from EMD (Gibbstown, NJ), and trifluoroacetic acid (TFA) was purchased from Halocarbon (River Edge, NJ). Acetonitrile (ACN) was purchased from Fisher (Fair Lawn, NJ), and deuterated phosphoric acid was from Cambridge Isotope Laboratories (Andover, MA).

A series of deuterated H/D exchange buffers (typically 20 mM phosphate, pD 7.4) was prepared with GdmCl concentration ranging from 0.3 to 7.4 M. Before used in SUPREX buffers, GdmCl and urea were deuterated through four cycles of dissolution in D_2O and lyophilization. The pD of each buffer was adjusted with sodium deuteroxide and deuterium chloride, and pH measurements were converted to pD measurements by adding 0.4 to the pH reading.[73] The final concentrations of GdmCl or urea in the
buffers were measured with a refractometer (Bausch and Lomb, Rochester, NY) as previously described.[74]

2.2 General methods and Instrumentation for Mass spectrometers

MALDI matrix solutions used contain 0.1% TFA, 45% ACN and two saturated matrix: HCCA for peptides smaller than 4 kDa and SA for peptides or proteins larger than 4 kDa. Three pairs of proteins were used as internal mass calibrants for MALDI analyses: insulin and bradykinin were used when HCCA matrix was utilized: insulin and trypsin inhibitor when SA matrix was utilized and sample’s size is smaller than 20 kDa; trypsin inhibitor and aldolase were used in SA matrix for proteins larger than 20 kDa. Internal mass calibrants were added directly to the matrix solution. MALDI mass spectra were collected on an Ultraflex II TOF/TOF mass spectrometer (Bruker Daltonics, Billerica, MA) equipped with Nd:YAG laser and 100 laser scans were summed. The following instrument parameters were used to obtain the average mass values of protein or peptide samples: ion source 1 voltage of 25.0 kV, ion source 2 voltage of 23.4 kV, a lens voltage of 6.5kV, and pulsed ion extraction of 100-130 ns.

For peptide sequencing performed on Bruker TOF-TOF, the instrument was externally calibrated using a calibration kit provided by Bruker Daltonics immediately before the peptide sequence analysis. The mono-isotopic mass values (MS1) for peptides of interest were determined with a reflector mode performed with the following
parameters: ion source 1 voltage of 25 kV, ion source 2 voltage of 21.55 kV, lens voltage of 9.55 kV, reflector 1 voltage of 26.30 kV, reflector 2 voltage of 13.77 kV and pulsed ion extraction of 30 ns. MS/MS signals for each selected peptide ion were obtained in a lift mode with parameters as follows: ion source 1 voltage of 8.03 kV, ion source 2 voltage of 7.23 kV, lens voltage of 3.60 kV, reflector 1 voltage of 29.63 kV, reflector 2 voltage of 13.86 kV, lift 1 voltage of 19.09 and lift 2 voltage of 3.60 kV.

LC-MS/MS analyses were performed on an Agilent 6520 Q-TOF LC/MS/MS mass spectrometer system equipped with a Chip Cube Interface. The following instrumental parameters were used: 350 ºC gas temperature, 6l/min drying gas, 175 V fragmentor, 65 V skimmer, 750 V OCT IRF V_pp, 3 spectra/s MS acquisition, and 2 spectra/s MS/MS acquisition. The maximum number of precursor ions selected per cycle was 3 for the analyses performed on peptides from the model protein mixture and 4 for the analyses performed on peptides from the cell lysates. A slope of 3.9 V/(100 Da) with -4.8 V offset was used for the collision energy in the analysis of peptides from the model protein mixture, and a slope of 3.9 V/(100 Da) with a 2.9 V offset was used for the collision energy in the analysis of the iTRAQ treated yeast cell lysate sample. An Agilent Chip with a 43 mm x 75 µM column packed with Zorbax 300SB-C18 (5 µM) was used in the LC separation of the peptides in the five-protein mixture, which employed a 7 minute gradient elution from 3-70 % buffer B. An Agilent Chip with a 150 mm x 75 µM column
packed with Zorbax 300SB-C18 (5 µM) was used for the LC separation of the peptides from the yeast cell lysate, which employed an 85 minute elution gradient from 2 - 45% buffer B, followed by a 10 minute gradient from 45 - 100 % buffer B. The LC flow rate in all the LC-MS/MS analyses was 400 nL/min.

### 2.3 High-Sensitivity SUPREX protocol

A high-sensitivity SUPREX protocol used to monitor intact protein ion signals using a MALDI readout was outlined in Figure 1 as described previously. [75] Briefly, protein samples were diluted 10-fold into 10 µl of each GdmCl-containing SUPREX buffer. After the specified H/D exchange time at room temperature, the H/D exchange reactions were quenched by adding TFA to a final concentration of 0.3% (v/v). A C18 ZipTip™ (Millipore Corporation, Billerica, MA) was used to desalt and concentrate individual protein samples according to the manufacturer’s instructions. The protein was eluted from the ZipTipTM with 10 µL saturated SA matrix prepared in aqueous buffer containing 45% v/v ACN and 0.1% v/v TFA. A 1 µL aliquot was spotted on a pre-deposited microcrystalline layer. All the steps after specified H/D exchange time need to be performed as quickly as possible to minimize back exchange of deuterons to protons.

### 2.4 SUPREX-digestion protocol

Incorporation of a protease digestion step into the SUPREX protocol (Figure 2) has been described previously.[76] Before SUPREX experiments, protein digestion
condition were typically optimized on an undeuterated sample by varying protease (i.e. pepsin) concentration, and digestion time to generate the greatest number of peptides in 1-10 kDa range while minimize back exchange issue by using the shortest digestion time at the same time. The specific protease digestion conditions were described in respect chapters. In SUPREX experiment, H/D exchange reactions were initiated by adding 1 μL equilibrated protein complex to 9 μL of each SUPREX buffer. After specified exchange time, the H/D exchange was quenched with HCl. A pepsin solution was immediately added to each quenched protein solution. The protease digestion was allowed to proceed for certain time period. Ultimately, 1 μL of the digestion solution was added to 9 μL matrix solution containing internal mass calibrants, and 1 μL of the resulting solution was spotted onto a MALDI target for mass spectral analysis.

2.5 SUPREX Data Acquisition and Analysis

MALDI-TOF analyses were performed to determine deuteron uptake (i.e., Δmass) in each sample. An in-house Microsoft Excel macro script was used to determine the mass-to-charge-ratio for MALDI ion signals from either intact protein or digested protein fragment following: 1) a 9-point floating average smoothing of peptide data (19-point floating for large sized protein data), 2) a two-point calibration using the internal standards, and 3) a center of mass calculation for the peak of interest. The Δmass value for each sample was calculated by subtracting its undeuterated mass from its deuterated
mass. Ten spectra were collected at each denaturant concentration, and the Δmass values derived from these spectra were averaged. The resulting Δmass_{avg} values were subsequently plotted against the GdmCl concentration to generate the SUPREX curves. [77] Data was fit to a four-parameter sigmoidal equation using a nonlinear regression routine in SigmaPlot (Systat Software, Inc., San Jose, CA), and this fitting was used to obtain C_{1/2}^{SUPREX} value, which was the denaturant concentration at the transition midpoint from each SUPREX curve.

SUPREX curves were collected at several exchange times if protein folding free energy (ΔG_f) and cooperativity (m-value) are required. The C_{1/2}^{SUPREX} values determined at the different H/D exchange times were used in Equation 2.1, the derivation of which is described elsewhere [45], to determine an m- and ΔG_f values for protein samples.

\[
-RT \ln \left( \frac{\frac{<k_{int}>}{0.693} t - 1}{n^n \frac{n!}{2^{n-1}} [P]^{n-1}} \right) = mC_{1/2}^{SUPREX} + ΔG_f
\]

Equation 2.1

In equation 2.1, R is the gas constant, T is the temperature in Kelvin, <k_{int}> is the average intrinsic exchange rate of an unprotected amide proton, which is typically 251 min^{-1} at pH 7.4 as calculated using the relationship <k_{int}> = 10^{pH-5} min^{-1} [77], t is the H/D exchange time, m is defined as δΔG_f/δDenaturant, ΔG_f is the folding free energy of the protein in the absence of denaturant, n is the number of protein subunits involved in the
folding reaction. $[P]$ is the protein concentration expressed in n-mer equivalents. The left side of the equality in equation 2.1 will hereafter be referred to as $\Delta G_{\text{app}}$.

Plots of $\Delta G_{\text{app}}$ vs. $C_1^{\text{SUPREX}}$ were generated for proteins. These plots were fit to Equation 2.1 using a linear regression routine, and the y-intercept and slope were taken as the $\Delta G_i$ and $m$-value (i.e., $\delta \Delta G_i / \delta [\text{denaturant}]$), respectively. The difference in folding free energy, $\Delta \Delta G_{\text{avg}}$ values which is also the binding free energy, before and after ligand binding to protein can be used to calculate for dissociation constants, $K_d$ value, which are commonly used to describe the affinity of a ligand binding with proteins.

### 2.6 Kd Value Determination

The dissociation constant $K_d$ value was calculated using equation 2.2. [78]

$$K_d = \frac{4L_{\text{total}}e^{-\Delta \Delta G / NRT} - 4P_{\text{total}}(e^{-\Delta \Delta G / NRT} - 1)}{(2e^{-\Delta \Delta G / NRT} - 1)^2 - 1}$$

Equation 2.2

In equation 2.2, $N$ is the number of independent and equivalent ligand-binding sites to protein, $[L]$ is the free ligand concentration, $\Delta \Delta G_i$ is the binding free energy, $P_{\text{total}}$ is the protein concentration expressed in n-mer equivalents and $R$ and $T$ are as defined above. When the ligand was more than a 10-fold excess protein concentration, the equation 2.2 can be simplified as

$$K_d = [L] / (e^{\Delta \Delta G / NRT} - 1)$$

Equation 2.3

The free ligand concentration, $[L]$, was taken as the total ligand concentration.
Figure 1: Schematic representation of the high-sensitivity SUPREX protocol

1. Dilute into SUPREX buffers for H/D exchange reaction

2. Add TFA (final concentration = 0.3-1%) to quench the reaction

3. Bind protein to C18 resin

4. Wash ziptips with 0.1% TFA

5. Elute protein with MALDI matrix containing 50% ACN

6. Spot eluted samples on a pre-deposited microcrystalline layer of MALDI matrix containing internal mass calibrants.

7. Apply to MALDI analysis to determine mass values
Figure 2: Schematic representation of the SUPREX-digestion protocol
3. Thermodynamic Analysis of a Molecular Chaperone Binding to Unfolded Protein Substrates

3.1 Introduction

Hsp33 functions as an ATP-independent molecular chaperone, which binds to a large number of unfolding proteins *in vivo* as well as to commonly studied chaperone substrate proteins, such as citrate synthase *in vitro* [79]. In its reduced form, Hsp33 is monomeric and inactive. The four conserved cysteine residues, located in the C-terminus of Hsp33, coordinate one zinc ion with very high affinity and constitute the so-called redox-switch domain of Hsp33. Upon oxidative activation of Hsp33, the four cysteine residues form two intramolecular disulfide bridges, and release their zinc ion. [80, 81] Disulfide bond formation and zinc release induce large conformational rearrangements in Hsp33’s C-terminus and cause Hsp33 to adopt a partially unfolded conformation. Two oxidized Hsp33 monomers then dimerize, and Hsp33 acquires its chaperone activity in which one Hsp33 dimer is thought to interact with one unfolded polypeptide substrate. [80, 82] Unfolding of Hsp33’s C-terminal domain appears to be essential for the exposure of large hydrophobic surface on the *N*-terminus of Hsp33, which has been suggested to serve as binding site for unfolded substrate proteins. [64] Light scattering experiments suggested that activated Hsp33 can inhibit aggregation of thermally unfolding citrate synthase (CS), malate dehydrogenase (MDH), lactate dehydrogenase (LDH), and aldolase. These four proteins not only differ in size,
oligomerization status and isoelectric point but also in the temperature that causes their unfolding and aggregation, which ranges from 43°C (CS) to 65°C (aldolase). [49]

Reported here is the quantitative in *vitro* analysis of Hsp33 binding interactions with CS, LDH, MDH and aldolase. One motivation for this is to determine if the binding affinities of a given chaperone to different substrate are the same or different. The experimental strategies employed to date, for evaluating the binding affinities of protein chaperones to their unfolded protein substrates have involved H/D exchange and NMR based methods,[69] fluorescence spectroscopy [70] and surface plasmon resonance. [71] Utilized here is an H/D exchange- and mass spectroscopy-based method, termed SUPREX (stability of unpurified proteins from rates of H/D exchange) to measure the binding affinity of Hsp33 to its protein substrates. In SUPREX, the denaturant dependence to the amide H/D exchange reaction is determined in order to obtain a measure of the thermodynamic parameters associated with the global unfolding/refolding reaction of the protein under study, the Hsp33 protein in this work. The thermodynamic parameters measured in the absence and in the presence of ligands (the unfolded protein substrates in this work), can ultimately be used to generate thermodynamic information about the binding affinity. This study is the first application of SUPREX to the thermodynamic analysis of chaperone-unfolded protein interactions.
3.2 Experimental Procedures

3.2.1 Materials

Soybean trypsin inhibitor and aldolase from rabbit muscle were purchased from Sigma-Aldrich (St. Louis, MO). Citrate synthase (CS) from pig heart, malate dehydrogenase (MDH) from pig heart and lactate dehydrogenase (LDH) from hog muscle were purchased from Roche (Indianapolis, IN).

SUPREX buffers were comprised of 20 mM phosphate (pD 7.4) containing different concentration of GdmCl.

3.2.2 Protein Expression and Purification

Purified Hsp33\textsubscript{ox} was provided by Dr. Ursula Jakob lab. Briefly, wild type Hsp33 and reduced inactive Hsp33\textsubscript{red} were prepared according to a protocol previously described [64]. Subsequently, to activate Hsp33, 50 μM Hsp33\textsubscript{red} was incubated with 500 μM NaOCl for 1 h at 30°C in 40 mM potassium phosphate buffer under constant shaking at 300 rpm. Excess NaOCl was removed using PD-10 desalting columns. The reduced Hsp33\textsubscript{red} and oxidized Hsp33\textsubscript{ox} were concentrated and stored at -20°C.

3.2.3 Chaperone Activity Measurements

To determine the activity of Hsp33\textsubscript{ox}, light scattering experiments were conducted by Sebastian Schmitt. All substrate proteins (CS, MDH, LDH, aldolase) were incubated in 40 mM HEPES, pH 7.5 in the absence or presence of Hsp33 with final concentrations
ranging from 100 nM (for aldolase) to 170 nM (for CS) at their individual temperatures (43°C for CS, 47°C for MDH, 57°C for LDH and 65°C for aldolase) that causes their unfolding and aggregation. Light scattering was monitored at $\lambda_{ex}/\lambda_{em}$ of 350 nm using a Hitachi F4500 fluorescence spectrophotometer equipped with thermostated cuvette holder and stirrer.

3.2.4 Sample Preparation

Concentrated solutions of the substrate proteins CS (376 μM in 50 mM Tris-HCl, 2 mM EDTA, pH 8.0), MDH (312 μM in 40 mM KH$_2$PO$_4$, pH 7.5), LDH (457 μM in 40 mM KH$_2$PO$_4$, pH 7.5), and aldolase (315 μM and 10 mM in ddH$_2$O) were prepared and centrifuged (13,000 rpm, 45 min, 4°C) to pellet already formed aggregates. The stock concentration of aldolase and LDH were determined using a Bradford Assay [83]. The concentration of CS and MDH were determined using their respective extinction coefficient at 280 nm: 1.78 ml/(mg cm) [84] and 0.85 ml/(mg cm), respectively. [85]. CS was stored at –80°C, while MDH, LDH and aldolase were stored at –20°C.

A solution of activated Hsp33$_{ox}$ (dimeric concentration of 216 μM) was pre-warmed in an Eppendorf tube at either 43°C (for CS), 47°C (for MDH), 57°C (for LDH) or 65°C (for aldolase). Concentrated stock solutions of the indicated substrate proteins were added to Hsp33$_{ox}$ in a step-wise fashion at the indicated temperatures under constant shaking at 330 rpm, until an Hsp33 dimer to substrate protein ratio of 1:1.5 was
reached (final Hsp33 dimer concentration ranging from 114 μM to 124 μM). During the titration, the solutions remained clear until a molecular ratio of 1:1 (Hsp33 dimer: substrate protein) was reached. Addition of excessive substrate proteins resulted in turbidity due to protein aggregation, indicating that the majority of Hsp33 substrate binding sites were saturated with substrate proteins. These aggregates were removed by centrifugation (13,000 rpm, 45 min, 4°C) and the supernatant was directly used for further experiments.

3.2.5 SUPREX Data Analysis and $K_d$ Value Determination

A high-sensitivity SUPREX protocol [75] followed by SUPREX data analysis were used in this study as described in Chapter 2. Equation 2.1 was used to determine the thermodynamic parameters, $\Delta G_f$ and $m$- values. In equation 2.1, Hsp33$_{ox}$ was treated as a dimer ($n = 2$). In our binding studies of Hsp33, the protein substrates were present in less than a 10-fold excess over Hsp33 concentration. Therefore, dissociation constants, $K_d$ values, were calculated using equation 2.2. In equation 2.2, $L_{total}$ is the final concentration of ligand in the H/D exchange reactions, $P_{total}$ is the concentration of Hsp33 dimer, $\Delta \Delta G_{f,avg}$ is the difference in $\Delta G_{f,avg}$ between the bound and unbound forms of the protein, and $N$ is the number of independent equivalent binding sites ($N = 1$ as expected for the Hps33 dimer).
3.3 Results

The stoichiometry of Hsp33 binding has been previously tested using thermally unfolded luciferase [86]. The results of these studies suggested that active Hsp33-dimers bind to substrate proteins in a 1:1 stoichiometry. Thus, the Hsp33-substrate complexes were prepared as ~1:1 complexes. No significant aggregation was visible at this 1:1 ratio of Hsp33 dimers to CS monomers. Additional titration of substrate proteins caused visible aggregation, suggesting that the Hsp33 dimer was saturated with substrate protein.

The release of protein substrates from Hsp33 appears to require the presence of reducing conditions and a functional DnaK-system (32). Thus, the pre-formed complexes between Hsp33 and the substrate proteins were expected to be sufficiently stable at room temperature, which was the temperature at which subsequent SUPREX analyses were performed. Shown in Figure 3 are representative SUPREX curves obtained for Hsp33 alone and Hsp33 complexed with the four different unfolded substrate proteins, CS (Figure 3A), MDH (Figure 3B), aldolase (Figure 3C) and LDH (Figure 3D). The midpoints of the SUPREX transitions (i.e., $C_{1/2}^{SUPREX}$ value) were shifted to a higher denaturant concentration when Hsp33 was in complex with the substrate proteins (Figure 3, compare open and closed circles). Such a $C_{1/2}^{SUPREX}$ shift is consistent with the ligand-induced stabilization of Hsp33. The $C_{1/2}^{SUPREX}$ are summarized in Table 1.
A total of four SUPREX curves were acquired on each Hsp33-substrate complex using different H/D exchange times that ranged from 2 minutes to 45 minutes. The \( C_{1/2}^{\text{SUPREX}} \) values for the protein and protein ligand complexes at the different H/D exchange times are summarized in Table 1. The H/D exchange times in this study were chosen such that they were (i) short enough to yield curves with sufficient amplitude to clearly distinguish pre-transition baseline values that were distinct, and (ii) long enough to yield significant \( C_{1/2}^{\text{SUPREX}} \) values shifts (i.e., greater than 0.1 M). This allowed for accurate determinations of \( \Delta G_f \) and m-values, which involved generating plots of \( \Delta G_{\text{app}} \) vs. \( C_{1/2}^{\text{SUPREX}} \) according to Equation 2.1 (Figure 4). Linear least squares analysis of the data in the resulting plots (Figure 4) indicated that the correlation coefficients for all the \( \Delta G_{\text{app}} \) vs. \( C_{1/2}^{\text{SUPREX}} \) plots were greater than 0.94. The \( \Delta G_f \) and m-values extracted from each plot are summarized in Table 2.

The SUPREX derived m-values for Hsp33 in the presence of the four different protein substrates were all larger than the m-value measured for Hsp33 in the absence of the substrate. Application of Peirce’s criterion[87] to the m-values in Table 2 confirmed that the apo-Hsp33 m-value in Table 2 could indeed be considered a statistically significant outlier compared to the other m-values in Table 2. The increased m-values obtained for the Hsp33 complexes imply that the Hsp33 protein folding/unfolding reaction becomes more cooperative in the presence of substrate, which is consistent with
large conformational changes in Hsp33 upon substrate binding. The binding interaction appears to involve the burial of hydrophobic surface area in the Hsp33 dimer and thus increased m-values. Based on the data in Myers et. al.[88], the average m-value increase observed in our experiments, ~1 kcal/(mol M), would correspond to ~600 Å² of hydrophobic surface area being buried in the Hsp33 dimer upon ligand binding.

A quantitative determination of ΔΔGᵩ values requires an accurate and precise m-value determination. Since the m-values of four different Holo-Hsp33 complexes were determined to be within the experimental error of each other, an average m-value for Hsp33 in its ligand bound state was determined from the Holo-Hsp33 m-values in Table 2. This average m-value was used in equation 2.1 to calculate replicate ΔGᵩ values directly from the C¹/²SUPREX values measured in this work. We have previously shown that this approach can greatly improve the accuracy and precision of ΔGᵩ values by SUPREX [41, 42]. The re-calculated ΔGᵩ values (i.e., the ΔGᵩavg, values) are shown in Table 2, along with the ΔΔGᵩavg, values calculated for each complex.

The ΔΔGᵩavg, values in Table 2 were used to determine the dissociation constant (i.e., Kᵩ) of each complex (Table 2). The Kᵩ values in Table 2 are reported as a range due to the uncertainty of the concentration of the protein substrates. As described in the Experimental Procedures Section, the Hsp33 complexes studied in this work were prepared by adding substrate to a heated Hsp33 solution and then separating the
soluble protein-complex from any insoluble substrate aggregates in a centrifugation step. Therefore, it was difficult to calculate the free ligand concentration due to the unknown amount of the discarded aggregates. The titration was stopped when the substrate-to-Hsp33 dimer concentration ratio was 1.5 to 1. Since each polypeptide chain of the unfolded protein substrate is expected to bind to one dimer of the Hsp33 molecule, the total ligand concentration used in the $K_d$ calculations was assumed to be as low as 1 and as high as 1.5 times that of the Hsp33 dimer concentration. It is noteworthy that only one population of H/D exchanged Hsp33 molecules was detected in the MALDI-readout of our SUPREX experiments on Hsp33. This is consistent with the Hsp33 dimer being saturated with protein substrates in our experiments, suggesting that an Hsp33 dimer does indeed bind substrate and that the ligand concentration in our experiments was at least that of the Hsp33 dimer.

3.4 Discussion

To obtain accurate $\Delta G_f$ and $m$-values for a protein folding reaction by SUPREX experiments, the protein under study must (i) have a two-state folding reaction (i.e., protein folding intermediates are not populated and only the unfolded or folded states of the protein dominate during the folding reaction) and (ii) exhibit EX2 exchange behavior (i.e., the folding reaction must be significantly faster than intrinsic exchange rate of an unprotected amide proton). To our knowledge, it is unknown whether the
folding reaction of Hsp33 is two-state. However, given its relatively large size and multi-domain structure, it is unlikely that folding of Hsp33 follows a two-state process. Therefore, the thermodynamic parameters reported here do not represent absolute thermodynamic values for the folding reactions of Hsp33. However, our interest is to measure the difference in folding free energy before and after ligand binding. If it is assumed that the partially folded intermediate state(s) populated in the folding reaction of the apo-protein are the same as those populated in the folding reaction of the holo-protein, then the SUPREX-derived folding free energy values can be used to obtain accurate binding free energies (i.e., $\Delta G_{\text{avg}}$ values). Also, if the protein is under non-EX2 (so-called EX1) exchange behavior in our SUPREX experiments, our previous studies demonstrate that the theoretical error introduced into the SUPREX analysis by assuming EX2 exchange was relatively small and can be ignored compared with the experiment error [89].

The SUPREX derived $\Delta G_f$ and $m$-values for Hsp33 were significantly increased upon ligand binding (shown in Table 2). The increased $\Delta G_f$ values for the Hsp33 complexes with CS, MDH, aldolase and LDH indicated that Hsp33 binds these unfolded proteins and forms a thermodynamically stable protein-substrate complex. These results are consistent with previous findings that Hsp33-substrate complexes are apparently very stable and require reducing conditions for their dissociation[86]. The increased $m$-
values suggest moreover that hydrophobic surface areas in Hsp33 are buried upon ligand binding. Like other heat shock proteins such as GroEL[90] and heat shock protein 16.5 [91], Hsp33 contains a large hydrophobic surface area located in the N-terminal domain [92], which becomes exposed upon the oxidative activation of Hsp33. This N-terminal region has been proposed to serve as the binding site of protein folding intermediates [92]. In agreement with this hypothesis, it was found that complex formation of Hsp33 with its denatured protein substrates causes indeed the burial of extensive hydrophobic surface areas making them much less solvent-accessible. This results in a more cooperative folding reaction of the Hsp33 complex. Our SUPREX results (i.e, the increased m-value) suggest that on the order of 600 Å² are buried at the Hsp33-substrate interface. This is a substantial part of the maximum surface area of 3600 Å² that has been postulated to potentially serve as substrate binding sites in the Hsp33 dimer (38).

The results reported here suggest that the Hsp33-LDH, Hsp33-aldolase and Hsp33-CS complexes have Kₐ values ranging from 15-300 nM. The ΔΔGf,avg value obtained for the Hsp33-MDH complex was slightly more negative compared to the other three complexes, and the lower bound of the reported Kₐ range for the Hsp33-MDH complex was approximately an order of magnitude smaller compared to the lower bounds of the other three complexes. The lower bound calculation of the Kₐ value
ranges in Table 2 assumes the excess protein substrates in our samples are aggregated and precipitated out of solution prior to our SUPREX analysis. The light scattering results, which were collected by my collaborators, suggest that such aggregation and subsequent precipitation is more likely to more readily occur with the CS, LDH, and aldolase substrates than with the MDH substrate, as the kinetics MDH aggregation is slower than with the other three substrates. Thus, the $K_d$ values of the Hsp33-LDH, Hsp33-aldolase and Hsp33-CS complexes are likely to be at the lower end of their reported ranges, whereas the $K_d$ value for the Hsp33-MDH complex is likely to be at the upper end of its reported range, where the $K_d$ calculation assumes no aggregation and precipitation of the unbound protein substrate prior to our SUPREX analysis.

If, as suggested above, the $K_d$ value of the MDH-Hsp33 is assumed to be near the lower bound reported in Table 2, and the $K_d$ values of the other three complexes are assumed to be near the lower bounds reported in Table 2, then the $K_d$ values determined for all the complexes in this work would fall in a range of 15 to 82 nM. This range is very similar to the range of previously reported binding affinities for other protein folding chaperones binding with their non-native protein substrates. For example, the molecular chaperones GroEL and SecB exhibited a $K_d$ value of 50 nM to denatured Barnase, which was determined using a H/D exchange and NMR based technique [69]. A complex of GroEL and a denatured pepsin substrate had a $K_d$ value of 3.6 nM [70].
Hsp60 was found to have a $K_d$ value of 1.5 nM to denatured recombinant prion protein [71]. Although protein chaperones differ in their primary amino acid sequence, in their three-dimensional structures and in their binding mechanisms, the combination of our new data together with existing data suggest that the binding affinities of protein folding chaperones to their denatured protein substrates fall in the low nanomolar range and cluster in a relatively narrow range (i.e., within 1-2 orders of magnitude).
Figure 3: Representative SUPREX curves obtained for Hsp33 (open circles) and its different complexes with (A) CS (filled circles), (B) MDH (filled circles), (C) Aldolase (filled circles) and (D) LDH (filled circles). Curves were obtained using the high sensitive SUPREX protocol when H/D exchange time was 15 minutes. Solid lines are the best fit of the data to a four-parameter sigmoidal equation in SigmaPlot, and the dotted arrows indicate the midpoint ($C_{1/2}^{\text{SUPREX}}$) extracted from each SUPREX curve.
Figure 4: Plots of $\Delta G_{\text{app}}$ versus $C_{1/2}^{\text{SUPREX}}$ for Hsp33 (open circles) and its different complexes with (A) CS (filled circles), (B) MDH (filled circles), (C) Aldolase (filled circles), and (D) LDH (filled circles). The resulting y-intercept and slope were taken as the $\Delta G$ and m-value, respectively. The H/D exchange time varied from 2 minutes to 45 minutes. The solid lines represent the linear least squares fitting of each data set. The correlation coefficient obtained for each data set was greater than 0.94.
Table 1: $C^{1/2}_{\text{SUPREX}}$ at different H/D exchange time ($t_{H/D}$) for Hsp33 and its complexes with different protein substrates.

<table>
<thead>
<tr>
<th>Protein</th>
<th>$t_{H/D} = 2$ min</th>
<th>$t_{H/D} = 6$ min</th>
<th>$t_{H/D} = 15$ min</th>
<th>$t_{H/D} = 45$ min</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C^{1/2}_{\text{SUPREX}}$ (M)$^a$</td>
<td>$C^{1/2}_{\text{SUPREX}}$ (M)$^a$</td>
<td>$C^{1/2}_{\text{SUPREX}}$ (M)$^a$</td>
<td>$C^{1/2}_{\text{SUPREX}}$ (M)$^a$</td>
</tr>
<tr>
<td>Apo-Hsp33</td>
<td>2.1</td>
<td>1.6</td>
<td>1.1</td>
<td>0.8</td>
</tr>
<tr>
<td>CS-Hsp33</td>
<td>1.9</td>
<td>1.8</td>
<td>1.5</td>
<td>1.3</td>
</tr>
<tr>
<td>MDH-Hsp33</td>
<td>2.4</td>
<td>2.0</td>
<td>1.8</td>
<td>1.5</td>
</tr>
<tr>
<td>Aldolase-Hsp33</td>
<td>2.3</td>
<td>1.7</td>
<td>1.5</td>
<td>1.3</td>
</tr>
<tr>
<td>LDH-Hsp33</td>
<td>2.3</td>
<td>2.0</td>
<td>1.7</td>
<td>1.5</td>
</tr>
</tbody>
</table>

$^a$Values are taken from the best fit of SUPREX curve. Errors were typically between ±0.1 and ±0.2.
Table 2: SUPREX-derived ΔG\text{f} , m-values and K\text{d} for Hsp33 and its complexes.

<table>
<thead>
<tr>
<th>Protein</th>
<th>ΔG\text{f} (kcal/mol)</th>
<th>m (kcal/(mol M))</th>
<th>ΔG\text{f,avg} (kcal/mol)</th>
<th>m\text{avg} (kcal/(mol M))</th>
<th>ΔΔG\text{f,avg} (kcal/mol)</th>
<th>K\text{d} (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apo-Hsp33</td>
<td>-12.9 ± 0.2</td>
<td>1.3 ± 0.1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CS-Hsp33</td>
<td>-15.5 ± 0.7</td>
<td>2.8 ± 0.4</td>
<td>-14.6 ± 0.2</td>
<td>2.2 ± 0.4</td>
<td>1.7 ± 0.2</td>
<td>30-302</td>
</tr>
<tr>
<td>MDH-Hsp33</td>
<td>-15.2 ± 0.3</td>
<td>2.1 ± 0.2</td>
<td>-15.3 ± 0.1</td>
<td>2.2 ± 0.4</td>
<td>2.4 ± 0.2</td>
<td>3-82</td>
</tr>
<tr>
<td>Aldolase-Hsp33</td>
<td>-13.8 ± 0.5</td>
<td>1.8 ± 0.3</td>
<td>-14.5 ± 0.3</td>
<td>2.2 ± 0.4</td>
<td>1.6 ± 0.3</td>
<td>60-296</td>
</tr>
<tr>
<td>LDH-Hsp33</td>
<td>-14.8 ± 0.1</td>
<td>2.2 ± 0.1</td>
<td>-14.8 ± 0.1</td>
<td>2.2 ± 0.4</td>
<td>1.9 ± 0.2</td>
<td>15-167</td>
</tr>
</tbody>
</table>

 superscript a The ΔG\text{f} and m-values were calculated from the linear least-squares analysis of the data using Equation 2.1. The reported errors are the fitting errors of the linear least-squares analyses.

 superscript b ΔG\text{f,avg} values were the average of folding free energy values determined using Equation 2.1 and an established m-value. The reported errors are the standard deviations of the folding free energy values calculated using Equation 2.1 and the established m-value.

 superscript c The m\text{avg} values were the average of SUPREX-derived m values of CS-Hsp33, MDH-Hsp33, LDH-Hsp33 and Aldolase-Hsp33.

 superscript d No m\text{avg} value was calculated for apo-Hsp33.

 superscript e K\text{d} values were calculated using Equation 2.2 when [Hsp33]\text{dimer} : [Substrate]\text{monomer} = 1:1

 superscript f K\text{d} values were calculated using Equation 2.2 when [Hsp33]\text{dimer} : [Substrate]\text{monomer} = 1:1.5
4. Application of the SUPREX-Protease Digestion Protocol to Thermodynamic Analysis of Hsp33 Binding to its Unfolded Protein Substrate

4.1 Introduction

Hsp33 is a highly conserved redox-regulated chaperone holdase, which prevents HOCl- and other stress-induced protein aggregation both in vitro and in vivo. NMR studies revealed that Hsp33 is activated and dimerized to bind protein unfolding intermediates upon the complete unfolding of its C-terminal redox-switch domain and its linker region (201-219), an event which is triggered by oxidative stress conditions such as elevated temperature and/or the presence of oxidants.[64, 82] This oxidative-induced unfolding event appears to be critical for exposure of substrate binding sites. Thus, Hsp33 is one of a growing number of proteins that carry out their protein function when they are intrinsically disordered. [93, 94] Once the cell returns to non-stress (i.e., reducing) conditions, the reduced Hsp33 transfers its protein substrate to other chaperone foldases from the Hsp70 family that ultimately refold the protein substrates to their native three-dimensional structure. At the final step of the cycle, substrate-free Hsp33 dimers return to their monomeric and inactive form. This redox-regulated chaperone network was identified in the reference [86] and outlined in Figure 5.

While the basic mechanism of Hsp33 chaperone action is known, detailed about Hsp33’s interaction with substrate proteins still remain unknown. There are questions
about how Hsp33 prevents binding with its unfolded domain and recognizes its partially unfolded protein substrates and how one substrate can be recognized by two different chaperone systems during the transfer from the chaperone holdase, Hsp33 to chaperone foldases such as DnaK/DnaJ/GrpE chaperone machinery. To address these questions, biophysical and computational studies have been coupled with mass spectrometry and cell biology methods, such as peptide arrays, SPR binding studies, limited proteolysis experiments and SUPREX analyses.[95] The work in this chapter focuses on the application of SUPREX with a protease digestion protocol to characterize Hsp33’s interactions with one of its unfolded substrate proteins, citrate synthase (CS), at different binding stages outlined in Figure 5 to obtain mechanistic insights.

4.2 Experimental Section

4.2.1 Sample Preparation and SUPREX-Protease Digestion Protocol

CS from pig heart was purchased from Roche (Indianapolis, IN). Concentrated solutions of the substrate proteins CS (394 μM in 25 mM phosphate buffer, pH 7.4) was prepared and provided by Dr. Ursula Jakob lab and centrifuged (13,000 rpm, 45 min, 4°C) to pellet already formed aggregates before the preparation of the Hsp33-CS complex. Reduced inactive Hsp33red and bleach-oxidized active Hsp33ox were prepared by Dr. Dana Reichmann from Dr. Ursula Jakob lab as described in the reference [64] and
The concentrated Hsp33<sub>red</sub> (220 μM in 25 mM phosphate buffer, pH 7.4) and Hsp33<sub>ox</sub> (285 μM in 25 mM phosphate buffer, pH 7.4) solutions were stored at -20 °C.

The Hsp33-CS complex was prepared daily fresh as described previously.[49]

Briefly, A 15 ml solution of 285 μM activated Hsp33<sub>ox</sub> was pre-warmed at for 10 minutes with constant shaking at 330 rpm. 1 μl aliquots from a 30 ml 394 μM solution of the CS substrate, also at 43 °C, were added at 5 minute intervals to the Hsp33<sub>ox</sub> solution. The solution became turbid, presumably due to the aggregation of the protein substrate, when the monomeric concentration of Hsp33<sub>ox</sub> to CS ratio was ~ 2 : 1. Additional CS solution was added until the ratio of 4:3 was reached. The precipitate (presumably CS) was spun down for 45 minutes at 13,000 rpm in a cold room (4 °C). The supernatant, containing the Hsp33-CS complex was used for the SUPREX experiments.

DTT-reduced Hsp33-CS was generated by incubating freshly Hsp33-CS complex with 5 mM (final concentration) DTT solution at 30 °C for 30 minutes. The reduced complex sample was immediately used for SUPREX experiment.

A series of H/D exchange buffers containing 20 mM phosphate, pD 7.4 and GdmCl concentrations ranging from 0.3 to 4.2 M were prepared using deuterated GdmCl.

The stock solution of protein or protein-substrate complex was diluted 10-fold into the series of deuterated SUPREX buffers. After the specified exchange time ranging
from 1 minute to 1.5 hours, the H/D exchange was quenched by adding 2 μL of each exchanged sample to 4 μL of aqueous HCl (pH 0.9). A 4 μL aliquot of 1.25 μM pepsin was added to each quenched protein solution to give a final pepsin concentration of 0.5 μM), and the protease digestion was allowed to proceed for 30 second. The molar ratios of protein to pepsin are typically 8:1 for protein complex and 4:1 for single protein. The final concentration of GdmCl in the digestion solutions is ranging from 0 to 0.8 M.

4.2.2. Peptide Mapping and Sequencing

Peptide mapping and sequencing were performed on undeuterated protein samples. A 117 μM CS, 113 μM Hsp33ox or 110μM Hsp33red solution was diluted 10 fold into a 20 mM phosphate buffer, pH 7.4. The diluted protein sample was digested by pepsin under the same conditions as described in the SUPREX protocols. The sequences of the peptides were determined using either the Bruker Ultraflex II TOF/TOF mass spectrometer or the Agilent 6520 Q-TOF.

The MS1 and MS/MS data generated in the MALDI-TOF-TOF were searched against two different SwissProt database: Escherichia CoLi for Hsp33 and Mammals for CS using online MASCOT software at http://www.matrixscience.com/search_form_select.html. The parameters used for the database search were as follows: Enzyme with no specific cleavage was used. Methionine oxidation was set as a variable modification and none of fixed modifications
were selected. Both peptide and MS/MS tolerance was ± 1.2 Da. The mono-isotopic mass with peptide charge of 1+ was used. Precursor mass was set with error tolerant checked according to the mono-isotopic mass value obtained using reflector mode.

For the LC-MS/MS analysis, 1 μL of 3 % TFA solution was quickly added to a 10 μL aliquot of the pepsin digested sample and the sample was desalted using C18 ZipTip™ (Millipore Corporation, Billerica, MA) according to the manufacturer’s instructions. Peptides were eluted using three buffers containing 0.1 % TFA with 20 %, 50% and 75 % ACN respectively. Each eluted solution was then diluted 20 fold with a 0.1% TFA solution. Ultimately, a 1 μL resulting solution was loaded to an Agilent Chip containing a 43 mm x 75 μM column packed with Zorbax 300SB-C18 (5 μM) and subject to LC-MS/MS analysis. Peptides were separated on the chip with a 7 minute gradient elution from 3-70 % buffer B. The mass spectral data generated in the LC-MS/MS analyses was searched against either Hsp33 or CS sequence using Spectrum Mill (Agilent Technologies, Santa Clara, CA). Prior to searching the product ion mass spectra derived from precursor ions with similar m/z values (i.e., ± 0.05 m/z) and similar retention times (i.e., ± 0.05 second) were merged together. The parameters used for the database search were that: Oxidation on methionine was set as a variable modification and none was set as fixed amino acid modifications. No enzyme was set as digest to
indicate the non-specific cleavage by pepsin. Mass tolerance was used for both precursor and product ion m/z values. Match filtering was disabled during the database search.

High quality sequence data could not be obtained for one peptide ion signal at 2202 Da in either the MALDI-TOF-TOF or the LC-MS/MS experiment. For this peptide the mono-isotopic masses obtained in the LC-MS instrument was matched against the protein sequence with mass tolerance of 0.5 Dalton to locate amino acid position using online software at http://prowl.rockefeller.edu/prowl-cgi/sequence.exe/.fsa.

4.3. Results

4.3.1. Pepsin digestion

The initial step in this study involved an examination of the pepsin digestion profile on undeuterated protein samples: Hsp33\textsubscript{red}, Hsp33\textsubscript{ox}, native-state-CS, Hsp33-CS complex, and DTT-reduced Hsp33-CS complex. A pepsin digestion protocol was developed in which the digestion was allowed to proceed for 30 seconds at room temperature, pH ~1.1, in the presence of GdmCl with final concentration of either 0 or 0.8 M, respectively, the same solution conditions for the deuterated samples that would be analyzed later analyses using SUPREX technique. The molar ratio of protein to pepsin was 4:1 for Hsp33 or CS alone and 8:1 for their complexes. The digested peptides were immediately mixed with two different MALDI matrix prior to the MALDI mass analysis: HCCA used for peptides smaller than 4 kDa and SA used for peptides or
proteins larger than 4 kDa. A total of 18 Hsp33-peptides ranging from 1.5 kDa to 3.7 kDa, (see Table 4), were consistently detected in the MALDI readout providing a peptide sequence coverage of 64% (Table 4). In the MALDI readout 13 CS-peptides ranging from 1.7 kDa to 8.6 kDa were consistently generated (see Table 3). An additional 4 CS-peptides, including a 2486 Da, 3129 Da, 3049 Da and 2361 Da peptide, were detected in all the samples with the exception of CS-Hsp33 complex. Ultimately, a total of 17 CS-peptides were monitored in the later SUPREX experiments, and 11 of them (see Table 3) were successfully sequenced providing a CS-sequence coverage of 44%.

Shown in Figure 6 (A-D) are representative MALDI mass spectra from each protein sample using HCCA as the matrix. It is noteworthy that the CS-peptide ion signal at 2202 Da (269-289 aa, zoomed peak in Figure 6) is associated with the different stages of the protein: native state (Figure 6A), complexed with Hsp33 (Figure 6B) and DTT-reduced complex (Figure 6C). The strong ion signal from native-CS significantly decreased in complex form, but partially recovered in reduced complex. In order to rule out the possibility of the signal suppression resulting in the reduction of the ion intensity for the protein complexes, a mixture of native-CS and Hsp33 without known interactions was subject to pepsin digestion under the same digestion condition. A mass spectrum of the generated peptides from the mixture (Figure 6E) showed that the ion signal at 2202 Da in the presence of Hsp33 peptides had similar intensity as in the
absence of Hsp33 (Figure 6A). The same phenomenon was observed for other CS-peptides, including 2486 Da (310-330 aa), 3129 Da (325-351aa), 3049 Da(38-66 aa), 2361 Da (437-457aa) and 1947 Da (290-309aa). Reduction or even disappearance of those CS peptide ion signals generated from Hsp33-CS is most likely due to the chaperone-substrate binding events. The solution conditions of pepsin digestion are apparently not sufficient to denature the non-covalent binding between CS and Hsp33, a nM interaction determined in Chapter 3. As a result, the region, from which the total of 6 CS-peptides were derived, is still under certain protection and thus, not solvent-accessible for pepsin cleavage within 30 seconds.

4.3.2. Thermodynamic analyses of CS using a peptide readout

Native-CS, the CS-Hsp33 complex, and the DTT-reduced-complex were subject to thermodynamic analyses using the SUPREX-digestion protocol. The DTT-reduced complex represented the stage at which the CS substrate is ready to be transferred to other chaperone machineries. The reduced complex was generated by incubation with 5 mM DTT for 30 minutes. Under this condition, two intra-molecular disulfide bridges from the Hsp33 C-terminal domain are reduced, leading to some conformational changes in Hsp33 but with the complex remains intact until the presence of the DnaK-system. [86] In additional, thiol-trapping experiments, which were preformed by Dr. Ursula Jakob’s Laboratory at the University of Michigan, on both the complex and the
reduced complex,[95] suggested no significant cysteine modification in CS and thus, rendered it unlikely that the DTT reagent itself altered the CS conformation. A thermal unfolding method described in the experimental section was used to generate Hsp33-CS complex. Note that native form of CS is not the actual protein substrate for Hsp33. The actual protein substrate is a thermally-unfolded CS structure, which is a transient state of a protein that either aggregated or being rescued by chaperone proteins.

The H/D exchange times in SUPREX experiment were chosen ranging from 1 minute to 1.5 hours such that the range was 1) wild enough to generate at least one intact sigmoidal curve (i.e. no missing pre- or post- transition lines within the range of GdmCl concentrations) for every monitored peptide and 2) exchange times were short enough before the degradation of the complex (presumably more than 4 hours). Multiple exchange times provided the possibility for checking the reproducibility of SUPREX results in term of difference in $C_{1/2}$ values (i.e. $\Delta C_{1/2}$) as well.

The SUPREX behaviors of the 17 CS-peptides monitored here can be grouped into three categories (see Figure 7). One category included those that exhibited the highest stability in native CS sample and were destabilized in the CS-Hsp33 complex sample. Further destabilization was detected after the DTT-treatment of the complex sample. Another category included those that were significantly destabilized in the DTT-reduced CS-Hsp33 complex samples and had missing signals from the CS-Hsp33
complex. Peptides in the third category had similar stabilities (i.e. the same SUPREX behavior within experimental error) in all three samples.

The quantitatively determined $\Delta C^{1/2}$ values obtained from high quality sigmoidal curves (i.e. curves with enough amplitude to accurately determine $C^{1/2}$ values) were calculated between pairs of samples (indicated in Table 3) and averaged $\Delta C^{1/2}$ values (summarized in Table 3) were calculated from different peptides that covers the same region and their $\Delta C^{1/2}$ values obtained at different H/D exchange times. For the calculation of the averaged $\Delta C^{1/2}$ values, an assumption was made that the different H/D exchanges times for each region should have no influence on $\Delta C^{1/2}$ values. In another word, m-value, which is equal to the slope of the linear fitting for plots of $\Delta G_{\text{app}}$ vs $C^{1/2}$ values obtained at different H/D exchange times (see Equation 2.1), should remain the same at each examined binding stages of protein. A total of six peptides from the first category exhibited smaller $C^{1/2}$ values from CS-Hsp33 complex sample compared with native-CS sample and the difference is ranging from -0.2 M to -1.8 M. However, upon DTT reduction, the stability of the regions, from which these six peptides were derived, became even less stable with the changes in $C^{1/2}$ values between DTT-reduced and non-reduced CS-Hsp33 complex at -0.5 ± 0.3 M. In some cases (see Figure 7A), under the H/D exchange condition described in experimental section), post-transition lines with no transition were generated.
The $\Delta C^{1/2}$ values and peptide locations listed in the Table 3 were used to generate Figure 8. The colors, which were used for labeling in Figure 8A, scaled 1, 2, 3, 4 and 5 that correspond to the $\Delta C^{1/2}$ ranging from 0-0.2 M, 0.2-0.6M, 0.6-1.2M 1.2-1.8 M and 1.8-2.1 M, respectively. The darkest color, scale 6, represents completely destabilization, showing that no $C^{1/2}$ values can be obtained (examples shown in Figure 7A, E-F). The thermodynamic stability from native-CS samples, which is the highest, was assigned with color of scale 1 regardless of the individual $C^{1/2}$ values for each region since this part of work focused on the evaluation of the stability changes from each binding state for CS. The determination of global or sub-global unfolding/refolding properties of native state of CS is not the goal of this project.

4.3.3. Thermodynamic analyses of Hsp33 using a peptide readout

Hsp33$_{\text{red}}$, Hsp33$_{\text{ox}}$, the Hsp33-CS complex, and the DTT-reduced Hsp33-CS complex represented the in-active, active, binding and ready-to-release stages of Hsp33, which are illustrated in Figure 5. Each sample was subjected to a SUPREX analysis and the SUPREX behaviors of a common set of 18 pepsin-generated peptides were monitored. The same method used to generate the $\Delta C^{1/2}$ values listed in Table 3 was used to generate the changes in the thermodynamic stabilities, i.e., $\Delta C^{1/2}$ values, for the 18 monitored Hsp33 peptides. The $\Delta C^{1/2}$ values determined here are summarized in Table 4. In Table 4, the $\Delta C^{1/2}_{\text{avg}}$ value for each peptide was calculated by averaging $\Delta C^{1/2}$
values from high quality sigmoidal curves. The standard deviations range from 0.1-0.4 M. when $\Delta C^{1/2}_{\text{avg}}$ values were 2× larger than their individual standard deviations, they were considered significant. On the contrast, if $\Delta C^{1/2}_{\text{avg}}$ values were small (i.e., less than 2-times of their individual standard deviation), they were considered as no changes or subtle changes that are not detectable within the resolution of the SUPREX method. The data in Table 4, suggests that all the regions of Hsp33red, from which the peptides were derived, were destabilized upon the oxidation reaction as the $\Delta C^{1/2}_{\text{avg}}$ values were between 0.4-0.8 M GdmCl. When CS binds to Hsp33ox, two peptides, which both covered the 174-192aa region, undergoes the most significant thermodynamic stability change, with $\Delta C^{1/2}_{\text{avg}}$ values equal to $+0.4 \pm 0.1$ M, suggesting ligand-induced stabilization. The next most significant thermodynamic change occurs in the 203-221aa region in which the peptides were destabilized ($\Delta C^{1/2}_{\text{avg}}$ values $0.5 \pm 0.1$ M [GdmCl]) upon oxidization and stabilized ($\Delta C^{1/2}_{\text{avg}}$ values $0.2 \pm 0.1$ M [GdmCl]) upon CS binding to Hsp33ox. Due to the fact that the standard deviation of this region was generated by 18 pairs of $\Delta C^{1/2}$ values (Table 4), the 0.2 M [GdmCl] shift was considered significant. These four peptides span approximately 64 % of Hsp33’s linker region (178-231aa), a ~ 50 aa flexible linker region that is part of C-terminal domain (Figure 10).

Figure 10 was generated based on the $\Delta C^{1/2}$ values listed in Table 4 in the same way that Figure 8 was generated with only difference being the definition of the color
scales: 1, 2, 3, 4 and 5 correspond to $\Delta C_{1/2}$ values ranging from 0-0.1 M, 0.1-0.2M, 0.2-0.4M, 0.4-0.6 M and 0.6-0.8 M [GdmCl], respectively. The thermodynamic stability from Hsp33$_{red}$ samples, which is the highest, was assigned with color of scale 1 regardless of the individual $C_{1/2}$ values for each region.

4.4. Discussion

4.4.1. Prerequisites for substrate binding and releasing

Thermodynamic studies on Hsp33-bound substrate protein revealed three categories of behaviors. Unchanged stability as the first category of behavior showed that certain regions of Hsp33 substrates were not affected by the thermal unfolding, under which condition protein unfolding intermediates were generated, and they remained significant amount of secondary and tertiary structure elements. This suggested that some folded structure on substrate is required for Hsp33 recognition and might be one of the reasons how Hsp33 prevents itself binding to own intrinsically unfolded regions.

The signature of the second behavior was the significant destabilization in chaperone-substrate complex form, suggesting that they are part of thermal-unfolded regions. Such regions involved three locations in CS crystal structure, helices A, L to M and T.[96] In the crystal structure, helices L (268-308aa) and M (310-324aa) form part of a largely hydrophobic, four-stranded antiparallel $\alpha$-helix sandwich. In addition, they
consist of the main component of the monomer-monomer interface of CS while helix T (451-459) forms a domain swap contributing to dimerization as well (Figure 8). The third category of peptides located from 310-351 aa, spanning helices N (327-344aa) and O (350-365aa), has no signals from Hsp33-CS complex. Pepsin-digestion profile created from a mixture of native-CS and Hsp33 indicated that peptides from helices L, M, N and O are solvent-inaccessible due to substrate binding to Hsp33 (shown in Figure 6). This is in consistent with the evidence provided by peptide array experiments, which were designed and performed in Ursula Jakob lab: helices A (23-55), L, M and parts of O have high affinities to Hsp33. [95] Both evidences strongly suggested that Hsp33 interacts with parts of CS’s hydrophobic dimerization interface.

After DTT was used to reduce the Hsp33-CS complex, complete destabilization was observed in both second and third categories of peptides, spanning helices A,L, M, N, O and T, exhibiting little or no transitions of SUPREX curves. Representative MALDI spectra from pepsin digestion (Figure 6) show increased intensity of ion signals from these peptides after DTT-treatment of Hsp33-substrate complex. This result agrees well with the result obtained from limited proteolysis analyses (1:200 trypsin to protein sample, 15 min) in combination of mass spectrometry, performed by Dr. Ursula Jakob lab. Their results demonstrated that DTT-treatment on Hsp33-CS complex exposed unique proteolytic sites that were solvent-inaccessible due to the protection of Hsp33 in
Hsp33-substrate complex. Both the exposure of proteolytic sites and the thermodynamic destabilization indicated increased substrate unfolding and the decreased Hsp33-substrate interactions. Such conformational rearrangements in CS, which were triggered by the DTT-induced refolding of Hsp33, were prerequisites for the successful transfer of substrate proteins to the other chaperone machineries.

4.4.2. Involvement of intrinsically unfolded region in Hsp33 for substrate binding

Previous studies revealed that oxidation of Hsp33 triggers conformational rearrangement in Hsp33, leading to unfolded Hsp33 C-terminal domain, involving a zinc center, as a redox-switch sensor, and a 54aa-linker region that connects N-terminal domain and the zinc center.[64] This marks Hsp33 as member of a new class of chaperones that require partial unfolding for full chaperone activity. In the SUPREX analysis of Hsp33 red and Hsp33ox at peptide level showed that most region of Hsp33 lost stabilities upon the oxidation. This is consistent with the SUPEX analysis of these proteins at intact protein level,[97] demonstrating that reduced form of Hsp33 is more stable than its oxidized form. The most destabilized region is from two peptides, which cover 174-192aa and are part of the linker region. This result agrees with earlier CD and fluorescence measurements, implying that the linker region became unfolded when Hsp33 was oxidized. [64, 98] Upon CS binding to the Hsp33, SUPREX analysis in this work showed that the same region of Hsp33 regained considerable thermodynamic
stabilities, which was consistent with the results collected by limited proteolysis analyses from Dr. Ursula Jakob’s lab: K198 became exposed upon oxidation of Hsp33, and less accessible to trypsin upon substrate binding. [95] All the evidences indicated that the intrinsically unfolded linker region serves as the binding sites for early unfolding intermediates. Involvement of intrinsically disordered region in substrate binding not only provides flexibility on substrate recognition with so called wide specificity, which is very common in chaperone binding, but also stables both binding partners by releasing the solvent molecules and thus, minimizing the overall entropy cost.

However, it still remains unclear that whether N-terminal domain involves the substrate interaction or not. With the exception of peptides from linker region (172-192aa and 203-221aa) and peptides from 24-47aa with standard deviation < 0.1 M, all peptides generated ΔC^{1/2} values with relatively bigger standard deviation 0.2-0.4M, making it hard to detect the stability changes upon ligand binding. The non-ideal standard deviation could be association with 1) the assumption of no changed m-values for data analysis (see above) or 2) back exchange issue, which depends on number of buried amide protons associated with peptide structure in Hsp33.

The fact of no thermodynamic changes detected for Hsp33 peptides after the DTT-treatment of Hsp33-CS complex might be due to the detection of limited numbers
of high quality curves for this protein sample or could also be reasoned that either the associated region has not been detected in the SUPREX analyses or the changes are more related with local fluctuation in protein structure that is not detectable by SUPREX.

4.5. Conclusions

A successful application of the SUPREX method was shown here to study a very complex macromolecular recognition by monitoring conformational changes of interacting partners simultaneously, while both binders are extensively unfolded. Each stage of binding including non-bound and bound form was examined to provide mechanism insights on how Hsp33 functions in multi-chaperone networks. The results suggested that Hsp33 uses intrinsically disordered regions to bind partially unfolded protein substrate and it converts the substrate proteins into less structured folding intermediates, the preferred substrates of the DnaK-system that is responsible for correctly refolding these proteins.
Figure 5: Schematic representation of Hsp33 network to prevent oxidative-induced protein aggregation. This includes: 1) Hsp33 is activated upon the oxidative conditions and binds with partially unfolded protein substrates (purple); 2) Once reducing condition restores, substrate is transferred to other chaperone proteins to be folded into its native structures and 3) Substrate-released Hsp33 returns to its inactive and reduced form.
Figure 6: Representative MALDI mass spectrum for peptic peptides obtained from A) native-CS, B) Hsp33-CS complex, C) DTT-reduced-Hsp33-CS complex, D) Hsp33 and E) mixture of CS and Hsp33. An ion signal at m/z equal to 2202 Da was zoomed in and shown as well.
Figure 7: Representative SUPREX curves for CS-peptides that can be grouped into three types of thermodynamic behavior when CS is at its different stages: native-CS (black lines or black filled circles), CS complex with Hsp33 (red lines or red filled circles) and DTT reduced CS complex (blue lines or blue filled circles). CS-peptides are A) 2202 Da (269-289), B) 3262 Da (38-69), C) 2673 Da (434-457), D) 1947 Da (290-309), E) 2486 Da (310-330), F) 3129 Da (325-351), G) 3307 Da (102-131), H) 1965 Da (182-197), I) 1681 Da (141-155) and J) 6054 Da. The line represents the best fit of the data to a four-parameter sigmoid equation using SigmaPlot. The dotted line marks the $C_{1/2}$ value. H/D exchange time was 1 minute for all the peptides with the exception of 2202 Da peptide (A, H/D = 5 minutes) and 1681 Da peptide (I, H/D = 20 minutes).
Table 3. Summary of the changes in thermodynamic stabilities for peptides from different stages of CS: native-state-CS (CS), CS complexed with Hsp33 (Com) and DTT-reduced CS-Hsp33 complex (DTTcom)

<table>
<thead>
<tr>
<th>Peptide Mass (Da)</th>
<th>Location</th>
<th>Difference in C1/2 values</th>
<th>ΔC1/2avg (M)</th>
<th>#Pairs of curves for averaging</th>
<th>#Observed destabilization/de-protection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CS→Com&lt;sup&gt;i&lt;/sup&gt;</td>
<td>-1.8±0.1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3049 &amp; 3262</td>
<td>38-66 (peptide 38-64 and 38-66)</td>
<td>CS→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>-2.1±0.1</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>-0.4±0.0&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I</td>
<td>2202</td>
<td>CS→Com&lt;sup&gt;i&lt;/sup&gt;</td>
<td>-0.2±0.3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>&lt;&lt;0&lt;sup&gt;d&lt;/sup&gt;</td>
<td>/</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>&lt;&lt;0&lt;sup&gt;d&lt;/sup&gt;</td>
<td>/</td>
<td>3</td>
</tr>
<tr>
<td>1947</td>
<td>290-309</td>
<td>CS→Com&lt;sup&gt;i&lt;/sup&gt;</td>
<td>-1±0.3&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>-1.4±NA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>-0.5±NA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2673 &amp; 2361</td>
<td>434-457 (pep434-457 and 437-457)</td>
<td>CS→Com&lt;sup&gt;i&lt;/sup&gt;</td>
<td>-1.7±0.1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>-2.2±0.2&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>-0.5±0.2&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>II</td>
<td>2486</td>
<td>CS→Com&lt;sup&gt;i&lt;/sup&gt;</td>
<td>NS&lt;sup&gt;e&lt;/sup&gt;</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>&lt;&lt;0&lt;sup&gt;d&lt;/sup&gt;</td>
<td>/</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>NS&lt;sup&gt;e&lt;/sup&gt;</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>3129</td>
<td>325-351</td>
<td>CS→Com&lt;sup&gt;i&lt;/sup&gt;</td>
<td>NS&lt;sup&gt;e&lt;/sup&gt;</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>&lt;&lt;0&lt;sup&gt;d&lt;/sup&gt;</td>
<td>/</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>NS&lt;sup&gt;e&lt;/sup&gt;</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Peptide Mass (Da)</td>
<td>Location</td>
<td>Difference in C1/2 values</td>
<td>ΔC1/2avg (M)</td>
<td>#Pairs of curves for averaging</td>
<td>#Observed Similar stability/protection&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>III</td>
<td>3307</td>
<td>CS→Com&lt;sup&gt;i&lt;/sup&gt;</td>
<td>0.0±0.2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>0.1±0.2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>0.1±0.0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>1965</td>
<td>182-197</td>
<td>CS→Com&lt;sup&gt;i&lt;/sup&gt;</td>
<td>0.0±NA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>0.2±0.2&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>-0.1±0.1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>1681</td>
<td>141-155</td>
<td>CS→Com&lt;sup&gt;i&lt;/sup&gt;</td>
<td>0.1±0.1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>-0.1±0.2&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>0.2±0.2&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5034</td>
<td>/&lt;sup&gt;g&lt;/sup&gt;</td>
<td>CS→Com&lt;sup&gt;i&lt;/sup&gt;</td>
<td>-0.2±0.0&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>-0.2±0.0&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTCom&lt;sup&gt;i&lt;/sup&gt;</td>
<td>-0.0±0.0&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Peptide Mass (Da)</td>
<td>Location</td>
<td>Difference in C^{1/2} values</td>
<td>ΔC^{1/2}_avg (M)</td>
<td># Pairs of curves for averaging</td>
<td># Observed similar stability/protection b</td>
</tr>
<tr>
<td>------------------</td>
<td>----------</td>
<td>------------------------------</td>
<td>-----------------</td>
<td>-------------------------------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>7085</td>
<td>/g</td>
<td>CS→Com i</td>
<td>0.1±0.1 c</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS→DTTcom i</td>
<td>0.1±0.0 c</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTcom i</td>
<td>0.1±0.1 c</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>8640</td>
<td>/g</td>
<td>CS→Com i</td>
<td>-0.1±0.1 c</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS→DTTcom i</td>
<td>-0.1±0.1 c</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTcom i</td>
<td>0.0±0.0 c</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6054</td>
<td>/g</td>
<td>CS→Com i</td>
<td>0.0±0.0 c</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS→DTTcom i</td>
<td>0.0±0.0 c</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTcom i</td>
<td>-0.0±0.0 c</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5284</td>
<td>/g</td>
<td>CS→Com i</td>
<td>0.1±0.1 c</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS→DTTcom i</td>
<td>0.1±0.0 c</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTcom i</td>
<td>0.0±0.1 c</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2072</td>
<td>/g</td>
<td>CS→Com i</td>
<td>0.2±0.1 c</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS→DTTcom i</td>
<td>-0.0±0.1 c</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTcom i</td>
<td>-0.2±0.1 c</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

a Destabilization/de-protection was assigned when 1) the calculated ΔC^{1/2}_avg values were significantly smaller than 0 or 2) when no transition lines were detected for certain stage of CS, which was resulting from the destabilization or de-protection on the region from which peptides were derive (examples shown in Figure 7A, 7E and 7F).

b Similar stability/protection was assigned when 1) the calculated ΔC^{1/2}_avg values were within 0.2 M or 2) when the Δ Mass values were overlapped for all the stages of CS if C^{1/2} values were not determined due to lack of either the post- or the pre-transition lines, (examples shown in Figure 7G and 7H).

c Standard deviations were calculated from two ΔC^{1/2}_avg values. When only one ΔC^{1/2}_avg value was obtained, standard deviations were shown as “NA”.

d “<<0” was assigned based on the observation (examples shown in Figure 7A, 7E and 7F).

e NS stands for no peptide ion signals detected for one stage of CS, usually when it is in its complex form.

f C^{1/2} comparison between two samples For example: “CS→Com” indicates that C^{1/2} values of peptides from native-state-CS were subtracted from those from CS complex.

g Peptide locations were difficult to be obtained from peptide mapping, especially for large peptides.
Figure 8. Thermodynamic stability changes in CS 2nd structure (A) and higher ordered structure (B) from PDB: 2CTS when peptides were from native-CS, CS-Hsp33 complex and reduced complex samples, respectively. Color labeling was based on the $\Delta C_{1/2}$ listed in Table 3. A start indicated that the standard deviation of $\Delta C_{1/2}$ at this region is relatively bigger (0.3 M) than the rest part of the protein. Gray color in B) indicated the region, from which no peptides were derived. CS helices that are most destabilized upon complex reduction are indicated by their letter codes (A, T, L, M, O and N)
Figure 9. Representative SUPREX curves for Hsp33-peptides from Hsp33\textsubscript{red} (black lines), Hsp33\textsubscript{ox} (pink lines), Hsp33-CS complex (red lines) and DTT-reduced-Hsp33-CS complex (blue lines). Hsp33-peptides are A) 2564 Da (26-47), B) 2142 Da (175-192) and C) 2112 Da (203-219). The line represents the best fit of the data to a four-parameter sigmoid equation using SigmaPlot. The dotted line marks the C\textsubscript{1/2} value. H/D exchange time was 1 minutes for 2142 Da peptide and 20 minutes for both 2112 Da and 2564 Da peptides.
Table 4. Summary of the changes in thermodynamic stabilities for peptides from different stages of Hsp33: Hsp33\textsubscript{red} (Red), Hsp33\textsubscript{ox} (Ox), Hsp33-CS complex (Com) and DTT-reduced Hsp33-CS complex (DTTcom)

<table>
<thead>
<tr>
<th>Peptide Mass (Da)</th>
<th>Location</th>
<th>Difference in C\textsubscript{1/2} values</th>
<th>ΔC\textsubscript{1/2avg} (M)</th>
<th>#Pairs of curves for averaging</th>
</tr>
</thead>
<tbody>
<tr>
<td>2219 and 2088</td>
<td>3-19 (Peptide 3-19 and 4-19)</td>
<td>Red→OX\textsuperscript{a}</td>
<td>-0.6±0.3</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ox→Com\textsuperscript{a}</td>
<td>+0.3±0.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTcom\textsuperscript{a}</td>
<td>-0.1±0.3</td>
<td></td>
</tr>
<tr>
<td>2764 and 2564</td>
<td>24-47 (Peptide 24-47 and 26-47)</td>
<td>Red→OX\textsuperscript{a}</td>
<td>-0.4±0.1</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ox→Com\textsuperscript{a}</td>
<td>+0.0±0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTcom\textsuperscript{a}</td>
<td>+0.0±0.1</td>
<td></td>
</tr>
<tr>
<td>3740, 3577 and 2907</td>
<td>80-114 (Peptide 80-114, 80-113 and 80-106)</td>
<td>Red→OX\textsuperscript{a}</td>
<td>-0.6±0.3</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ox→Com\textsuperscript{a}</td>
<td>+0.2±0.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTcom\textsuperscript{a}</td>
<td>+0.2±0.1</td>
<td></td>
</tr>
<tr>
<td>2695 and 2532</td>
<td>114-138 (Peptide 114-138 and 115-138)</td>
<td>Red→OX\textsuperscript{a}</td>
<td>-0.6±0.3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ox→Com\textsuperscript{a}</td>
<td>+0.3±0.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTcom\textsuperscript{a}</td>
<td>+0.0±0.3\textsuperscript{b}</td>
<td>2</td>
</tr>
<tr>
<td>1785 and 1541</td>
<td>143-156 (Peptide 143-156 and 145-156)</td>
<td>Red→OX\textsuperscript{a}</td>
<td>-0.4±0.2</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ox→Com\textsuperscript{a}</td>
<td>+0.1±0.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTcom\textsuperscript{a}</td>
<td>-0.4±0.3\textsuperscript{b}</td>
<td>2</td>
</tr>
<tr>
<td>1818 and 1705</td>
<td>157-174 (Peptide 157-174 and 157-173)</td>
<td>Red→OX\textsuperscript{a}</td>
<td>-0.4±0.2</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ox→Com\textsuperscript{a}</td>
<td>+0.1±0.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTcom\textsuperscript{a}</td>
<td>+0.1±0.3</td>
<td></td>
</tr>
<tr>
<td>2255 and 2142</td>
<td>174-192 (Peptide 174-192 and 175-192)</td>
<td>Red→OX\textsuperscript{a}</td>
<td>-0.8±0.1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ox→Com\textsuperscript{a}</td>
<td>+0.4±0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTcom\textsuperscript{a}</td>
<td>+0.0±0.2</td>
<td></td>
</tr>
<tr>
<td>2312 and 2112</td>
<td>203-221 (Peptide 203-221 and 203-219)</td>
<td>Red→OX\textsuperscript{a}</td>
<td>-0.5±0.1</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ox→Com\textsuperscript{a}</td>
<td>+0.2±0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Com→DTTcom\textsuperscript{a}</td>
<td>+0.1±0.1</td>
<td>12</td>
</tr>
<tr>
<td>1732</td>
<td>289-294</td>
<td>Red, Ox Com DTTcom</td>
<td>Completely unfolded in all the</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}C\textsubscript{1/2} comparison between two samples. For example: “Red→Ox” means that C\textsubscript{1/2} values of peptides from reduced Hsp33 were subtracted from those from oxidized Hsp33 to obtain Δ C\textsubscript{1/2}. \textsuperscript{b} Standard deviations were calculated from two ΔC\textsubscript{1/2avg} values.
Figure 10: Thermodynamic stability changes in Hsp33 2nd structure (A) and higher ordered structure (B) from PDB: 1VZY when peptides were from Hsp33_{red}, Hsp33_{ox}, Hsp33-CS complex and reduced complex samples, respectively. Color labeling was based on the $\Delta C^{1/2}$ listed in Table 4. A start indicated that the standard deviation of $\Delta C^{1/2}$ at this region is relatively bigger (0.3 M) than the rest part of the protein. Gray color in B) indicated the region, from which no peptides were derived.
5. A Methionine Enrichment Protocol for SPROX Analysis of Protein-Ligand Binding on the Proteomic Scale

5.1 Introduction

Methionine residues, with approximate natural occurrence of only 2.37%, according to the SWISS-PROT database [72], exist in the primary structure of 95.8% of *E. coli* proteins [99] (the initiator methionine was not counted). As a result, isolation of the methionine containing peptides (met-peptides) can minimize the number of peptides to be identified, while still retaining comprehensive proteome coverage. Reported approaches for isolating methionine peptides in LC-MS/MS based proteomic methods include diagonal-chromatography-based [99] and solid phase-based strategies. [100-102] In the first approach, peptides were fractionated based on their retention time from a reverse-phase HPLC run and subsequent oxidation specifically on methionine side chain were applied to each peptide fractions to shift the elution time of met-peptides. The newly merged oxidized species were collected and analyzed in the LC-MS. This requires at least two HPLC-based separations and more than one LC-MS step. Therefore, the second approach was incorporated into SPROX technique as outlined in Figure 11. The solid-phase-based approach, which is based on a reaction described in the reference [103], utilizes commercially available glass beads derivatized with bromoacetyl groups
to selectively react with thioether group from methionine side chain at pH 2-3. This covalent reaction allows for removal of non-binders including methionine-free peptides and oxidized-met-peptides. Subsequently the met-peptides are released by the reducing reagent, β-mercaptoethanol (BME), at pH 8.5-8.8 (Figure 11). Previous reports on the application of this strategy to simple peptide mixtures\cite{100}, tryptic digested model proteins \cite{100, 101} and an Escherichia coli lysate system \cite{101} using MALDI- or SLEDI-TOF as a readout demonstrated the selectivity of the methionine-reactive beads in isolation of the met-peptides and its ability to detect proteins at low abundance level (i.e. 500 ppm). Some met-free peptides were observed in the MALDI spectra, resulting from non-covalently binding with the matrix of the beads. However, this can be eliminated by optimizing loading amount of the digested sample, increasing the efficiency of the washing steps and reducing / alkylating cysteine residues. \cite{101} In this work, this solid-phase based approach in combination with SPROX technique was used to isolate iTRAQ-labeled met-peptides from yeast cell lysate for globally profiling of protein thermodynamic unfolding/refolding properties in drug-motivated ligand binding studies.
5.2. Experimental Section

5.2.1 Materials

The commercially available methionine-reactive beads, named Pi™ methionine reagent, and C18 silica resin were purchased from The Nest Group, Inc. (Southboro, MA). The following materials were purchased from Sigma-Aldrich (St. Louis, MO): ammonium bicarbonate (NH₄HCO₃), triethylammonium bicarbonate, SDS, proteomics grade trypsin, S-methyl methanethiosulfonate (MTMS), H₂O₂, free L-methionine and BME. Methanol and acetic acid were purchased from VWR International, LLC (Radnor, PA) and Acros Organic N.V. (Geel, Belgium), respectively. Tris(2-carboxyethyl)phosphine, the Better Bradford Assay™ reagent and protease inhibitor reagent, which contains 1 mM AEBSF, 50 μM bestatin, 15 μM E64, 20 μM leupeptin, and 10 μM peptstatin A, were purchased from Thermo Scientific (South Logan, UT).

5.2.2 Sample Preparation

Preparation of the NAD(+) glutamate dehydrogenase (YDL215C) over-expressed yeast cell lysate, for the SPROX experiment and the iTRAQ labeling steps were performed by Patrick Dearmond. Briefly, Yeast cells were grown according to the previous reported protocol.[104] Yeast pellets from 250 mL cultures of the YDL215C yeast over-expression strains were lysed in 0.5 mL of 20 mM phosphate buffer (pH 7.4) containing protease inhibitors. The cell lysis step included 10 cycles of 20 s disruption
using 0.5 mm glass beads with 1 min intervals on ice in between. Lysates were centrifuged at 14,000 rpm for 5 min, and supernatants obtained were saved for later analysis. The total protein concentration in the final YDL215C supernatant was determined to be ~10-20 mg/mL using the Better Bradford Assay™ reagent.

SPROX buffers containing various concentrations of urea were prepared in 20 mM phosphate, pH 7.4. A 20 uL of yeast lysate sample was added into each of eight 75 uL SPROX buffers and equilibrated for 30 minutes. Protein oxidation was initiated by adding 5 uL 9.8 M H₂O₂ into the equilibrated lysate samples with final urea concentration of 0, 1.0, 2.0, 2.7, 3.3, 4.0, 4.9, and 6.0 M, respectively. After 6 minute oxidation, reaction was quenched by 950 μL 300 mM L-methionine solutions. TCA precipitation by adding 200 μL ice-cold 1 g/mL TCA to each sample was subsequently performed to remove urea and L-methionine compound. After incubated on ice overnight, protein pellets were collected by centrifuging at 8,000 x g for 30 min at 4°C and then washed by 300 uL ice-cold ethanol twice. Ultimately, pellets were obtained by removal of residue ethanol using a Savant SPD SpeedVac.

Each protein pellet was reconstituted in iTRAQ reaction buffers consisting of 30 uL 0.5 M triethylammonium bicarbonate and 1.5 μL 2% (w/w) SDS, pH 8.5. A series of vortexing and sonicating steps were used to facilitate dissolution of the protein pellets. After dissolution, 5 μL of each sample was aliquoted for a Bradford assay to
determine protein recovery following the TCA precipitation. The remainder of each sample were reduced with tris(2-carboxyethyl)phosphine, alkylated with MMTS, digested with trypsin, and labeled with one of the eight isobaric mass tags in the iTRAQ 8-plex. The manufacturer’s protocol was followed with the exception that 0.5 units of each labeling reagent were used instead of 1.0 unit. The 113, 114, 115, 116, 117, 118, 119, and 121 iTRAQ mass tags were used to label the tryptic peptide samples from SPROX reactions at the eight different urea concentrations (i.e., 0, 1.0, 2.0, 2.7, 3.3, 4.0, 4.9, and 6.0 M, respectively).

A 10 uL 0.6-0.7 mg/ml sample labeled with each iTRAQ mass tag was pooled together into an 80 uL single sample to serve as a control, which was not applied to met-peptide enhancement protocol. A 40 uL of the same sample from each iTRAQ labeling were pooled together into a 320 uL solution and subject to the enhancement step.

5.2.3. Met-peptide Enhancement

The methionine-reactive beads were treated with 200 uL methanol so the beads swelled and exposed the reactive groups, bromoacetyl groups. The SpeedVac was used to evaporate isopropanol from the pooled sample and reduce the volume to around 50-75 uL. Approximately 35 uL 100 % HOAc was added to the sample to adjust pH to 2-3. The resulting peptide solution was added to the pretreated beads to isolate met-peptide according to the manufacturer’s protocol. Briefly, peptides reacted with the attached
bromoacetyl groups on beads in a vortex manner for 1.5 hours at pH 2-3. 100 μL 0.2 BME in 25% acetic acid was subsequently added to the sample to remove non-covalent binders by vortex for 30 min. Beads were then extensively washed with four repeats of each washing solvent: 0.2 M BME followed by 70 % ACN/0.1% TFA and H2O. Met-peptides were released by adding 14 μL BME in an 86 μL 1M NH₄HCO₃ buffer at pH 9.0. The releasing step was performed in a vortex manner for 2 hours. Beads were then washed with 50 μL 20% ACN/0.1% TFA by vortexing for 15 minutes followed by 400 μL of 2 % TFA to maximize met-peptide recovery. All of eluted solution were combined into a 550 μL sample and diluted by adding 400 μL of 2 % TFA solution.

The resulting total of 950 μL met-enriched solution and non-met-enriched sample were treated with C18 silica resin prior to the LC-MS/MS analyses according to the manufacture protocol. Peptides were eluted from C18 silica by 50 μL 70 % ACN/0.1% TFA. The elution step was repeated 3 times to ensure no remaining peptides non-covalently attached to the silica. All the eluted solutions were combined together and concentrated to around 30 μL by removal the ACN using SpeedVac. Each sample was diluted by 120 μL 0.1% TFA and ultimately, was analyzed by LC-MS/MS to determine the amino acid sequence and iTRAQ reporter ion intensities for each labeled peptide.
5.2.4 LC-MS/MS-Based Proteomics

Samples were analyzed multiple times on the Agilent Q-TOF mass spectrometer with chip cube interface to generate six runs on met-enhanced sample and six runs on non-met-enhanced sample. MS/MS spectra were acquired in a data-dependent mode with HPCL gradient and instrumental parameters described in Chapter 2. Spectrum Mill software was used to merge the mass spectra derived from precursor ions with similar m/z values (i.e., ± 0.05 m/z) and similar retention times (i.e., ± 0.05 second) together. The resulting mass spectral data from enhanced data set was searched against tryptic met-peptides from the NCBI *saccharomyces cerevisiae* database with fixed amino acid modifications on MMTS modification of cysteine residues and iTRAQ8-plex on both lysine and N-termini. Oxidation on methionine and deamidation on asparagine and glutamine were set as variable modifications. The maximum number of trypsin missed cleavage sites allowed was 3. The maximum ambiguous precursor charge was 7. A 20 ppm mass tolerance was used for both precursor and product ion m/z values. Match filtering was disabled during the database search. The non-met-enhanced data set was analyzed using the same parameters described above with the exception of database searching against tryptic peptides.
5.2.5 Determination of protein expression level

Protein expression levels were manually searched on websites listed below: 1) http://www.ncbi.nlm.nih.gov/protein to locate the locus numbers for all proteins and 2) according to the locus numbers, protein expression level were searched from http://www.uniprot.org/

5.3 Results and Discussion

In typical bottom-up proteomics experiments, the confidence of peptide and eventually protein identification were indicated by the score values, which are based on two factors: 1) how close the mass values of a precursor ion and its product ions compared to the predicted mass values from the sequence database and 2) how many product ions for each precursor were identified. In Spectra Mill, a software used in this study, higher score indicates a more confident identification. Score cutoff values were determined according to the false positive rate (i.e. identification confidence), which suggests how many false identification in every 100 IDs. The false positive rate was obtained by searching mass spectral data against regular database and reversed sequence database, which was created by reversing the peptide sequence, resulting in the switch of N-terminal and C-terminal. For each score, its corresponding false positive rate is the number of hits identified by the reversed database divided by the number of hits identified by the regular database.
In SPROX experiment with LC-MS-based proteomics readout using isobaric mass tags, the intensity of each iTRAQ reporter ions for an identified peptide is associated with the relative amount of this peptide at corresponding denaturant concentration. If a peptide contains globally protected methionine residues, this peptide disappeared as the denaturant concentration increases. As a result, the intensity of each iTRAQ reporter ions reflects denaturant-dependant both disappearance of such peptide and appearance of its oxidized product. In order to obtain high quality denaturation curves, peptides that contain a total iTRAQ intensity sum of greater than 1000 counts and no more than one zero-intensity were included in the data analysis.

A total of 294 peptides were identified in the LC-MS/MS analyses with less than 1% false positive rate after the met-enrichment protocol. Approximately 12% of 294 peptides did not contain methionine residues and most likely were coming from the non-covalently binding to the met-resin. A total of 86 % of 294 were met-containing peptides including 52 oxidized-met-containing peptides. These oxidized met-containing peptides were either present because they non-specifically interacted with the resin or they were the oxidized products during the met-enhancement process. The remaining 208 un-oxidized-met-containing peptides were useful probes in a SPROX experiment.

Shown in Figure 12 are numbers of useful met-peptides and their corresponding proteins detected by SPROX with and without the met-enrichment protocol using LC-
MS/MS as readout when false positive rate less than 1%. The useful met-peptides are un-oxidized-met-peptides from enhanced sample and both oxidized and un-oxidized-met-peptides from non-enhanced samples. Note that 1.5-2 times more protein and peptides were detected after the met-enrichment protocol. It is noteworthy that using SPROX technique alone still generated 31 unique met-peptides that could not be identified by SPROX in the combination of the met-enrichment protocol. This phenomenon might be due to: 1) signal suppression, in which appearance of one peptide suppress the appearance of other peptides, and 2) steric restrictions, in which residues near the methionine residue prevent binding on the met-resin.

The expression levels of proteins that detected using SPROX before and after the combination of met-enhancement protocol were examined and the distribution in occurrence was shown in Figure 13. Figure 13A shows the distribution of protein expression level for 124 proteins detected after the enhancement protocol and 70 proteins detected before the protocol. Note that there is no significant difference in protein expression level between these two protocols. However, Figure 13B shows the protein expression level on unique proteins detected before and after using the enrichment protocol. Clearly, the proteins that were detected from both protocols are most abundant proteins (i.e. ~10⁶ molecules /cell in log phase SD medium). Compared with the expression level on overlapped proteins, the expression levels of unique
proteins detected before and after met-enrichment protocol shifted toward lower abundance with difference of $10^{4.5}$. Results indicated that the additional proteins which are useful for thermodynamic analysis using SPROX detected before and after the enhancement protocol are less abundant. However, only 25 unique proteins were detected in SPROX before the met-enrichment protocol and as a result, the determined distribution of expression level from non-enhanced samples might not be statistically significant.

Overall, the results reported here indicated that SPROX in combination of met-enhancement protocol can increase the number of peptides or proteins identified by using SPROX technique alone. It provides unique information on less abundant proteins as well. In addition, using SPROX without combination of met-enhancement protocol not only generated non-met-peptides, from which the iTRAQ intensity can be used for generation of normalization factors, but also formed unique met-peptides that can be not detected after the met-enrichment protocol.
Figure 11: Schematic representation of the SPROX in combination of solid-phase-based met-peptide-enrichment protocol
Figure 12: Number of met-peptides and proteins detected by SPROX method with and without combination of met-enhancement protocol. Peptide and protein identification was used with false positive rate less than 1% for each peptide, higher than 1000 on summed iTRAQ intensity and no more than one zero intensity from each iTRAQ report ion.
Figure 13: Distribution of protein expression level from proteins detected using met-enhancement protocol (open squares), proteins detected without using met-enhancement protocol (filled squares), unique proteins that were detected only before (filled circles) and after (open circles) using the enrichment protocol and proteins (open triangles) that were commonly detected in both.
6. Mass Spectrometry- and Lysine Amidination-Based Protocol for Protein Folding and Ligand Binding Analyses

6.1 Introduction

The introduction to this dissertation in Chapter 1 shows the importance of the protein-ligand binding studies on the proteomics scale to understanding the roles that proteins play in its biological context. SPROX has shown promise for the analysis of proteins in complex biological mixtures. [23, 24] In SPROX the chemical denaturant dependent oxidation rates of methionine side chains in a protein are used to determine the thermodynamic properties of the protein’s global and/or subglobal unfolding/refolding reactions. [23] A limitation of SPROX is the requirement for the protein(s) of interest to contain methionine residues at buried (i.e., solvent inaccessible) positions within its three-dimensional structure. The primary motivation for this work is to expand the scope of SPROX by developing a SPROX-like protocol using s-methyl thioacetimidate (SMTA) to measure the chemical denaturant dependent amidination rates of lysine side chains in a protein and evaluate its global and/or subglobal unfolding/refolding properties.

The utility of SMTA as a chemical probe of protein structure has been previously demonstrated. [105-108] In these earlier studies, which utilized continuous labeling protocols, SMTA was shown to be a useful probe of higher order protein structure. Reported here is an SMTA labeling protocol that can be used to probe the global
unfolding/refolding properties of proteins. Like the previously reported SPROX protocol, [23, 24] the SMTA labeling protocol utilizes a mass spectrometry readout to generate thermodynamic information about the chemical denaturant-induced equilibrium unfolding/refolding properties of proteins in solution. The main difference between the two protocols is that the SPROX protocol relies on the oxidation rates of globally protected methionine residues, whereas the SMTA labeling protocol relies on the amidination rates of globally protected lysine residues to generate such thermodynamic information. Because the two protocols rely on different modification reactions, they can provide complementary information, which can be particularly useful in proteomic-scale experiments.

The SMTA labeling protocol developed in this work is applied to a series of model protein systems in order to evaluate the protocol’s ability to generate quantitative thermodynamic information on protein folding and ligand binding interactions. The SMTA labeling protocol is also demonstrated to be compatible with three different mass spectrometry-based readouts, including an intact protein readout using MALDI, a gel-based proteomics readout using MALDI, and an LC-MS-based proteomics readout using isobaric mass tags.
6.2 Experimental Section

6.2.1 Materials

The following materials were purchased from Sigma-Aldrich (St. Louis, MO): des-Ac-α-Melanocyte Stimulating Hormone (lysine-containing peptide), ubiquitin from bovine red blood cells, carbonic anhydrase II from bovine erythrocytes (BCA II), cytochrome C from bovine heart, lysozyme from chicken egg white, ribonuclease A from bovine pancreas (RNase A), sequencing-grade trypsin, N-Acetyl-D-glucosamine (GlcNAc), thioacetamide, iodomethane, acetone, Anhydrous ethyl ether were purchased from EMD (Gibbstown, NJ). Sequencing-grade modified Glutamic-C endoproteinase was purchased from Princeton Separations (Freehold, NJ). Laemmli’s SDS-sample buffer (non-reducing), 15% gradient Tris-HCl ready gels and the Mini-Protean Tetra Cell used for gel electrophoresis were purchased from Bio-Rad Laboratories (Hercules, CA).

S-methyl thioacetimidate (SMTA) was synthesized from thioacetamide as described previously.[109] The lyophilized SMTA was stored under nitrogen at -20 °C. Stock solutions of SMTA (4 M) were prepared in water immediately prior to use.

Purified 4-oxalocrotonate tautomerase (4-OT) was kindly provided by Professor Christian P. Whitman (University of Texas-Austin). The 4-OT protein was over-expressed in E. coli and purified as described elsewhere.[110]
The yeast cell lysate sample was derived from two yeast GAL1 over expression strains: one in which cyclophilin A (YDR155C) was over expressed and one in which calcineurin A1 (YLR433C) was over expressed. Both of the over expression strains were purchased from Open Biosystems. Each over expression strain was cultured and lysed according to standard protocols. Briefly, yeast were grown overnight in 5 mL of a 2% dextrose SC-Ura solution. A 2 mL portion of this solution was used to inoculate 50 mL of a 2% raffinose SC-Ura solution and incubated overnight. This solution was diluted into 400 mL of 2% raffinose SC-Ura to give an O.D.600 of 0.3 and incubated until the O.D.600 of the solution was between 0.8 and 1.2 at which time 200 mL of 6% galactose in 3xYP solution was added. After six hours of incubation yeast pellets were obtained by centrifuging 250 mL fractions of this final solution.

Yeast pellets from 250 mL cultures of the YDR155C and YLR433C yeast over-expression strains were lysed in 0.5 mL of 20 mM phosphate buffer (pH 7.4) containing 1mM AEBSF, 50μM Bestatin, 15μM E64, 20μM Leupeptin, and 10μM Pepstatin A. In each case, the cell lysis was accomplished using glass beads (0.5mm) at 4°C with 20 s of disruption ten times with 1 min intervals on ice in between. Lysates were centrifuged at 14,000 rpm for 5 min, and supernatants obtained from a given strain were combined and saved for later analysis. The total protein concentrations in the final YDR155C and
YLR433C supernatants were determined to be ~10 mg/mL using the Better Bradford Assay™ reagent (from Thermo Scientific).

6.2.2 SMTA Labeling Protocol.

The SMTA labeling protocol outlined in Figure 14 was used to analyze 8 different protein samples including: six purified proteins, a five-component model protein mixture, and a yeast cell lysate. The six purified protein samples included ubiquitin, BCAII, RNaseA, 4OT, and lysozyme with, and without GlcNAc. Stock solutions containing approximately 2.5 mg/ml of each purified protein sample were prepared in H₂O and stored at -20 °C. The concentration of the GlcNAc ligand in the holo-lysozyme sample was 1.25 M, and the holo-lysozyme sample was heated to 45 °C for 30 minutes in order to ensure proper dissolution prior to analysis. The stock solution of the five-component model protein mixture that contained ubiquitin, lysozyme, BCAII, cytochrome C, and RNaseA, was prepared in water such that each protein was present at a concentration of 1 mg/ml. Equal volumes of the YDR155C and YLR433C lysate samples (see above) were combined to make the yeast cell lysate sample.

The SMTA labeling protocol was initiated by dilution of the above protein stock solutions into a series of SMTA reaction buffers which were comprised of 20 mM phosphate buffer (pH 7.4) and increasing concentrations of GdmCl for all the analyses except the RNaseA analysis in which the buffer was comprised of 20 mM HEPES (pH
7.4) and increasing concentrations of GdmCl. Stock solutions of each purified protein and the stock solution of the five-component protein mixture were diluted 10-fold into 50 µl volumes of the SMTA reaction buffers such that each reaction buffer contained ~12 µg of protein for the purified protein analyses and ~25 µg of total protein for the mixture analysis. The yeast lysate sample was diluted 5-fold into 50 µl volumes of the SMTA reaction buffers such that each reaction buffer contained ~100 µg of total protein. The pH of all the protein-containing SMTA reaction buffers was measured and adjusted, where needed, to pH 7.4 using NaOH.

The protein samples in the SMTA reaction buffers were equilibrated for 30 min at room temperature, with exception of the 4-OT samples, which were equilibrated overnight. The SMTA modification reaction was initiated with the addition of a 5 µl aliquot of a 4 M SMTA stock solutions. After a 66 min modification reaction time, which was the same for all samples, the protein samples were prepared for their respective mass spectral analyses, which ultimately terminated the amidination reaction. The purified protein samples were prepared for MALDI-TOF analysis by combining 1 µl of each protein-containing SMTA reaction buffer with 9 µl of a MALDI matrix solution and depositing 1 µl of the resulting solution on a stainless steel MALDI sample target and the solvent was evaporated.
The five-protein model mixture and the yeast cell lysate were prepared for MS analysis by first performing an acetone precipitation which involved the addition of 230 µl of H2O and 1.6 ml of cold (-20 °C) acetone to the protein sample in each SMTA reaction buffer before each sample was centrifuged at 14,000 x g for 20 min to precipitate the protein. The resulting protein pellets were washed three times with 2 mL cold (-20 °C) acetone. The protein pellets from the 5-protein model mixture samples were analyzed by the gel-based proteomics readout described below; and the protein pellets from the yeast lysate samples were analyzed by the LC-MS/MS-based proteomics readout, also described below.

SA was the MALDI matrix used for the MALDI analysis of the purified protein samples and the in-gel digestion products of BCA II. HCCA was the MALDI matrix used for the MALDI analysis of the in-gel digestion products of ubiquitin, RNase A, and lysozyme. Insulin and trypsin inhibitor were used as internal mass calibrants for the MALDI analyses in which the SA matrix was used. Insulin and bradykinin were used as internal mass calibrants for the MALDI analyses in which the HCCA matrix was used.

**6.2.3 Intact Protein Readout by MALDI-TOF.**

The extent of amidination following SMTA-modification was determined by taking the weighted average mass \( \text{mass}_{wt,av} \) of the singly charged protein ion signals corresponding to modified and wild-type protein species that were observed in the
MALDI mass spectrum. The weighted mass average was calculated from the following equation,

$$m/z_{wt,av} = \frac{\sum m_i I_i}{\sum I_i}$$  \hspace{1cm} \text{Equation 6.1}

where $I_i$ is the intensity of the ion signal of each species, and $m_i$ is the individual mass value for a singly charged protein peak. The weighted average, $mass_{wt,av}$, was calculated from a 9-point floating average smoothing of the data by an in-house Excel macro script that used a 2-point mass calibration based on the ion signals of internal mass standards. Ten spectra were collected at each GdmCl concentration. $mass_{wt,av}$ values derived from these spectra were averaged. The extent of modification (i.e., $\Delta mass_{wt,av}$) was determined by subtracting the mass of wild type protein from $mass_{wt,av}$. The calculated $\Delta mass_{wt,av}$ values were plotted against the GdmCl concentration. The data were then fit to a sigmoidal curve using a four-parameter sigmoid equation using a nonlinear regression routine in SigmaPlot (Systat Software, Inc., San Jose, CA) in order to determine $C^{1/2}$ value, which was the denaturant concentration at the transition midpoint.

The $\Delta mass_{wt,av}$ versus [GdmCl] data were also fit to equation 6.2

$$\Delta Mass_{wt,av} = \Delta M_\infty + (\Delta M_0 - \Delta M_\infty) e^{-[k_{int}/(1+K_{fold})]t}$$  \hspace{1cm} \text{Equation 6.2}

In equation 6.2, $\Delta M_0$ is the weighted average mass value of the pre-transition region of the sigmoidal curve, $\Delta M_\infty$ is the weighted average mass value of the post-transition region of the sigmoidal curve, $k_{int}$ is the intrinsic pseudo –first order rate constant for the
reaction of SMTA with an unprotected amine group in the protein, and $K_{\text{fold}}$ is defined by the expression, $$e^{\frac{\Delta G_f + n[GdmCl]}{RT}},$$ where $\Delta G_f$ is the protein folding free energy in the absence of chemical denaturant, $m$ is $\frac{\delta \Delta G_f}{\delta [GdmCl]}$, $R$ is the ideal gas constant, and $T$ is the temperature in Kelvin. In fitting the $\Delta \text{mass}_{\text{wt,av}}$ versus $[\text{GdmCl}]$ plots to equation 6.2 $\Delta M_0$, $\Delta M_\infty$, $\Delta G_f$, and $m$ were allowed to float and $k_{\text{int}}$ was assigned a value based on the exact SMTA concentration used in the labeling protocol and an experimentally determined second-order rate constant for the amidination reaction of an unprotected amine group with SMTA (see below).

### 6.2.4 Gel-Based Proteomics Readout

The protein pellets from the SMTA-labeling reactions involving the five component model mixture were reconstituted in 30 µL of Laemmli’s SDS-sample buffer containing 5% (v/v) β-mercaptoethanol. The proteins samples from each SMTA labeling reaction were separated by SDS-PAGE on 15% gradient Bis/Tris ready gels. Each liter of electrophoresis running buffer contained 3.03 g Tris base, 14.4 g Glycine and 1 g SDS. The gel was stained with Coomassie blue and the protein bands in the gel were excised and washed three times with 25 mM ammonium bicarbonate in 50 % ACN. The gel pieces from the gel bands were dehydrated by treatment with ACN followed by solvent removal using a Savant SPD SpeedVac (Thermo Scientific, South Logan, UT). The gel pieces were rehydrated in 25 mM phosphate buffer, reduced in 10 mM DTT, and
alkylated in 55 mM iodoacetamide according to the manufacturer’s protocol. The
protein-containing gel-pieces were digested with Glu-C protease and the resulting
peptides were subjected to a MALDI analysis to determine the extent of the SMTA
modification.

In the MALDI readout, ion signals that disappeared at high concentrated GdmCl
and appeared at the same denaturant concentration with the mass difference of 41 Da
were grouped together. The $\Delta\text{mass}_{\text{wt,av}}$ of these peaks was calculated according to
equation 6.1 and subsequently used to generate plots of $\Delta\text{mass}_{\text{wt,av}}$ versus GdmCl
concentration. The data was fit to a four-parameter sigmoidal equation using a nonlinear
regression routine in SigmaPlot (Systat Software, Inc., San Jose, CA), and the fit was
used to extract a $C_1/2$ value (i.e., the concentration of denaturant at the transition
midpoint) from each data set.

Peptide ion signals of interest were initially matched to corresponding peptide
sequences using ExPASy peptideMass at http://ca.expasy.org/tools/peptide-mass.html,
which matched the experimentally determined mass values of the peptide fragments
generated by Glu-C digestion to the theoretical mass values of those expected in the Glu-
C digestion of BCAII, ubiquitin, cytochrome C, lysozyme and RNase A in phosphate
buffer. Anticipated amino modifications included the oxidation of methionine, the
alkylation of cysteine, and the amidination of lysine or N-termini. Peptides extracted
from the protein gel bands at the lowest and highest concentrations of GdmCl were also analyzed on the Agilent 6520 Q-TOF LC/MS/MS mass spectrometer system (Agilent Technologies, Santa Clara, CA) in order to confirm the amino acid sequences of all peptide ion signals of interest that were identified in the MALDI-TOF readout.

6.2.5 LC-MS/MS-Based Proteomics Readout using Isobaric Mass Tags

The protein pellets from the SMTA-labeling reactions involving the yeast cell lysate were dissolved in 30 μL 0.5 M triethylammonium bicarbonate buffer containing 0.1% SDS, pH 8.5. The eight solutions were heated to 60 ºC for 10 minutes to facilitate dissolution of the protein pellets. The eight protein samples derived from the eight SMTA labeling reactions at the eight different GdmCl concentrations used in the experiment (i.e., 0.2, 0.8, 1.2, 1.7, 2.2, 2.8, 3.5, and 4.1 M GdmCl) were reduced, alkylated, digested with trypsin, and labeled with one of the eight isobaric mass tags in the iTRAQ 8-plex, all according to the manufacturer’s protocol with the exception that 0.5 units of each labeling reagent was used instead of 1.0 unit. The 113, 114, 115, 116, 117, 118, 119, and 121 iTRAQ mass tags were used to label the tryptic peptide samples from the SMTA labeling reactions at the eight different GdmCl concentrations (i.e., 0.2, 0.8, 1.2, 1.7, 2.2, 2.8, 3.5, and 4.1 M, respectively). A 10 μl aliquot of each iTRAQ-labeled sample was ultimately pooled into a 80 μL single sample. A Savant SPD speedvac was used to evaporate isoproponal from the pooled sample and reduce the volume to 10 μL. A 40 μL
 aliquot of 0.1% TFA solution was added to the sample. The sample was subsequently treated with TopTips™ using polySDS™ material (PolyLC Inc, Columbia, MD) to remove the SDS according to the manufacturer’s protocol. The SDS-depleted pooled sample was then analyzed by LC-MS/MS to determine the amino acid sequence and iTRAQ reporter ion intensities for each labeled peptide.

The pooled sample was analyzed three times using the data-dependent acquisition method described above. Peptides that were sequenced 3 or more times were used to generate an exclusion list. The sample was re-run and MS/MS data were acquired using the same data-dependent acquisition method except the exclusion list was used to avoid redundant peptide sequencing.

The mass spectral data generated in the LC-MS/MS analyses was searched against the NCBI *saccharomyces cerevisiae* database using Spectrum Mill (Agilent Technologies, Santa Clara, CA). Prior to searching the product ion mass spectra derived from precursor ions with similar m/z values (i.e., ± 0.05 m/z) and similar retention times (i.e., ± 0.05 second) were merged together. The parameters used for the database search where as follows: MMTS modification of cysteine residues and iTRAQ8-plex on both lysine and N-termini were set as fixed amino acid modifications. Oxidation on methionine was set as a variable modification. The maximum number of trypsin missed cleavage sites allowed was 3. The maximum ambiguous precursor charge was 7. A 20
ppm mass tolerance was used for both precursor and product ion m/z values. Match filtering was disabled during the database search. However, only peptides with Spectrum Mill scores higher than 6 were taken into the further data analysis described below. A Spectrum Mill score of greater than 6 yielded a false positive rate of less than 4%, as determined by the number of reversed database hits with a score of greater than 6.

Spectrum Mill was used to extract the iTRAQ reporter ion intensities of each identified peptide. The extracted iTRAQ reporter ion intensities were normalized using the same strategy employed in our previous SPROX study that utilized a similar isobaric mass tagging strategy. Briefly, the eight iTRAQ reporter ion intensities in each product ion mass spectra were averaged and the raw intensity value of each reporter ion in the product ion mass spectra was divided by the average intensity value obtained for that spectra. This process is referred to as the first normalization, N1. The N1-normalized intensities for the non-lysine-containing peptides were then used to generate so-called N2-normalization factors for the lysine-containing peptides. N2-normalization factors were generated by averaging all N1-normalized intensity values obtained for the non-lysine-containing peptides and each isobaric mass tag. The N2-normalization factors for each tag (and corresponding denaturant concentration) were 1.5, 1.2, 0.4, 1.0, 1.0, 1.1, 0.9 and 1.0 for the 113, 114, 115, 116, 117, 118, 119, and 121 reporter ions.
(respectively) in the iTRAQ 8-plex. Ultimately, these N2-normalization factors were used in a second so-called N2-normalization, in which the N1-normalized intensity values obtained for the lysine-containing peptide reporter ions were divided by the appropriate N2-normalization factor.

Between 1 and 3 sets of N2-normalized iTRAQ reporter ion intensity values were obtained for each of the 126 peptides listed in Table 7. In cases where multiple product-ion spectra were obtained for a given peptide (e.g., the peptide was sequenced in multiple LC-MS/MS runs), the N2-normalized intensities for each iTRAQ reporter ion were averaged to obtain a single set of data for the peptide. This averaging step was performed using an AWK script developed in-house.

The distributions of the N2-normalized intensities obtained for the 113 and 121 reporter ions from the lysine-containing peptides (Figure 22) were used to determine an N2 normalized reporter ion intensity that best separated the pre- and post transition baselines in the SMTA-derived denaturation curves. This N2 normalized reporter ion intensity was 1.1 (Figure 22). Lysine containing peptides with more than one N2-normalized reporter ion intensity that was not >1.1 or < 1.1 in the pre- and post-transition baselines (respectively) were deemed uninterpretable and not assigned a transition midpoint. In cases where only a single N2 normalized reporter ion intensity was inconsistent with the expected pre- and post baseline values, the intensity value was
not used for the denaturation curve construction and the transition midpoint assignment.

6.2.6 Amidination Rate Determinations

A lysine-containing peptide derived from des-Ac-α-melanocyte stimulating hormone and ubiquitin were used to determine the second-order rate constant for the amidination of unprotected free amine groups in protein. A series of des-Ac-α-melanocyte stimulating hormone solutions (0.05 mg/ml) were prepared in 20 mM phosphate buffers with concentrations of GdmCl ranging from 1 to 6 M. A solution of ubiquitin (23 μM) was also prepared in 20 mM phosphate buffer (pH 7.4) containing 6 M GdmCl. SMTA was added to the above ubiquitin and des-Ac-α-melanocyte stimulating hormone solutions such that the final SMTA concentration in each solution was 0.4 M SMTA. The time course of each amidination reaction was determined using a MALDI-TOF readout in which a Δmasswt,av value (see above) was determined at each time point (Figure 15). Pseudo-first-order reaction rate constants, (i.e. \( k_{\text{int}} \)) were extracted from the time course data by fitting the data to equation 6.3:

\[
\Delta\text{Mass}_{\text{wt,av}} = \Delta\text{Mass}_{\text{wt,av},0} + A (1-e^{-kt})
\]

Equation 6.3

In equation 6.3, \( \Delta\text{Mass}_{\text{wt,av},0} \) is the \( \Delta\text{Mass}_{\text{wt,av}} \) at \( t=0 \), \( A \) is the amplitude of the curve and \( k \) is the intrinsic pseudo-first-order rate constant of amidination at 0.4 M SMTA.
Eventually, the k value was divided by the [SMTA] to generate $k_{\text{int}}$, the second-order rate constant for the amidination of unprotected free amine groups in protein.

### 6.2.7 SUPREX Analysis

Two 1 mg/ml stock solutions of 4OT were prepared in 20 mM phosphate buffer (pH 7.4), one contained 0 M GdmCl and one contained 3 M GdmCl. The stock solutions were equilibrated overnight at room temperature to yield a stock solution of 4-OT in its folded and unfolded states. The folded and unfolded 4OT samples were amidinated with 0.4 M SMTA for 1 hour at room temperature and stored for three days at 4 °C during which time the excess SMTA reagent hydrolyzed. The two 4-OT samples were analyzed by SUPREX exactly as described previously.[111]

### 6.2.8 Kd Value Determination

The $K_d$ value for the lysozyme-GlcNAc complex was calculated using equation 2.3. In equation 2.3, N was assigned a value of 1, and $\Delta\Delta G_f$ was calculated by subtracting the folding free energy of the ligand bound protein from its unbound forms. The GlcNAc ligand concentration in our experiments (125 mM) was in excess of the lysozyme protein concentration (20 µM), therefore the free ligand concentration, [L], was taken as the total ligand concentration.
6.3 Results

6.3.1 Protocol Development

The initial step in the protocol developed here involves reacting equal aliquots of a protein sample with SMTA at a range of different denaturant concentrations. The specific SMTA concentration and reaction time used in the protocol, 0.4 M and 66 minutes (respectively), were chosen such that the pseudo-first order amidination reaction of an unprotected amine would proceed for approximately 6 half-lives. As part of this work a pseudo-first order reaction rate constant for the SMTA amidination of an unprotected amine, a $k_{int}$ value, was evaluated under the reaction conditions employed in our labeling protocol (20 mM phosphate buffer, pH 7.4, containing 0.4 M SMTA) using two different model systems: one including chemically denatured ubiquitin, one including an unstructured peptide (des-Ac-a-melanocyte stimulating hormone).

The average $k_{int}$ value determined from three replicate time course measurements on the ubiquitin sample was $0.118 \pm 0.003 \text{ min}^{-1}$ (Figure 15B) and this value was used to determine the 66 min reaction time in our SMTA labeling protocol. The pseudo-first-order rate constants determined for the peptide hormone at 1, 2, 3, 4, 5, and 6 M GdmCl were 0.29, 0.33, 0.27, 0.21, and 0.20 min$^{-1}$, respectively (Figure 15A). The consistency of these values demonstrates that the amidination reaction rate of an unprotected amine is not denaturant dependent.
6.3.2 Intact Protein Analysis

Five model protein systems (including zinc-loaded BCA II, ubiquitin, RNaseA, lysozyme, and 4-OT) were each analyzed using the SMTA labeling protocol and the intact protein MALDI readout outlined in Figure 14. Plots of $\Delta \text{Mass}_{\text{wt,av}}$ versus $[\text{GdmCl}]$ were generated for each protein Figure 16, and these plots were used to generate $C^{1/2}$, $\Delta G_r$, and $m$-values for each (Table 5). The lysozyme protein sample was analyzed both in the absence and in the presence of GlcNAc, which is known to bind lysozyme with $K_d$ values ranging from 20-60 mM at room temperature depending on the pH of the solution. [112-120] The amplitudes of the denaturation curves obtained for ubiquitin, BCA II, RNaseA, 4OT, and lysozyme (i.e., ~40, 80, 30, 20, 50 Da, respectively) are consistent with these proteins containing 1, 2, 1, 1, and 1 buried and/or partially buried reaction sites. X-ray crystallographic data available for these proteins suggest that K27 in ubiquitin (PDB 3M3J), K165 and K222 in BCA II (PDB 1V9E), K97 RNaseA (PDB 1Z6S), K6 in 4OT (PDB 3ABF), and the amino terminus in lysozyme (PDB 3IJU) are the buried sites in these proteins. The peptide data (see Gel-Based Proteomics section below) is also consistent with these sites being buried.

The $\Delta \text{Mass}_{\text{wt,av}}$ versus $[\text{GdmCl}]$ plots obtained for lysozyme in the absence and presence of GlcNAc (Figure 16E and F, respectively) show that the $C^{1/2}$ value increased by 0.5 M in the presence of the GlcNAc ligand. This increase in the presence of ligand is
consistent with the ligand-induced stabilization of lysozyme. The binding free energy, \( \Delta \Delta G_f \) value, which is the difference in \( \Delta G_{f,\text{avg}} \) between the ligand-bound and unbound forms of the protein was found to be 1 kcal/mol. This \( \Delta \Delta G_f \) value was subsequently used in equation 2.3 to calculate a \( K_d \) value of 27 mM for the lysozyme-GlcNAc complex, which is in reasonable agreement with the value of 40 mM that can be determined from the data in reference [112].

The \( \Delta G_f \) and \( m \)-values obtained for ubiquitin, lysozyme and RNase A using the SMTA labeling protocol were in reasonable agreement within 20% of corresponding values previously reported in the literature (Table 5). [121-123] However, in the case of 4OT and BCA II, the \( \Delta G_f \) and \( m \)-values obtained using the SMTA labeling protocol were significantly different from the previously reported values (Table 5), which were obtained using other techniques. [41, 111] One potential reason for the discrepancy is that the lysine amidination reaction may perturb the equilibrium unfolding reaction of these proteins and alter their thermodynamic stability. In order to evaluate the extent of such a perturbation in 4OT, SUPREX experiments were performed on two SMTA modified 4OT samples. One sample was prepared by reacting 4OT with SMTA in 20 mM phosphate buffer at pH 7.4, a condition in which the protein is known to be folded in its native three-dimensional structure. The second sample was prepared by reacting
the protein with SMTA in 20 mM phosphate buffer containing 3 M GdmCl at pH 7.4, a condition in which the protein is known to be unfolded.

The SMTA-modified 4-OT samples, which contained approximately 2 and 3 SMTA-modifications per monomer unit for the folded and unfolded sample (respectively), were analyzed by SUPREX according to established protocols.[111] The \( C^{1/2}_{\text{SUPREX}} \) value obtained for the 4OT sample subject to SMTA-modification under native solution conditions (i.e., the sample in which only exposed amine groups were amidinated) was 1.5 M (Figure 17A). This value is 0.6 M smaller than the value of 2.1 M that was previously obtained from the SUPREX experiment performed on wild-type 4OT under the same conditions.[111] No SUPREX transition (Figure 17B) was detected for the 4OT sample modified under denaturing solution conditions, in which all three free amine groups in the protein were amidinated. This suggests that the amidination of all the free amines in 4OT substantially destabilized the protein’s folded three-dimensional structure.

6.3.3 Gel-based Proteomics Analysis

A model protein mixture containing ubiquitin, BCA II, lysozyme, cytochrome C, and RNase A was analyzed using the gel-based proteomics readout and SMTA-labeling protocol outlined in Figure 14. As part of this protocol the protein samples from each denaturant-containing SMTA-labeling buffer were subject to a gel electrophoresis
separation step after the SMTA labeling reaction. Four gel bands were resolved in the separation (Figure 18); and these four bands were excised from each of the 12 lanes of the gel. They were each destained, reduced, alkylated, and treated with protease (Glu-C), in accordance with established gel-based proteomics protocols. The proteolytic peptides were extracted from the gel and analyzed by MALDI-TOF mass spectrometry.

A series of MALDI mass spectra were generated for the peptide samples obtained from the four protein bands in each of the gel lanes (Figure 18). The spectra were visually inspected to identify peptide ion signals that either disappeared or appeared as a function of the denaturant concentration (Figure 19A-C). Disappearing and appearing peptide ion signals that differed by 41 m/z were identified for analysis, as the expected SMTA modification mass was 41 amu. One pair of appearing/disappearing peptide ion signals was identified for each protein band, with the exception of the BCA II band, for which two pairs of appearing/disappearing ion signals were detected. The amino acid sequences of the peptide ions detected in each pair were determined initially by mapping their MALDI derived molecular weights to specific sequences in the five proteins in the test mixture (Table 6). These sequences were also confirmed via LC-MS/MS analyses of the peptides extracted from the protein bands in the lanes corresponding to the lowest and highest concentration of GdmCl.
The ion signals observed for the peptides in each group were used to obtain weighted mass average values for each peptide group at each denaturant concentration. The $\Delta$Mass$_{wt,av}$ values were plotted against the denaturant concentration, and the data for each peptide group (Figure 19D and Figure 20) were fit to a four-parameter sigmoid equation using a nonlinear regression routine in SigmaPlot to extract a $C_{1/2}$ value for each peptide (Table 6). The $C_{1/2}$ values extracted from the peptide analysis were within experimental error (i.e., ± 0.4 M) of the $C_{1/2}$ values extracted from the intact protein analysis (Table 5).

6.3.4 LC-MS-Based Proteomic Analysis using Isobaric Mass Tags

The proteins in a yeast cell lysate were analyzed using the SMTA labeling protocol and LC-MS-based proteomics readout outlined in Figure 14. As part of the LC-MS-based proteomics analysis, the protein pellets obtained from each SMTA-labeling buffer were redissolved in the iTRAQ reaction buffer, reduced, alkylated, and digested with trypsin, which cleaves peptides at arginine residues and non-SMTA-modified lysine residues. In this way, the lysine-containing peptides identified in the bottom-up proteomics strategy were primarily those derived from protected (i.e., solvent-inaccessible) regions of protein structure. Thus, the concentration of a given lysine-containing peptide was expected to be relatively high in the samples at low denaturant concentration (i.e. the amine group was protected from SMTA modification), but
relatively low at high denaturant concentration (i.e., the amine group was reacted with SMTA). Ultimately, the denaturant dependence of the concentration of a specific lysine-containing peptide reflects the global or sub-global stability of the protein structure from which it was derived.

The relative concentration of a specific lysine-containing peptide derived from the SMTA-modified protein samples in the eight-denaturant containing buffers was determined using an 8-plex of isobaric mass tags. The peptide fragments generated from the protein samples in the eight denaturant containing buffers were labeled with an 8-plex of iTRAQ reagents according to the manufacturer’s protocol. The iTRAQ reporter ion intensities generated in subsequent LC-MS/MS analyses were then used to quantify the degree of lysine modification in the protein samples originally reacted with SMTA in the denaturant containing buffers. This was accomplished by quantifying the disappearance of non-SMTA modified lysine-containing peptides detected in the proteomics readout as a function of the denaturant concentration in the SMTA reaction buffer. Because the SMTA-modification reaction was performed on the intact proteins in the denaturant-containing buffers prior to the protease digestion step, the degree of SMTA-modification on the lysine-containing peptides detected in the LC-MS/MS-based proteomics readout provides biophysical information on the proteins from which they were derived. Ultimately, the LC-MS/MS-based proteomics readout identified 126
different lysine-containing peptides, providing biophysical information on the folding behavior of 77 different proteins.

The iTRAQ data obtained for 112 of the 126 different lysine-containing peptides detected in this work yielded denaturation curves with clearly defined pre- and post transition baselines (see Experimental) and easily assigned transition midpoint values (e.g., Figure 21A). The distribution of transition midpoints recorded for these 112 peptides (Figure 23) ranged from 0 (i.e., all the normalized iTRAQ reporter ion intensities obtained for a given peptide were ~1) to >3 M. The other 14 lysine-containing peptides yielded iTRAQ data that did not have the expected structure of a denaturation curve (i.e., a pre- and/or post-transition baseline could not be established) (Figure 21B).

6.4 Discussion

6.4.1 Evaluation of Thermodynamic Parameters

The SMTA labeling protocol developed here is closely related to SPROX. Like SPROX, the SMTA labeling protocol can be used to evaluate the $\Delta G$ and $m$-values associated with protein folding reactions. As with SPROX, there are several assumptions associated with the evaluation of these values using the SMTA-based protocol. These assumptions are that: (1) the protein of interest must exhibit reversible, two-state folding/unfolding behavior such that only the folded or unfolded states are significantly populated in the chemical denaturant-induced equilibrium
unfolding/refolding reaction; (2) the unfolding/refolding reaction rate must be greater than the rate of the amidination reaction; (3) the amidinated protein must retain the thermodynamic stability of the wild-type protein; (4) the protein under study must contain a lysine residue and/or N-termini that is globally protected (i.e., solvent inaccessible) in the protein’s native three-dimensional structure. Use of the MALDI-MS readout in the SMTA labeling protocol also requires an assumption that the ionization efficiency of the amidinated protein/peptide does not differ significantly from that of the wild-type protein/peptide.

The first assumption is necessary for the determination of $\Delta G_f$ and $m$-values in any chemical denaturation experiment, even those that utilize more conventional optical spectroscopy-based methods. Many proteins, especially those that are large and multi-domains, do not exhibit two-state folding behavior. The SMTA labeling protocol cannot be used to generate meaningful $\Delta G_f$ and $m$-value for such proteins. However, in many cases the protocol can still be useful for evaluating the free energy associated with protein-ligand binding interactions (i.e., a $\Delta \Delta G_f$ value), just as SUPREX and SPROX techniques have proven useful for the measurement of $\Delta \Delta G_f$ values.[23, 25, 41] The main requirement for such $\Delta \Delta G_f$ measurements is that the ligand must interact only with the native state of the protein and not with the partially folded intermediate state(s) that may also be populated.
The gel- and LC-MS-based readouts are especially useful for the analysis of large multi-domain proteins using the SMTA labeling protocol described here. This is because the peptide readout enables the analysis of the protein-folding and ligand-binding behavior of individual domains to be probed. Because the amidination reaction is performed on the intact protein in the SMTA-labeling buffers prior to protease digestion, the amidination levels of the detected lysine-containing peptides provide biophysical information on the protein domain from which the peptides were derived. In the case of small single-domain proteins that display two-state folding/unfolding, all peptides containing lysines at buried positions in the protein structure will exhibit the same thermodynamic properties (i.e., have the same $C_{1/2}$ value) as the intact protein. This was observed, for example, in the intact protein and gel-based analyses of ubiquitin, BCAII, lysozyme, and RNaseA (Table 6). In the case of large multi-domain proteins, however, peptides from different domains may display different $C_{1/2}$ values. [24, 76] This was observed for some proteins in the yeast cell lysate. For example, one peptide from translational elongation factor EF-1α, EF-1α(225-245) of sequence TLLEAIDAIEQPSRPTDKPLR, yielded a $C_{1/2}$ value of 1.5 M, while another peptide from this protein, EF-1α(265-288) of sequence VETGVIKPGMVVFAPAGVTTEVK, yielded a $C_{1/2}$ value of 0.5 M (respectively).
It is important that the amidination modification not perturb the protein folding or the ligand binding interaction that is being probed. One attractive feature of the amidination reaction for studying the conformational properties of proteins is that it incorporates a relatively conservative modification that preserves the native charge state of the unmodified residue.\cite{124} Three of the five model protein systems analyzed here using the SMTA-labeling protocol (i.e., ubiquitin, lysozyme and RNase A) yielded $\Delta G_f$ values that were similar to the corresponding literature values (Table 5). The $\Delta G_f$ values determined by the SMTA-labeling protocol for BCA II and 4-OT were significantly lower than the corresponding literature values (Table 5). This discrepancy is most likely due to destabilizing effects associated with the amidination of these proteins. Indeed, our SUPREX analysis of amidinated 4-OT constructs showed that the amidination reaction perturbed the equilibrium unfolding reaction of the protein and lowered its thermodynamic stability.

The evaluation of $\Delta G_f$ values using equation 6.2 requires the assignment of the rate constant associated with the amidination reaction of the unprotected free amine groups in a protein. This rate constant (i.e., $k_{int}$) was assigned a value of 0.118 min$^{-1}$ on the basis of the reaction rate data obtained from unfolded ubiquitin. It is possible that $k_{int}$ values may vary as a function of amino acid sequence. Indeed, the same measurements made on the lysine containing peptide studied in this work yielded rate
constants that were ~2-fold higher than that measured for ubiquitin. However, we note that such small changes in $k_{int}$ (i.e., < 5-fold differences) would have minimal impact on the $\Delta G_f$ values calculated using equation 6.2, as the $k_{int}$ term is in the ln-term.

### 6.4.2 Complementarity to SPROX

We recently reported on the combination of SPROX and an LC-MS-based proteomics platform to make highly-multiplexed, thermodynamic measurements of protein folding and ligand binding in complex biological mixtures such as cell lysates.[24] The multiplex capabilities of SPROX were limited by the number of proteins that could be identified with a methionine-containing peptide in the LC-MS based proteomics readout. Ultimately, about one-third of the total numbers of proteins identified in the proteomics readout were amenable to thermodynamic analysis by SPROX. The main motivation for developing the SMTA labeling protocol described here was to increase the number of peptides that could be identified and used to probe the thermodynamic properties of proteins in such large-scale analyses of proteins in complex cell lysates. The yeast cell lysate results presented here suggest that the SMTA labeling protocol is indeed complementary to SPROX. A total of 99 of the 126 peptide probes detected in the analysis of the yeast cell lysate by SMTA-labeling did not contain a methionine residue and would not have been useful probes in a SPROX experiment. Moreover, approximately 80% of the 80 lysine-containing peptides that yielded
denaturation curves with transition midpoints >0 M in our SMTA-labeling experiment also did not contain a methionine residue.

In addition to identifying the unique peptides not previously identified by SPROX, the SMTA labeling facilitates the comparative thermodynamic analysis of proteins identified under both the SMTA- and SPROX-based protocols. A total of 13 of the 112 lysine-containing peptides that yielded high quality denaturation curves after SMTA-labeling in the yeast cell lysate sample also contained a methionine residue and had been detected in our earlier SPROX experiment that was performed on a similar yeast cell lysate sample. Of these 13 peptides shown in Table 7 from the supplemental material, 7 exhibited transition midpoints in that were reasonably consistent across both experiments (i.e., within 0.5 M units of each other). The more significant discrepancies observed between the transition midpoints of the other 6 overlapping peptides are likely due to the differential effects of the lysine amidination and methionine oxidation on the folding equilibrium of the proteins from which the peptides were derived.

The 126 lysine-containing peptides detected in the proteomic analysis of the yeast cell lysate represent only a fraction of the number of lysine-containing peptides that would be expected in a more conventional proteomic analysis using a bottom up approach. The only lysine-containing peptides expected to be identified in the proteomic analysis of our SMTA labeled lysate are those that contain lysine residues that
are either buried or at least partially buried in the three-dimensional structures of the proteins from which they are derived. A previous analysis of 418 protein crystal structures suggested that lysine residues are found to occur at buried and partially buried positions at a rate of approximately 6 and 26%.[125] Thus, the 126 lysine-containing peptide detected in the proteomic analysis of our SMTA labeled yeast cell lysate are likely to represent approximately 32% of the total number of lysine-containing peptides in a more conventional proteomics analysis.

6.5 Conclusion

Here we demonstrate that an SMTA modification protocol can be used to evaluate the thermodynamic parameters associated with protein folding and ligand binding interactions (e.g., $\Delta G_i$, $m$, and $K_d$ values) with reasonable accuracy, at least in ideal cases. The intrinsic chemical stability of the amidination products formed in the SMTA labeling protocol allows the protocol to be coupled with commonly used mass spectrometry-based proteomic strategies including both gel and gel-free platforms. The results of our cell lysate study also demonstrate that the SMTA labeling protocol developed here is indeed complementary to the methionine labeling protocol in SPROX. The combined use of SPROX and the SMTA labeling protocol in future thermodynamic analysis of protein folding and ligand binding on the proteomic scale can potentially expand the numbers of proteins and peptides that can be probed.
Figure 14: Schematic representation of the SMTA labeling protocol developed here.
Figure 15: Pseudo-first-order reaction curves obtained by reacting 0.4M SMTA with (A) des-Ac-a-Melanocyte stimulating hormone at different concentrations of GdmCl ranging from 1M to 6 M and (B) Ubiquitin in 6 M GdmCl. The lines represent the fitting of the data to a three-parameter single-exponential rise to max equation using SigmaPlot.
Figure 16: Representative data obtained using the SMTA-labeling protocol to analyze the thermodynamic stability of (A) ubiquitin, (B) zinc-loaded BCA II, (C) RNase A, (D) 4OT, (E) Apo-lysozyme, (F) lysozyme complex form with GlcNAc. The error bars represent +/- 1 standard deviation of the data determined from 10 replicate mass spectra. The lines represent the best fit of the data in each curve to a four-parameter sigmoid equation using SigmaPlot. The dotted lines mark the C_{1/2} value for each curve.
Table 5: Thermodynamic parameters derived from intact protein analysis. All values derived as part of this work represent the average and standard deviation of three replicate measurements.

<table>
<thead>
<tr>
<th>Proteins</th>
<th>$C_{1/2}$ (M)</th>
<th>$\Delta G_f$ (kCal/mol)</th>
<th>$m$ (kCal/(molM))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ubiquitin</td>
<td>3.1 ±0.1</td>
<td>-7.9±0.8</td>
<td>2.2±0.2</td>
</tr>
<tr>
<td>BCA II</td>
<td>1.3±0.1</td>
<td>-6±1</td>
<td>4±1</td>
</tr>
<tr>
<td>RNase A</td>
<td>3.0 ±0.1</td>
<td>-8±2</td>
<td>2.6±0.9</td>
</tr>
<tr>
<td>4OT</td>
<td>0.9 ±0.1</td>
<td>-3.4±0.4</td>
<td>3.0±0.4</td>
</tr>
<tr>
<td>Lysozyme</td>
<td>4.3±0.1</td>
<td>-10±3</td>
<td>2.0±0.6</td>
</tr>
<tr>
<td>Lysozyme + GlcNAc</td>
<td>4.8±0.1</td>
<td>-11±2</td>
<td>2.0±0.6</td>
</tr>
</tbody>
</table>

$^a$Value from reference [121]; $^b$Value from reference (X).[41] $^c$Value from reference [122]. $^d$Value from reference [111] $^e$Value from reference [123].
Figure 17: Representative SUPREX results on 4OT samples, which were reacted with the SMTA labeling reagent (0.4 M) at pH 7.4 in 20 mM phosphate buffer containing 0 and 3 M GdmCl in (A) and (B), respectively. The line in (A) represents the best fit of the data to a four-parameter sigmoid equation using SigmaPlot. The dotted line marks the $C^{1/2}$ value. The error bars represent +/- 1 standard deviation of the data determined from 10 replicate mass spectra.
Figure 18: SDS-PAGE gel obtained in the analysis of the five protein model mixture using the SMTA labeling protocol and the gel-based proteomics readout. Loaded in the 12 lanes are the protein samples that were labeled in the 12 denaturant containing buffers used in the SMTA labeling protocol.
Figure 19: Gel-based proteomics analysis of lysozyme (1-7) using the SMTA modification method. Representative MALDI mass spectra obtained are shown from our control experiment (A) and after SMTA modification in reaction buffers containing [GdmCl] of 0.8 and 6.1 M in (B) and (C), respectively. (D) The $\Delta$Mass$_{wt,av}$ versus [GdmCl] plot for the typical sigmoidal curves was generated for this peptide. The error bars represent +/- 1 standard deviation of the data determined from 10 replicate mass spectra. The dotted line marks the $C^{1/2}$ value, which is listed in Table 6.
Figure 20: Representative data obtained using the SMTA labeling protocol to analyze the five protein mixture using the gel-based proteomics readout. Shown are the denaturation curves obtained for (A) the ubiquitin (25-34) peptide, (B) the RNase A(87-111) peptide, (C) the BCAII(204-236) peptide, and (D) the BCAII(204-236) peptide with an oxidized methionine. The lines represent the best fit of the data to a four-parameter sigmoid equation using SigmaPlot. The error bars represent $\pm 1$ standard deviation of the values determined from 10 replicate mass spectra. The dotted lines mark the C$_{1/2}$ values which are summarized in Table 6.
Table 6: $C_{1/2}$ values determined for five peptides in the gel-based proteomic analysis of five-protein model mixture using SMTA labeling protocol.

<table>
<thead>
<tr>
<th>Protein (AA#)</th>
<th>Amino Acid Sequence$^a$</th>
<th>Average Mass (Da)</th>
<th>$C_{1/2}$ (M)$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Calculated</td>
<td>Observed</td>
</tr>
<tr>
<td>Ubiquitin (25-34)</td>
<td>NVKAKIQDKE</td>
<td>1173</td>
<td>N/O</td>
</tr>
<tr>
<td></td>
<td>[NVKAKIQDKE]$^*$</td>
<td>1214</td>
<td>N/O</td>
</tr>
<tr>
<td></td>
<td>[NVKAKIQDKE]$^{**}$</td>
<td>1255</td>
<td>1255±1</td>
</tr>
<tr>
<td></td>
<td>NVK<em>AK</em>IQDK*E</td>
<td>1296</td>
<td>1298±1</td>
</tr>
<tr>
<td>BCAII (204-236)</td>
<td>SVTWIVLKEPISVSSQQMLKFRTLNFNAEPE</td>
<td>3780</td>
<td>3780±2</td>
</tr>
<tr>
<td></td>
<td>[SVTWIVLKEPISVSSQQMLKFRTLNFNAEPE]$^*$</td>
<td>3821</td>
<td>3822±2</td>
</tr>
<tr>
<td></td>
<td>SVTWIVL<em>KEPISVSSQQMLK</em>FRTLNFNAEPE</td>
<td>3862</td>
<td>3864±2</td>
</tr>
<tr>
<td>BCAII (204-236)</td>
<td>SVTWIVLKEPISVSSQQmLKFRTLNFNAEPE</td>
<td>3796</td>
<td>3796±2</td>
</tr>
<tr>
<td></td>
<td>[SVTWIVLKEPISVSSQQmLKFRTLNFNAEPE]$^*$</td>
<td>3837</td>
<td>3837±2</td>
</tr>
<tr>
<td></td>
<td>SVTWIVL<em>K</em>EPISVSSQQmL<em>K</em>FRTLNFNAEPE</td>
<td>3878</td>
<td>3879±2</td>
</tr>
<tr>
<td>RNase A (87-111)</td>
<td>TGSSKYPncAYKTQANKHIIVAcE</td>
<td>2843</td>
<td>2838±1</td>
</tr>
<tr>
<td></td>
<td>[TGSSKYPncAYKTQANKHIIVAcE]$^*$</td>
<td>2884</td>
<td>2883±1</td>
</tr>
<tr>
<td></td>
<td>[TGSSKYPncAYKTQANKHIIVAcE]$^{**}$</td>
<td>2925</td>
<td>2924±1</td>
</tr>
<tr>
<td></td>
<td>TGSS<em>K</em>YPncAYK<em>TQANK</em>KHIIVAcE</td>
<td>2966</td>
<td>2965±1</td>
</tr>
<tr>
<td>Lysozyme (1-7)</td>
<td>KVFGRcE</td>
<td>896</td>
<td>895±1</td>
</tr>
<tr>
<td></td>
<td>K*VFGRcE</td>
<td>937</td>
<td>936±1</td>
</tr>
<tr>
<td></td>
<td><em>K</em>VFGRcE</td>
<td>978</td>
<td>978±1</td>
</tr>
</tbody>
</table>

$^a$K$^*$ indicates an amidinated lysine residue. A * symbol at the beginning of a peptide sequence indicates an amidinated N-terminus, and a bracketed peptide sequence with a * symbol(s) outside the brackets represents an amidinated peptide in which the exact site(s) of the lysine modification are unknown. A lowercase “m” represents an oxidized methionine, and a lowercase “c” represents an alkylated cysteine. $^b$ $C_{1/2}$ Values in parentheses are from the intact protein analyses summarized in Table 5.
Figure 21: Representative iTRAQ data obtained for 126 non-SMTA-modified-lysine peptides detected in yeast cell lysate. (A) A peptide yielding a denaturation curve with clearly defined pre- and post-transition baselines. (B) A peptide yielding iTRAQ data without a defined denaturation curve.
Figure 22: Distribution on normalized reporter ion intensity values from the lysine-containing peptides detected in the analysis of the yeast cell lysate sample using the LC-MS-based proteomics analysis with isobaric mass tags. The open and closed circles represented the data from the 113 and 121 tags, respectively. The dotted line marks the intensity cut-off value of 1.1 that was used to discriminate pre- and post-transition lines.
Figure 23: Pie chart with distribution of $C^{1/2}$ values obtained for 112 non-SMTA-modified-lysine peptides detected in yeast cell lysate that yield well-defined denaturation curves. In cases where the transition midpoint analysis generated two possible $C^{1/2}$ values, the higher value was used in the pie chart distribution shown.
Table 7: $C^{1/2}$ values determined in the LC/MS-based proteomic analysis for non-SMTA-modified-lysine peptides from an over-expressed yeast lysate sample modified by SMTA reagent

<table>
<thead>
<tr>
<th>Peptide Sequence</th>
<th>Protein Accession Number</th>
<th>Score</th>
<th>I-113</th>
<th>I-114</th>
<th>I-115</th>
<th>I-116</th>
<th>I-117</th>
<th>I-118</th>
<th>I-119</th>
<th>I-121</th>
<th>$C^{1/2}$ (M)</th>
<th>$C^{1/2}$ (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K)VVVFEDAPAGIAAGK(A)</td>
<td>557815</td>
<td>6.77</td>
<td>1.94</td>
<td>1.04</td>
<td>0.87</td>
<td>0.66</td>
<td>0.72</td>
<td>0.73</td>
<td>0.78</td>
<td>0.70</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>(K)AKGEVVLPGDFIIADAFSADANTK(T)</td>
<td>10383781</td>
<td>6.04</td>
<td>0.93</td>
<td>0.31</td>
<td>1.47</td>
<td>0.25</td>
<td>1.40</td>
<td>1.00</td>
<td>1.31</td>
<td>1.72</td>
<td>U</td>
<td>0.5</td>
</tr>
<tr>
<td>(K)ALENTRPFLAILGGAK(V)</td>
<td>10383781</td>
<td>7.44</td>
<td>2.11</td>
<td>0.74</td>
<td>0.86</td>
<td>0.66</td>
<td>0.68</td>
<td>0.84</td>
<td>0.76</td>
<td>0.73</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>(K)ALLDEVKK(S)</td>
<td>10383781</td>
<td>6.56</td>
<td>2.00</td>
<td>0.90</td>
<td>0.74</td>
<td>0.86</td>
<td>0.86</td>
<td>0.81</td>
<td>0.64</td>
<td>0.55</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>(K)ELPGVAELSEK(K)</td>
<td>10383781</td>
<td>8.86</td>
<td>2.42</td>
<td>0.74</td>
<td>0.57</td>
<td>0.77</td>
<td>0.81</td>
<td>0.49</td>
<td>0.62</td>
<td>0.64</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>(K)GVEVLPVFDDAFAIDSAMTOK(T)</td>
<td>10383781</td>
<td>9.86</td>
<td>0.33</td>
<td>0.93</td>
<td>1.17</td>
<td>0.81</td>
<td>1.20</td>
<td>1.08</td>
<td>0.96</td>
<td>1.80</td>
<td>U</td>
<td>0.5</td>
</tr>
<tr>
<td>(K)SSAAGLTVIIGGGDTATVAK(K)</td>
<td>10383781</td>
<td>9.65</td>
<td>1.49</td>
<td>0.88</td>
<td>1.12</td>
<td>0.77</td>
<td>0.55</td>
<td>1.33</td>
<td>0.80</td>
<td>0.87</td>
<td>NT*</td>
<td>NT*</td>
</tr>
<tr>
<td>(K)TVWNGPPGVFEFEK(P)</td>
<td>10383781</td>
<td>8.00</td>
<td>1.54</td>
<td>0.61</td>
<td>0.31</td>
<td>0.29</td>
<td>0.89</td>
<td>1.57</td>
<td>1.29</td>
<td>0.94</td>
<td>U</td>
<td>0.5</td>
</tr>
<tr>
<td>(R)JAAAGFLLEK(E)</td>
<td>10383781</td>
<td>6.39</td>
<td>2.16</td>
<td>0.89</td>
<td>0.60</td>
<td>0.76</td>
<td>0.76</td>
<td>0.65</td>
<td>0.68</td>
<td>0.71</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>(R)JVAALPTIK(Y)</td>
<td>10383781</td>
<td>8.32</td>
<td>1.35</td>
<td>0.90</td>
<td>0.62</td>
<td>0.81</td>
<td>0.93</td>
<td>1.29</td>
<td>1.04</td>
<td>0.73</td>
<td>0.5*</td>
<td>0.5</td>
</tr>
<tr>
<td>(R)VDFNVPDLGK(K)</td>
<td>10383781</td>
<td>8.62</td>
<td>1.16</td>
<td>1.14</td>
<td>0.75</td>
<td>1.01</td>
<td>0.95</td>
<td>1.12</td>
<td>0.84</td>
<td>0.79</td>
<td>1*</td>
<td>1*</td>
</tr>
<tr>
<td>(K)SIIGAITEFISK(L)</td>
<td>6321977</td>
<td>6.01</td>
<td>1.63</td>
<td>1.13</td>
<td>0.60</td>
<td>0.88</td>
<td>0.87</td>
<td>0.90</td>
<td>0.70</td>
<td>0.74</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>(R)GALDWPWPK(N)</td>
<td>6323887</td>
<td>8.38</td>
<td>1.23</td>
<td>1.26</td>
<td>1.82</td>
<td>1.13</td>
<td>0.56</td>
<td>0.72</td>
<td>0.97</td>
<td>0.58</td>
<td>1.9</td>
<td>1.9</td>
</tr>
<tr>
<td>(R)GINSYSYAVVWNTGK(A)</td>
<td>6323887</td>
<td>8.18</td>
<td>1.41</td>
<td>0.92</td>
<td>0.71</td>
<td>0.63</td>
<td>1.09</td>
<td>1.09</td>
<td>0.83</td>
<td>0.95</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>(R)VAPEHPPVLEAMPNPK(S)</td>
<td>3328</td>
<td>6.10</td>
<td>1.25</td>
<td>0.83</td>
<td>1.33</td>
<td>0.72</td>
<td>0.70</td>
<td>1.15</td>
<td>1.12</td>
<td>0.97</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>(R)LAAYHAQTEPIDFMYK(K)</td>
<td>3341</td>
<td>6.48</td>
<td>1.21</td>
<td>1.16</td>
<td>0.70</td>
<td>0.93</td>
<td>0.96</td>
<td>0.95</td>
<td>0.79</td>
<td>1.00</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(R)MVIGPGGAGK(G)</td>
<td>3341</td>
<td>9.71</td>
<td>1.50</td>
<td>1.05</td>
<td>0.63</td>
<td>0.81</td>
<td>0.84</td>
<td>1.03</td>
<td>0.84</td>
<td>0.84</td>
<td>0.5</td>
<td>1.3M</td>
</tr>
<tr>
<td>(K)EVSPTKPLDLVFLAASNLR(A)</td>
<td>6324514</td>
<td>10.08</td>
<td>1.85</td>
<td>0.70</td>
<td>0.98</td>
<td>0.74</td>
<td>1.04</td>
<td>0.72</td>
<td>0.71</td>
<td>0.83</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>(R)QETDVPPLVVPNPEHDLVAQQK(L)</td>
<td>6320362</td>
<td>8.13</td>
<td>1.17</td>
<td>1.17</td>
<td>0.89</td>
<td>0.99</td>
<td>0.97</td>
<td>0.76</td>
<td>0.89</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(K)LDSKQDSELVGLGGSTIR</td>
<td>6320044</td>
<td>7.46</td>
<td>1.71</td>
<td>0.92</td>
<td>1.03</td>
<td>0.84</td>
<td>0.75</td>
<td>1.08</td>
<td>0.66</td>
<td>0.65</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>(K)NEFTQQISMVFLGK(M)</td>
<td>6320044</td>
<td>8.05</td>
<td>1.06</td>
<td>1.36</td>
<td>0.81</td>
<td>0.54</td>
<td>0.57</td>
<td>0.47</td>
<td>1.04</td>
<td>0.60</td>
<td>1</td>
<td>1.4M</td>
</tr>
<tr>
<td>(R)GVPQIETFDVDSNGLNSVAEKG(G)</td>
<td>6320044</td>
<td>7.49</td>
<td>0.57</td>
<td>1.48</td>
<td>3.31</td>
<td>1.07</td>
<td>ND</td>
<td>0.42</td>
<td>ND</td>
<td>0.38</td>
<td>1.5*</td>
<td>1.5*</td>
</tr>
<tr>
<td>(R)VINPAAIAIYGLDK(K)</td>
<td>6320044</td>
<td>11.03</td>
<td>1.33</td>
<td>1.11</td>
<td>1.53</td>
<td>0.90</td>
<td>0.85</td>
<td>0.99</td>
<td>0.75</td>
<td>0.76</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>(R)VINPAAIAIYGLDK(G)</td>
<td>6320044</td>
<td>6.68</td>
<td>1.45</td>
<td>0.86</td>
<td>1.06</td>
<td>0.95</td>
<td>0.75</td>
<td>0.98</td>
<td>0.77</td>
<td>0.98</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>(R)GISLQIYPAVPLDSLK(S)</td>
<td>6322581</td>
<td>9.80</td>
<td>1.24</td>
<td>0.75</td>
<td>0.60</td>
<td>1.17</td>
<td>0.97</td>
<td>0.92</td>
<td>1.10</td>
<td>0.96</td>
<td>0.5*</td>
<td>0.5</td>
</tr>
<tr>
<td>(K)LVSITIEVAPGVTK(H)</td>
<td>6324328</td>
<td>11.02</td>
<td>1.90</td>
<td>1.16</td>
<td>1.15</td>
<td>0.52</td>
<td>1.45</td>
<td>0.24</td>
<td>0.26</td>
<td>0.86</td>
<td>1.5/2.5*</td>
<td>1.5/2.5*</td>
</tr>
<tr>
<td>(R)VINPAAAIAYGQLAGK(G)</td>
<td>6319972</td>
<td>9.47</td>
<td>1.44</td>
<td>0.95</td>
<td>0.71</td>
<td>0.92</td>
<td>0.76</td>
<td>1.07</td>
<td>0.86</td>
<td>0.92</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>(R)QRELXYVASHQTVTDPVLSK(L)</td>
<td>6319972</td>
<td>6.02</td>
<td>1.19</td>
<td>0.81</td>
<td>0.43</td>
<td>1.78</td>
<td>0.96</td>
<td>0.92</td>
<td>0.84</td>
<td>0.69</td>
<td>0.5*</td>
<td>0.5</td>
</tr>
<tr>
<td>(R)HSEFVAYPIQQLVTK(E)</td>
<td>6323840</td>
<td>8.91</td>
<td>1.19</td>
<td>1.00</td>
<td>1.28</td>
<td>0.89</td>
<td>0.71</td>
<td>1.10</td>
<td>1.01</td>
<td>0.86</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>(K)LAELKPVFQ(-)</td>
<td>6320720</td>
<td>8.43</td>
<td>2.26</td>
<td>0.90</td>
<td>0.30</td>
<td>0.39</td>
<td>0.61</td>
<td>0.78</td>
<td>0.77</td>
<td>0.99</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Peptide Sequence</td>
<td>Protein Accession Number</td>
<td>Score</td>
<td>I-113</td>
<td>I-114</td>
<td>I-115</td>
<td>I-116</td>
<td>I-117</td>
<td>I-118</td>
<td>I-119</td>
<td>I-121</td>
<td>C^12</td>
<td>C^13/14</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>------</td>
<td>---------</td>
</tr>
<tr>
<td>(K)HVFGGEVDGYDV</td>
<td>6320359</td>
<td>7.05</td>
<td>1.74</td>
<td>0.96</td>
<td>0.82</td>
<td>0.56</td>
<td>0.73</td>
<td>0.83</td>
<td>0.90</td>
<td>0.99</td>
<td>0.5</td>
<td>1.7</td>
</tr>
<tr>
<td>(K)AEVSMPTLLFY</td>
<td>6321648</td>
<td>7.15</td>
<td>1.90</td>
<td>1.02</td>
<td>0.80</td>
<td>0.88</td>
<td>0.82</td>
<td>0.72</td>
<td>0.66</td>
<td>0.61</td>
<td>0.5</td>
<td>1.4</td>
</tr>
<tr>
<td>(K)NVSAIPMPLLF</td>
<td>6320372</td>
<td>6.59</td>
<td>1.62</td>
<td>1.00</td>
<td>0.88</td>
<td>0.76</td>
<td>0.67</td>
<td>1.07</td>
<td>0.68</td>
<td>0.91</td>
<td>0.5</td>
<td>1.4</td>
</tr>
<tr>
<td>(K)VGAPNPAIQ</td>
<td>6320372</td>
<td>8.44</td>
<td>1.51</td>
<td>0.83</td>
<td>0.55</td>
<td>0.68</td>
<td>0.89</td>
<td>0.96</td>
<td>1.08</td>
<td>1.05</td>
<td>0.5</td>
<td>1.4</td>
</tr>
<tr>
<td>(K)TNACKNLQLVY</td>
<td>6323263</td>
<td>6.00</td>
<td>2.72</td>
<td>0.95</td>
<td>0.68</td>
<td>0.47</td>
<td>0.79</td>
<td>0.32</td>
<td>0.51</td>
<td>0.48</td>
<td>0.5</td>
<td>1.4</td>
</tr>
<tr>
<td>(R)TAYAGFLLA</td>
<td>6320593</td>
<td>8.51</td>
<td>1.36</td>
<td>1.00</td>
<td>1.07</td>
<td>1.01</td>
<td>1.03</td>
<td>0.83</td>
<td>0.76</td>
<td>0.78</td>
<td>0.5</td>
<td>1.4</td>
</tr>
<tr>
<td>(R)JETVEESSQ</td>
<td>6320593</td>
<td>7.69</td>
<td>1.39</td>
<td>1.15</td>
<td>0.73</td>
<td>0.79</td>
<td>0.91</td>
<td>1.01</td>
<td>0.98</td>
<td>0.70</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(K)AVYAGENFH</td>
<td>6321968</td>
<td>8.29</td>
<td>1.04</td>
<td>1.01</td>
<td>0.48</td>
<td>0.72</td>
<td>0.75</td>
<td>0.89</td>
<td>0.76</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(K)GVMNAN</td>
<td>6321968</td>
<td>6.74</td>
<td>0.80</td>
<td>1.27</td>
<td>0.20</td>
<td>0.99</td>
<td>0.91</td>
<td>0.47</td>
<td>1.65</td>
<td>1.44</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(R)KNRT</td>
<td>6324989</td>
<td>6.52</td>
<td>1.16</td>
<td>0.53</td>
<td>2.95</td>
<td>0.86</td>
<td>1.25</td>
<td>0.81</td>
<td>0.79</td>
<td>0.59</td>
<td>0.5</td>
<td>1.4</td>
</tr>
<tr>
<td>(K)IGNE</td>
<td>6322790</td>
<td>6.47</td>
<td>1.33</td>
<td>0.83</td>
<td>1.13</td>
<td>0.96</td>
<td>1.27</td>
<td>0.90</td>
<td>0.66</td>
<td>0.80</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(R)MP</td>
<td>6324993</td>
<td>6.45</td>
<td>2.12</td>
<td>0.83</td>
<td>0.50</td>
<td>0.69</td>
<td>0.54</td>
<td>0.96</td>
<td>0.68</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(K)KVTAPSSA</td>
<td>3724</td>
<td>8.54</td>
<td>3.01</td>
<td>0.71</td>
<td>0.56</td>
<td>0.48</td>
<td>0.41</td>
<td>0.49</td>
<td>0.37</td>
<td>0.56</td>
<td>0.5</td>
<td>0.75</td>
</tr>
<tr>
<td>(K)LVPELQGK</td>
<td>3724</td>
<td>8.00</td>
<td>3.25</td>
<td>0.82</td>
<td>0.58</td>
<td>0.69</td>
<td>0.71</td>
<td>0.65</td>
<td>0.59</td>
<td>0.72</td>
<td>0.5</td>
<td>0.75</td>
</tr>
<tr>
<td>(K)VVITAPSSA</td>
<td>3724</td>
<td>13.44</td>
<td>2.31</td>
<td>0.69</td>
<td>0.52</td>
<td>0.61</td>
<td>0.69</td>
<td>0.65</td>
<td>0.87</td>
<td>0.70</td>
<td>0.5</td>
<td>0.75</td>
</tr>
<tr>
<td>(R)TASGNIP</td>
<td>3724</td>
<td>7.42</td>
<td>1.73</td>
<td>0.92</td>
<td>0.61</td>
<td>0.69</td>
<td>0.89</td>
<td>0.98</td>
<td>0.80</td>
<td>0.81</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>(R)VVDL</td>
<td>3724</td>
<td>9.60</td>
<td>1.20</td>
<td>0.12</td>
<td>0.61</td>
<td>0.55</td>
<td>0.84</td>
<td>1.31</td>
<td>1.02</td>
<td>1.10</td>
<td>1.05</td>
<td>1</td>
</tr>
<tr>
<td>(K)VDLVEHVA</td>
<td>3724</td>
<td>6.45</td>
<td>1.85</td>
<td>0.85</td>
<td>0.51</td>
<td>0.63</td>
<td>0.80</td>
<td>0.97</td>
<td>0.98</td>
<td>0.75</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(R)PVTPDVS</td>
<td>3724</td>
<td>8.99</td>
<td>2.39</td>
<td>0.79</td>
<td>0.57</td>
<td>0.64</td>
<td>0.99</td>
<td>0.55</td>
<td>0.59</td>
<td>0.54</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>(R)DPANLP</td>
<td>3724</td>
<td>7.31</td>
<td>0.89</td>
<td>1.06</td>
<td>1.54</td>
<td>0.22</td>
<td>0.66</td>
<td>1.18</td>
<td>0.26</td>
<td>1.66</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(K)LIPS</td>
<td>6319673</td>
<td>10.05</td>
<td>2.29</td>
<td>1.63</td>
<td>1.15</td>
<td>0.54</td>
<td>0.45</td>
<td>0.35</td>
<td>0.38</td>
<td>0.51</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(K)VVDPET</td>
<td>6319673</td>
<td>8.10</td>
<td>2.02</td>
<td>1.41</td>
<td>0.68</td>
<td>0.53</td>
<td>0.62</td>
<td>0.49</td>
<td>0.82</td>
<td>0.70</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(R)ANKPM</td>
<td>6319673</td>
<td>8.10</td>
<td>1.44</td>
<td>0.50</td>
<td>0.85</td>
<td>0.99</td>
<td>1.06</td>
<td>0.64</td>
<td>1.49</td>
<td>0.82</td>
<td>0.5</td>
<td>1.4</td>
</tr>
<tr>
<td>(R)FLIA</td>
<td>6320662</td>
<td>3.02</td>
<td>1.09</td>
<td>1.03</td>
<td>1.46</td>
<td>0.84</td>
<td>1.02</td>
<td>0.89</td>
<td>1.04</td>
<td>0.82</td>
<td>0.5</td>
<td>1.4</td>
</tr>
<tr>
<td>(K)DFFNGKEP</td>
<td>6320950</td>
<td>7.81</td>
<td>0.68</td>
<td>0.64</td>
<td>0.89</td>
<td>0.95</td>
<td>0.94</td>
<td>1.36</td>
<td>1.26</td>
<td>1.45</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(K)KGGNP</td>
<td>14318578</td>
<td>8.78</td>
<td>2.39</td>
<td>1.15</td>
<td>0.31</td>
<td>0.75</td>
<td>0.43</td>
<td>0.38</td>
<td>0.82</td>
<td>0.70</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(K)KGGNP</td>
<td>6321184</td>
<td>6.02</td>
<td>1.91</td>
<td>0.97</td>
<td>1.48</td>
<td>1.30</td>
<td>0.49</td>
<td>0.21</td>
<td>0.75</td>
<td>0.66</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(K)FDP</td>
<td>3793</td>
<td>6.49</td>
<td>1.87</td>
<td>0.81</td>
<td>1.07</td>
<td>0.85</td>
<td>0.81</td>
<td>0.66</td>
<td>0.74</td>
<td>0.78</td>
<td>0.5</td>
<td>0.75</td>
</tr>
<tr>
<td>(K)TITLE</td>
<td>3722899</td>
<td>7.79</td>
<td>1.77</td>
<td>1.40</td>
<td>0.54</td>
<td>0.79</td>
<td>0.78</td>
<td>1.01</td>
<td>0.53</td>
<td>0.48</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(R)TE</td>
<td>3722899</td>
<td>6.34</td>
<td>1.41</td>
<td>1.21</td>
<td>0.32</td>
<td>0.82</td>
<td>0.90</td>
<td>1.17</td>
<td>0.72</td>
<td>0.85</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(R)KEGL</td>
<td>6323613</td>
<td>8.02</td>
<td>1.41</td>
<td>0.94</td>
<td>0.82</td>
<td>0.83</td>
<td>0.95</td>
<td>0.88</td>
<td>1.12</td>
<td>0.85</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Peptide Sequence</td>
<td>Protein Accession Number</td>
<td>Score (b)</td>
<td>I-113 (c)</td>
<td>I-114 (c)</td>
<td>I-115 (c)</td>
<td>I-116 (c)</td>
<td>I-117 (c)</td>
<td>I-118 (c)</td>
<td>I-119 (c)</td>
<td>I-121 (c)</td>
<td>C(^{22}) (M)</td>
<td>C(^{2}d) (M)</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------------------</td>
<td>-------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>(K)PE1LVSTDLESQK(T)</td>
<td>6320215</td>
<td>6.46</td>
<td>1.32</td>
<td>0.77</td>
<td>0.68</td>
<td>1.76</td>
<td>0.96</td>
<td>0.57</td>
<td>0.47</td>
<td>1.12</td>
<td>U</td>
<td>0.5</td>
</tr>
<tr>
<td>(R)EAAEYQLLA(R)</td>
<td>6319658</td>
<td>6.90</td>
<td>1.38</td>
<td>0.93</td>
<td>0.82</td>
<td>0.88</td>
<td>0.83</td>
<td>1.06</td>
<td>1.01</td>
<td>0.83</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>(K)VPAEDGGAGAGAHAALQK(R)</td>
<td>3793</td>
<td>10.96</td>
<td>2.16</td>
<td>0.74</td>
<td>0.58</td>
<td>0.58</td>
<td>0.28</td>
<td>0.79</td>
<td>0.81</td>
<td>0.59</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>(R)YNVQIQAGAGAHAATAGVQV(L)</td>
<td>6322599</td>
<td>7.50</td>
<td>1.38</td>
<td>1.23</td>
<td>1.22</td>
<td>1.28</td>
<td>0.58</td>
<td>0.79</td>
<td>0.81</td>
<td>0.59</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>(K)GIDLTVLPTDPTYSK(A)</td>
<td>4199</td>
<td>7.26</td>
<td>1.27</td>
<td>0.67</td>
<td>1.15</td>
<td>0.61</td>
<td>1.27</td>
<td>0.87</td>
<td>1.15</td>
<td>1.00</td>
<td>NT</td>
<td></td>
</tr>
<tr>
<td>(R)LALMGLENPNTPLK(S)</td>
<td>6321429</td>
<td>6.65</td>
<td>1.10</td>
<td>0.98</td>
<td>0.98</td>
<td>0.84</td>
<td>1.22</td>
<td>0.99</td>
<td>1.05</td>
<td>0.78</td>
<td>NT</td>
<td>0.25OX</td>
</tr>
<tr>
<td>(K)KLDLTPFAFVTMPMK(G)</td>
<td>6323073</td>
<td>9.71</td>
<td>2.29</td>
<td>0.75</td>
<td>1.27</td>
<td>0.51</td>
<td>0.69</td>
<td>0.80</td>
<td>0.51</td>
<td>0.63</td>
<td>0.5/1.5*</td>
<td>0.9OX</td>
</tr>
<tr>
<td>(K)LIDLTQFPAFVTMPMK(G)</td>
<td>6323073</td>
<td>10.29</td>
<td>1.46</td>
<td>0.79</td>
<td>1.21</td>
<td>0.79</td>
<td>0.65</td>
<td>0.57</td>
<td>0.87</td>
<td>1.26</td>
<td>NT</td>
<td></td>
</tr>
<tr>
<td>(R)TYTYQGRPVYVLPLNVLNVPK(L)</td>
<td>6323073</td>
<td>9.83</td>
<td>1.43</td>
<td>0.89</td>
<td>0.81</td>
<td>0.62</td>
<td>0.83</td>
<td>1.02</td>
<td>1.07</td>
<td>1.04</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>(K)LINENEKVPDPK(Q)</td>
<td>6323341</td>
<td>7.37</td>
<td>1.33</td>
<td>0.87</td>
<td>0.53</td>
<td>1.12</td>
<td>0.82</td>
<td>0.90</td>
<td>0.96</td>
<td>1.07</td>
<td>0.5*</td>
<td></td>
</tr>
<tr>
<td>(K)VTGAGGGIGQPLSLLK(L)</td>
<td>6322765</td>
<td>6.30</td>
<td>1.39</td>
<td>1.08</td>
<td>0.90</td>
<td>0.52</td>
<td>0.90</td>
<td>0.65</td>
<td>1.12</td>
<td>1.20</td>
<td>NT</td>
<td></td>
</tr>
<tr>
<td>(R)DVIESFDVFDLFK(S)</td>
<td>6322766</td>
<td>6.76</td>
<td>1.42</td>
<td>1.31</td>
<td>0.67</td>
<td>0.76</td>
<td>0.82</td>
<td>0.99</td>
<td>0.78</td>
<td>0.60</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(R)AQFETLAPVLK(R)</td>
<td>6322505</td>
<td>7.11</td>
<td>1.29</td>
<td>1.31</td>
<td>0.67</td>
<td>0.67</td>
<td>0.97</td>
<td>0.73</td>
<td>1.08</td>
<td>0.94</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(R)QAVNVPLETLP(K)</td>
<td>6322505</td>
<td>7.11</td>
<td>1.29</td>
<td>1.31</td>
<td>1.09</td>
<td>0.91</td>
<td>0.56</td>
<td>1.63</td>
<td>0.61</td>
<td>1.12</td>
<td>NT*</td>
<td></td>
</tr>
<tr>
<td>(K)LVLESSGEIAKP(Q)</td>
<td>6319517</td>
<td>6.52</td>
<td>0.90</td>
<td>0.95</td>
<td>2.59</td>
<td>1.21</td>
<td>0.98</td>
<td>0.55</td>
<td>0.75</td>
<td>0.89</td>
<td>NT*</td>
<td></td>
</tr>
<tr>
<td>(R)FKNGLITYGHVLNSVNMFD(D)</td>
<td>914198</td>
<td>6.47</td>
<td>0.32</td>
<td>1.53</td>
<td>3.33</td>
<td>1.43</td>
<td>0.29</td>
<td>1.29</td>
<td>0.10</td>
<td>1.07</td>
<td>U</td>
<td></td>
</tr>
<tr>
<td>(R)SAYLAAPVIAALIK(T)</td>
<td>6320702</td>
<td>11.07</td>
<td>1.74</td>
<td>1.78</td>
<td>0.84</td>
<td>0.93</td>
<td>0.64</td>
<td>1.05</td>
<td>0.77</td>
<td>0.72</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>(K)TRKAAIARQKVASLTVPTVILLAG(F)</td>
<td>6323567</td>
<td>11.07</td>
<td>1.01</td>
<td>1.44</td>
<td>4.18</td>
<td>ND</td>
<td>0.61</td>
<td>0.59</td>
<td>1.57</td>
<td>0.50</td>
<td>U</td>
<td></td>
</tr>
<tr>
<td>(K)NFLTVELQVGLK(N)</td>
<td>6321303</td>
<td>7.97</td>
<td>1.31</td>
<td>1.22</td>
<td>0.55</td>
<td>0.88</td>
<td>0.74</td>
<td>1.26</td>
<td>0.82</td>
<td>0.82</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(R)AAERFDPK(S)</td>
<td>6323777</td>
<td>6.80</td>
<td>0.74</td>
<td>1.03</td>
<td>2.01</td>
<td>1.08</td>
<td>1.19</td>
<td>0.83</td>
<td>0.88</td>
<td>0.86</td>
<td>NT*</td>
<td></td>
</tr>
<tr>
<td>(K)SKDFVVIFIAQ(D)</td>
<td>6320814</td>
<td>7.03</td>
<td>1.49</td>
<td>0.79</td>
<td>1.14</td>
<td>0.69</td>
<td>1.10</td>
<td>0.78</td>
<td>0.98</td>
<td>0.88</td>
<td>0.5/1.5*</td>
<td></td>
</tr>
<tr>
<td>(R)JSPDGATTPFPK(R)</td>
<td>6324194</td>
<td>6.96</td>
<td>0.92</td>
<td>1.21</td>
<td>0.41</td>
<td>0.86</td>
<td>1.12</td>
<td>1.16</td>
<td>1.05</td>
<td>0.91</td>
<td>NT*</td>
<td></td>
</tr>
<tr>
<td>(R)GSSLYTYLINDAGK(A)</td>
<td>6323357</td>
<td>6.83</td>
<td>1.10</td>
<td>1.27</td>
<td>1.12</td>
<td>1.14</td>
<td>0.83</td>
<td>0.48</td>
<td>0.67</td>
<td>1.34</td>
<td>1.9*</td>
<td></td>
</tr>
<tr>
<td>(R)PFAASGDVMV(N)</td>
<td>6319384</td>
<td>9.01</td>
<td>1.50</td>
<td>1.17</td>
<td>1.00</td>
<td>0.88</td>
<td>0.99</td>
<td>0.68</td>
<td>0.68</td>
<td>0.65</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(K)ASLGAGVGLPLSVEITGP(N)</td>
<td>6322171</td>
<td>8.08</td>
<td>1.31</td>
<td>0.83</td>
<td>1.13</td>
<td>0.75</td>
<td>0.69</td>
<td>1.18</td>
<td>0.44</td>
<td>1.32</td>
<td>NT*</td>
<td></td>
</tr>
<tr>
<td>(R)REIIFIANEVHTGQFIYAGK(K)</td>
<td>6322171</td>
<td>7.73</td>
<td>0.91</td>
<td>0.88</td>
<td>0.55</td>
<td>0.80</td>
<td>1.52</td>
<td>0.98</td>
<td>0.81</td>
<td>1.36</td>
<td>NT*</td>
<td></td>
</tr>
<tr>
<td>(R)DGKAPERAEVLINAAATIFQPATDVEAR(A)</td>
<td>6323791</td>
<td>10.93</td>
<td>1.19</td>
<td>0.84</td>
<td>1.15</td>
<td>1.02</td>
<td>1.12</td>
<td>0.81</td>
<td>0.82</td>
<td>1.06</td>
<td>NT</td>
<td></td>
</tr>
<tr>
<td>(R)YLTVDFEAFK(S)</td>
<td>6321798</td>
<td>6.63</td>
<td>1.07</td>
<td>1.24</td>
<td>1.05</td>
<td>0.92</td>
<td>1.24</td>
<td>0.86</td>
<td>0.93</td>
<td>0.64</td>
<td>NT</td>
<td></td>
</tr>
<tr>
<td>(K)TLEAIDAIEQPSRDKPLR(L)</td>
<td>6319594</td>
<td>9.15</td>
<td>1.55</td>
<td>1.22</td>
<td>1.17</td>
<td>0.78</td>
<td>1.00</td>
<td>0.94</td>
<td>0.49</td>
<td>0.59</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>(R)AIPPIEGERDDAQQGSGTGK(T)</td>
<td>6322323</td>
<td>7.86</td>
<td>1.28</td>
<td>1.06</td>
<td>1.10</td>
<td>0.94</td>
<td>1.01</td>
<td>0.84</td>
<td>0.89</td>
<td>0.78</td>
<td>0.5</td>
<td>1.4 M</td>
</tr>
</tbody>
</table>

\(b\), \(c\), \(d\) = Scores, and \(U\) = Unidentified.
<table>
<thead>
<tr>
<th>Peptide Sequence</th>
<th>Protein Accession Number&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Score&lt;sup&gt;b&lt;/sup&gt;</th>
<th>I-113&lt;sup&gt;c&lt;/sup&gt;</th>
<th>I-114&lt;sup&gt;c&lt;/sup&gt;</th>
<th>I-115&lt;sup&gt;c&lt;/sup&gt;</th>
<th>I-116&lt;sup&gt;c&lt;/sup&gt;</th>
<th>I-117&lt;sup&gt;c&lt;/sup&gt;</th>
<th>I-118&lt;sup&gt;c&lt;/sup&gt;</th>
<th>I-119&lt;sup&gt;c&lt;/sup&gt;</th>
<th>I-121&lt;sup&gt;c&lt;/sup&gt;</th>
<th>C&lt;sup&gt;1/2&lt;/sup&gt; (M)</th>
<th>C&lt;sup&gt;1/2d&lt;/sup&gt; (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R)VETGVIKPGVMVFAPAGVTEVK(S)</td>
<td>6319594</td>
<td>12.00</td>
<td>1.59</td>
<td>0.72</td>
<td>0.59</td>
<td>0.73</td>
<td>0.77</td>
<td>0.80</td>
<td>0.99</td>
<td>1.19</td>
<td>0.5*</td>
<td>1&lt;sub&gt;M&lt;/sub&gt;</td>
</tr>
<tr>
<td>(K)IQDKEGIPPDQQR(L)</td>
<td>6322989</td>
<td>6.16</td>
<td>1.38</td>
<td>1.17</td>
<td>0.46</td>
<td>0.63</td>
<td>0.91</td>
<td>1.22</td>
<td>0.62</td>
<td>1.10</td>
<td>1&lt;sup&gt;*&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>(R)YYDLEAPVQIVK(R)</td>
<td>6321772</td>
<td>7.60</td>
<td>1.25</td>
<td>0.93</td>
<td>1.11</td>
<td>1.04</td>
<td>1.13</td>
<td>0.92</td>
<td>0.72</td>
<td>0.81</td>
<td>NT</td>
<td></td>
</tr>
<tr>
<td>(R)GLSAVAQAESMK(F)</td>
<td>6324151</td>
<td>6.29</td>
<td>1.27</td>
<td>0.76</td>
<td>1.21</td>
<td>1.18</td>
<td>1.10</td>
<td>0.81</td>
<td>0.95</td>
<td>0.70</td>
<td>1.9*</td>
<td>1.4&lt;sub&gt;M&lt;/sub&gt;</td>
</tr>
<tr>
<td>(R)GGISIVASPQVAK(K)</td>
<td>6321315</td>
<td>6.76</td>
<td>1.31</td>
<td>0.99</td>
<td>0.90</td>
<td>0.94</td>
<td>0.99</td>
<td>0.79</td>
<td>0.72</td>
<td>0.5*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R)GTDWQSEGKANGMNRAK(N)</td>
<td>6321919</td>
<td>6.55</td>
<td>0.82</td>
<td>0.85</td>
<td>3.09</td>
<td>0.76</td>
<td>1.43</td>
<td>0.74</td>
<td>0.76</td>
<td>0.68</td>
<td>U</td>
<td></td>
</tr>
<tr>
<td>(R)KNRRyMNPGEVSFK(S)</td>
<td>6323780</td>
<td>6.32</td>
<td>1.54</td>
<td>0.92</td>
<td>0.38</td>
<td>0.25</td>
<td>1.99</td>
<td>1.10</td>
<td>0.46</td>
<td>0.73</td>
<td>0.5*</td>
<td></td>
</tr>
<tr>
<td>(K)gLNPMPMEATQAAAPTIF(K)</td>
<td>6323208</td>
<td>8.25</td>
<td>1.30</td>
<td>0.66</td>
<td>0.95</td>
<td>0.88</td>
<td>1.31</td>
<td>0.98</td>
<td>0.91</td>
<td>0.86</td>
<td>0.5*</td>
<td></td>
</tr>
<tr>
<td>(R)AVDITWNFNIKT(T)</td>
<td>6324694</td>
<td>8.11</td>
<td>1.17</td>
<td>0.85</td>
<td>1.00</td>
<td>1.38</td>
<td>1.03</td>
<td>0.84</td>
<td>0.67</td>
<td>0.96</td>
<td>0.5*</td>
<td></td>
</tr>
<tr>
<td>(K)GMEILGRLGKIIIFGAWGIIR(E)</td>
<td>6322677</td>
<td>7.61</td>
<td>1.38</td>
<td>0.79</td>
<td>0.88</td>
<td>1.24</td>
<td>0.52</td>
<td>1.43</td>
<td>0.80</td>
<td>0.69</td>
<td>NT*</td>
<td></td>
</tr>
<tr>
<td>(K)QVQFmYKTRNQHQQEKK(K)</td>
<td>6320924</td>
<td>6.31</td>
<td>1.34</td>
<td>0.50</td>
<td>0.84</td>
<td>3.11</td>
<td>1.19</td>
<td>1.45</td>
<td>0.63</td>
<td>1.47</td>
<td>U</td>
<td></td>
</tr>
<tr>
<td>(R)GDLGIEPAPEVLAVQK(K)</td>
<td>6319279</td>
<td>13.97</td>
<td>1.71</td>
<td>1.11</td>
<td>0.92</td>
<td>0.89</td>
<td>0.72</td>
<td>1.19</td>
<td>0.80</td>
<td>0.72</td>
<td>0.63</td>
<td>0.5*</td>
</tr>
<tr>
<td>(K)ETAPVTDLAAGYK(I)</td>
<td>132847</td>
<td>6.41</td>
<td>1.69</td>
<td>1.14</td>
<td>0.67</td>
<td>0.91</td>
<td>0.70</td>
<td>0.88</td>
<td>0.80</td>
<td>0.66</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(R)IPNVPVVK(A)</td>
<td>132847</td>
<td>8.26</td>
<td>1.43</td>
<td>1.11</td>
<td>1.05</td>
<td>0.86</td>
<td>0.86</td>
<td>0.87</td>
<td>0.67</td>
<td>0.91</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(K)VVGLSTLPEIYK(M)</td>
<td>1168350</td>
<td>8.93</td>
<td>1.54</td>
<td>0.78</td>
<td>0.58</td>
<td>0.76</td>
<td>0.96</td>
<td>1.09</td>
<td>0.98</td>
<td>0.85</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>(R)SJGVEVFDTKE(K)</td>
<td>1168350</td>
<td>6.65</td>
<td>1.75</td>
<td>0.62</td>
<td>0.62</td>
<td>0.90</td>
<td>0.90</td>
<td>0.81</td>
<td>0.95</td>
<td>0.76</td>
<td>0.88</td>
<td>0.5</td>
</tr>
<tr>
<td>(K)LVEDPOQIAFLGK(L)</td>
<td>119180</td>
<td>7.52</td>
<td>0.88</td>
<td>1.24</td>
<td>1.89</td>
<td>1.42</td>
<td>0.31</td>
<td>1.09</td>
<td>0.91</td>
<td>0.73</td>
<td>1.9*</td>
<td></td>
</tr>
<tr>
<td>(K)AVDPLSLDQANK(T)</td>
<td>119336</td>
<td>9.25</td>
<td>1.52</td>
<td>1.02</td>
<td>0.95</td>
<td>0.80</td>
<td>0.66</td>
<td>0.84</td>
<td>0.73</td>
<td>0.72</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>(R)GNPTVEELITTEK(G)</td>
<td>119336</td>
<td>10.50</td>
<td>1.03</td>
<td>0.65</td>
<td>1.58</td>
<td>0.99</td>
<td>1.29</td>
<td>0.85</td>
<td>0.95</td>
<td>0.95</td>
<td>NT*</td>
<td></td>
</tr>
<tr>
<td>(R)JALVEHIIQAK(A)</td>
<td>602897</td>
<td>7.04</td>
<td>1.22</td>
<td>0.66</td>
<td>1.01</td>
<td>0.99</td>
<td>1.01</td>
<td>1.16</td>
<td>0.86</td>
<td>1.00</td>
<td>0.5*</td>
<td></td>
</tr>
<tr>
<td>(R)AIQVFEGTSIDQVK(K)</td>
<td>6319603</td>
<td>6.09</td>
<td>1.11</td>
<td>1.15</td>
<td>1.12</td>
<td>0.61</td>
<td>1.34</td>
<td>0.61</td>
<td>0.57</td>
<td>3.1*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R)LEPKRWWWYTFLAVLTFR(C)</td>
<td>6324623</td>
<td>6.52</td>
<td>1.37</td>
<td>1.31</td>
<td>0.77</td>
<td>0.95</td>
<td>0.53</td>
<td>0.82</td>
<td>1.10</td>
<td>0.84</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(K)GVETLAEAAATLPGK(G)</td>
<td>6323288</td>
<td>9.21</td>
<td>0.85</td>
<td>0.63</td>
<td>1.68</td>
<td>0.75</td>
<td>0.97</td>
<td>1.13</td>
<td>1.09</td>
<td>1.36</td>
<td>NT*</td>
<td></td>
</tr>
<tr>
<td>(R)AAEVEGILPGGTTALVK(A)</td>
<td>6323288</td>
<td>8.72</td>
<td>1.14</td>
<td>0.91</td>
<td>1.06</td>
<td>1.77</td>
<td>1.01</td>
<td>0.77</td>
<td>0.90</td>
<td>0.90</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(R)GFI5YPFITDPK(S)</td>
<td>6323288</td>
<td>7.92</td>
<td>1.06</td>
<td>0.94</td>
<td>1.12</td>
<td>1.13</td>
<td>1.03</td>
<td>0.75</td>
<td>0.92</td>
<td>1.09</td>
<td>NT</td>
<td></td>
</tr>
<tr>
<td>(K)LNFVGGWDDVSK(L)</td>
<td>632697</td>
<td>6.75</td>
<td>1.83</td>
<td>0.82</td>
<td>0.61</td>
<td>0.98</td>
<td>0.66</td>
<td>0.78</td>
<td>0.82</td>
<td>0.88</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>(R)AJAHTANIKL(A)</td>
<td>632697</td>
<td>7.56</td>
<td>1.35</td>
<td>0.87</td>
<td>0.74</td>
<td>1.00</td>
<td>0.98</td>
<td>0.96</td>
<td>0.91</td>
<td>0.90</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>(R)LJLPYQDVQIK(D)</td>
<td>632697</td>
<td>9.24</td>
<td>1.79</td>
<td>1.09</td>
<td>0.69</td>
<td>0.66</td>
<td>0.80</td>
<td>0.88</td>
<td>0.80</td>
<td>0.71</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>(R)SFDPVRPPIDASSPSQK(G)</td>
<td>632697</td>
<td>12.22</td>
<td>1.14</td>
<td>1.11</td>
<td>0.70</td>
<td>0.83</td>
<td>1.08</td>
<td>1.09</td>
<td>1.10</td>
<td>0.73</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(R)SVHSTERFACmFCEKFSR(S)</td>
<td>6322788</td>
<td>6.09</td>
<td>1.32</td>
<td>1.49</td>
<td>0.80</td>
<td>0.55</td>
<td>0.94</td>
<td>0.52</td>
<td>1.28</td>
<td>0.82</td>
<td>1&lt;sup&gt;*&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

---

<br><sup>a</sup> Protein names can be obtained from NCBI database according to their individual accession numbers;  
<sup>b</sup> Values were Spectrum Mill score for each peptide with false positive rate less than 4%.  
<sup>c</sup> Values were N2-normalized reporter ion intensities obtained for each isobar mass tag.  
<sup>d</sup> Values were obtained from the reference [24]; A “M” in Column N indicate an unoxidized methionine-containing.
peptide was detected in the SPROX experiment, and a “Ox” in Column N represents an oxidized methionine-containing peptide was detected in the SPROX experiment. A “U” in Column M indicates N2-normalized reporter ion intensity data that was un-interpretable (see text), and a “**” symbol represents a peptide data set in which one of the eight N2-normalized reporter ion intensities was not used for the denaturation curve construction and the transition midpoint assignment. In cases were the eliminated N2-normalized reporter ion intensity complicated the C1/2 value assignment, the two possible midpoints are listed (see text). In some cases no transition was detected (i.e., all the N2-normalized reporter ion intensities were within 45% of 1). These cases are noted in Column M with an “NT”. ‘ND’ stands for no data due to the lack of the report ion signals for the corresponding isobar mass tags.
Appendix A. Extension of a SUPREX with Protease Digestion Protocol to a Large Sized Protein

A.1 Research Objective

SUPREX – protease digestion protocol has previously been shown to be effective for the analysis of proteins with well-defined domains in which each domain exhibits cooperative folding behavior.[39] To examine the scope of this technique, I applied the SUPREX-protease digestion protocol to analyze ligand binding for a large, multidomain protein: glycogen phosphorylase b (Phb; ~98 kDa/subunit).

A.2 Introduction

In a typical H/D exchange experiment, a fully folded protein sample is allowed to undergo H/D exchange. After a specified amount of time, the exchange is quenched by lowering pH and temperature, and the protein is digested with pepsin, a protease with high activity under acidic conditions. The same experiment is performed on the ligand-bound protein, and comparisons are drawn between the uptake of deuterons for the apo and holo forms of the protein. While these experiments are valuable for obtaining information about ligand binding sites, they are low throughput and do not provide any quantitative information about ligand binding affinities. Recently, we reported on an H/D exchange- and MALDI mass spectrometry-based technique that combines SUPREX (stability of unpurified proteins by rates of H/D exchange) with
protease digestion. Unlike conventional H/D exchange and protease digestion techniques, this SUPREX-protease digestion protocol is capable of providing quantitative information about ligand binding affinities. Furthermore, this technique is amenable to higher throughput analyses.

To examine the scope of this technique, I used the SUPREX-protease digestion protocol to analyze ligand binding for a large, multidomain protein: glycogen phosphorylase b (Phb; ~98 kDa/subunit). The SUPREX-protease digestion protocol has previously been shown to be effective for the analysis of proteins with well-defined domains in which each domain exhibits cooperative folding behavior. However, this work represents the first application of this protocol for the analysis of proteins with poorly defined domains. In fact, each subunit Phb contains two domains; however, these domains are closely associated with one another, and the linear amino acid sequence of one domain interrupts the sequence of the other domain (Figure 24). In this work, I sought to apply the SUPREX-protease digestion protocol to large proteins with poorly defined domains to help us gain insight into the generality of the technique. Note that unlike the use of the SUPREX-protease digestion protocol to obtain domain-specific stability information as in our previous work, the goal of this study was to assess the ability of this technique to detect and quantify ligand binding in large, multidomain proteins.
Here, we demonstrate the ability of the SUPREX-protease digestion protocol to determine protein-ligand binding affinities for a multidomain, homodimeric protein. Since the protease digestion step takes place after the H/D exchange reaction is complete, this technique reports on the thermodynamic stability of the intact protein. We expect this technique to be applicable to a wide variety of proteins, and the successful evaluation of the additional protein systems in this work provides further evidence of the generality of the technique.

**A.3 Experimental Section**

**A.3.1 Materials**

The following reagents were purchased from Sigma-Aldrich (St. Louis, MO): N-(2-hydroxyethyl)piperazine-N'-2-ethanesulfonic acid (HEPES), ethylenediaminetetraacetic acid disodium salt (EDTA), adenosine 3'-monophosphate from yeast (AMP), D-glucose-6-phosphate potassium salt (G6P) and glycogen phosphorylase b (Phb) from rabbit muscle. Urea were purchased from EMD (Gibbstown, NJ).

**A.3.2 Sample Preparation and Pepsin Digestion Conditions**

A 50 mM HEPES, 100 mM KCl and 1 mM EDTA (pH 7.4) buffer was used to prepare the stock solutions of the apo-Phb, AMP and G6P. The concentration of the protein was 178 μM and was determined using E1%278 = 13.2[127] and a molecular
weight of 97.5 kDa. Protein-ligand complexes were prepared by combining Phb with AMP or G6P and incubating at room temperature for at least 3 hours.

The deuterated SUPREX buffers used to analyze Phb contained 50 mM HEPES, 100 mM KCl, 1 mM EDTA (pD 7.4), and between 0.1 and 6.5 M urea. For use in SUPREX buffers, urea was deuterated using four cycles of dissolution in D2O and lyophilization.

The stock solution of protein or protein-ligand complex was diluted 10-fold into the series of deuterated SUPREX buffers. After the specified exchange time, a 1.5 μL aliquot of the exchanged sample was transferred into 6 μL of the quenching solution, which contained 5 μL HCl (pH 1.6) and 1 μL 8 μM pepsin stock (final pepsin concentration equal to 1.1 μM). Pepsin digestion of Phb was allowed to proceed for 2 min. 1 μL resulting solution was mixed with 9 μL of SA matrix using trypsin inhibitor and insulin as internal mass calibrants. Ultimately, 1 μL aliquot was subject to MALDI analysis.

A.3.3 SUPREX Data Analysis and Kd Value Determination

SUPREX data analysis was described in Chapter 2. Equation 2.1 was used to determine ΔGf and m- values. In equation 2.1, Phb was treated as a dimer (n= 2). In our binding studies, the concentration of ligand is more than a 10-fold excess over protein concentration. Therefore, dissociation constants, Kd values, were calculated using equation 2.3. In equation 2.3, \textit{L_{total}} is the final concentration of ligand in the H/D exchange
reactions, \( P_{\text{total}} \) is the concentration of Ph\( b \) dimer, \( \Delta \Delta G_{\text{f,avg}} \) is the difference in \( \Delta G_{\text{f,avg}} \) between the bound and unbound forms of the protein, and \( N \) is the number of independent equivalent binding sites (\( N = 2 \) as expected for the Ph\( b \) dimer).

### A.4 Results

Two different ligands were used to form Ph\( b \)-ligand complexes in this study: G6P and AMP (with final concentrations in the SUPREX buffers of 67 mM and 0.47 mM, respectively). Using various concentrations of urea as denaturant, a 2 min pepsin digestion (pH~2) of Ph\( b \) in the presence or absence of ligands consistently generated 10 peptic fragments. SUPREX analysis of six fragments (fragments 3489, 4062, 4266, 4505, 4589, and 6168) yielded sigmoidal curves with discernible transition midpoints (\( C_{1/2}^{\text{SUPREX}} \) values). This suggests these 6 fragments contain buried amide protons. SUPREX analysis of peptides 3710, 3962, 4375, and 4218 generated straight, horizontal lines rather than sigmoidal curves. For these peptides, the \( \Delta \text{mass} \) values remained constant as the concentration of urea was varied (data not shown). This suggests that either these 4 peptides are located in unstructured regions of the protein or that their H/D exchange behaviors are dominated by local unfolding/refolding events. Shown in **Figure 25** are representative SUPREX curves for fragments 6168, 4505, and 4266 in the absence and in the presence of ligands when the H/D exchange was performed for 4 minutes at room temperature. Note that the \( C_{1/2}^{\text{SUPREX}} \) value for peptide 6168 (**Figure 25A**) shifted to a
higher urea concentration in the presence of ligands. In contrast, for fragment 4505, the difference in $C_{1/2}^{\suprex}$ values between apo-Ph$b$ and Ph$b$-AMP (Figure 25B) was very small (~ 0.2 M) and within experimental error. However, the midpoint for this same fragment was significantly shifted to the left in the presence of 67 mM G6P, indicating destabilization of this region of the protein upon G6P binding. The $C_{1/2}^{\suprex}$ value for peptide 4266 (Figure 25C) remained the same in the presence and absence of AMP. For all peptides, sigmoidal curves were also generated using H/D exchange times ranging from 4 to 225 minutes at room temperature. All the $C_{1/2}^{\suprex}$ values obtained from SUPREX curves for the six fragments at the different H/D exchange times are summarized in the Table 8.

When the H/D exchange time was increased, all of the $C_{1/2}^{\suprex}$ values shifted toward lower denaturant concentrations as expected. The $C_{1/2}^{\suprex}$ values were plotted as a function of H/D exchange time according to Equation 2.1, and the data were analyzed using a linear least-squares analysis. The linear fits of all of the SUPREX data were generally good, with correlation coefficients greater than 0.90 (Figure 26). The $\Delta G_i$ and $m$-values extracted from the data are summarized in Table 9. According to the measured thermodynamic parameters, the peptides from Ph$b$ could be grouped into three categories: 1) peptides 6168 and 4589; 2) peptides 4505 and 3489; and 3) peptides 4266 and 4062. For each of the three categories of SUPREX behavior, the measured $m$-
value did not change upon ligand binding (values were within ~30% of each other; see Discussion section). Since no change in $m$-value was detected, an average $m$-value ($m_{\text{avg}}$) was calculated for each peptide. The $m_{\text{avg}}$ values were then used to calculate $\Delta G_{\text{f,avg}}$ values (Table 9). The $\Delta G_{\text{f,avg}}$ values for peptides 6168 and 4589 became more negative (indicating an increase in stability) in the presence of ligands. $\Delta G_{\text{f}}$ values for peptides 4505 and 3489 remained the same in the presence and absence of AMP. However, these values became less negative (indicating a decrease in stability) in the presence of 67 mM G6P. To ensure that the decrease in stability was not due to a change in ionic strength of the solution, we also measured $\Delta G_{\text{f,avg}}$ for Phb in the presence of 67 mM KCl. The presence of KCl had no effect on the stabilities of these two peptides (data not shown), suggesting that the stability change can indeed be attributed to G6P binding. For peptides 4062 and 4266, we did not detect a change in the thermodynamic parameters in the presence of each ligand.

The results of our SUPREX-protease digestion analysis on Phb suggest the presence of three sub-global unfolding/refolding behaviors with different thermodynamic parameters. The signature of one behavior involved stabilization by both AMP and G6P binding as evidenced by a decreased (i.e., more negative) $\Delta G_{\text{f,avg}}$ value. $\Delta G_{\text{f,avg}}$ values for a second unfolding behavior were increased (i.e., less negative) in the presence of 67 mM G6P, suggesting a sub-global destabilization of a portion of the
protein upon G6P binding. Peptides in the third behavioral category showed no
difference in stability upon binding of G6P and AMP.

A.5 Discussion

To calculate accurate ΔGᵢ and m-values for a protein folding reaction, the
reaction must be two-state (i.e., only the unfolded or folded states of the protein can be
significantly populated during the folding reaction). Phb is a known non-two-state
folder, as indicated by the observation of three distinct transitions in its denaturation
profile.[128] Because two-state folding is a required assumption when determining ΔGᵢ
and m-values by SUPREX, the thermodynamic parameters reported here should not be
taken as an absolute representation of the folding reactions of Phb. However, if we
assume that the folding reaction of the protein does not change upon ligand binding, we
can use these values to calculate ΔΔGᵢ,avg values, which in turn can be used to calculate
Kₐ values. Thus, the likely non-two-state nature of these proteins did not preclude
quantitation of ligand binding, which was the goal of this study.

Glycogen phosphorylase is a key enzyme in the regulation of glycogen
metabolism. It catalyses the following reaction:

Glycogen (n+1) + Pᵢ → Glycogen (n) + G-1-P

Since glycogen serves as a metabolic fuel reservoir for glucose, inhibiting
glycogen phosphorylase can reduce glycogenolysis. Therefore, glycogen phosphorylase
is a target for finding new drugs for the treatment of type 2 diabetes.[129] The Phb we studied here is the inactive form of glycogen phosphorylase. It can be activated by AMP and inhibited by G6P.

Due to its large molecular weight (98 kDa/subunit), the detection of a sufficiently intense and well-resolved ion signal for Phb is difficult using our MALDI instrument. This difficulty in obtaining an ion signal precludes analysis of Phb using the conventional SUPREX protocol. However, performing a quick digestion of Phb after the H/D exchange step allows us to generate easily detectable low molecular weight fragments, which allows for the examination of the thermodynamic parameters of Phb using SUPREX.

In our Phb ligand binding studies, AMP and G6P were used to form two different Phb-ligand complexes. The SUPREX-protease digestion analysis of apo-Phb revealed three different sets of thermodynamic parameters from six different peptic fragments. Both AMP and G6P stabilized the Phb around its fragment 6168 and fragment 4589 regions. This suggests that both fragments are either close to or involved with the AMP and G6P binding sites. This observation is in agreement with the crystal structures for Phb-AMP and Phb-G6P, which reveal that the G6P binding site is located near the AMP binding site.[130]
The thermodynamic stabilities of Phb measured in the region(s) of fragments 4505 and 3489 were not affected by AMP binding. However, both fragments were destabilized upon G6P binding, and their thermodynamic parameters became more similar to the SUPREX-derived values for fragments 6168 and 4589. The observed destabilization suggests that a conformational change has occurred that has destabilized one or more local regions of the protein structure. A conformational change has indeed been observed in the crystal structure of Phb-G6P (relative to apo-Phb).[130] The observed conformational change occurs on the subunit-subunit interface, affecting the interactions between the two subunits and making the structure of the Phb-G6P complex more tightly packed. However, the same change was not observed in the crystal structure of the Phb-AMP complex.[131] These observations are in agreement with the thermodynamic behavior of peptides 4505 and 3489, as differences in stability were detected upon G6P binding, but not upon AMP binding.

In contrast with the first two categories of thermodynamic behavior, peptides 4062 and 4266 did not exhibit any change in stability upon ligand binding. This indicates that ligand binding had no discernible effect on the sub-global stability of these regions of the protein. This observation could indicate that these two peptides are located in a region distant from the AMP and G6P binding sites. However, our goal was to detect
and quantify ligand binding rather than to map binding sites, so we focused our further analysis on the fragments that exhibited changes in stability upon ligand binding.

$K_d$ values for both ligands were determined based on the local $\Delta \Delta G_{f,\text{avg}}$ values of peptides 6168 and 4589 using Equation 2.3. These two peptides were chosen based on their observed stabilization upon ligand binding. Our SUPREX-derived $K_d$ value for AMP binding (0.6 mM) is in good agreement with literature values (0.1 mM – 0.7 mM), which were determined using calorimetric, fluorescence, and spin-label experiments.[130, 132-134] However, our $K_d$ values for G6P binding were approximately 100-fold higher than the values reported in the literature.[135, 136]

There are several possible explanations for this discrepancy in $K_d$ values for G6P binding. One potential cause for the aberrantly high $K_d$ value is that the ligand may interact with the denatured state of Phb. This interaction would stabilize the denatured state, leading to inaccurate $\Delta G_f$ measurements for the protein-ligand complex (since the difference in stability between the native and denatured states would be decreased). This effect would cause the measured $\Delta G_f$ value for the protein-ligand complex to be less negative, leading to a smaller difference in $\Delta G_f$ values between the bound and unbound forms of the protein. The inaccuracy of the measured $\Delta \Delta G_{f,\text{avg}}$ value would lead to an aberrantly high $K_d$ value. While this effect would reduce the accuracy of the
measured $K_d$ value, it does not preclude our ability to qualitatively detect ligand binding.

A second possible reason for the differences in $K_d$ values is related to the protein sample. The commercially available Ph$b$ contains an unknown amount of AMP, but the protein preparation was subjected to the SUPREX-protease digestion analysis without further purification. In contrast, the $K_d$ values reported in the literature were measured in the absence of the AMP. The presence of AMP affects G6P binding affinity to Ph$b$, which may account for the higher measured $K_d$ value.

Another potential reason for the discrepancy is related to the sub-global character of our SUPREX-protease digestion measurements. As evidenced by the observation of three discrete SUPREX behaviors in Ph$b$, the local $\Delta \Delta G_{f,\text{avg}}$ value does not necessarily represent the $\Delta \Delta G_{f,\text{avg}}$ values for the intact protein. We believe that the $\Delta \Delta G_{f,\text{avg}}$ values for the intact protein should reflect the net effect of all sub-global stabilizations and destabilizations. However, measuring this net effect is not particularly feasible, as it would likely require complete peptide coverage without overlapping of fragments.

The majority of the proteins previously studied with the SUPREX-protease digestion protocol were relatively low in molecular weight, which reduces the probability of observing multiple stabilities in different regions of the protein. The
exception was the analysis of transferrin,[39] a relatively large protein (~80 kDa) with two distinct domains. However, transferrin differs from the proteins in this study in that its domains are clearly defined, and each domain exhibits fairly cooperative folding behavior. The cooperativity of the domain-specific folding reactions allowed for separate thermodynamic parameters to be measured for each domain. Our data highlights the complexity of the ligand binding interactions of Phb, in which the domains are less clearly defined, and suggests that this method may be more useful for the measurement of sub-global ligand-induced changes in stability rather than global or domain-specific changes in stability. We also note that in cases where multiple SUPREX behaviors are observed, this protocol can still be used for the qualitative detection of ligand binding, even when the ligand binding interactions are complicated as in the case of Phb.
Figure 24: Crystal structure of phosphorylase b from rabbit muscle form (PDB: 3MQF)
Figure 25: Representative SUPREX curves obtained for peptic fragments from Phb. (A) Peptide 6168 Da, (B) peptide 4505 Da and (C) peptide 4266 Da were obtained using the SUPREX-digestion protocol from Apo-Phb (filled circles), AMP-Phb with final concentration of AMP equal to either 0.5 mM (open circles) in SUPREX buffer and G6P-Phb (filled triangles) with final concentration of G6P equal to 67 mM. The H/D exchange of each sample was performed with 4 minutes at room temperature. The lines represent the best fit of the data in each SUPREX curve to a four-parameter sigmoid equation using SigmaPlot. The dotted lines mark the $C_{1/2}^{SUPREX}$ value for each curve.
Table 8 $C^{1/2}_{SUPREX}$ at different H/D exchange time ($t_{H/D}$) for peptic peptides from Phb

<table>
<thead>
<tr>
<th>Phb Fragment</th>
<th>$t_{H/D} = 4$ min</th>
<th>$t_{H/D} = 14$ min</th>
<th>$t_{H/D} = 35$ min</th>
<th>$t_{H/D} = 65$ min</th>
<th>$t_{H/D} = 95$ min</th>
<th>$t_{H/D} = 125$ min</th>
<th>$t_{H/D} = 225$ min</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C^{1/2}_{SUPREX}$ (M)$^a$</td>
<td>$C^{1/2}_{SUPREX}$ (M)$^a$</td>
<td>$C^{1/2}_{SUPREX}$ (M)$^a$</td>
<td>$C^{1/2}_{SUPREX}$ (M)$^a$</td>
<td>$C^{1/2}_{SUPREX}$ (M)$^a$</td>
<td>$C^{1/2}_{SUPREX}$ (M)$^a$</td>
<td>$C^{1/2}_{SUPREX}$ (M)$^a$</td>
</tr>
<tr>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6168-apoPhb</td>
<td>1.3</td>
<td>1.0</td>
<td>0.9</td>
<td>0.7</td>
<td>0.4</td>
<td>ND$^b$</td>
<td>ND$^b$</td>
</tr>
<tr>
<td>6168-AMP</td>
<td>1.7</td>
<td>ND$^b$</td>
<td>ND$^b$</td>
<td>0.9</td>
<td>1.0</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>6168-G6P</td>
<td>2.7</td>
<td>ND$^b$</td>
<td>2.2</td>
<td>2.0</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4589-apoPhb</td>
<td>1.5</td>
<td>1.1</td>
<td>0.6</td>
<td>0.6</td>
<td>ND$^b$</td>
<td>ND$^b$</td>
<td></td>
</tr>
<tr>
<td>4589-AMP</td>
<td>1.6</td>
<td>ND$^b$</td>
<td>ND$^b$</td>
<td>ND$^b$</td>
<td>0.9</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>4589-G6P</td>
<td>2.7</td>
<td>ND$^b$</td>
<td>2.2</td>
<td>2.0</td>
<td>2.0</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>III</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4505-apoPhb</td>
<td>3.6</td>
<td>3.3</td>
<td>3.0</td>
<td>2.9</td>
<td>2.8</td>
<td>2.7</td>
<td>2.6</td>
</tr>
<tr>
<td>4505-AMP</td>
<td>3.6</td>
<td>ND$^b$</td>
<td>3.0</td>
<td>2.9</td>
<td>2.8</td>
<td>ND$^b$</td>
<td>ND$^b$</td>
</tr>
<tr>
<td>4505-G6P</td>
<td>2.7</td>
<td>ND$^b$</td>
<td>2.2</td>
<td>2.0</td>
<td>1.9</td>
<td>1.9</td>
<td>1.7</td>
</tr>
<tr>
<td>3489-apoPhb</td>
<td>3.8</td>
<td>3.6</td>
<td>3.3</td>
<td>3.0</td>
<td>2.9</td>
<td>ND$^b$</td>
<td>ND$^b$</td>
</tr>
<tr>
<td>3489-AMP</td>
<td>3.9</td>
<td>3.1</td>
<td>3.0</td>
<td>2.9</td>
<td>2.8</td>
<td>ND$^b$</td>
<td>ND$^b$</td>
</tr>
<tr>
<td>3489-G6P</td>
<td>3.0</td>
<td>2.5</td>
<td>2.2</td>
<td>2.1</td>
<td>ND$^b$</td>
<td>ND$^b$</td>
<td>ND$^b$</td>
</tr>
<tr>
<td>IV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4062-apoPhb</td>
<td>3.1</td>
<td>2.3</td>
<td>2.2</td>
<td>2.0</td>
<td>ND$^b$</td>
<td>1.6</td>
<td>0.9</td>
</tr>
<tr>
<td>4062-AMP</td>
<td>3.0</td>
<td>ND$^b$</td>
<td>2.2</td>
<td>1.8</td>
<td>ND$^b$</td>
<td>1.7</td>
<td>1.4</td>
</tr>
<tr>
<td>4062-G6P</td>
<td>2.7</td>
<td>ND$^b$</td>
<td>2.1</td>
<td>1.9</td>
<td>1.8</td>
<td>1.7</td>
<td>1.7</td>
</tr>
<tr>
<td>4266-apoPhb</td>
<td>3.3</td>
<td>3.2</td>
<td>2.6</td>
<td>2.5</td>
<td>ND$^b$</td>
<td>ND$^b$</td>
<td>ND$^b$</td>
</tr>
<tr>
<td>4266-AMP</td>
<td>3.0</td>
<td>ND$^b$</td>
<td>2.5</td>
<td>2.2</td>
<td>1.8</td>
<td>ND$^b$</td>
<td>ND$^b$</td>
</tr>
<tr>
<td>4266-G6P</td>
<td>3.0</td>
<td>ND$^b$</td>
<td>2.1</td>
<td>2.0</td>
<td>1.8</td>
<td>1.8</td>
<td>1.7</td>
</tr>
</tbody>
</table>

$^a$Values are taken from the best fit of SUPREX curve. Errors were typically between ±0.1 and ±0.2.

$^b$ND stands for no data resulting from either the disappearance of the peptide ion signals or low quality of SUPREX curves.
Figure 26: Plots of $\Delta G_{\text{app}}$ vs. $C^{1/2}_{\text{SUPREX}}$ for (A) peptide 6168, (B) peptide 4505, and (C) peptide 4266. The resulting $y$-intercept and slope were taken as the $\Delta G_f$ and $m$-value, respectively. These peptide fragments were obtained using the SUPREX-protease digestion protocol from Apo-Phb (filled circles), Phb-AMP with a final AMP concentration of 0.5 mM (open circles), and Phb-G6P with a final G6P concentration of 67 mM (filled triangles). H/D exchange times varied from 4 minutes to 225 minutes. The lines represent the results of our linear least-squares analysis of each data set.
Table 9: SUPREX-derived $\Delta G_f$ and $m$-values for Ph$\beta$ and its complexes.

<table>
<thead>
<tr>
<th>Ph$\beta$ Fragment</th>
<th>$\Delta G_f^a$ (kcal mol$^{-1}$)</th>
<th>$m^a$ (kcal mol$^{-1}$ M$^{-1}$)</th>
<th>$\Delta G_{f,avg}^b$ (kcal mol$^{-1}$)</th>
<th>$m_{avg}^c$ (kcal mol$^{-1}$ M$^{-1}$)</th>
<th>$\Delta G_{f,avg}^d$ (kcal/mol)</th>
<th>SUPREX-derived $K_d$ (mM)</th>
<th>Literature $K_d$ (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6168-apoPh$\beta$</td>
<td>-13.6 ± 0.3</td>
<td>2.0 ± 0.3</td>
<td>-13.2 ± 0.2</td>
<td>2.2 ± 0.2</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>6168-Ph$\beta$+AMP</td>
<td>-14.5 ± 0.2</td>
<td>2.2 ± 0.2</td>
<td>-14.3 ± 0.1</td>
<td>-0.9</td>
<td>0.6 mM</td>
<td>0.1-0.7$^d$</td>
<td></td>
</tr>
<tr>
<td>6168-Ph$\beta$+G6P</td>
<td>-17.2 ± 0.4</td>
<td>2.4 ± 0.2</td>
<td>-16.2 ± 0.2</td>
<td>-3.0</td>
<td>5.5 mM</td>
<td>0.3-0.9$^e$</td>
<td></td>
</tr>
<tr>
<td>4589-apoPh$\beta$</td>
<td>-13.4 ± 0.3</td>
<td>1.8 ± 0.3</td>
<td>-13.4 ± 0.5</td>
<td>2.2 ± 0.2</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>4589-Ph$\beta$+AMP</td>
<td>14.5 ± 0.2</td>
<td>2.3 ± 0.2</td>
<td>-14.3 ± 0.1</td>
<td>-0.9</td>
<td>0.6 mM</td>
<td>0.1-0.7$^d$</td>
<td></td>
</tr>
<tr>
<td>4589-Ph$\beta$+G6P</td>
<td>-17.2 ± 0.5</td>
<td>2.4 ± 0.2</td>
<td>-16.4 ± 0.2</td>
<td>-3.0</td>
<td>5.5 mM</td>
<td>0.3-0.9$^e$</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4505-apoPh$\beta$</td>
<td>-18.7 ± 0.4</td>
<td>2.2 ± 0.1</td>
<td>-18.7 ± 0.1</td>
<td>2.4 ± 0.2</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>4505-Ph$\beta$+AMP</td>
<td>-19.3 ± 0.3</td>
<td>2.4 ± 0.1</td>
<td>-18.6 ± 0.1</td>
<td>-0</td>
<td>2.0</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>4505-Ph$\beta$+G6P</td>
<td>-17.3 ± 0.2</td>
<td>2.4 ± 0.1</td>
<td>-16.7 ± 0.1</td>
<td>2.0</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>3489-apoPh$\beta$</td>
<td>-18.4 ± 0.6</td>
<td>2.0 ± 0.2</td>
<td>-17.1 ± 0.2</td>
<td>2.2 ± 0.2</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>3489-Ph$\beta$+AMP</td>
<td>-17.6 ± 0.4</td>
<td>1.8 ± 0.1</td>
<td>-17.0 ± 0.1</td>
<td>-0</td>
<td>2.0</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>3489-Ph$\beta$+G6P</td>
<td>-16.9 ± 0.5</td>
<td>2.0 ± 0.2</td>
<td>-15.7 ± 0.2</td>
<td>1.4</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>III</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4062-apoPh$\beta$</td>
<td>-14.3 ± 0.3</td>
<td>1.1 ± 0.2</td>
<td>-14.7 ± 0.4</td>
<td>1.6 ± 0.4</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>4062-Ph$\beta$+AMP</td>
<td>-15.0 ± 0.2</td>
<td>1.4 ± 0.1</td>
<td>-15.0 ± 0.2</td>
<td>-0</td>
<td>2.0</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>4062-Ph$\beta$+G6P</td>
<td>-16.6 ± 0.2</td>
<td>2.2 ± 0.1</td>
<td>-14.9 ± 0.2</td>
<td>-0</td>
<td>2.0</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>4266-apoPh$\beta$</td>
<td>-15.0 ± 0.7</td>
<td>1.2 ± 0.2</td>
<td>-15.4 ± 0.3</td>
<td>1.6 ± 0.2</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>4266-Ph$\beta$+AMP</td>
<td>-15.7 ± 0.7</td>
<td>1.6 ± 0.3</td>
<td>-15.2 ± 0.2</td>
<td>-0</td>
<td>2.0</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>4266-Ph$\beta$+G6P</td>
<td>-15.6 ± 0.3</td>
<td>1.7 ± 0.1</td>
<td>-15.0 ± 0.2</td>
<td>-0</td>
<td>2.0</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

$^a$The $\Delta G_f$ and $m$-values were taken from the linear least-squares analysis of the data using Equation 2.1. Errors reported for the $\Delta G_f$ and $m$-values are the fitting errors of the linear least-squares analyses. $^b$$\Delta G_{f,avg}$ values were the average of folding free energy values calculated using Equation 2.1 and an established $m$-value. The errors reported for the $\Delta G_{f,avg}$ values are the standard deviations of the folding free energy values calculated using Equation 2.1 and the established $m$-value (i.e. $m_{avg}$). $^c$ $m_{avg}$ values were calculated by averaging the m values obtained.
from apo-Phb, AMP-Phb and G6P-Phb for individual peptides. *Literature values were obtained from refs [130, 131]. *Literature values were obtained from refs [137-139].
References


167


Biography

Ying Xu got her bachelor’s degree in organic chemistry in year 1997 from Fudan University, China. After her graduation, she worked for Sanofi-Aventis Inc. at Shanghai branch before she came to the United States. In 2003, she was accepted by chemistry department of University of North Caroline at Charlotte and finished her master’s degree in bio-analytical chemistry in 2005. She was awarded Thomas Walsh Fellowship from 2003-2005 and her thesis was selected as UNC Charlotte Distinguished Master’s Thesis in the area of Life Sciences in 2005. In 2006, Ying joined Dr. Michael C. Fitzgerald’s group to pursue her Ph.D degree in bio-analytical chemistry at Duke University. She was awarded Duke University Gross Fellowship in year 2010 and won 2011 Poster Award at the 7th Annual Meeting of US Human Proteome Organization. Ying is expected to receive her Ph.D degree in May, 2011.

Publications

1) Xu, Y.; Schmitt, S.; Tang, L. and Fitzgerald, M.C. “Thermodynamic Analysis of a Molecular Chaperone Binding to Unfolded Protein Substrate”, Biochemistry, 2010, 49(6), 1346-1353
