Mass Spectrometry-Based Strategies for Multiplexed Analyses of Protein-Ligand Binding Interactions

by

Patrick Dean DeArmond

Department of Chemistry
Duke University

Date:_______________________

Approved:

___________________________
Michael C. Fitzgerald, Supervisor

___________________________
Alvin L. Crumbliss

___________________________
Stephen L. Craig

___________________________
Edward F. Patz, Jr.

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

Mass Spectrometry-Based Strategies for Multiplexed Analyses of Protein-Ligand Binding Interactions

by

Patrick Dean DeArmond

Department of Chemistry
Duke University

Date: _______________________

Approved: _______________________

Michael C. Fitzgerald, Supervisor

___________________________

Alvin L. Crumbliss

___________________________

Stephen L. Craig

___________________________

Edward F. Patz, Jr.

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

The detection and quantitation of protein-ligand binding interactions is important not only for understanding biological functions but also for the characterization of novel protein ligands. Because protein ligands can range from small molecules to other proteins, general techniques that can detect and quantitate the many classes of protein-ligand interactions are especially attractive. Additionally, the ability to detect and quantify protein-ligand interactions in complex biological mixtures would more accurately represent the protein-ligand interactions that occur in vivo, where differential protein expression and protein complexes can significantly affect a protein’s ability to bind to a ligand of interest.

The work described in this dissertation is focused on the development of new methodologies for the detection and measurement of protein-ligand interactions in complex mixtures using multiplex analyses. Methodologies for two types of multiplexed analyses of protein-ligand binding interactions are investigated here. The first type of multiplex analysis involves characterizing the binding of one protein target to many potential ligands, and the second type involves characterizing the binding of one ligand to many proteins. The described methodologies are derived from the SUPREX (stability of unpurified proteins from rates of H/D exchange) and SPROX (stability of proteins from rates of oxidation) techniques, which are chemical modification strategies that measure thermodynamic stabilities of proteins using a relationship between a protein’s folding equilibrium and the extent of chemical modification. These two techniques were utilized
in the development and application of several different experimental strategies designed to multiplex the analysis of protein-ligand interactions.

The first strategy that was developed involved a pooled compound approach for making SUPREX-based measurements of multiple ligands binding to a target protein. Screening rates of 6 s/ligand were demonstrated in a high-throughput screening project that involved the screening of two chemical libraries against human cyclophilin A (CypA), a protein commonly overexpressed in types of cancer. This study identified eight novel ligands to CypA with micromolar dissociation constants. Second, an affinity-based protein purification strategy was developed for the detection and quantitation of specific protein-ligand binding interactions in the context of complex protein mixtures. It involved performing SPROX in cell lysates and selecting the protein of interest using immunoprecipitation or affinity tag purification. A third strategy developed here involved a SPROX-based stable isotope labeling method for measuring protein-ligand interactions in multi-protein mixtures. This strategy was used in a proof-of-principle experiment designed to detect and quantify the indirect binding between yeast cyclophilin and calcineurin in a multi-component protein mixture. Finally, a quantitative proteomics platform was developed for the detection and quantitation of protein-ligand binding interactions on the proteomic scale. The platform was used to profile interactions of the proteins in a yeast cell lysate to several ligands, including the bioactive small molecules resveratrol and manassantin A, the cofactor nicotinamide adenine dinucleotide (NAD+), and two proteins, phosphoglycerate kinase (Pgk1) and pyruvate kinase (Pyk1). The
above approaches should have broad application for use as discovery tools in the development of new therapeutic agents.
## Contents

Abstract ........................................................................................................................................ iv  
List of Tables .................................................................................................................................. xi  
List of Figures ................................................................................................................................. xii  
List of Abbreviations ...................................................................................................................... xvii  
1. Introduction ............................................................................................................................... 1  
   1.1 Background .............................................................................................................................. 1  
   1.2 Equilibrium Unfolding Properties of Proteins ......................................................................... 7  
   1.3 H/D Exchange and Chemical Modification Theory ............................................................... 8  
   1.4 Development of SUPREX and SPROX .................................................................................. 11  
      1.4.1 SUPREX and SPROX Data Analysis Method 1 ................................................................. 11  
      1.4.2 SUPREX and SPROX Data Analysis Method 2 ................................................................. 14  
   1.5 Measurement of Protein-Ligand Interactions Using SUPREX and SPROX .................... 16  
   1.6 Overview of Research Objectives ........................................................................................... 18  
2. Discovery of Novel Cyclophilin A Ligands Using an H/D Exchange- and Mass  
   Spectrometry-Based Strategy ................................................................................................... 21  
   2.1 Introduction ............................................................................................................................. 21  
   2.2 Experimental .......................................................................................................................... 22  
      2.2.1 Materials .......................................................................................................................... 22  
      2.2.2 SUPREX Buffer Preparation ............................................................................................ 24  
      2.2.3 Mass Spectrometry .......................................................................................................... 25  
      2.2.4 Library Screening (Single-Point SUPREX Protocol) ......................................................... 25  
      2.2.5 Binding Affinity Measurements (Full SUPREX Analyses) ............................................... 28
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.6 Theoretical SUPREX Curve Construction</td>
<td>30</td>
</tr>
<tr>
<td>2.2.7 CypA Inhibition Studies</td>
<td>30</td>
</tr>
<tr>
<td>2.3 Results and Discussion</td>
<td>31</td>
</tr>
<tr>
<td>2.3.1 General Strategy</td>
<td>31</td>
</tr>
<tr>
<td>2.3.2 Library Screening</td>
<td>33</td>
</tr>
<tr>
<td>2.3.3 Throughput and Efficiency</td>
<td>46</td>
</tr>
<tr>
<td>2.3.4 Identification of Novel CypA Ligands</td>
<td>49</td>
</tr>
<tr>
<td>3. Thermodynamic Analysis of Selective Proteins in Complex Mixtures</td>
<td>53</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>53</td>
</tr>
<tr>
<td>3.2 Experimental</td>
<td>54</td>
</tr>
<tr>
<td>3.2.1 SPROX</td>
<td>54</td>
</tr>
<tr>
<td>3.2.2 Immunoprecipitation Pull-Down</td>
<td>55</td>
</tr>
<tr>
<td>3.2.3 Affinity Tag Pull-Down with Whole Protein Readout</td>
<td>55</td>
</tr>
<tr>
<td>3.2.4 Affinity Tag Pull-Down with Peptide Readout</td>
<td>57</td>
</tr>
<tr>
<td>3.2.5 Mass Spectrometry Analyses</td>
<td>59</td>
</tr>
<tr>
<td>3.2.6 Data Analysis</td>
<td>60</td>
</tr>
<tr>
<td>3.3 Results and Discussion</td>
<td>62</td>
</tr>
<tr>
<td>3.3.1 BCA Immunoprecipitation</td>
<td>62</td>
</tr>
<tr>
<td>3.3.2 Affinity Purified Cpr1</td>
<td>64</td>
</tr>
<tr>
<td>3.4 Conclusions</td>
<td>70</td>
</tr>
<tr>
<td>4. Stable Isotope Labeling Strategy for Protein-Ligand Binding Analysis in Multi-Component Protein Mixtures</td>
<td>74</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>74</td>
</tr>
</tbody>
</table>
4.2 Experimental ............................................................................................................. 75
  4.2.1 Protein Sample Preparation ............................................................................. 75
  4.2.2 PrSUIT Analyses ............................................................................................. 77
  4.2.3 LC-MS Analyses ............................................................................................. 80
  4.2.4 Calculation of %^{18}O Labeling ...................................................................... 82
  4.2.5 Construction of SPROX Curves from PrSUIT Plots of Peptide Hits ............... 84
  4.2.6 Binding Affinity Measurements ....................................................................... 84
4.3 Results .................................................................................................................. 86
  4.3.1 General Protocol ............................................................................................. 86
  4.3.2 PrSUIT Analysis of a Model Protein Mixture ............................................... 89
  4.3.3 PrSUIT Analyses of Purified Cpr1 and Purified Cna1 ................................. 97
  4.3.4 Binding Affinity Measurements ....................................................................... 100
4.4 Discussion ............................................................................................................. 101

5. Protocol for SPROX Analysis of Protein-Ligand Binding on the Proteomic Scale ... 109
  5.1 Introduction .......................................................................................................... 109
  5.2 Experimental ..................................................................................................... 110
    5.2.1 Yeast Growth ................................................................................................. 110
    5.2.2 NAD^+ Cofactor Binding Study .................................................................... 110
    5.2.3 Resveratrol Binding Study ............................................................................. 111
    5.2.4 Manassantin A Binding Study ....................................................................... 112
    5.2.5 Protein-Protein Interactions Studies ............................................................. 112
    5.2.6 General Proteomics Procedure ..................................................................... 112
    5.2.7 iTRAQ Labeling ........................................................................................... 113
List of Tables

Table 1. Summary of the four LOPAC ligands selected in this work. .......................... 41
Table 2. Summary of the five DIVERSet ligands selected in this work ..................... 42
Table 3. SPROX-derived thermodynamic parameters for BCA and Cpr1 in the presence of known ligands .............................................................. 66
Table 4. List of identified peptides from Cpr1 .............................................................. 73
Table 5. List of identified methionine-containing peptides ....................................... 94
Table 6. Met peptide selection efficiency ................................................................. 123
Table 7. Assay statistics ......................................................................................... 123
Table 8. List of peptides selected as hits from NAD⁺ binding SPROX analysis .......... 132
Table 9. List of peptides selected as hits from resveratrol binding SPROX analysis .... 140
Table 10. Proteins identified as targets of the proteins studied in this work .......... 144
List of Figures

Figure 1. Plot of ΔMass as a function of [Den] in the SUPREX analysis of the nanobody VHH122. Highlighted are the pre-transition region of the curve, the post-transition region of the curve, and the midpoint of the curve, or $C_{\text{SUPREX}}^{1/2}$, which in this plot is located at 2.7 M. 

Figure 2. Example of a $\Delta G_{\text{app}}$ vs. $C_{\text{SUPREX}}^{1/2}$ plot for empirically determining the free energy of the protein in the absence of denaturant, or $\Delta G_f$, and the $m$-value. The $\Delta G_f$ is the y-intercept of the plot, and the $m$-value is the slope. Here, the $\Delta G_f$ and $m$-value for the VHH122 nanobody are $-12.1 \pm 0.4 \text{ kcal/mol}$ and $2.1 \pm 0.1 \text{ kcal M}^{-1}\text{mol}^{-1}$, respectively. The reported uncertainties are the standard fitting errors of the data.

Figure 3. Apo periplasmic binding protein (apoPBP) in the presence of different metals. In order of increasing tightness of binding (from left to right), the curves represent apoPBP in the absence of any ligand, apoPBP in the presence of $\text{Zn}^{2+}$, apoPBP in the presence of $\text{FeNTA}$, apoPBP in the presence of $\text{Co}^{2+}$, and apoPBP in the presence of $\text{Ni}^{2+}$. Error bars associated with replicate measurements of each data point are not shown.

Figure 4. A. Theoretical SUPREX curves generated for CypA in the presence and in the absence of a hypothetical CypA binding ligand ($K_d$ value = 10 µM). The arrow indicates the [GdmCl] used in the single point SUPREX experiments in this work. B. Representative MALDI mass spectra of CypA from the screening assay. The top spectrum shows the CypA ion signal generated in the MALDI readout of a selected pool of LOPAC compounds that did not contain a hit, while the bottom spectrum shows the CypA ion signal generated in the MALDI readout of the pool containing Compound 3, which was identified as a hit.

Figure 5. Schematic representation of the screening methodology used in this work. Library compounds were screened as 10-compound pools to improve throughput, and CypA was added and allowed to exchange. The mass increase was then used to determine whether a binding event had occurred. The bar graph plot represents a portion of actual data collected in this work.

Figure 6. Distribution of ΔMass values recorded for the 1089 library pools (grey bars), 126 positive controls (black bars), and 125 negative controls (open bars) analyzed in the 1st tier screening of the two libraries used in this work. The solid lines represent the fit of each distribution to a normal Gaussian curve.

Figure 7. The ΔMass values from the positive (●) and negative (○) control samples analyzed with the LOPAC library compounds (A) and the DIVERSet library compounds...
(B). The solid lines represent the average \( \Delta \text{Mass} \) values of the control data and the dashed line represents the cut-off values used in this work. The negative control data point indicated with an arrow was determined to be an outlier and was not used to determine the cut-off value. Note that a 9-point moving average was used to determine the cut-off values in B. Shown in C is the distribution of \( Z' \)-factors observed in both library screenings.

Figure 8. Full SUPREX curves obtained for CypA in the absence (●) and presence (○) of the nine selected hit-compounds. The solid and dashed arrows indicate the transition midpoints obtained in the absence and presence of ligand, respectively. Eight of the nine had measurable binding due to a shift in the SUPREX curve. Compounds 5 and 8 did not produce a full SUPREX curve (see text), but the data points collected suggested a transition midpoint of at least 1.6 and 2.3 M, respectively. A measurable transition midpoint shift was not detected with Compound 6.

Figure 9. Results obtained in CypA inhibition studies. The closed bars and left y-axis represent the initial velocities determined for the chymotrypsin-catalyzed hydrolysis of the peptide substrate \( N \)-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide in the presence and absence of CypA and in the presence of CypA and different ligands, including CsA and the nine hit compounds identified in this study. The initial velocity values are the average of four to six determinations and the error bars represent 1 standard deviation. The open bars and the right y-axis indicate the fraction of CypA protein expected to be free in solution in the assay (i.e., unbound to the target ligands) based on the CypA protein concentration in the assay and the measured \( K_d \) value (see Table 1 and Table 2).

Figure 10. MALDI mass spectra from BCA immunoprecipitation. The peak at approximately 17,000 \( m/z \) is myoglobin, the internal standard used in this experiment. A. Initial flowthrough following capture of BCA/Anti-BCA on Protein G. B. Spectrum from a wash of the Protein G resin with PBS. C. Spectrum from an elution fraction showing BCA was present.

Figure 11. A. SPROX curves for pure BCA in the absence (●) and presence (○) of CBS. B. SPROX curves for BCA spiked into a lysate and subsequently immunoprecipitated with Anti-BCA in the absence (●) and presence (○) of CBS. Error bars represent the standard deviation from 10 spectra.

Figure 12. A. MALDI mass spectrum of yeast lysate overexpressing Cpr1. The arrow denotes where the Cpr1 was expected to be based on the mass of Cpr1 with the addition of the affinity tag. B. MALDI mass spectrum of affinity purified Cpr1 from the SPROX experiment. Peaks at approximately 10,000 and 20,000 \( m/z \) belong to the \([M+H]^+\) and \([M+2H]^{2+}\) peaks of soybean trypsin inhibitor, which was used as an internal standard in the SPROX experiment.
Figure 13. A. SPROX curve for whole protein MALDI readout of Cpr1 purified from yeast lysate. Error bars represent one standard deviation of the average mass from 10 replicate spectra each. B. SPROX curve of peptide-level Cpr1 in the absence (●) and presence (○) of CsA, which was also purified from yeast lysate following SPROX. No error bars are present because each data point is taken from one LC-MS run.

Figure 14. Schematic representation of the PrSUIT protocol developed in this work.

Figure 15. Theoretical structure of PrSUIT data and its relationship to SPROX data. A. Hypothetical PrSUIT data (i.e., %18O versus [Den] plots) expected for proteins that do not (Case I) and do (Cases II or III) have their thermodynamic stability altered in the presence of the ligand. Cases II and III correspond to situations in which the thermodynamic stability of the identified protein is stabilized and destabilized, respectively, as a result of ligand binding. B. Schematic representation of the molecular 16O and 18O oxidized species generated in the 1st and 2nd oxidation reactions in the PrSUIT protocol at three different points in a protein’s SPROX curves generated in the absence, (-), and in the presence, (+), of a binding ligand.

Figure 16. Relative ion signal intensities observed (black bars) and theoretically calculated (grey bars) for the isotopologues of the methionine-containing peptide, Cpr1(54-74) of sequence VIPDFMLQGGDFTAGNGTGGK, that were detected in the PrSUIT analysis of the model protein mixture. The data obtained at two different denaturant concentrations, 3.5 and 2.0 M GdmCl, are shown in A and B, respectively. In each case M represents the monoisotopic mass of the peptide, 2096.9837 Da. The theoretical distributions of the relative ion signal intensities for the isotopologues in A and B are those expected for the oxidized Cpr1(54-74) peptide with 41 and 11% 18O labeling, respectively, in the PrSUIT protocol.

Figure 17. Representative PrSUIT results obtained in the CsA binding analysis using the model protein mixture (closed symbols) and the purified proteins (open symbols). A. PrSUIT plot obtained for a methionine-containing peptide, Gdh2(920-943) of sequence GGVTSSSMEVLASLALNDNDFVHK, derived from a protein with no CsA-induced interactions. B. and C. PrSUIT plots obtained for two methionine-containing peptides, Cpr1 (54-74) of sequence VIPDFMLQGGDFTAGNGTGGK and Cna1(332-354) of sequence VTGFPSLITMFSAPNYLDTYHNK, respectively, each derived from proteins with known CsA-induced interactions. The CsA binding interaction with Cna1 is also known to require Cpr1. Note that the purified protein plots were constructed with %18O labeling values that were not normalized.

Figure 18. Distribution of all the normalized %18O labeling values at all the denaturant concentrations (black bars) obtained on the 27 methionine-containing peptides analyzed in the PrSUIT experiment performed on the model protein mixture, as well as the
contribution of normalized $^{18}$O values from just the negative control peptides (white bars).

Figure 19. SPROX curves generated from PrSUIT data collected in this work on Cpr1 and Cna1 in the model mixture analysis. A. SPROX curves generated for the Cpr1(54-74) peptide of sequence VIPDFMLQGDFTAGNGTGGK using the PrSUIT data shown in Figure 17B. B. SPROX curves generated for the Cna1(332-354) peptide of sequence VTGFPSLTMFSAPNYLDTYHNK using the PrSUIT data shown in Figure 17C.

Figure 20. Schematic representation of the protocol utilized in this study. SPROX was performed on yeast lysates, samples were labeled with iTRAQ, and Met peptides were isolated using immobilized bromoacetyl functional groups, enhancing the frequency of identified Met peptides in subsequent LC-MS/MS analyses.

Figure 21. Three of the ligands investigated in this study. A. NAD$^+$ B. Resveratrol C. Manassantin A

Figure 22. A. Numbers of peptides identified with ≥ 99% confidence in the NAD$^+$ binding assay. B. Number of proteins from the peptides in (A).

Figure 23. Distributions of iTRAQ intensities at low (dashed line) and high (solid line) denaturant for (A) NAD$^+$ binding, (B) manassantin A binding, (C) resveratrol, (D) and protein-protein interaction experiments.

Figure 24. A. Representative SPROX curve from the peptide IADTVGKDVVQLYQEQLSAQGMPMIK from glucokinase-1. B. One of the selected peptides that displayed NAD$^+$ binding, GDIESISDKTMYK from NAD$^+$-dependent glutamate dehydrogenase. The first iTRAQ intensity (0 M GdmCl) in the curve in the presence of NAD$^+$ is an example of an overlooked outlier intensity.

Figure 25. Methionine-containing peptides from eIF4A in the SPROX iTRAQ resveratrol-binding experiment. Arrows denote transition midpoints. A. (52-72) AIMPIIIEIGHDVLAQAQSGTGK B. (90-100) APQALMLAPTR C. (24-35) FDDMELDENLLR D. (357-373) KGVAINFVTNEDVGAMR. All peptides display similar midpoint transitions with and without resveratrol.

Figure 26. SPROX curves from purified eIF4A in presence (○) and absence (●) of resveratrol. A-D correspond to the same peptides found in Figure 25A-D. The N-terminal peptides found in A-C exhibited a slight change in SPROX behavior in the presence of resveratrol.

Figure 27. Potential hits of resveratrol. A. Cytosolic aldehyde dehydrogenase (QQFDTIMNYIDIGK) B. Elongation factor 3A (MPELIPVLSETMWDTK) C. 60S
ribosomal protein L8-A (MGVPYAIVK) D. 40S ribosomal protein S9-A (QIVNIPSFMVR).................................................................................................................. 139

Figure 28. Representative SPROX curves from detected PPIs. A. Pyruvate decarboxylase with Pgk1. B. Glyceraldehyde-3-phosphate dehydrogenase with Pgk1. C. Acetohydroxyacid reductoisomerase with Pyk1. D. A peptide from Pyk1 in the presence of excess Pyk1........................................................................................................................................ 146
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACN</td>
<td>acetonitrile</td>
</tr>
<tr>
<td>BME</td>
<td>2-mercaptoethanol</td>
</tr>
<tr>
<td>$C_{SPROX}^{1/2}$</td>
<td>denaturant concentration at the transition midpoint of a SPROX curve</td>
</tr>
<tr>
<td>$C_{SUPREX}^{1/2}$</td>
<td>denaturant concentration at the transition midpoint of a SUPREX curve</td>
</tr>
<tr>
<td>Cna1</td>
<td>yeast calcineurin A</td>
</tr>
<tr>
<td>CsA</td>
<td>cyclosporin A</td>
</tr>
<tr>
<td>CypA</td>
<td>human cyclophilin A</td>
</tr>
<tr>
<td>Cpr1</td>
<td>yeast cyclophilin A</td>
</tr>
<tr>
<td>[Den]</td>
<td>chemical denaturant concentration</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulfoxide</td>
</tr>
<tr>
<td>$\Delta G_{app}$</td>
<td>apparent Gibbs folding free energy at specific solution conditions</td>
</tr>
<tr>
<td>$\Delta G_f$</td>
<td>Gibbs free energy of protein folding in the absence of denaturant</td>
</tr>
<tr>
<td>$\Delta$Mass</td>
<td>Mass change as a result of chemical labeling (oxidation or $^2$H)</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>ESI</td>
<td>electrospray ionization</td>
</tr>
<tr>
<td>FA</td>
<td>formic acid</td>
</tr>
<tr>
<td>[GdmCl]</td>
<td>guanidinium chloride concentration</td>
</tr>
<tr>
<td>H/D exchange</td>
<td>hydrogen/deuterium exchange</td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatography</td>
</tr>
</tbody>
</table>
HTS  high-throughput screening
IDA  iodoacetamide
$K_d$  dissociation constant
LEM  linear extrapolation method
LC  liquid chromatography
MALDI  matrix-assisted laser desorption/ionization
Met peptide  methionine-containing peptide
MS  mass spectrometry
$m$-value  $\delta \Delta G/\delta [\text{Den}]$
$m/z$  mass to charge ratio of an ion
Pgk1  3-phosphoglycerate kinase
PPI  protein-protein interaction
Pyk1  pyruvate kinase
SPROX  stability of proteins from rates of oxidation
SUPREX  stability of unpurified proteins from rates of H/D exchange
TAP  tandem affinity purification
TCA  trichloroacetic acid
TFA  trifluoroacetic acid
Q-TOF  quadrupole time-of-flight
Acknowledgements

Graduate school has been part of a special season of my life, unlike any other period of my life. I cannot begin to describe all of the failures, accomplishments, and lessons learned in a couple pages, so I will just start thanking everyone that made this time transpire in a positive light. First, I must thank my advisor, Professor Michael Fitzgerald, for his role in shaping me as a scientist. I have appreciated his optimism and his guidance over the past few years. I’ve learned so much working under him, and am very grateful for his interest in my research.

Second, I would like to thank my committee members, Dr. Stephen Craig, Dr. Alvin Crumbliss, and Dr. Edward Patz, for monitoring my research progress during my time as a graduate student.

I would like to acknowledge both current and past Fitzgerald lab members. Specifically, Dr. Erin Hopper and Dr. Graham West for their friendship, patience, and guidance through my first few years. Additionally, Ying Xu, what a fun ride it has been over the last 5 years. It has been great working with you from the beginning, and I’m glad we ended up in the same lab. Finally, Hai-Tsang “Hubert” Huang, what a terrific undergraduate researcher you have been. It truly has been a pleasure working with you the last couple years, and I wish you the best as you now begin your career as a scientist. I hope I have given you an honest snapshot of what graduate school is like, and I know you will excel wherever you end up.
Special thanks to collaborators Dr. Claire Parker Siburt and Dr. Mike Campa. Also thanks to Jennifer Hawk, Dana Peles, and Julie Pollock for allowing me to teach the freshman seminar course with them. I would love to thank individually all the friendships I have developed through these graduate studies, but there is just not enough space (if you do not see your name above, consider yourself acknowledged). Also, special thanks to my parents for their support in my pursuit of education.

Last and most importantly, my wife Bethany for her love and support. Though she would deny it, I truly could not have done this without her. I still consider convincing her to marry me my greatest accomplishment.
1. Introduction

1.1 Background

Proteins perform a wide variety of different biological functions within living organisms. These functions often require the non-covalent association of proteins with other ligand molecules. Ligand molecules represent a broad spectrum of chemical classes, ranging from small molecules [1, 2], to metals [3-5], peptides [6], nucleic acids [7], and other proteins [8, 9]. Some important examples of protein-ligand interactions that are involved in important biological processes include the interactions of proteins with nucleic acid sequences during the process of gene regulation and expression [10], transmembrane G-protein coupled receptors initiating signal transduction pathways due to ligand binding [11], or the activation of Ca$^{2+}$/calmodulin-dependent protein kinases after interaction with the protein calmodulin (which itself is activated by the divalent Ca$^{2+}$) [12]. The ability to characterize protein interactions with other molecules is crucial to understanding how proteins function at both microscopic and macroscopic levels. Additionally, the effects of therapeutic agents in biological systems can only be understood once the proteins that interact with the therapeutic agent are identified, in addition to whether any protein signal cascade is triggered. Such drug mode-of-action studies are a critical part of drug discovery and development.

The ability to detect interactions between proteins and ligands in the context of complex biological mixtures, where protein-protein interactions (PPIs) might influence ligand binding, has eluded researchers for years. Traditionally, the detection of protein targets of drugs required immobilization of potential ligands onto a necessary solid
support, which was problematic for two reasons. First, the methodology was not general to all ligands (i.e., each ligand required unique and often cumbersome immobilization methods), and second, the biophysical properties of the interactions could be affected by the immobilization of the ligand. Additionally, making measurements of the strength of binding required large amounts of purified protein, which experimentally is an inconvenient challenge. Moreover, the majority of current ligand-binding and protein target assays are limited to direct interactions between protein and ligand. However, if the interaction between a protein and ligand induces and/or precludes interactions with other proteins, these indirect binding events (i.e., “off-target” effects) can result in unpredicted and undesirable side effects. A general technique that can detect and measure these indirect binding events is presently lacking.

Extensive research has been dedicated to the discovery of protein targets of drugs [13-16] or of other proteins (i.e., PPI networks [8, 9, 17, 18]). The biological responses of drugs can be explained by their interactions with proteins, both direct and indirect. Therapeutic agents work to inhibit or activate specific enzymes, disrupt protein-protein interfaces, or trigger a protein-mediated chemical chain reaction. PPIs are responsible for numerous cellular processes, and their misregulation often results in disease such as cancer, HIV, and diabetes [19]. As a result, PPIs have themselves recently become therapeutic targets [20-22]. However, identifying the proteins that make up the PPI network is a necessary prerequisite before therapeutic targets can be identified. The ability to detect both types of targets (i.e., of both drugs and proteins) is invaluable to recognizing how ligands exert their effects in intricate biological systems. While it is still
common practice to select a particular protein target (or protein-protein interface) of interest and develop ligands specific to that target (e.g., rational drug design), these approaches have their own limitations and often do not lead to significant or desired in vivo activity [23, 24]. Consequently, much emphasis is placed on first identifying desired phenotypic responses from drug candidates and then identifying the protein targets of those drugs.

Multiplexed assays of protein-ligand binding, the development and application of which are the focus of the work in this dissertation, can be classified into two categories. The first category involves protein-ligand binding assays in which the goal is to characterize the binding of one protein target to many potential ligands. The ready availability of natural and synthetic compound libraries of peptides, oligonucleotides, and small molecules (MW <1000 Da) has created a new route to the discovery of novel ligands that bind to specific protein targets. This route, which typically involves the use of high-throughput screening (HTS) strategies to rapidly screen library members for protein-ligand binding, can be useful for the identification of new protein-targeted diagnostic and/or therapeutic agents.

The second category involves protein-ligand binding assays in which the goal is to characterize the binding of one ligand to many proteins. This category of protein-ligand binding assays requires the analytical ability to detect and quantify protein-ligand binding interactions in multi-component protein mixtures. Currently there are very few protein-ligand binding assays that can be used for the detection and quantitation of protein-ligand binding interactions in multi-component mixtures. Even fewer techniques
exist for analysis of both the direct and indirect binding events that can result from protein-ligand binding interactions. Conventional protein-ligand binding techniques involving spectroscopic or calorimetric methods can only be applied to the analysis of highly purified proteins. While experimental approaches, such as the yeast two-hybrid assay [25-28], have been developed for the large scale and high-throughput analysis of protein-ligand binding interactions, such approaches are generally not quantitative and largely limited to the direct analysis of binary interactions involving two different protein components. Indirect interactions (e.g., those that are induced or precluded as a result of another direct ligand binding interaction) and protein interactions with non-protein ligands are more difficult to assay using the yeast two-hybrid experiments.

Several mass spectrometry-based methods for the analysis of protein-ligand binding interactions have also been developed [29-31]. Mass spectrometry-based proteomic methods used in conjunction with protein complex purification techniques have yielded useful information about protein-protein binding interactions in complex biological mixtures. However, like the yeast two-hybrid binding assay, they have been largely limited to the analysis of protein-protein interactions, and they do not generally yield quantitative information about protein-ligand binding affinities. More recently, amide H/D exchange and covalent labeling methods have been used in combination with mass spectrometry to detect and quantify protein-ligand binding interactions [32-35]. While these approaches have proven useful for the analysis of a wide variety of protein-ligand binding interactions, they are limited to the analysis of ligand binding interactions in relatively simple protein mixtures.
Common strategies for the detection of protein targets of drugs include affinity chromatography, protein microarrays [36], yeast three-hybrid systems [37], or global proteomic strategies such as SILAC [38], ICAT [39], and DIGE [40]. PPIs are often discovered through the use of yeast two-hybrid screening [41], phage display [42], chemical cross-linking [43], tandem affinity purification (TAP) tags [44], or genetic interactions [45]. Current strategies for target identification are limited by the reality that most can only detect direct binding interactions, most are highly specific for certain ligands/systems, low abundant proteins often go undetected, many only detect tight-binding interactions, some are relegated to relatively simple mixtures, and most suffer from high false positive rates. Overall, there is no general methodology that can detect targets of both drugs and proteins that is appropriate for indirect interactions, low affinity interactions, and low abundance proteins within the context of the complex proteome.

A large part of the work in this dissertation is focused on the development of new methodologies for the detection and measurement of protein interactions in complex mixtures, and it concerns the discovery of either molecular targets of proteins or protein targets of small molecules. The methodologies discussed in later chapters are based on the SUPREX (stability of unpurified proteins from rates of H/D exchange) and SPROX (stability of proteins from rates of oxidation) techniques developed in the Fitzgerald lab [46, 47], which have been developed to circumvent the previously mentioned limitations in addition to the inadequacies of traditional methods for measuring protein-ligand interactions. Traditional methods for the measurement of protein-ligand interactions include calorimetric (e.g., differential scanning (DSC) and isothermal (ITC)) methods,
spectroscopic (e.g., fluorescence, absorbance, NMR, Raman, and circular dichroism) methods, and surface plasmon resonance (SPR). DSC and ITC measure the heat evolved from thermal denaturation or molecular association reactions, respectively [48]. Spectroscopic methods generally probe structural or vibrational differences [49] as well as the degrees to which proteins are unfolded using thermal or chemical denaturation [50]. SPR measures the kinetic on and off rates of ligand association using the optical phenomenon associated with changes in refractive index near a biospecific surface under conditions of total internal reflection [51]. However, these traditional methods for protein-ligand binding measurements require large amounts of purified protein. Fluorescent techniques often require the presence of a native fluorophore. Additionally, DSC methods and spectroscopic thermal denaturation techniques determine binding parameters at temperatures where the proteins are thermally denatured.

To get around these limitations, SUPREX and SPROX were developed as an alternative means to measure stabilities of proteins and their non-covalent interactions with ligands. These techniques involve the use of chemical modification strategies analogous to traditional chemical denaturation techniques that employ spectroscopic readouts to make thermodynamic measurements of protein folding and stability. Such chemical denaturation techniques connect a protein’s stability with the degree to which the protein is unfolded at increasing concentrations of chemical denaturant ([Den]). SUPREX and SPROX utilize either MALDI and/or ESI mass spectrometry as the final readout of mass increases due to the chemical modification, and as a result can manage well with low levels of unpurified protein.
1.2 Equilibrium Unfolding Properties of Proteins

The protein folding/unfolding equilibrium of reversible, two-state folding proteins can be described by the following equilibrium [52]:

\[ N \rightleftharpoons U \]  \hspace{1cm} \text{Equation 1-1}

\[ K = \frac{[N]}{[U]} \]  \hspace{1cm} \text{Equation 1-2}

In equations 1-1 and 1-2, \( N \) is the fully folded native structure, \( U \) is the disordered, unfolded state, and \( K \) is the equilibrium constant describing that relationship. The Gibbs free energy between the two states is given by \( \Delta G_f \), the free energy of folding, and has commonly been measured using a combination of spectroscopy and denaturation. In chemical denaturation studies, the protein sample is equilibrated in buffers containing increasing [Den] and the degree to which the protein is unfolded at each [Den] is determined, typically using a conventional spectroscopic technique such as CD, UV-Vis, or fluorescence. Ultimately, the \( \Delta G_{app} \) can be obtained from the resulting denaturation curves using the following equations [53]:

\[ K = \frac{y_N - y_{obs}}{y_{obs} - y_U} \]  \hspace{1cm} \text{Equation 1-3}

\[ \Delta G_{app} = -RT \ln K \]  \hspace{1cm} \text{Equation 1-4}

In equation 1-3, \( y_{obs} \) is the observed value of the extent of unfolding, and \( y_N \) and \( y_U \) are the values of \( y \) for the fully folded and fully unfolded protein, respectively. In equation 1-4, \( \Delta G_{app} \) is the apparent free energy at a specific [Den], \( R \) is the universal gas constant, \( T \) is the temperature in kelvins, and \( K \) is the folding equilibrium constant.
During initial denaturation curve studies, it was observed that log $K$ varied linearly with the log of the denaturant concentration ([Den]) and that $\Delta G$ also varied linearly with [Den] [54, 55]. From this, an analysis of denaturation curves using the linear relationship known as the linear extrapolation method (LEM) was established, which is shown in equation 1-5 [53].

$$\Delta G_{\text{app}} = \Delta G_f + m[\text{Den}]$$  \hspace{1cm} \text{Equation 1-5}

In equation 1-5, $\Delta G_f$ is the free energy of protein folding in the absence of denaturant, and $m$ is simply a measure of the dependence of $\Delta G$ on the [Den]. It is assumed in this equation that the relationship between $\Delta G$ and [Den] continues to 0 M denaturant. The $m$-value was later shown to strongly correlate to the change in a protein’s solvent-accessible surface area upon unfolding [56].

### 1.3 H/D Exchange and Chemical Modification Theory

SUPREX and SPROX are based on the rates of certain modification reactions in proteins as the protein’s folding equilibrium is shifted using chemical denaturants. SUPREX is based on protein amide H/D exchange [57], while SPROX uses methionine oxidation with hydrogen peroxide [58]. Specifically, these techniques probe the global folding/refolding reactions in proteins as shown in equations 1-6 and 1-7:

\begin{align*}
\text{SUPREX} & \quad P_F^H \xrightleftharpoons[k_f]{k_u} P_U^H \xrightleftharpoons[k_{\text{int}}]{k_u} P_U^D \xrightleftharpoons[k_f]{k_u} P_F^D \\
\text{SPROX} & \quad P_F^{\text{Met}} \xrightleftharpoons[k_f]{k_u} P_U^{\text{Met}} \xrightleftharpoons[k_{\text{ox}}]{k_u} P_U^{\text{Met-O}} \xrightleftharpoons[k_f]{k_u} P_F^{\text{Met-O}}
\end{align*}

\hspace{1cm}  \text{Equation 1-6 and 1-7}
In equations 1-6 and 1-7, $P_F^H$ and $P_F^{Met}$ represent fully folded proteins that are either protonated or unoxidized, respectively, and $P_U^H$ and $P_U^{Met}$ represent the unfolded versions of those proteins. $P_F^D$ and $P_F^{Met-O}$ are the fully folded proteins that are deuterated and oxidized, respectively, and $P_U^D$ and $P_U^{Met-O}$ represent the deuterated and oxidized versions, respectively, after the protein has unfolded. The rate constants $k_u$ and $k_f$ are the rates for the global unfolding and refolding of the protein, respectively, $k_{int}$ is the intrinsic, chemical exchange rate for an amide proton in H/D exchange, and $k_{ox}$ is the pseudo-first order rate constant for the oxidation of a methionine side chain. Equations 1-6 and 1-7 characterize cases in which the protein is well modeled by a two-state folding process (i.e., the protein predominantly exists in fully folded or fully unfolded states, and no partially folded intermediates are present) [52]. One assumption in these equations is that a globally protected amide proton or methionine side chain will only be labeled if it is first exposed to the bulk solvent. Additionally, it is important to note that the oxidation of a protein can occasionally affect the thermodynamic stability of a protein, which usually results in shifting the right side of equation 1-7 towards the unfolded state. However, the kinetics of the unfolding and folding reactions for the unoxidized protein will not change, and so being able to monitor the unperturbed protein can become an experimental advantage (e.g., see how isobaric mass tags are used to do this in Chapter 5).

Exactly how the stability of a protein is connected to these rates can be explained by classical H/D exchange theory, which is also applicable to the oxidation reaction involved with SPROX. Both the deuteration of an amide proton or the oxidation of a
methionine side chain will henceforth be referred to as “labeling”. From equation 1-6, the observed rate of labeling \( (k_{ex}) \) can be derived as the following quadratic equation [57]:

\[
k_{ex} = \frac{k_u + k_f + k_{int} - \sqrt{(k_u + k_f + k_{int})^2 - 4k_u k_{int}}}{2}
\]

Equation 1-8

Under native conditions where \( k_f \gg k_u \), equation 1-8 can be simplified into a second order reaction as follows:

\[
k_{ex} \approx \frac{k_u k_{int}}{k_f + k_{int}}
\]

Equation 1-9

There exist two kinetic limits in this model of H/D exchange. The first, termed EX1 [57], occurs when \( k_f << k_{int} \), and as a result, \( k_{ex} \) equals \( k_u \). In other words, when the rate of labeling is much faster than the refolding rate, the protein will be labeled every time the protein opens itself up to the bulk solvent. Under exchange conditions termed EX2 [57], \( k_f >> k_{int} \), and the protein must unfold many times before the protein is labeled. Under EX2 conditions, equation 1-8 can be written as the following first order reaction:

\[
k_{ex} = k_{int} \frac{k_u}{k_f} = k_{int} K_u
\]

Equation 1-10

In equation 1-10, \( K_u \) is the equilibrium constant for the protein’s unfolding reaction, assuming that the unfolding reaction described in equation 1-10 is dominated by a global unfolding reaction. Therefore under EX2 conditions, the apparent free energy of the protein’s structural unfolding reaction can be determined by using the relationship found between \( K \) and \( \Delta G \) in equation 1-4 using techniques that can measure \( k_{ex} \), such as NMR. This assumes that \( k_{int} \) is calculated accurately using known relationships between an amino acid’s \( k_{int} \) and p\( \text{H} \), temperature, and the location of the amino acid within the
protein [59, 60]. The H/D exchange technique has been used to accurately measure the $\Delta G_f$ of many proteins that match values obtained from spectroscopic and calorimetric methods [61].

1.4 Development of SUPREX and SPROX

Through the use of the LEM given by equation 1-5, SUPREX and SPROX combine H/D exchange and Met oxidation, respectively, with solvent denaturation curve methodologies to accurately assess the stabilities of proteins. The SUPREX technique involves the dilution of the protein of interest into a series of chemical denaturant-containing buffers (SUPREX buffers are deuterated). The labeling reaction is initiated and proceeds for a specified amount of time before quenching the reaction. As the increasing [Den] shifts the protein’s folding equilibrium (equation 1-1) towards the right, the increased solvent exposure of the globally protected backbone amide protons or buried Met residues results in an increased frequency of labeling. The increase in mass as a result of the deuterium uptake or Met oxidation, which is measured using a mass spectrometry readout, results in a typical plot shown in Figure 1. The curve in Figure 1 looks very similar to a chemical denaturation curve measured with spectroscopic techniques [50].

1.4.1 SUPREX and SPROX Data Analysis Method 1

To extract meaningful stability parameters from the curve shown in Figure 1, equation 1-5, which plots the protein’s $\Delta G_{app}$ as a function of [Den] to obtain $\Delta G_f$, can be used in two different data analysis methods. Equation 1-10 shows how the protein’s
folding equilibrium is related to the observed rate of labeling. Because a folded protein has $K_u << 1$, equation 1-10 can be rewritten as [46, 62]:

$$k_{ex} = \frac{K_u k_{int}}{K_u + 1}$$  

Equation 1-11

Substituting $1/K_f$ for $K_u$ and rearranging gives:

$$K_f = \frac{k_{int}}{k_{ex}} - 1$$  

Equation 1-12

Using data analysis method 1 [46], the increase in mass of the protein ($\Delta$Mass) as a result of labeling is estimated by the following first-order rate equation:

$$\Delta$Mass = $\Delta$M$_{\infty}$ + ($\Delta$M$_0$ - $\Delta$M$_{\infty}$)$e^{-k_{ex}t}$  

Equation 1-13

In equation 1-13, $\Delta$M$_0$ is the $\Delta$Mass before global labeling, $\Delta$M$_{\infty}$ is the $\Delta$Mass after global labeling, and $t$ is the time of the labeling reaction. Combining equations 1-4, 1-5, 1-12, and 1-13 results in equation 1-14.

$$\Delta$Mass = $\Delta$M$_{\infty}$ + ($\Delta$M$_0$ - $\Delta$M$_{\infty}$)$e^{-\left(\frac{\langle k_{int} \rangle}{1+e^{-(\Delta G_f + m[Den])/RT}}\right)t}$  

Equation 1-14

In equation 1-14, $\langle k_{int} \rangle$ represents the average of the $k_{int}$ values for all globally exchanging amide protons, and in the case of SPROX, $\langle k_{int} \rangle$ is replaced with $k_{ox}$. Therefore, the plot of $\Delta$Mass as a function of [Den] can be fit to equation 1-14 to extract $\Delta G_f$ and $m$-values. However, because the $m$-value depends very much on the number of data points found within the transition region of the curve, the $m$-value is often estimated based on the protein’s size [56]. This has led to an alternative method of data analysis, which determines both $\Delta G_f$ and $m$-values empirically from SUPREX or SPROX data.
Figure 1. Plot of ΔMass as a function of [Den] in the SUPREX analysis of the nanobody VHH122. Highlighted are the pre-transition region of the curve, the post-transition region of the curve, and the midpoint of the curve, or $C^{1/2}_{\text{SUPREX}}$, which in this plot is located at 2.7 M.
1.4.2 SUPREX and SPROX Data Analysis Method 2

At the midpoint of a SUPREX curve ($C_{SUPREX}^{1/2}$), the H/D exchange reaction has progressed to one half-life under those specific denaturing conditions, and as a result, the observed exchange rate is equal to:

$$k_{ex} = \frac{\ln 2}{t}$$  \hspace{1cm} \text{Equation 1-15}

Again, $t$ here is the time of the labeling reaction used. Substitution of equation 1-15 into equation 1-12 yields an expression for $K_f$ in terms of $k_{int}$ and $t$, shown in equation 1-16.

$$K_f = \frac{(k_{int})t}{\ln 2} - 1$$ \hspace{1cm} \text{Equation 1-16}

The behavior of monomeric proteins within the SUPREX technique can then be described by the following equation, resulting from the substitution of equation 1-16 into equations 1-4 and 1-5.

$$-RT\ln \left(\frac{(k_{int})t}{\ln 2} - 1\right) = mC_{SUPREX}^{1/2} + \Delta G_f$$ \hspace{1cm} \text{Equation 1-17}

In equation 1-17, $C_{SUPREX}^{1/2}$ is substituted for [Den] only because the $C_{SUPREX}^{1/2}$ is merely the [Den] at which the midpoint of the SUPREX transition occurs. Equation 1-17 also assumes that the protein under study exchanges under EX2 conditions and that the protein is well modeled by two-state folding. Plotting the $\Delta G_{app}$ (i.e., the left side of equation 1-17) versus the $C_{SUPREX}^{1/2}$ for multiple H/D exchange times results in an empirically derived $m$-value and $\Delta G_f$ (see Figure 2).
Figure 2. Example of a $\Delta G_{\text{app}}$ vs. $C_{\text{SUPREX}}^{1/2}$ plot for empirically determining the free energy of the protein in the absence of denaturant, or $\Delta G_f$, and the $m$-value. The $\Delta G_f$ is the y-intercept of the plot, and the $m$-value is the slope. Here, the $\Delta G_f$ and $m$-value for the VHH122 nanobody are $-12.1 \pm 0.4$ kcal/mol and $2.1 \pm 0.1$ kcal M$^{-1}$mol$^{-1}$, respectively. The reported uncertainties are the standard fitting errors of the data.
1.5 Measurement of Protein-Ligand Interactions Using SUPREX and SPROX

Because the SUPREX and SPROX methodologies measure the stability of proteins, they have the inherent capability of measuring protein-interactions because the stability of a protein changes as a result of ligand binding. Shown in equations 1-18 to 1-20 are derivations for the free energy of binding [63].

\[
P_U \xrightleftharpoons[k_{f1}]{k_{u1}} P_F, \quad \Delta G_{f1} = -RT \ln \left( \frac{k_{f1}}{k_{u1}} \right) \quad \text{Equation 1-18}
\]

\[
P_U + L \xrightleftharpoons[k_{f2}]{k_{u2}} P_F L, \quad \Delta G_{f2} = -RT \ln \left( \frac{k_{f2}}{k_{u2}} \right) \quad \text{Equation 1-19}
\]

\[
\Delta G_B = \Delta \Delta G_f = \Delta G_{f2} - \Delta G_{f1} \quad \text{Equation 1-20}
\]

In equations 1-18 to 1-20, \( L \) represents ligand, \( \Delta G_B \) is the free energy of binding, and \( \Delta \Delta G_f \) is the change in the folding free energy as a result of ligand binding. These equations show that the free energy of binding is equal to the difference in the protein’s folding free energy in the presence and absence of excess ligand. Because the free energy changes as a result of ligand binding, this affects the midpoint of the SUPREX and SPROX curves. Shown in Figure 3 is an example of a protein binding to various metals with different affinities. The \( C_{\text{SUPREX}}^{1/2} \) shifts to different extents based on the tightness of the binding.

The dissociation constant, \( K_d \), of a protein-ligand interaction can then be given by equation 1-21, assuming a ten-fold excess of ligand [63]:

\[
K_d = \frac{[P_F] [L]}{[P_U]}
\]
Figure 3. Apo periplasmic binding protein (apoPBP) in the presence of different metals. In order of increasing tightness of binding (from left to right), the curves represent apoPBP in the absence of any ligand, apoPBP in the presence of Zn$^{2+}$, apoPBP in the presence of FeNTA, apoPBP in the presence of Co$^{2+}$, and apoPBP in the presence of Ni$^{2+}$. Error bars associated with replicate measurements of each data point are not shown.
\[ K_d = \frac{[L]}{\left( e^{-\Delta G_f/nRT} - 1 \right)} \]  

Equation 1-21

In equation 1-21, \([L]\) is the concentration of free ligand, and \(n\) is the number of independent binding sites on the protein.

Since its introduction, SUPREX and SPROX have proven quite useful in many biophysical applications, including the determination of stabilities of different protein mutants [46, 64], protein small-molecule interactions [65-67], protein-DNA interactions [32, 68], protein-peptide interactions [62], protein-protein interactions [69-71], multimeric proteins [72], and ligand binding in complex mixtures [73-75].

1.6 Overview of Research Objectives

The purpose of this work was to develop methodologies for the multiplex analysis of protein-ligand binding interactions. The new protocols developed in this work are derived from the basic SUPREX and SPROX techniques described above. Two types of protocols were developed in this work including one type that was aimed at identifying binding interactions in systems containing one protein in the presence of many potential ligands and another aimed at identifying the binding of one ligand to many potential protein targets. Chapter 2 is focused on a one protein-many ligand approach that deals with determining small molecules that bind to the protein cyclophilin A from pools of potential target. Chapters 3-5 discuss the development and application of several different strategies for studying the binding of one ligand with many different proteins.
Described in Chapter 2 is a high-throughput screening (HTS) method based on the SUPREX technique. The protein of interest in this study was human cyclophilin A (CypA), which has been shown to be overexpressed in lung cancer tumor cells and appears to play an important role in lung tumor growth. The goal of this work was to identify novel CypA ligands that could potentially be useful in the development of new CypA-targeted molecular imaging agents for the diagnosis of lung cancer. Two chemical libraries were screened against CypA using the SUPREX-based HTS method, and a compound-pooling strategy was used to increase the screening rate of the SUPREX protocol to less than 10 seconds/ligand.

Discussed in Chapter 3 is the use of protein purification strategies for simplifying the mass spectrometry readouts following the SPROX analysis of proteins in a complex mixture. Specifically, the use of antibodies and affinity tags were used to demonstrate the methodology using several different model protein-ligand systems. The protein bovine carbonic anhydrase II (BCA) was purified from a complex protein mixture using antibodies specific to BCA following the SPROX reaction, and binding to the small molecule 4-carboxybenzenesulfonamide (CBS) was measured. Additionally, the use of a recombinant affinity tag on yeast cyclophilin A (Cpr1) was used to measure its affinity to the immunosuppressive drug cyclosporin A (CsA) in the context of a yeast cell lysate.

The work in Chapter 4 involves the initial development and application of a new stable isotope labeling strategy termed PrSUIT (Probing with SPROX Using Isotope Tags) for the detection and quantitation of protein-ligand binding interactions by SPROX in multi-protein mixtures. The strategy involves performing SPROX analyses on a
protein mixture using differentially labeled hydrogen peroxide in the oxidation reactions in the form of H$_2^{18}$O$_2$ and H$_2^{16}$O$_2$. A quantitative LC-MS based proteomics readout was used to identify binding proteins, both direct and indirect, of the target ligand CsA. The PrSUIT technique was applied to a model protein mixture in order to evaluate its ability to detect binding between CsA and Cpr1 as well as the indirect interaction of CsA with calcineurin that involves the Cpr1-CsA complex.

The work in Chapter 5 involves development and application of a SPROX-based platform for the detection of binding targets on a global proteomics scale. As part of this work a selective isolation strategy for the enhancement of methionine-containing peptides was used to expand the scope of the existing platform. Combined with a quantitative proteomic strategy, this facilitated the thermodynamic analysis of an increased number of proteins in the multiplexed assay. The strategy was initially used to identify NAD$^+$ binding proteins in a yeast cell lysate as part of a proof-of-principle experiment. It was also employed in experiments designed to identify the protein targets of two small molecules (resveratrol and manassatin A), as well as the detection and quantitation of protein-protein interaction networks in *S. cerevisiae* lysates.
2. Discovery of Novel Cyclophilin A Ligands Using an H/D Exchange- and Mass Spectrometry-Based Strategy

2.1 Introduction

This chapter describes the use of a SUPREX-based assay to screen two moderately sized chemical libraries against a protein targeted for its possible diagnostic and/or therapeutic potential. This work describes a pooled compound approach for higher throughput than previously attained with the same assay. The protein target in this work, CypA, is a peptidyl-prolyl cis-trans isomerase that has been shown to be overexpressed in lung cancer tumor cells [76, 77], and as such is an attractive target for potential diagnostic imaging probes. It also appears to play an important role in lung tumor growth [78] and thus could be an important therapeutic target as well. A goal of this work was to identify novel CypA ligands that might be useful in the development of new CypA-targeted molecular imaging agents for the diagnosis of lung cancer. To this end, small molecule compounds in two chemical libraries were screened for CypA binding. One library was the Library of Pharmacologically Active Compounds, or LOPAC library, which consists of 1280 pharmacologically active compounds that span a range of cell signaling and neuroscience areas and that include a number of already marketed drugs. The other library was the 9600-member DIVERSet library, which consists of a highly diverse collection of drug-like small molecules selected through the use of a 3D pharmacophore analysis method to ensure broad coverage among the various types of interactions between target and ligand.
The high-throughput screening assay used in this work relies on a variation of SUPREX, termed single-point SUPREX, to detect the increase in a protein’s thermodynamic stability upon ligand binding. Initially, the single-point SUPREX assay was developed in the context of a model protein-peptide system [79-81], and the assay was used to screen the 880 compounds in the Prestwick Chemical Library for binding to CypA [80]. While no new CypA ligands were identified in the Prestwick Chemical Library screening, that work did establish a 3 minute/ligand throughput of the single-point SUPREX protocol. As part of the current work, a compound-pooling strategy is described that increases the screening rate of the single-point SUPREX protocol to less than 10 seconds/ligand. The results demonstrate the feasibility of the new methodology for screening 1,000-10,000 member libraries, and identify eight novel ligands to CypA. The work in this chapter was published in the Journal of Biomolecular Screening\(^1\).

2.2 Experimental

2.2.1 Materials

The human CypA used in this work was obtained from T7 Express E. coli transformed with pGEX-CypA, a plasmid encoding the glutathione S-transferase (GST) fusion protein of CypA. Briefly, the E. coli was grown up in LB-Amp media (sodium chloride, tryptone, yeast extract, and ampicillin) at 37°C. Isopropyl β-D-1-\(^1\)

thiogalactopyranoside (Mediatech, Inc., Mannassas, VA) was added to a final concentration of 0.125 mM to induce overexpression when the OD$_{600}$ of the culture reached 0.6-0.8 units. After 3.5 h, the E. coli was pelleted by centrifuging at 4000 x g for 10 min. Cells were lysed with the addition of BugBuster reagent (EMD Chemicals, Inc., Gibbstown, NJ) and the insoluble material was pelleted at 17,000 x g for 20 min. CypA was purified from the supernatant using GST•Bind Resin (EMD Chemicals, Inc.), and it was concentrated with a Vivaspin 20 (MWCO 10,000) centrifugal concentrator (Sartorius, Bohemia, NY). Thrombin (EMD Chemicals, Inc.) was added to remove the GST tag. The GST tag was subsequently removed via the GST•Bind Resin, and the thrombin was removed with a HiTrap Benzamidine FF column (GE Healthcare Bio-Sciences Corp., Piscataway, NJ). The CypA was further concentrated to about 100 µM. All CypA concentrations were determined using absorbance measurements at 280 nm and an extinction coefficient of 8730 M$^{-1}$ cm$^{-1}$ as calculated through ProtParam using the protein’s amino acid sequence (http://ca.expasy.org/tools/protparam.html).

The LOPAC library (Sigma-Aldrich, St. Louis, MO) was obtained from the Small Molecule Synthesis and Screening Facility in the Center for Chemical Biology at Duke University. The library compounds were provided as 10 mM solutions in dimethyl sulfoxide (DMSO) and were pipetted into 128 10-member pools for a resulting concentration of 1 mM of each compound. The 9600-compound DIVERSet library was from ChemBridge Corp. (San Diego, CA) and was provided as 10 mM solutions in DMSO. The DIVERSet library compounds were combined into 960 10-member pools for a resulting concentration of 1 mM of each compound. 3-[5-(4-Fluorobenzylidene)-4-
oxo-2-thioxo-1,3-thiazolidin-3-yl]propanoic acid, N-(3-chloro-1,4-dioxo-1,4-dihydro-2-naphthalenyl)-N-cyclohexylacetamide, 1-(3,5-dichlorophenyl)-1H-pyrrole-2,5-dione, 4-(2-methoxy-4-nitrophenoxy)-1-methyl-10-oxa-4-azatricyclo[5.2.1.0~2,6~]dec-8-ene-3,5-dione, and 1-[2-(3,4-dimethoxyphenyl)ethyl]-1H-pyrrole-2,5-dione were from ChemBridge Corp. (San Diego, CA). Cyclosporin A (CsA) was purchased from LKT Laboratories (St. Paul, MN). Deuterium oxide (D$_2$O; 99.9% atom D), sodium deuteroxide (40 wt.% in D$_2$O, 99.9% atom D), sinapic acid (98% titration), and cis-diammineplatinum (II) chloride (cisplatin) were from Sigma-Aldrich (St. Louis, MO). (E)-3-(4-t-butylphenylsulfonyl)-2-propenenitrile (Bay 11-7085) and 2-(α-naphthoyl)ethyltrimethylammonium iodide (α-NETA) were purchased from Alexis Biochemicals (Lausen, Switzerland). 3-(N-benzyl-N-isopropyl)amino-1-(naphthalen-2-yl)propan-1-one hydrochloride (ZM 39923) was from Tocris Bioscience (Ellisville, MO). DMSO was from Acros Organics (Geel, Belgium), and phosphoric acid-d$_3$ was from Cambridge Isotope Laboratories, Inc. (Andover, MA). Guanidinium chloride (GdmCl) was purchased from EMD Chemicals, Inc. (Gibbstown, NJ), and trifluoroacetic acid (TFA) was from Halocarbon (River Edge, NJ). Acetonitrile (ACN) was purchased from Fisher (Fair Lawn, NJ). The chymotrypsin, N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide, and trifluoroethanol were all from Sigma-Aldrich (St. Louis, MO).

2.2.2 SUPREX Buffer Preparation

A deuterated solution of 20 mM sodium phosphate buffer, pD 7.4, containing 1.5 M GdmCl was used for the single point SUPREX experiments. A series of solutions containing 20 mM phosphate buffer, pD 7.4, containing GdmCl concentrations ranging
from 0 to 4 M was prepared for the full SUPREX analyses performed on CypA in the presence of the hit compounds. The GdmCl used for buffer preparation was deuterated by repeated dissolution and lyophilization (4 times total) of 19 g GdmCl in 25 mL D$_2$O. The pD of the buffers were adjusted by adding sodium deuteroxide and monitoring the pH using a Jenco microcomputer pH-Vision 6072 pH meter equipped with a Futura calomel pH electrode from Beckman (Fullerton, CA). The pD of the buffers was obtained by adding 0.4 to the pH reading to correct for isotope effects [82]. The GdmCl concentration in the buffers was measured using a refractometer (Bausch & Lomb, Rochester, NY) utilizing the method reported by Nozaki [83].

2.2.3 Mass Spectrometry

MALDI mass spectra were acquired on an Ultraflex II TOF/TOF from Bruker Daltonics (Billerica, MA). Spectra were the sum of 100 laser shots collected in the linear mode using a 100 Hz Nd:YAG laser. Positive ion mass spectra were manually collected using the following parameters: an ion source 1 voltage of 25 kV, an ion source 2 voltage of 23.4 kV, a lens voltage of 6.5 kV, and a delay time of 100 ns. The MALDI matrix was sinapic acid, and it was prepared as a saturated solution containing 25% water and 75% ACN with 0.1% TFA.

2.2.4 Library Screening (Single-Point SUPREX Protocol)

For the screening assay, 1 µL of each 10-member pool containing 1 mM of each library compound in DMSO was combined with 9 µL of the deuterated exchange buffer containing 1.5 M GdmCl. A 1 µL aliquot of protonated CypA solution (i.e., CypA dissolved in H$_2$O-containing buffer), which was ~100 µM in CypA, was added to the
resulting 10 µL volume of the D$_2$O-containing compound/exchange buffer, and the H/D exchange reaction proceeded for 35 minutes before it was quenched by transferring 1 µL of the exchange reaction to 9 µL of ice-cold matrix solution. A 1-2 µL aliquot of the quenched reaction mixture was spotted onto a MALDI sample target and dried with N$_2$. Five replicate MALDI mass spectra were collected to determine an average molecular mass of the deuterated CypA for each sample, and this was used to calculate an average ΔMass value. Average ΔMass values were obtained by subtracting the expected average molecular mass of the fully protonated CypA from the average molecular mass measured for each sample. Samples were prepared and analyzed in sets of twelve using a 12-channel pipette. Each set of twelve samples included ten pools of 10 compounds, one positive control containing 1 mM CsA ($K_d$ ~30-200 nM [73, 84-87]) and one negative control containing only DMSO. The H/D exchange reactions were staggered at 10 min intervals, so while one sample set was undergoing H/D exchange, MALDI mass spectra were being collected on the previous sample set.

The average molecular weight of CypA was determined from the MALDI-TOF mass spectra using either an Excel macro or MATLAB (The MathWorks, Inc., Natick, MA) script, both of which were developed in-house. Each takes the spectral data in text file form containing m/z and intensity values for each data point and performs a 19-point floating average smoothing of the data, a calibration using the internal standard, and a center of mass calculation for the CypA peak.

In the LOPAC library screening, the ΔMass values of the control data were averaged, and the cut-off value was determined by subtracting 2.5 standard deviations
from the average negative control $\Delta$Mass value. In the DIVERSet library screening, the $\Delta$Mass values of the control data from the screen were smoothed using a 9-point central moving average. An average $\Delta$Mass value (i.e., a $\Delta$Mass$_{av}$ value) for each negative control was calculated using the 9-point moving average and was subsequently used to calculate a cut-off value for each set of ten pools. The average of all the 9-point standard deviations was calculated, multiplied by 2.5, and then subtracted from the $\Delta$Mass$_{av}$ value to determine the cut-off values. Samples with $\Delta$Mass value less than or equal to the cut-off value were classified as preliminary hits (i.e., one of the compounds in the pool of 10 binds to CypA).

$Z'$-factors were calculated using equation 2-1 [88].

$$Z' = 1 - \frac{(3\sigma_{c+} + 3\sigma_{c-})}{|\mu_{c+} - \mu_{c-}|}$$

Equation 2-1

In equation 2-1, $\sigma_{c+}$ is the standard deviation of the positive control, $\sigma_{c-}$ is the standard deviation of the negative control, $\mu_{c+}$ is the mean of the positive control, and $\mu_{c-}$ is the mean of the negative control.

The preliminary pools classified as hits were subjected to a “second tier” analysis, in which the initial hit pools were re-screened using the single-point SUPREX protocol described above. The two-tier screening approach has previously been shown to reduce the occurrence of false positives in single-point SUPREX [80]. Each hit-pool was re-screened under the same conditions as the initial screen. The individual compounds from the pools that were confirmed as hit-pools in the second tier screen were individually screened using the single-point SUPREX protocol, just as described above for the
compound-pools, except that the ligand stock solution only contained a single ligand at 1 mM.

### 2.2.5 Binding Affinity Measurements (Full SUPREX Analyses)

Full SUPREX analyses were performed on CypA in the absence and in the presence of the individual compounds selected as hits. In these full SUPREX analyses 1 µL of a 1 mM DMSO solution of the hit ligand (or just DMSO in the case of the CypA analyses performed in the absence of ligand) was added to 9 µL of the SUPREX buffers, which in the full SUPREX analyses contained [GdmCl] concentrations ranging from 0 to 4 M (see above). A 1 µL aliquot of protonated CypA solution, which was ~100 µM in CypA, was added to each 10 µL volume of compound/exchange buffer, and the H/D exchange reaction proceeded for 35 minutes before it was quenched by transferring 1 µL of the exchange reaction to 9 µL of ice-cold matrix solution. A 1-2 µL aliquot of the quenched reaction mixture was deposited onto a MALDI sample target and dried with N₂. Ten replicate MALDI mass spectra were collected to determine an average molecular mass of the deuterated CypA for each sample, and this was used to calculate an average ΔMass value at each denaturant concentration for each ligand.

Plots of ΔMass vs. [GdmCl] were constructed for the CypA in the presence and absence of the hit compounds. The data in the plots were fit to the following four-parameter sigmoidal equation using a nonlinear regression in SigmaPlot (Systat Software Inc., San Jose, CA).

\[
\Delta \text{Mass} = \Delta M_0 + \frac{a}{1 + e^{-\left(\frac{[\text{GdmCl}] - C_{[\text{GdmCl}]\text{MAX}}}{h}\right)}}
\]

Equation 2-2
In equation 2-2, $\Delta M_0$ is the pre-transition baseline (i.e., the change in mass of CypA before global exchange), $a$ is the amplitude of the curve in Da, $[\text{GdmCl}]$ is the molar concentration of GdmCl, $C^{1/2}_{\text{SUPREX}}$ is the $[\text{GdmCl}]$ at the transition midpoint of the curve, and $b$ is a parameter describing the steepness of the transition. These parameters were typically allowed to float when fitting the data to equation 2-2.

The $C^{1/2}_{\text{SUPREX}}$ values obtained from equation 2-2 were used in equation 2-3 to determine the magnitude of the increase in thermodynamic stability (i.e., decrease in folding free energy), assuming the $m$-value, which is a measure of a protein’s folding cooperativity, is constant in the presence and absence of ligand.

$$\Delta \Delta G_f = -m \times C^{1/2}_{\text{SUPREX}}$$  \hspace{1cm} \text{Equation 2-3}

In equation 2-3, $\Delta \Delta G_f$ is the change in the protein’s folding free energy (i.e., the protein’s free energy of binding), $m$ is the $m$-value, and $C^{1/2}_{\text{SUPREX}}$ is the change in the $C^{1/2}_{\text{SUPREX}}$ value of the CypA upon binding to a ligand. Equation 2-3 was derived from the general form of the SUPREX equation found in reference [32]. The $m$-value was assigned a value of 3.7 kcal mol$^{-1}$ M$^{-1}$ based on an average of the values reported for CypA in reference [73].

The following equation was used to calculate dissociation constants, $K_d$, for the protein-ligand complexes [63].

$$K_d = \frac{[L]}{\left( e^{-\Delta G_f / nRT} - 1 \right)}$$  \hspace{1cm} \text{Equation 2-4}

In equation 2-4, $[L]$ is the molar concentration of the free compound, $n$ is the number of independent binding sites (equal to 1), $R$ is the universal gas constant, and $T$ is
the temperature in kelvins. Because the ligand concentration was present at 10-fold excess over the CypA, the total ligand concentration was used for the free ligand concentration.

2.2.6 Theoretical SUPREX Curve Construction

The theoretical SUPREX curves generated in this work (Figure 4) were generated exactly as described in reference [80], with the exception that the pre- and post-transition baselines were set to 45 and 65 Da, respectively, in order to more accurately represent the pre- and post-transition baselines observed in the current work.

2.2.7 CypA Inhibition Studies

The ability of the hit compounds to inhibit the peptidyl-prolyl cis-trans isomerase activity of CypA was determined using a chymotrypsin-coupled enzyme assay that has been previously described [89, 90]. In these assays, 875 µL of ice-cold buffer (50 mM HEPES, 100 mM NaCl, 1% DMSO, pH 8.0) containing 5 nM CypA and 10 µM of the hit-compound was combined with 100 µL of an ice-cold chymotrypsin solution (30 mg/mL in 0.001 HCl) in a quartz cuvette. A 25 µL aliquot of a stock solution of the peptide substrate N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide, which was prepared in trifluoroethanol at a concentration of 3 mM, was added to the cuvette. The contents of the cuvette were rapidly mixed and the absorbance at 390 nm was measured over a period of 3 minutes in order to monitor the formation of p-nitroaniline upon chymotrypsin cleavage of the substrate peptide, which only occurs when the proline residue is in the trans-configuration. CypA catalyzes the cis-trans isomerization of proline residues, thus the
catalytic activity of CypA can be indirectly determined from the rate at which chymotrypsin cleaves the peptide substrate.

In the inhibition assay it was possible to calculate the fraction of CypA that was free in solution (i.e., unbound to the test compounds). The calculation involved the use of equations 2-5 and 2-6:

\[
[L] = L_{total} - \frac{P_{total} + L_{total} + K_d - \sqrt{(P_{total} + L_{total} + K_d)^2 - 4P_{total}L_{total}}}{2}
\]  
Equation 2-5

\[
\%Free = 100 \times \left( \frac{L_{total} - [L]}{P_{total}} \right)\times 100
\]  
Equation 2-6

In equations 2-5 and 2-6, \([L]\) is the concentration of free ligand, \(P_{total}\) is the sum of the free and bound protein concentration, \(L_{total}\) is the sum of the free and bound ligand concentration, \(K_d\) is the dissociation constant of the CypA complex with the hit compound, which was determined in the binding affinity measurements (see above), and \(\%Free\) is the fraction of CypA protein that is free in solution (i.e., not bound to ligand). Equation 2-5 was from reference [91].

\[2.3 \text{ Results and Discussion}\]

\[2.3.1 \text{ General Strategy}\]

Single-point SUPREX is an abbreviated version of the SUPREX technique, which exploits the H/D exchange properties of globally protected amide protons in a protein to
measure the increase in a protein’s thermodynamic stability upon ligand binding in solution. In a full SUPREX analysis, the extent to which a protein, or protein complexed with ligand, undergoes amide H/D exchange during a specified amount of time is determined as a function of denaturant concentration (see Figure 4). Ultimately, the denaturant concentration at the transition midpoint is used to ascertain the thermodynamic stability of the protein or protein-ligand complex. Proteins complexed with ligand have a SUPREX transition midpoint that is shifted to a higher denaturant concentration than that of the protein in the absence of ligand, and the magnitude of the shift is related to the binding affinity.

In single-point SUPREX, the extent to which a protein, or a protein in the presence of ligand, undergoes amide H/D exchange during a specified amount of time is determined at a single denaturant concentration. The denaturant concentration is chosen such that there is a maximum difference between the $\Delta$Mass values expected for the protein and for the protein complexed with ligands having the minimum binding affinity desired in the selection. Based on the SUPREX curves expected for CypA in the presence and absence of a hypothetical ligand that binds CypA with a $K_d$ value of 10 µM (see Figure 4), a 1.5 M denaturant concentration was chosen for the single-point SUPREX experiments in this work. This allowed for the selection of library compounds with $K_d$ values $\leq$ 10 µM. The assumption was made that ligands with $K_d$ values up to ~10 µM would be useful lead compounds in the search for CypA-targeted diagnostic and imaging agents.
The 1280 and 9600 compounds in the LOPAC and DIVERSet libraries, respectively, were screened for binding to CypA using the protocol outlined in Figure 5. The protocol involved pooling ten library compounds per well of a microtiter plate prior to initiating the single-point SUPREX experiment. The library pools were prepared in 10 µl of deuterated buffer containing 1.5 M GdmCl and 10% DMSO, and the library compounds in each pool were present at a concentration of ~100 µM per compound. The single-point SUPREX protocol was initiated upon addition of CypA to each library pool. The CypA was allowed to undergo H/D exchange for 35 min in the presence of each library pool. Ultimately, the H/D exchange reactions were quenched upon the addition of a MALDI matrix solution, which also prepared the CypA protein in each sample for a MALDI-TOF analysis. The MALDI-TOF analysis was used to determine the number of amide protons in CypA that were exchanged with solvent deuterons (i.e., the ΔMass value) during the 35 min H/D exchange reaction. The magnitude of the ΔMass value was used as described below to determine whether or not a given pool of library compounds contained binding ligand(s).

2.3.2 Library Screening

The pools of 10 library compounds, along with two sets of control samples, were screened for CypA binding using the single-point SUPREX protocol described above. The distributions of ΔMass values obtained in the analysis of the 1089 pools and 251 control samples are shown in Figure 6. The control samples included 126 positive controls that contained CsA, which is a known tight binding ligand to CypA, and 125 negative controls that did not contain a ligand. The data obtained on these control
Figure 4. A. Theoretical SUPREX curves generated for CypA in the presence and in the absence of a hypothetical CypA binding ligand ($K_d$ value = 10 µM). The arrow indicates the [GdmCl] used in the single point SUPREX experiments in this work. B. Representative MALDI mass spectra of CypA from the screening assay. The top spectrum shows the CypA ion signal generated in the MALDI readout of a selected pool of LOPAC compounds that did not contain a hit, while the bottom spectrum shows the CypA ion signal generated in the MALDI readout of the pool containing Compound 3, which was identified as a hit.
Figure 5. Schematic representation of the screening methodology used in this work. Library compounds were screened as 10-compound pools to improve throughput, and CypA was added and allowed to exchange. The mass increase was then used to determine whether a binding event had occurred. The bar graph plot represents a portion of actual data collected in this work.
samples, which were analyzed with every set of 10 pools, were used to generate the cut-off values for hit selection and to help establish the false positive and false negative rates of the strategy employed in this work. Figure 7 shows the ΔMass values obtained for the positive and negative controls obtained during the analysis of each library.

In the LOPAC library screening, a cut-off value for the selection of hits in the first screening was set at 49.0 Da, which was 2.5 standard deviations below the average ΔMass value obtained for the negative controls. The ΔMass value of one negative control in the LOPAC library (see Figure 7) was determined using the Q-test at the 95% level to be an outlier [92], and it was not used in the cut-off value calculation. The ΔMass value cut-offs used for the selection of hits in the DIVERSet library screening were calculated by subtracting 2.5 standard deviations from a 9-point moving average of the negative control data. Thus, a specific ΔMass value cut-off was determined for each set of ten pools that were analyzed in the DIVERSet library screening. It was previously shown that the use of such a moving average in the calculation of ΔMass cut-off values is useful in single-point SUPREX screening experiments that are performed over the course of multiple days, as was done in the DIVERSet library screening [80].

Ultimately, a list of hit-pools (i.e., those that yielded ΔMass values below the cut-off values) was generated. A total of 8 out of the 131 LOPAC library pools and 13 out of 958 DIVERSet Library pools were identified “hit-pools” in this first tier of screening. The hit-pools from each library were rescreened in a second tier of screening. It has previously been demonstrated that such a two-tier screening strategy significantly reduces the false positive rate using the single-point SUPREX assay [80]. The two tier screening
Figure 6. Distribution of $\Delta$Mass values recorded for the 1089 library pools (grey bars), 126 positive controls (black bars), and 125 negative controls (open bars) analyzed in the 1st tier screening of the two libraries used in this work. The solid lines represent the fit of each distribution to a normal Gaussian curve.
Figure 7. The ΔMass values from the positive (●) and negative (◇) control samples analyzed with the LOPAC library compounds (A) and the DIVERSet library compounds (B). The solid lines represent the average ΔMass values of the control data and the dashed line represents the cut-off values used in this work. The negative control data point indicated with an arrow was determined to be an outlier and was not used to determine the cut-off value. Note that a 9-point moving average was used to determine the cut-off values in B. Shown in C is the distribution of Z'-factors observed in both library screenings.
strategy is useful because the assay is subject to random error, which is reflected in the Gaussian distributions observed for the ΔMass values recorded here (see Figure 6). The main sources of random error include the mass measurement uncertainty and the differential back-exchange (see below). After the second tier of screening, five hit-pools were identified from the LOPAC library, and five hit-pools were identified from the DIVERSet library.

The pooling strategy described here ultimately requires that separate analyses be performed on the individual compounds in the selected pools in order to determine which compound(s) yielded the “hit.”, i.e., were responsible for the low ΔMass value. Thus, the 100 compounds from the ten hit-pools identified from the LOPAC and DIVERSet libraries were individually screened for binding to CypA using the single-point SUPREX protocol. Ultimately, nine compounds were identified as CypA binding ligands in these individual screens (see Table 1 and Table 2). This included one compound from each of the ten hit-pools with the exception of one hit-pool that did not yield an individual hit. The binding affinity of these nine newly identified CypA ligands were then measured using the conventional SUPREX protocol.

Full SUPREX analyses were performed on CypA in the presence of the selected compounds in order to determine their affinity for CypA (Figure 8). The individual data points in a full SUPREX curve are subject to the same ΔMass value uncertainties as those in the single-point SUPREX protocol. However, protein-ligand binding analyses using the full SUPREX protocol are much less sensitive to this error because multiple ΔMass values are used to define a full SUPREX curve (see Figure 8). The full SUPREX analyses
indicated that 8 of the 9 selected compounds yielded measurable binding interactions with CypA, i.e., a shift in the transition midpoint of the SUPREX curves. The SUPREX-derived $K_d$ values of these 8 compounds ranged from 0.200-37 µM, with the majority being in the low micromolar range (Table 1 and Table 2).

Dissociation constants for two hit compounds, 5 and 8 from the DIVERSet library, could not be determined due to difficulties associated with obtaining MALDI ion signals for the CypA in the presence of these compounds at high [GdmCl] concentrations. At these elevated concentrations of denaturant, a MALDI ion signal for CypA was not observed and multiple peaks at $m/z$ 200-1000 Da higher than that expected for CypA were detected. However, the $\Delta$Mass values obtained at the lower denaturant concentrations used in the full SUPREX analyses of CypA in the presence of these individual compounds were consistent with those expected for a pre-transition baseline, suggesting that binding was, in fact, detected, and that the transition midpoint of the curve was at least 1.6 and 2.3 M for 5 and 8, respectively. Binding was not detected in the full SUPREX analyses of CypA in the presence of one of the selected ligands, 6 (i.e., the SUPREX transition midpoints for CypA were similar in the presence and absence of this ligand), which appeared to be a false positive.

The nine newly identified CypA ligands were also tested for their ability to inhibit CypA’s peptidyl-prolyl isomerase activity. Initial velocities of the chymotrypsin-catalyzed hydrolysis of the peptide substrate N-succinyl-Ala-Ala-Pro-Phe-$p$-nitroanilide were determined in the presence and absence of CypA and then in the presence of CypA and each newly identified CypA ligand. The substrate peptide is cleaved by
Table 1. Summary of the four LOPAC ligands selected in this work.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Structure</th>
<th>$K_d$ (µM)$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound 1</td>
<td><img src="compound1.png" alt="Structure" /></td>
<td>4$^b$</td>
</tr>
<tr>
<td>Compound 2</td>
<td><img src="compound2.png" alt="Structure" /></td>
<td>16$^b$</td>
</tr>
<tr>
<td>Compound 3</td>
<td><img src="compound3.png" alt="Structure" /></td>
<td>4$^b$</td>
</tr>
<tr>
<td>Compound 4</td>
<td><img src="compound4.png" alt="Structure" /></td>
<td>37$^b$</td>
</tr>
</tbody>
</table>

$^a$ Determined from data collected in the full SUPREX analyses performed in this work.

$^b$ Inhibition of CypA’s isomerase activity was detected (see Figure 9).
Table 2. Summary of the five DIVERSet ligands selected in this work.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Structure</th>
<th>$K_d$ (µM)$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compound 5</td>
<td><img src="image1" alt="Structure" /></td>
<td>≤10$^b$</td>
</tr>
<tr>
<td>Compound 6</td>
<td><img src="image2" alt="Structure" /></td>
<td>No binding$^c$</td>
</tr>
<tr>
<td>Compound 7</td>
<td><img src="image3" alt="Structure" /></td>
<td>2$^b$</td>
</tr>
<tr>
<td>Compound 8</td>
<td><img src="image4" alt="Structure" /></td>
<td>≤0.2$^c$</td>
</tr>
<tr>
<td>Compound 9</td>
<td><img src="image5" alt="Structure" /></td>
<td>4$^b$</td>
</tr>
</tbody>
</table>

$^a$Determined from data collected in the full SUPREX analyses performed in this work.

$^b$Inhibition of CypA’s isomerase activity was detected (see Figure 9).

$^c$No inhibition of CypA’s isomerase activity was detected (see Figure 9).
Figure 8. Full SUPREX curves obtained for CypA in the absence (●) and presence (○) of the nine selected hit-compounds. The solid and dashed arrows indicate the transition midpoints obtained in the absence and presence of ligand, respectively. Eight of the nine had measurable binding due to a shift in the SUPREX curve. Compounds 5 and 8 did not produce a full SUPREX curve (see text), but the data points collected suggested a transition midpoint of at least 1.6 and 2.3 M, respectively. A measurable transition midpoint shift was not detected with Compound 6.
chymotrypsin when the Ala-Pro peptide bond is in the trans configuration. CypA increases the apparent rate that chymotrypsin hydrolyzes the scissile bond in this substrate by catalyzing the conversion of the cis isomer to the trans isomer. The degree to which CsA and the nine hit compounds identified in this study impact the rate of this hydrolysis reaction in the presence of CypA was determined from initial rate measurements (see Figure 9).

The chymotrypsin-coupled enzyme assays revealed that seven of the nine compounds (including Compounds 1-5, 7, and 9) inhibited CypA’s isomerase activity, and the degree to which each of these ligands inhibited CypA was generally consistent with the amount of free protein calculated based on each ligand’s measured $K_d$ value. An exception was Compound 8 that appears to bind CypA with relatively high affinity but have minimal effect on CypA’s isomerase activity. It is not surprising that such a compound was identified in this work, as the single-point SUPREX assay does not select for compounds with specific activities, rather it only selects for binding interactions. The variety of compounds that can be selected in the single-point SUPREX assay (e.g., both those that alter a specific protein function and those that do not) can be an advantage of the methodology over other HTS approaches for certain applications, such as the one described here to identify molecular imaging agents. Interestingly, Compound 6, which was not found to have a measurable CypA binding interaction in the full SUPREX analysis ($K_d$ value > 40 µM), did not significantly inhibit the isomerase activity of CypA isomerase. Compound 6 is most likely a false positive.
Figure 9. Results obtained in CypA inhibition studies. The closed bars and left y-axis represent the initial velocities determined for the chymotrypsin-catalyzed hydrolysis of the peptide substrate N-succinyl-Ala-Ala-Pro-Phe-\(p\)-nitroanilide in the presence and absence of CypA and in the presence of CypA and different ligands, including CsA and the nine hit compounds identified in this study. The initial velocity values are the average of four to six determinations and the error bars represent 1 standard deviation. The open bars and the right y-axis indicate the fraction of CypA protein expected to be free in solution in the assay (i.e., unbound to the target ligands) based on the CypA protein concentration in the assay and the measured \(K_d\) value (see Table 1 and Table 2).
2.3.3 Throughput and Efficiency

The first tier of the 1280-compound LOPAC library screening was accomplished by researchers working over the course of one day for 3.5 h, which was the time it took to perform the H/D exchange reactions and collect the mass spectral data. This equates to approximately 1.5 minutes per pool or less than 10 seconds/compound. The first tier of the 9600-compound Chembridge DIVERSet library screening was accomplished by two researchers working over the course of four days for a total of 22 h, which again was the time it took to perform the H/D exchange reactions and collect the mass spectral data. This equates to a screening rate of approximately 1 min/pool or about 6 seconds/compound.

The overall throughput achieved in this work, 6 seconds/compound, is about 30 times faster than that previously reported for the single-point SUPREX protocol [80]. In theory, the pooling strategy used here should increase the throughput of the original single-point SUPREX protocol by 10-fold. The additional 3-fold increase in throughput realized in this work is likely due to the use of a high-throughput MALDI-TOF instrument, which was equipped with a high repetition rate laser.

The false positive and false negative rates of a high-throughput screening assay can provide a measure of its efficiency. A total of 126 positive controls and 125 negative controls were analyzed during the single-point SUPREX screening experiments described here. Using the ΔMass cut-off values described above, false positive and false negative rates of 0% and 9% were determined from the control data, which are both identical to
the rates previously observed for the control data in the Prestwick Chemical Library screening, which were 0 and 9%, respectively [80].

During the screening of the pooled library samples, false negatives go undetected, and false positives are only detected after subsequent screens and/or analyses. A total of 10 hit-pools were initially identified after analysis of the 1,089 10-compound pools. Subsequent analyses of the individual ligands from these 10 hit-pools identified hit-compounds from all but two pools, suggesting a false positive rate of 20%. Interestingly, one of the two false positive hit pools consistently produced low ΔMass values upon re-analysis, even though it failed to produce a hit-compound when the ligands from the pool were individually analyzed for CypA binding. This suggests that the higher occurrence of false positives in the pooled samples (i.e., ~20%) compared to the false positives in the controls (i.e., 0%), which were not pooled, may be a result of the pooling strategy. This is also supported by the observation that the 20% false positive rate determined from the pooled library compounds in this work was larger than the 0% false positive rate determined from the library compounds in the earlier screen of the Prestwick Chemical Library, which did not employ a pooling strategy [80]. The 20% false positive rate in this work is in line with typical pharmaceutical screens, which have been estimated to have false positive rates of about 40% [93].

A complicating issue in the pooling strategy is the potential for higher false positive rates due to promiscuous aggregation and non-specific binding of library compounds [94, 95]. There have been reports of small molecules aggregating into particles with diameters ranging from 95 to 400 nm [94]. Separate studies have shown
that enzymes can be inhibited through reversible adsorption of enzyme onto the surface of such molecular aggregates [95]. This phenomenon is worse at higher concentrations of small molecules relative to protein concentration. Sequestration of the protein of interest in a screening assay by aggregate particles would thus increase the protein’s stability due to non-specific binding. This could explain why some pools (with total small molecule concentrations of 1 mM in the analysis buffers) were selected as positive hits, but then none of the individual compounds from the pools would be selected as binding ligands when they were analyzed individually (with the total small molecule concentrations on the order of 100 µM).

A measure of assay robustness is the $Z'$-factor. The $Z'$-factor is maximized by increasing the difference between the $\Delta$Mass values of hits and non-hits and by minimizing the standard deviations of the $\Delta$Mass values. The difference between hit and non-hit $\Delta$Mass values is largely a property of the specific protein under study, and it is directly related to the number of amide protons that are globally protected in the protein’s three-dimensional structure. The standard deviations of the $\Delta$Mass values are limited by the mass spectrometer. The standard deviations obtained using the MALDI TOF mass spectrometer in this work averaged 3.7 Da during the screening of the LOPAC library and 3.4 Da in the DIVERSet. Generally, HTS assays with $Z'$-factors $\geq 0$ are useful for screening large combinatorial libraries [88]. The distribution of $Z'$-factors determined from the positive and negative controls in this work is shown in Figure 7C.

The $Z'$-factors varied in this work mainly because the amplitude of the assay signal (i.e., $\Delta$Mass value difference between the positive and negative controls) varied
over the course of the screening (see Figure 7B). This was a result of differential back-exchange (a random effect of not being able to determine the ΔMass values in the MALDI-TOF experiment at exactly the same time after they are quenched) and the fact that the standard deviations for the mass measurements of some controls were large and clearly outliers. While the moving average procedure accounted for this differential back-exchange to set the cut-off values in the screen, it is clear that the back-exchange reaction negatively impacts the Z’-factor. While there was a large range of Z’-factors (i.e., -1.3-0.7), the large majority of Z’-factors were positive with most clustered between 0.1 and 0.5 (see Figure 7C) using the protocol described here. The distribution of Z’-factors observed here using the pooling strategy is very similar to that observed in the previous application of the single-point SUPREX assay that did not utilize a pooling strategy [80]. In theory, the Z’-factors in the single-point assay should be normally distributed. The higher than expected number of low Z’-factors here (Figure 7C), and in previous work [80], is likely due to the back-exchange reaction, which compromise Z’-factors by reducing the separation between hits and non-hits. These problems associated with the back-exchange reaction can be minimized by recording ΔMass values of the H/D exchanged protein samples in the MALDI-TOF experiment as soon after they are quenched as is possible.

2.3.4 Identification of Novel CypA Ligands

Full SUPREX analyses of the nine hit-compounds selected in the screen (Figure 8) revealed that eight of the nine selected compounds bound CypA with $K_d$ values in the low micromolar to high nanomolar range (Table 1 and Table 2). These eight hit-
compounds are all novel CypA ligands. It is interesting to find that CypA binds to (cisplatin) since this drug is a commonly used anti-tumor agent. CypA has been found to be overexpressed in many cancer types including non-small cell lung cancer [76, 77], pancreatic cancer [96, 97], endometrial carcinoma [98], and oral cancer [99]. In addition, CypA has been shown to be associated with the progression of breast cancer [100]. The mode of action for cisplatin-induced apoptosis is not completely understood but is at least partly through the generation of reactive oxygen species [101]. It has previously been observed that cells overexpressing CypA have been more resistant to cisplatin-induced death by minimizing stress-induced apoptosis, while offering no additional resistance to other anti-cancer drugs [102]. From the HTS results, one hypothesis is that this may be the result of CypA binding to cisplatin, thus decreasing the amount of cisplatin available to the cell to generate reactive oxygen species.

It has also been shown that Compound 2 (Bay 11-7085) inhibits the tumor necrosis factor-α (TNFα)-induced phosphorylation of IκBα, the inhibitor form of nuclear factor kappa B (NFκB), which serves to decrease expression of the inflammatory molecules ICAM-1, VCAM-1, and E-selectin [103]. The specific protein target of Bay 11-7085 was not determined in the study. However, it was speculated that the phosphorylation inhibition might be due to inhibition of a signaling protein upstream of the IκBα. It has been demonstrated that CypA has the ability to induce phosphorylation of IκBα in inflammation sites caused by rheumatoid arthritis [104], and thus it is feasible to hypothesize that the mode of action of Bay 11-7085 involves CypA being inhibited by Bay 11-7085, which in turn decreases the phosphorylation of IκBα and in turn the
activation of NFκB-activated expression of these inflammatory molecules. The other selected compounds from the LOPAC library have known roles. Compound 3 (ZM 39923 HCl) is a selective Janus Kinase 3 (JAK3) inhibitor, binding to JAK3’s ATP binding site [105]. It has also been found to inhibit transglutamase during a HTS assay [106]. Compound 1 (α-NETA) is a selective, fluorescent inhibitor of choline acetyltransferase and is used for the investigation of acetylcholine synthesis [107]. To date, the hits from the Chembridge DIVERSet library have no previously known function.

This study demonstrated the use of single-point SUPREX for small (<10,000 compounds) libraries. Using a pooling strategy, the throughput was increased more than ten-fold compared to what has been previously reported. However, one drawback of the pooling strategy was a higher false positive rate of 20%. While the pooling strategy did not specifically impact Z’-factors of the single-point SUPREX assay, it is important to recognize that the Z’-factors determined in this work were generally between 0 and 0.5, with approximately one-third of the values being less than 0. As a result of these relatively low Z’-factors, the assay requires at least a two-tier screening strategy. This was feasible for the relatively small (1000-10,000) compound libraries studied here. However, the relatively low Z’-factors of the assay may limit applications to larger compound libraries. The results of this study confirmed the assay’s capacity to identify novel ligands to a target protein. In this case, eight novel CypA ligands were discovered, with one compound estimated to have a $K_d$ of less than 200 nM. Additionally, the use of a traditional biochemical assay showed that seven of the eight novel ligands also
inhibited the isomerase activity of CypA. The assay is unique among high-throughput screening assays in its generality, i.e., the ligand selection is based solely on the increase of structural stability and not necessarily on the modulation of a protein’s specific biological activity.
3. Thermodynamic Analysis of Selective Proteins in Complex Mixtures

3.1 Introduction

Mass spectrometry (MS) is commonly used as an approach to study protein-ligand interactions because of its speed of measurements and its high sensitivity, which is a tremendous advantage over more traditional spectroscopic and calorimetric techniques. Different approaches have been taken, including native ESI [108-111], H/D exchange [33, 112, 113], and covalent modification strategies [114-116]. However, these techniques are still generally limited to studying protein interactions using purified proteins or relatively simple mixtures. The ability to make measurements of protein interactions in the context of complex protein mixtures was the motivation behind the development of the SUPREX and SPROX methodologies [46, 47].

However, occasionally problems arise during the thermodynamic analysis of proteins of interest in complex mixtures due to signal suppression. Ion suppression can occur due to the differing acidity/basicity of other analytes (i.e., charge competition), hydrophobicity of the analyte and its ability to crystallize with matrix (in the context of MALDI-MS), gas-phase proton affinities of differing analytes, and nonvolatile salts or analytes (in the context of ESI-MS). As a result, the ion signal for the protein of interest may be suppressed by other analytes [117-120]. To be able to analyze a certain protein, it may be advantageous to perform a ligand binding study within the context of the complex mixture and then purify the protein of interest from the mixture. To be generally
applicable, this would most likely be relevant using covalent modification strategies that allow the permanent modification of proteins.

Described in this chapter is the application of a covalent modification- and MS-based technique coupled with standard affinity purification strategies for the quantitation of protein-ligand binding within complex mixtures. Using SPROX, which utilizes the denaturant-dependent H$_2$O$_2$-mediated oxidation of methionine residues to study protein thermodynamics, an antibody-based purification and recombinant affinity tag-based purification was used to quantify the interactions in two model systems. The interaction of bovine carbonic anhydrase II (BCA) and 4-carboxybenzenesulfonamide was studied after spiking BCA into a cell lysate. In addition, cyclophilin A (Cpr1) was expressed in yeast, and its interaction with cyclosporin A (CsA) was investigated within the context of the cell lysate.

3.2 Experimental

3.2.1 SPROX

Lyophilized carbonic anhydrase II from bovine erythrocytes (BCA), catalase from bovine liver, L-methionine, CBS, ZnCl$_2$, and 30 wt.% hydrogen peroxide (H$_2$O$_2$) were from Sigma-Aldrich. IgG Sepharose™ 6 FastFlow was from GE Healthcare Life Sciences (Piscataway, NJ). The NR6M cell lysate was a gift from Dr. Michael J. Campa in the Patz laboratory at Duke University Medical Center.

The denaturant-containing SPROX buffers were comprised of 20 mM phosphate buffer (pH 7.4) and GdmCl (EMD Chemicals, Inc) concentrations that ranged from 0 to 8
M. The specific [GdmCl] in each buffer was determined by measuring the refractive index of the buffer using the method described by Nozaki [83].

3.2.2 Immunoprecipitation Pull-Down

Either 25 µL of solution containing 150 µM BCA and 225 µM ZnCl₂ or containing 150 µM BCA, 225 µM ZnCl₂ and 450 µM CBS was added to 100 µL 0.8 mg/mL NR6M (3T3 fibroblast cell line overexpressing mutant EGFRvIII) cell lysate. SPROX was performed by combining 36.7 µL SPROX buffer, 5 µL lysate sample, and 8.3 µL 9.8 M H₂O₂ together and allowing the oxidation reaction to occur for 1 min. The reaction was quenched by adding 50 µL 375 mM L-methionine containing 25 units/µL catalase, where 1 unit is defined as the amount of catalase that decomposes 1 µmol H₂O₂ per min at a pH of 7.0 at 25°C as the H₂O₂ concentration falls from 10.3 mM to 9.2 mM. After quenching, 100 µL phosphate buffered saline (PBS) was added to the samples, followed by 10 µL 10 mg/mL rabbit polyclonal anti-BCA antibody (Millipore). The immune complexes were allowed to equilibrate for 2 h at room temperature. The samples were added to immobilized protein G resin and allowed to equilibrate for 10 min. The resin was washed 3 times with 400 µL PBS, followed by elution of the BCA with 200 µL 0.1 M glycine, pH 2. Eluates were desalted using C4 ZipTips (Millipore), and eluted off the tip onto a stainless steel MALDI target using 2 µL saturated sinapic acid in 75:25:0.1 ACN/H₂O/TFA.

3.2.3 Affinity Tag Pull-Down with Whole Protein Readout

A Saccharomyces cerevisiae Y258 strain (MATα pep4-3, his4-580, ura 3-53, leu2-3,112) that overexpressed yeast cyclophilin A (Cpr1, YDR155C) under control of
the GAL1 promoter in the form of a C-terminal fusion protein was used for these experiments (Open Biosystems, Huntsville, AL) [121]. The overexpressed yeast fusion protein contained a 19 kDa purification tag containing an HA epitope, a 3C protease cleavage site, a ZZ domain from Protein A, and a 6xHis-tag.

Yeast were grown overnight at 30°C in 2% dextrose SC-Ura, and 2 mL of this solution was used to inoculate 50 ml of 2% raffinose SC-Ura solution, which was then incubated overnight. The solution was diluted into 500 ml of 2% raffinose SC-Ura to give an OD$_{600}$ of 0.3 and was incubated at 30°C until the OD$_{600}$ of the solution was 0.8-1.2. A 250 mL volume of 3x YP-Gal (3% yeast extract, 6% peptone, 6% galactose) was added to induce overexpression. After 6 h of incubation, yeast cell pellets were obtained by centrifuging 150 mL culture at 1,000 x g for 10 min, and pellets were stored at -20°C until further use.

Three yeast pellets overexpressing Cpr1 were lysed with acid-washed 425-600 μm glass beads (Sigma-Aldrich) at 4°C in 0.5 mL (per pellet) of CE extraction buffer (50 mM Tris, 1 mM EDTA, 4 mM MgCl$_2$, 5 mM DTT, 10% glycerol, 0.75 M NaCl, pH 7.5 and HALT protease inhibitor cocktail (Pierce)). The pellets were vortexed 10 times at 20 s each, with 1 min on ice between vortexing. The insoluble fractions were pelleted by centrifuging at 14,000 x g for 5 min, the lysates from all 3 pellets were combined, and 150 μL lysate was aliquoted into 12 different microcentrifuge tubes. The samples were reduced to dryness in a SpeedVac concentrator and subsequently reconstituted in 30 μL water. Following dissolution, 3 μL DMSO was added to the samples. After 5 min, 60 μL SPROX buffers were added. The SPROX buffers that were used among the 12 samples
ranged from 0.25-6.0 M GdmCl prior to sample dilution. The SPROX reaction was initiated by adding 18.5 µL 9.8 M H₂O₂ and oxidizing for 1 min. The reaction was quenched by adding 200 µL IPP0 buffer (10 mM Tris, 0.1% Tergitol type NP-40, pH 8.0) containing 10 units/µL catalase. Approximately 200 µL IgG Sepharose (50% slurry) was added to each quenched SPROX solution, and 550 µL IPP0 buffer was added to the samples. The samples were nutated at 4°C for 2 h.

The resin in each sample was centrifuged, the supernatant was discarded, and the resin was washed with 0.5 mL IPP-150 (10 mM Tris, 0.1% Tergitol type NP-40, 150 mM NaCl, pH 8.0) twice. The resin was then washed twice with 400 µL 3C cleavage buffer (10 mM Tris, 0.1% Tergitol type NP-40, 150 mM NaCl, 0.5 mM EDTA, 1 mM DTT, pH 8.0). After discarding the wash solutions, 250 µL 3C cleavage buffer was added to the resin, and 1 µL 2 mg/mL HRV 3C protease was added to each sample. The samples were nutated overnight at 4°C. After digestion, the supernatant was saved for analysis. The resin was washed twice more with 100 µL 3C cleavage buffer without the NP-40 detergent, and these washes were added to the supernatant. The volume of the sample was reduced in the SpeedVac for 1 h. C18 ZipTips were used to clean and desalt the samples, and samples were eluted onto a stainless steel MALDI target with 2 µL saturated sinapic acid in 75:25:0.1 ACN/H₂O/TFA. The samples were deposited onto preexisting dried sinapic acid spots.

3.2.4 Affinity Tag Pull-Down with Peptide Readout

A yeast pellet from 150 mL culture was lysed with glass beads in 0.4 mL of CE extraction buffer using the same lysing procedure as Section 3.2.3. The lysate was split
into two 225 µL aliquots. A total of 25 µL 10 mM CsA (LKT Laboratories, St. Paul, MN) in DMSO was added to one aliquot, and 25 µL DMSO was added to the other. These 2 samples, with and without CsA, were incubated for 15 min at room temperature, and the lysate samples were then allowed to equilibrate in the denaturant-containing SPROX buffers by combining 20 µL lysate sample to 25.8 µL SPROX buffer. Twelve different denaturant concentrations were used for each sample, i.e., with and without CsA. SPROX was performed by adding 4.2 µL 9.8 M H$_2$O$_2$, oxidizing for 3 min, and then quenching the reaction with the addition of 1 mL 300 mM methionine.

Each SPROX sample (24 total) was added to 20 µL IgG Sepharose (50% slurry), followed by the addition of 4 mL 10 mM Tris, pH 8.0. The pull-down reactions were nutated for 2 h at 4°C, and the resin was subsequently spun down at approximately 1,250 x g for 3 min. The supernatant was poured off, and the resin was transferred with buffer consisting of 10 mM Tris, 150 mM NaCl, pH 8.0 (wash buffer) to spin cups containing a cellulose acetate filter. The spin cup was placed inside a microcentrifuge tube, and the resin was washed by adding the wash buffer and centrifuging to remove the wash buffer. The affinity-tagged Cpr1 was eluted off the resin with the addition of 0.1 M glycine, pH 2 (total volume of eluate was 85 µL). Buffer exchange was performed using Zeba Desalting Columns (Thermo Scientific) into 50 mM ammonium bicarbonate, pH 8.5. This volume was reduced to approximately 50 µL using a Thermo Savant SpeedVac concentrator.

RapiGest SF Surfactant (Waters, Milford, MA) was added to a final concentration of 0.1%, and the samples were heated at 40°C for 10 min while shaking at 500 rpm in a
Thermoshaker (Eppendorf). The samples were then reduced by adding DTT to a final concentration of 10 mM, and heating at 80°C for 15 min while shaking. Iodoacetamide was added to a concentration of 20 mM, and samples incubated in the dark for 30 min. Approximately 50 ng trypsin (Sigma) was added to each sample, and samples were allowed to digest for 16 h at 37°C. The digestion was quenched by adding 0.7 µL 100% formic acid (FA) and 1.4 µL ACN. The peptide samples were heated at 60°C for 2 h, and then centrifuged at 15,000 x g for 5 min. The samples were pipetted into autosampler vials and stored at 4°C until analysis.

3.2.5 Mass Spectrometry Analyses

BCA and affinity tag pull-down protein-level Cpr1 samples were analyzed using MALDI mass spectrometry. Mass spectra were collected on an Ultraflex II TOF/TOF from Bruker Daltonics, and the method parameters were the same as those described in Section 2.2.3. A total of 10 spectra (100 laser shots/spectrum) were collected for each denaturant concentration.

The trypsinized Cpr1 samples were analyzed using nanoLC-MS involving an Agilent 6520 Q-TOF mass spectrometer system equipped with a Chip Cube Interface. The HPLC Chip used in this experiment contained a 43 mm x 75 µm column with Zorbax 300SB-C18 5 µm packing. The gradient progressed from 3-5% Buffer B in 0-0.1 min, and then 5-40% Buffer B from 0.1-40 min. Buffer A was 0.1% FA in water, and Buffer B was 0.1% FA in ACN. The flow rate was 0.4 µL/min. The capillary voltage was 1875 V. The drying gas (N₂) was 350°C at a flow rate of 5 L/min. The skimmer and fragmentor were set to 65 V and 175 V, respectively. Six runs containing data-dependent
product ion spectra were collected to identify sequences of Cpr1 peptides. For MS/MS spectra, the collision energy was 3.5 V/100 Da with a -4.8 V offset, and the inclusion window width for precursor ions was 4 m/z units. The scan rate was 4 scans per second in the mass spectra and 3 scans per second in the product ion mass spectra, and 3 precursors were selected for fragmentation per cycle. For SPROX runs, no tandem mass spectrometry was performed.

### 3.2.6 Data Analysis

The average molecular weight of BCA or protein-level Cpr1 was determined from the MALDI-TOF mass spectra using an Excel macro as described in Section 2.2.4.

The Cpr1 peptides analyzed in this work were identified from the product ion mass spectra using Agilent’s Spectrum Mill MS Proteomics Workbench software, Rev A03.03.084 SR4. In the Spectrum Mill searches, the precursor and product ion mass tolerances were set to 20 ppm, the protein cleavage chemistry was set for trypsin with 3 maximum missed cleavages, and the maximum ambiguous precursor charge was set to 7. Carbamidomethylation of cysteines was set as a fixed modification, and oxidation of methionines as well as the deamidation of asparagines and glutamines were set as variable modifications. The peptide fragment products and peptide precursor masses were searched against the NCBInr database for *S. cerevisiae*.

The weighted average ΔMass (ΔMassWtAv) for the peptide-level Cpr1 was determined by first extracting chromatograms corresponding to the m/z of both the unoxidized and oxidized species of methionine-containing peptides from Cpr1 using Agilent’s Qualitative Analysis software. The areas under the chromatographic peaks for
each species at each denaturant were determined, and the $\Delta$Mass<sub>WtAV</sub> was calculated using equation 3-1:

$$\Delta\text{Mass}_{WtAV} = \left( \frac{m_{ox} A_{ox} + m_{unox} A_{unox}}{A_{ox} + A_{unox}} \right) \times z - z m_H$$  \hspace{1cm} \text{Equation 3-1}

In equation 3-1, $m_{ox}$ and $m_{unox}$ are the $m/z$ of the oxidized and unoxidized species of the peptide, respectively, and $A_{ox}$ and $A_{unox}$ are the chromatographic peak areas for the oxidized and unoxidized species of the peptide, respectively. The symbol $z$ is the charge of the peptide and $m_H$ is the mass of a proton. The $\Delta$Mass<sub>WtAV</sub> was plotted as a function of [GdmCl].

The $\Delta$Mass<sub>WtAV</sub> vs. [GdmCl] plots were fit to the following sigmoidal equation to yield $\Delta G_f$ and $m$-values [47]:

$$\Delta\text{Mass}_{WtAV} = \Delta M_\infty + (\Delta M_0 - \Delta M_\infty)e^{-\frac{t}{k_{ox}(1+K_{fold})}}$$  \hspace{1cm} \text{Equation 3-2}

In equation 3-2, $\Delta$Mass<sub>0</sub> is the $\Delta$Mass<sub>WtAV</sub> before global oxidation, $\Delta$Mass<sub>\infty</sub> is the $\Delta$Mass<sub>WtAV</sub> after global oxidation, $t$ is the time of oxidation in seconds, $k_{ox}$ is the average pseudo-first order rate constant for the oxidation of an unprotected methionine side chain (0.014 s<sup>-1</sup> x [H<sub>2</sub>O<sub>2</sub>] [47]), $K_{fold} = e^{-\frac{\Delta G_f + m[Den]}{RT}}$, $\Delta G_f$ is the protein’s folding free energy in the absence of denaturant, [Den] is the denaturant concentration, $m$ is $\delta\Delta G/\delta[Den]$, $R$ is the universal gas constant, and $T$ is the temperature in kelvins.

The binding free energy ($\Delta\Delta G_f$) was calculated using equation 2-3. $K_d$ values for Cpr1 and CsA were determined using equation 2-4. Because CBS was not at least 10 times greater in concentration than BCA, $K_d$ values for BCA and CBS were calculated using the following equation, which has been derived previously [65]:
\[ K_d = \frac{4L_{\text{total}}e^{-\Delta G_f/nRT} - 4P_{\text{total}} \left( e^{-\Delta G_f/nRT} - 1 \right)} {\left( 2e^{-\Delta G_f/nRT} - 1 \right)^2 - 1} \]  

Equation 3-3

In equation 3-3, \( L_{\text{total}} \) and \( P_{\text{total}} \) are the concentration of ligand and protein, respectively, and \( n \) is the number of independent equivalent binding sites.

### 3.3 Results and Discussion

#### 3.3.1 BCA Immunoprecipitation

Initially, BCA was studied as a model protein for immunoprecipitation in conjunction with SPROX. Its SUPREX and SPROX behavior has previously been studied [47, 65] in addition to its interaction with CBS [47, 65, 122]. Rabbit polyclonal anti-BCA was used to capture BCA from a mixture of proteins, which was then incubated with immobilized protein G and subsequently eluted off as a pure protein. Shown in Figure 10 are MALDI mass spectra following a BCA immunoprecipitation.

Figure 10A shows the spectrum of the initial flowthrough after incubation of the protein mixture containing BCA with anti-BCA and immobilized protein G. BCA, which has a mass of approximately 29,000 Da, was not present, nor was it present in Figure 10B, which shows a spectrum of a wash of the protein G resin to elute off nonspecific interacting proteins. However, a peak for BCA appears in Figure 10C, which is a mass spectrum of an elution fraction of the protein. This demonstrated that the antibody clearly had affinity for BCA and specifically targeted it out of this mixture.

Next, the immunoprecipitation was performed in conjunction with SPROX. Shown in Figure 11 are SPROX curves of BCA and BCA complexed with CBS.
Figure 10. MALDI mass spectra from BCA immunoprecipitation. The peak at approximately 17,000 m/z is myoglobin, the internal standard used in this experiment. A. Initial flowthrough following capture of BCA/Anti-BCA on Protein G. B. Spectrum from a wash of the Protein G resin with PBS. C. Spectrum from an elution fraction showing BCA was present.
Figure 11A is a SPROX curve of pure BCA in the absence and presence of CBS without the immunoprecipitation step. The $K_d$ for the BCA-CBS interaction has previously been measured in the range of 0.76 to 3.2 µM [47, 65, 122], and so the shift in the $C_{SPROX}^{1/2}$ value in the presence of CBS would not be expected to be very large due to the relatively weak interaction. Indeed, the measured shift is only 0.2 M (see Table 3), which corresponds to a calculated $K_d$ of approximately 2 µM. Figure 11B shows a SPROX curve for BCA in the presence and absence of CBS after being spiked into a lysate. Following the immunoprecipitation, the SPROX curves were essentially identical, with the same magnitude of $C_{SPROX}^{1/2}$ value shift. The measured $K_d$ values agreed well with literature values for this interaction [47, 65, 122].

### 3.3.2 Affinity Purified Cpr1

The second strategy for protein purification following SPROX involved the use of an affinity tag expressed in yeast. The protein Cpr1 was overexpressed in yeast as a fusion protein containing a 19 kDa purification tag, which contained an HA epitope, a 3C protease cleavage site, a ZZ domain from Protein A, and a 6xHis-tag. No methionines are present in the tag’s primary sequence. According to its expression level [121], Cpr1 was approximately 10 µM in yeast lysate. However, Cpr1 was not visible in a mass spectrum of the lysate (see Figure 12A) because lower mass biomolecules dominated the spectrum. Two approaches were taken for the thermodynamic analysis of Cpr1. Both involved SPROX followed by protein purification and mass spectrometry analysis, but one utilized a whole protein readout while the other utilized a peptide readout. The first approach involved performing SPROX in the context of the lysate, capturing Cpr1 via the
Figure 11. A. SPROX curves for pure BCA in the absence (●) and presence (○) of CBS
B. SPROX curves for BCA spiked into a lysate and subsequently immunoprecipitated with Anti-BCA in the absence (●) and presence (○) of CBS. Error bars represent the standard deviation from 10 spectra.
Table 3. SPROX-derived thermodynamic parameters for BCA and Cpr1 in the presence of known ligands

<table>
<thead>
<tr>
<th>Protein</th>
<th>$C^{1/2}_{SPROX}$ in the Absence of Ligand (M)</th>
<th>$C^{1/2}_{SPROX}$ in the Presence of Ligand (M)</th>
<th>$K_d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCA (+/- CBS)</td>
<td>1.5 ± 0.1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.7 ± 0.1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.5 µM&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>BCA (+/- CBS) (immunoprecipitation)</td>
<td>1.6 ± 0.1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.8 ± 0.1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.3 µM&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Cpr1 (whole protein readout)</td>
<td>1.2 ± 0.1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cpr1 (+/- CsA) (peptide readout)</td>
<td>1.0 ± 0.1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2.1 ± 0.1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>180 nM&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> Reported uncertainty is standard deviation of replicate measurements.

<sup>b</sup> Used an $m$-value of 9.0 ± 2.7 kcal mol$^{-1}$ M$^{-1}$, which was the average of the values from the curves in Figure 11. Uncertainty is standard curve fitting error.

<sup>c</sup> Reported uncertainty is standard curve fitting error.

<sup>d</sup> Used an $m$-value of 4.1 ± 0.9 kcal mol$^{-1}$ M$^{-1}$, which was the average of the values from the curves in Figure 13B. Uncertainty is standard curve fitting error.
protein A tag using immobilized IgG, and reading out the mass of Cpr1 using MALDI-MS following elution off of the resin. Shown in Figure 12B is a mass spectrum of the purified Cpr1 following elution from the SPROX experiment. Shown in Figure 13A is a SPROX curve for Cpr1 purified from the lysate at the protein level. The $c^{1/2}_{SPROX}$ was 1.2 M, which agrees with the previously studied SPROX behavior of human cyclophilin (CypA) [47], even though human CypA contains 4 methionines while Cpr1 contains 2 methionines. However, Cpr1 and CypA do share 66% sequence identity, and an additional 11% of the sequence contains positive substitutions (e.g., an aspartate substituted for a glutamate) as apparent from UniProt’s sequence alignment tool (www.uniprot.org), so it is not surprising that they might share similar biophysical properties. Additionally, making the assumption that CypA and Cpr1 have approximately the same thermodynamic stability, the presence of the affinity tag does not appear to affect the stability of Cpr1.

A ligand binding analysis between Cpr1 and CsA was next performed using a peptide readout. This involved performing SPROX on the lysate in the absence and presence of CsA, purifying the protein A-tagged Cpr1 with immobilized IgG, and digesting Cpr1 with trypsin prior to LC-MS analysis. Tandem mass spectrometry was first performed on the peptides to identify peptide sequences. The sequence coverage of Cpr1 was 82% (see Table 4). Two methionine-containing peptides were identified from the product ion spectra, which are the number of methionines found in Cpr1 if the N-terminal starter methionine is not counted. One methionine-containing peptide, Cpr1(189-123) HHDRPGLLSMANAGPNTNGSQFFITTVPCPWLDGK, was a large
Figure 12. A. MALDI mass spectrum of yeast lysate overexpressing Cpr1. The arrow denotes where the Cpr1 was expected to be based on the mass of Cpr1 with the addition of the affinity tag. B. MALDI mass spectrum of affinity purified Cpr1 from the SPROX experiment. Peaks at approximately 10,000 and 20,000 m/z belong to the [M+H]$^+$ and [M+2H]$^{2+}$ peaks of soybean trypsin inhibitor, which was used as an internal standard in the SPROX experiment.
peptide with a mass of 3834.8 and wasn’t highly abundant in the chromatogram (see Table 4). As a result of its low abundance, it couldn’t be used for SPROX analysis. The other methionine-containing peptide, Cpr1(54-74) VIPDFMLQGGDFTAGNGTGK, was very abundant in the chromatogram and was used for SPROX analysis. SPROX curves were constructed using the fraction oxidized of Cpr1(54-74) based on the area of the extracted ion chromatogram of both oxidized and unoxidized species, which were then converted to $\Delta$Mass$_{\text{WtAv}}$ values (equation 3-1). Shown in Figure 13B are SPROX curves for Cpr1 in the absence and presence of CsA in the context of the cell lysate, using the methionine-containing peptide Cpr1(54-74). The $C_{SPROX}^{1/2}$ for Cpr1 in the absence of CsA using the peptide readout was the same as that for the whole protein readout (see Figure 13A), which indicates that the peptide Cpr1(54-74) comes from a domain that exhibits approximately the same stability as the whole protein assuming similar $m$-values, or that Cpr1 is well modeled as a single-domain protein. Because the structure of Cpr1 is predominantly a globular $\beta$ barrel-type structure (PDB 1IST), it is probable that Cpr1 naturally unfolds as a single domain. However, the $m$-value of the whole protein SPROX curve was slightly less than what was found from the peptide readout (2.8 vs. 4.1 kcal mol$^{-1}$ M$^{-1}$, respectively), indicating a lower degree of folding cooperativity. This could be explained by the presence of the purification tag, which would act as a separate domain and likely affect the cooperativity of the folding/unfolding reaction of Cpr1 (i.e., the folding/unfolding of the whole protein would not be as concerted with the addition of the purification tag). In the whole protein readout, the measured $m$-value is the combination of the folding properties from both methionines. Cpr1(54-74) does not
come from a region near the affinity tag, which could explain a higher $m$-value measured at the peptide level. Regardless, the affinity between Cpr1 and CsA was easily detected, as is seen with the 1.1 M shift in $C_{SPROX}^{1/2}$ values (see Figure 13B and Table 3). The magnitude of this shift corresponds to a $K_d$ of approximately 30 nM, which agrees well with literature values for CypA ($K_d$~30-200 nM, [73, 84-87]). Additionally, the affinity between Cpr1 and CsA has previously been shown to be as tight as that for human CypA and CsA [123].

### 3.4 Conclusions

Described here was the application of protein purification techniques in combination with the SPROX technique to measure the folding and binding behavior of proteins in complex mixtures. BCA was used as a model system for SPROX coupled with an immunoprecipitation to purify proteins from complex biological mixtures. Using antibodies, BCA was pulled out of a complex mixture and its relatively weak interaction with CBS was detected and quantified. It was shown that this protein could clearly be captured, even after oxidation, and its SPROX behavior measured following immunoprecipitation was identical to that of pure BCA.

Additionally, the Cpr1-CsA interaction was studied in the context of the yeast lysate. Cpr1 was a high expressing protein, but this could easily be performed on lower expressing proteins as well, especially because the peptide readout with LC-MS is much more sensitive than the MALDI whole protein readout. In the peptide readout, only one
Figure 13. A. SPROX curve for whole protein MALDI readout of Cpr1 purified from yeast lysate. Error bars represent one standard deviation of the average mass from 10 replicate spectra each. B. SPROX curve of peptide-level Cpr1 in the absence (●) and presence (○) of CsA, which was also purified from yeast lysate following SPROX. No error bars are present because each data point is taken from one LC-MS run.
methionine-containing peptide was necessary to measure the binding interaction between Cpr1 and CsA, and it was sufficient to accurately quantify the interaction.

The described purification strategies can be used as a general strategy that could be applied to many MS-based covalent modification approaches used in studying protein-ligand binding if the protein of interest is known. Immunoprecipitation and affinity chromatography were used here, but other types of protein purification strategies would also be applicable, such as size exclusion chromatography, hydrophobic interaction chromatography, or ion exchange chromatography.
Table 4. List of identified peptides from Cpr1

<table>
<thead>
<tr>
<th>Peptide Sequence</th>
<th>Amino Acids</th>
<th>Retention Time (min)</th>
<th>[Mass+H]+</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPDENFK</td>
<td>81-87</td>
<td>11.2</td>
<td>896.415</td>
</tr>
<tr>
<td>FPDENFKK</td>
<td>81-88</td>
<td>9.3</td>
<td>1024.51</td>
</tr>
<tr>
<td>GFGYAGSPFHR</td>
<td>43-53</td>
<td>13.9</td>
<td>1195.564</td>
</tr>
<tr>
<td>HHDRPGLLSMANAGPNNTNGSQFFITTVPCPWLDGK</td>
<td>89-123</td>
<td>28.4</td>
<td>3835.843</td>
</tr>
<tr>
<td>HVVFGEVVDGYDIVK</td>
<td>124-138</td>
<td>22.2</td>
<td>1675.869</td>
</tr>
<tr>
<td>HVVFGEVVDGYDIVKK</td>
<td>124-139</td>
<td>19.1</td>
<td>1803.964</td>
</tr>
<tr>
<td>KHHDRPGLLSMANAGPNNTNGSQFFITTVPCPWLDGK</td>
<td>88-123</td>
<td>26.0</td>
<td>3963.938</td>
</tr>
<tr>
<td>KVESLGSPSGATK</td>
<td>139-151</td>
<td>6.5</td>
<td>1260.679</td>
</tr>
<tr>
<td>LYNDIVPK</td>
<td>22-29</td>
<td>13.6</td>
<td>961.535</td>
</tr>
<tr>
<td>SIYGGK</td>
<td>75-80</td>
<td>4.2</td>
<td>624.335</td>
</tr>
<tr>
<td>SIYGGKFPDENFK</td>
<td>75-87</td>
<td>16.2</td>
<td>1501.732</td>
</tr>
<tr>
<td>SIYGGKFPDENFKK</td>
<td>75-88</td>
<td>13.8</td>
<td>1629.827</td>
</tr>
<tr>
<td>VESLGSPSGATK</td>
<td>140-151</td>
<td>7.2</td>
<td>1132.584</td>
</tr>
<tr>
<td>VIPDFMLQGGDFTAGNGTGGK&lt;sup&gt;a&lt;/sup&gt;</td>
<td>54-74</td>
<td>23.3</td>
<td>2081.996</td>
</tr>
<tr>
<td>VVFKLYNDIVPK</td>
<td>18-29</td>
<td>20.7</td>
<td>1434.835</td>
</tr>
</tbody>
</table>

<sup>a</sup>Peptide used for SPROX analysis
4. Stable Isotope Labeling Strategy for Protein-Ligand Binding Analysis in Multi-Component Protein Mixtures

4.1 Introduction

The SPROX technique has been demonstrated in its ability to detect and quantify protein-ligand binding interactions [47, 74, 124]. It is compatible with proteomic fractionation methods such as one- and two-dimensional chromatography, and it can be used to screen complex protein mixtures for ligand binding interactions [74]. An important requirement for SPROX experiments on the proteomic scale is that the extent of protein oxidation must be accurately quantified in the proteomics readout (i.e., the extent of methionine oxidation must be quantified in each of the denaturant-containing SPROX buffers). The use of isobaric mass tags for such quantitative measurements using a MudPIT approach [74] has recently been demonstrated.

Here, the use of a stable isotope labeling strategy termed PrSUIT (Probing with SPROX Using Isotope Tags) is described for the detection and quantitation of protein-ligand binding interactions by SPROX using a mass spectrometry-based proteomics readout. The strategy involves performing two SPROX analyses on a protein mixture (one on the protein mixture in the absence of ligand and one on the protein mixture in the presence of ligand) using differentially labeled hydrogen peroxide in the oxidation reactions (e.g., H$_2^{18}$O$_2$ and H$_2^{16}$O$_2$). Ultimately, a quantitative LC-MS based proteomics readout is used to identify binding proteins (both direct and indirect) of the target ligand. The described protocol utilizes the O$^{18}$/O$^{16}$ ratio in the oxidized protein samples to quantify the ligand-induced protein stability changes. The ratio is determined using the
isotopic distributions observed for the methionine-containing peptides used for protein identification in the LC-MS-based proteomics readout. In the proof-of-principle study described here, the PrSUIT technique is applied to a model protein mixture in order to evaluate its ability to detect and quantify the direct binding of the immunosuppressant drug, cyclosporin A (CsA), to yeast cyclophilin A (Cpr1) and the indirect binding interaction of CsA with yeast calcineurin that involves the Cpr1-CsA complex. The work in this chapter has been published in the *Journal of the American Society for Mass Spectrometry*.

### 4.2 Experimental

#### 4.2.1 Protein Sample Preparation

*Saccharomyces cerevisiae* overexpression strains (Open Biosystems, Huntsville, AL) in which each yeast open reading frame (ORF) is expressed under control of the GAL1 promoter [121] were used in these experiments. The specific overexpression strains used in this work include those that overexpressed cyclophilin A (Cpr1), calcineurin A (Cna1), calcineurin B (Cnb1), glycogen synthase (Gsy2), pyruvate kinase (Pyk1), protein Tma108 (Tma108), and glutamate dehydrogenase (Gdh2). The UniProt accession numbers for these yeast proteins were: P14832 (Cpr1), P23287 (Cna1), P25296 (Cnb1), P40462 (Tma108), P33327 (Gdh2), P27472 (Gsy2), and P00549 (Pyk1). Each of the overexpressed yeast proteins in the overexpression strains also contained a 19 kDa

---

purification tag sequence at its C-terminus. The C-terminal tag contained an HA epitope, a 3C protease cleavage site, a ZZ domain from Protein A, as well as a 6xHis-tag that was exploited in the purification protocol employed here.

Yeast were grown overnight at 30°C in 2% dextrose SC-Ura, and 2 mL of this solution was used to inoculate 50 mL of 2% raffinose SC-Ura solution, which was then incubated overnight. The solution was diluted into 500 mL of 2% raffinose SC-Ura to give an OD$_{600}$ of 0.3 and was incubated at 30°C until the OD$_{600}$ of the solution was 0.8-1.2. A 250 mL volume of 3x YP solution containing 6% galactose was added to induce overexpression. After 6 h of incubation, yeast cell pellets were obtained by centrifuging 250 ml culture at 1,000 x g for 10 min. Cell pellets were lysed in 500 μL of buffer containing 25 mM HEPES, 500 mM NaCl, 10% glycerol, 2 mM mercaptoethanol (pH 7.5) and Halt protease inhibitors (Pierce). Cell lysis was accomplished using 0.5 mm glass beads (Sigma) with 20 s of disruption ten times and 1 min intervals on ice in between. The samples were then centrifuged at 14,000 x g for 5 min to pellet the insoluble material, the supernatants generated for a given overexpressed protein were combined, and the overexpressed protein was isolated using immobilized metal affinity chromatography (IMAC). The number of pellets that were lysed to generate the seven target proteins depended on the overexpression level of each protein and was as follow: 7 pellets for Cpr1, 12 pellets for Cna1, 12 pellets for Cnb1, 1 pellet for Pyk1, 1 pellet for Gsy2, 1 pellet for Tma108, and 1 pellet for Gdh2.

Talon Metal Affinity Resin (Clontech, Mountain View, CA) was used according to the manufacturer’s directions to isolate the soluble proteins in the supernatants. In
these IMAC experiments, the lysate was incubated with the Talon resin for 2 h. The resin was washed 3 times with 0.5 mL of 50 mM phosphate containing 300 mM NaCl (pH 7), and proteins were eluted off the resin with 450 µL of buffer containing 50 mM phosphate, 300 mM NaCl, and 150 mM imidazole (pH 7). Vivaspin 500 centrifugal filter units (Sartorius Stedim Biotech, Aubagne, France) were used according to the manufacturer’s instructions in order to buffer exchange the IMAC-purified protein samples into a 20 mM phosphate buffer (pH 7.4). The concentration of each IMAC-isolated protein was determined using a Bradford Assay [125] in which the total protein concentration was determined. The IMAC-isolated Cpr1 and Cna1 samples, which were determined to be 40 and 20 µM, respectively, in total protein, were used directly in the PrSUIT experiments on these two isolated proteins. The model protein mixture used in this work was prepared by combining aliquots of the seven IMAC-isolated proteins (including Cpr1, Cna1, Cnb1, Pyk1, Gdh2, Gsy2, and Tma108) into a single mixture. In this mixture, the final concentrations of Cpr1, Cna1, Cnb1, Pyk1, Gdh2, Gsy2, and Tma108 in the mixture were approximately 20, 2, 3, 4, 2, 0.6, and 0.6 µM, respectively. These concentrations were based on the results of Bradford Assays of the total purified protein concentrations obtained for each IMAC-purified protein.

4.2.2 PrSUIT Analyses

The denaturant-containing buffers used for the PrSUIT analyses were comprised of 20 mM phosphate buffer (pH 7.4) and guanidinium chloride (GdmCl) (EMD Chemicals, Inc) concentrations that ranged from 0 to 8 M. The specific [GdmCl] in each
buffer was determined by measuring the refractive index of the buffer using the method as described elsewhere [83].

In the PrSUIT experiments, each protein sample (i.e., the model protein mixture, the isolated Cpr1 sample, or the isolated Cna1 sample) analyzed here was split into two fractions. An aliquot of a concentrated stock solution of CsA (LKT Laboratories, St. Paul, MN) prepared in DMSO was added to one fraction such that the final concentration of CsA in the protein fraction was 1 mM and the final concentration of DMSO was 10% (v/v). An aliquot of DMSO was added to the other fraction such that the final concentration of DMSO in this second fraction (i.e., the sample without ligand) was also 10%. The two fractions (i.e., the one with and the one without ligand) generated for each protein sample analyzed here were equilibrated at room temperature for 30 minutes. A 20 µL aliquot of each protein sample was combined with 25 µL of each denaturant-containing buffer. The final GdmCl concentrations in the 10 denaturant-containing buffers used in this work were 0.5, 1.0, 1.3, 1.7, 2.0, 2.3, 2.6, 3.0, 3.5, and 4.0 M. The protein samples in each denaturant-containing buffer were equilibrated for 15 minutes before a 5 µL aliquot of a 0.8 M hydrogen peroxide solution was added to each buffer to initiate the methionine oxidation reaction. The protein samples without CsA were oxidized with H$_2^{16}$O$_2$ (Sigma) and the protein samples with CsA were oxidized with H$_2^{18}$O$_2$ (90% enriched, Isotec, Miamisburg, OH). In each case the oxidation reaction in each denaturant-containing buffer was allowed to proceed for 30 minutes before it was quenched with the addition of 100 µL of a 300 mM L-methionine (Sigma) solution.
An aqueous solution of 1 g/mL TCA was added to each denaturant-containing buffer such that the final concentration of TCA was approximately 20% (w/v), and the samples were incubated overnight on ice to precipitate the protein. The samples were centrifuged at 8,000 x g at 4°C for 30 min, the supernatants were removed, and the resulting protein pellets were washed three times with 300 µL of ice-cold ethanol. Residual ethanol was removed using a Thermo Savant SpeedVac concentrator. The protein pellets in each sample were redissolved in 60 µL of buffer containing 50 mM ammonium bicarbonate (pH 8.5) and 0.1% RapiGest SF surfactant (Waters, Milford, MA). The samples were heated at 40°C for 10 minutes while shaking. Dithiothreitol (Thermo) was added to a final concentration of 10 mM and the sample was heated for 15 min at 80°C. Iodoacetamide (Thermo) was added to a final concentration of 20 mM and the protein samples were incubated at room temperature for 30 minutes in the dark. A total of 0.5 µg of trypsin (Sigma) was added to give an enzyme:protein ratio of approximately 1:50, and the samples were incubated overnight at 37°C while shaking. The trypsin digestion reaction was quenched with the addition of TFA (Halocarbon, River Edge, NJ) and acetonitrile (ACN) such that the final concentrations of TFA and ACN were 1 and 2%, respectively.

The samples were heated at 60°C for 2 h to cleave the RapiGest, and the peptides were subjected to a second round of oxidation. In this second round of oxidation, a 5 µL aliquot of 0.8 M H$_2^{18}$O$_2$ was added to the protein samples without ligand and a 5 µL aliquot of 0.8 M H$_2^{16}$O$_2$ was added to the protein samples with ligand. The second oxidation reaction was allowed to proceed for 2 h, before the samples were evaporated to
Dryness using the SpeedVac concentrator. The samples were redissolved in 50 µL 98:2:0.1 H₂O/ACN/FA. After centrifuging each sample at 15,000 x g for 5 min, the samples from the same denaturant-containing buffers but from the with and without ligand fractions were combined to give a set of 10 peptide-containing samples for each PrSUIT analysis.

4.2.3 LC-MS Analyses

The set of 10 peptide-containing samples generated in each PrSUIT analysis was analyzed using an Agilent 6520 Q-TOF mass spectrometer system equipped with a Chip Cube Interface. The HPLC Chip used in this experiment contained a 43 mm x 75 µm column with Zorbax 300SB-C18 5 µm packing. The tryptic peptides were eluted using a linear gradient from 5% to 50% Buffer B over 30 min and then to 80% Buffer B over 4 min. Buffer A was 0.1% FA in water and Buffer B was 0.1% FA in ACN. The flow rate was 0.4 µL/min. The capillary voltage was 1800 V. The drying gas (N₂) was 350°C at a flow rate of 6 L/min. The skimmer and fragmentor were set to 65 V and 175 V, respectively. The collision energy was 3.5 V/100 Da with a -4.8 V offset. The inclusion window width for precursor ions was 4 m/z units. The scan rate was 3 scans per second in the mass spectra and 2 scans per second in the product ion mass spectra, and there were 4 precursors selected for fragmentation per cycle.

The peptide sequences and proteins in this work were initially identified in five MS/MS analyses performed on a sample composed of 15 µL aliquots from the set of ten samples generated in the PrSUIT analysis of the model protein mixture. For these five MS/MS analyses a total of 20 µL of the combined sample was loaded on column, which
corresponds to an estimated 2-4 µg of total protein, assuming the protein recovery from the protein precipitation step in the PrSUIT protocol was 25-75%. The methionine-containing peptides analyzed in this work were identified from the product ion mass spectra using Agilent’s Spectrum Mill MS Proteomics Workbench software, Rev A03.03.084 SR4. In the Spectrum Mill searches, the precursor and product ion mass tolerances were set to 20 ppm, the protein cleavage chemistry was set for trypsin with 3 maximum missed cleavages, and the maximum ambiguous precursor charge was set to 7. Carbamidomethylation of cysteine and oxidation of methionine were set as fixed modifications, and deamidation was set as a variable modification. The peptide fragment products and peptide precursor masses were searched against the NCBI nr database for *S. cerevisiae*. Separate Spectrum Mill searches were also performed to specifically identify peptide sequences with oxidation and dioxidation of cysteine and methionine residues (respectively). In these searches, no such peptide sequences were identified with Spectrum Mill scores >8, which was the minimum score of the peptides identified in the other Spectrum Mill searches.

Agilent’s Bioconfirm was also used to identify some methionine-containing peptides. This software takes the protein’s sequence and searches the mass spectra for tryptic peptides with the expected m/z and charge state. Only methionine-containing peptides that were not identified in the MS/MS runs were inspected, and these were validated based on the appearance of $^{18}$O peaks in the isotopic distribution. Extracted ion chromatograms based on the monoisotopic ion signal of each identified methionine-
containing peptide (identified with either Spectrum Mill or Bioconfirm) were generated using Agilent’s Qualitative Analysis software.

Ultimately, the %$^{18}$O labeling of each methionine-containing peptide was determined from the data in a single LC-MS analysis of each of the 10 peptide-containing samples generated in each PrSUIT analysis. The chromatographic conditions used for these LC-MS runs were identical to those described above for the LC-MS/MS runs. A total of 20-30 µL of each peptide-containing sample was loaded on column for these LC-MS runs. This corresponded to ~2-4 µg of total protein for the model protein mixture analysis, and 1.25 and 2 µg for the isolated protein analyses of CPRI and Cna1, respectively, again assuming a sample losses in the protein precipitation and redissolution steps in the above PrSUIT protocol were between 25-75%.

4.2.4 Calculation of %$^{18}$O Labeling

Extracted ion chromatograms were generated for each methionine-containing peptide analyzed in this work. The ion signals from the isotopologues detected in the top ~50% of the peak observed in the extracted ion chromatogram for a given methionine-containing peptide were used to calculate an experimentally derived weighted average molecular weight for each methionine-containing peptide. A theoretically derived weighted average molecular weight of each oxidized methionine-containing peptide containing no $^{18}$O enrichment was also determined using the natural abundance of each element in the peptide. In the case of peptides containing one methionine residue, the theoretically derived value was subtracted from the experimentally derived value to give a number between 0 and 2, which corresponded to between 0 and 100% $^{18}$O labeling,
respectively, of the methionine-containing peptides in the PrSUIT protocol. In the case of peptides containing two methionines, the same procedure was applied, except that the theoretically derived value was subtracted from the experimentally derived value to give a number between 0 and 4, which corresponded to between 0 and 100% $^{18}$O labeling, respectively. This treatment of multiple methionine residues assumes that the multiple methionine residues in a given peptide have the same level of global protection in the protein’s three-dimensional structure.

The $^{18}$O labeling for each peptide in the model mixture analysis was normalized to account for buffer variations (e.g., different TCA precipitation and redissolution efficiencies in the different samples). For this normalization, the $^{18}$O labeling for all the methionine-containing peptide generated at a specific denaturant concentration were averaged, and the average $^{18}$O labeling at each denaturant concentration was divided by the average $^{18}$O labeling at the 4 M denaturant concentration to get a set of normalization factors for each denaturant concentration. These normalization factors were 0.68, 0.70, 0.76, 0.78, 0.78, 0.79, 0.81, 0.87, 0.93, and 1.00 for the 0.5, 1.0, 1.3, 1.7, 2.0, 2.3, 2.6, 3.0, 3.5, and 4.0 M denaturant concentrations, respectively. The $^{18}$O labeling of each methionine-containing peptide at each denaturant concentration was divided by the specific normalization value at that denaturant concentration to yield a normalized $^{18}$O labeling value.
4.2.5 Construction of SPROX Curves from PrSUIT Plots of Peptide Hits

The PrSUIT plot of a peptide “hit” (i.e., a peptide derived from a protein with a stability change in the presence of ligand) was divided into two sections, including one section comprising the data at and below the denaturant concentration of the lowest %$^{18}$O value in the plot and a second section comprising the data at and above the denaturant concentration of the lowest %$^{18}$O value in the plot. The %$^{18}$O values from the first section were used to construct the SPROX curve transition expected in the absence of CsA, and those values from the second section were used to construct the SPROX curve transition in the presence of CsA. The transitions of the SPROX curves (i.e., *Fraction Oxidized* values versus [Den] plots) expected with and without ligand were constructed with *Fraction Oxidized* values that were calculated using equations 4-1 and 4-2, respectively.

\[
\text{Fraction Oxidized} = 1 - \frac{\%^{18}O - \%^{18}O_{\text{low}}}{\%^{18}O_{\text{ave}} - \%^{18}O_{\text{low}}} \quad \text{Equation 4-1}
\]

\[
\text{Fraction Oxidized} = \frac{\%^{18}O - \%^{18}O_{\text{low}}}{\%^{18}O_{\text{ave}} - \%^{18}O_{\text{low}}} \quad \text{Equation 4-2}
\]

In equations 4-1 and 4-2, %$^{18}$O is the normalized %$^{18}$O value at each denaturant concentration, %$^{18}$O$_{\text{low}}$ is the lowest %$^{18}$O value observed for the peptide, and %$^{18}$O$_{\text{ave}}$ is the average %$^{18}$O value observed for all the methionine-containing peptides, which was 39% in this work.

4.2.6 Binding Affinity Measurements

The $K_d$ values for the Cpr1-CsA complex and for the Cpr1-CsA-Cna1 complex were determined using equation 4-3.
\[ K_d = \frac{[L]}{e^{-\Delta \Delta G_f/nRT} - 1} \quad \text{Equation 4-3} \]

In equation 4-3 (identical to equation 2-4), \([L]\) is the concentration of free ligand, \(n\) is the number of independent equivalent binding sites, \(\Delta \Delta G_f\) is the binding free energy, \(R\) is the ideal gas constant, and \(T\) is the temperature. In these experiments \(n=1\) and the concentration of free ligand was estimated as the total ligand concentration because both the CsA and Cpr1-CsA ligands in this work were present in a large excess over their protein targets. The free CsA and Cpr1-CsA concentrations were estimated as 400 and 8 \(\mu\)M (respectively), with their protein targets, Cpr1 and Cna1 having estimated concentrations of 8 and 0.8 \(\mu\)M (respectively).

The \(\Delta \Delta G_f\) values used in binding affinity measurements were calculated from \(\Delta G_f\) values derived from the SPROX curve transitions constructed from the PrSUIT data on the Cpr1(54-74) and Cna1(332-354) peptide “hits.” The \(\Delta G_f\) values were obtained by fitting the SPROX curve transitions (i.e., Fraction Oxidized versus [Den] plots) to equation 4-4.

\[
\text{Fraction Oxidized} = FrOx_{\infty} + (FrOx_0 - FrOx_{\infty})e^{-(k_{OX}/(1 + K_{fold}))t} \quad \text{Equation 4-4}
\]

In equation 4-4, \(K_{fold} = e^{(\Delta G_f + m[Den])/RT}\), \(FrOx_0\) is the fraction oxidized before the first oxidation reaction in PrSUIT, \(FrOx_{\infty}\) is the fraction oxidized after the second oxidation reaction in PrSUIT, \(t\) is the time of the first oxidation in seconds, \(k_{OX}\) is the average pseudo-first order rate constant for the oxidation of an unprotected methionine residue in the protein, \(\Delta G_f\) is the free energy of folding in the absence of denaturant, \([\text{Den}]\) is the denaturant concentration, \(m\) is \(\delta \Delta G_f/\delta [\text{Den}]\), \(R\) is the gas constant, and \(T\) is
the temperature in kelvins. In fitting the Fraction Oxidized versus [Den] plots to equation 4-4, the \( FrOx_0 \) and \( FrOx_\infty \) were assigned values of 0 and 1 (respectively), \( \Delta G_f \) and \( m \) were allowed to float, and \( k_{OX} \) was assigned a value of 4.1 hr\(^{-1}\), which was based on the 80 mM \( \text{H}_2\text{O}_2 \) concentration used in the first oxidation reaction and the second order rate constant for the oxidation of unprotected methionine residues in proteins that were previously determined to be approximately 50.4 \( \text{M}^{-1}\text{hr}^{-1} \) [47]. Equation 4-4 is identical to equation 3-2 that was previously described to analyze SPROX data with the exception that the \( \Delta \text{Mass}_{\text{WtAv}} \), \( \Delta M_\infty \), and \( \Delta M_0 \) were replaced with Fraction Oxidized, \( FrOx_\infty \), and \( FrOx_0 \), respectively.

### 4.3 Results

#### 4.3.1 General Protocol

The stable isotope labeling strategy developed in this work is outlined in Figure 14. In this strategy, hydrogen peroxide isotopically enriched with \( ^{18}\text{O} \) (\( \text{H}_2^{18}\text{O}_2 \)) and hydrogen peroxide containing naturally occurring \( ^{16}\text{O} \) (\( \text{H}_2^{16}\text{O}_2 \)) were used to selectively oxidize the methionine residues in a protein sample pre-equilibrated with and without ligand, respectively, at a series of different chemical denaturant concentrations. The oxidation reactions were quenched and the protein sample in each denaturant-containing buffer was subjected to an LC-MS-based proteomics analysis using a bottom-up approach that involved generating tryptic peptides of the proteins from each denaturant-containing buffer. Prior to the LC-MS readout, the tryptic peptides from the proteins in each denaturant containing buffer were subjected to a second round of oxidation to
quantitatively oxidize all of the methionine side chains in the samples to methionine sulfoxide.

This second round of oxidation oxidized the methionine side chains that were globally protected (i.e., buried in a protein’s folded three-dimensional structure) and not oxidized during the first round of oxidation. The second round oxidation reaction conditions were tuned such that the methionine sulfoxide formed in the first round oxidation reactions were not further oxidized to the sulfone. The second round oxidation reactions also involved the use of $\text{H}_2^{18}\text{O}_2$ and $\text{H}_2^{16}\text{O}_2$. However, in the second round of oxidation, the peptide mixtures derived from the protein samples originally incubated with and without ligand were reacted with $\text{H}_2^{16}\text{O}_2$ and $\text{H}_2^{18}\text{O}_2$, respectively (i.e., the opposite of that done in the first round). This created isotopic diversity between the methionine-containing peptides derived from the protein samples that were originally in the presence and absence of ligand, and permits the samples prepared with and without ligand to be paired and combined (see Figure 14) such that they can be simultaneously analyzed in the same LC-MS run.

Ultimately, the chemical denaturant dependence of the $^{18}\text{O}$ enrichments observed for the methionine-containing peptides detected in the LC-MS readout were used to detect and quantify binding interactions of the target ligand with different proteins in the mixture (Figure 15). Proteins that do not interact with the target ligand should have the same $^{18}\text{O}$ enrichment at all denaturant concentrations. This is because the chemical denaturant dependence observed in the first round of oxidation reactions performed in the presence and absence of ligand will be the same for proteins that do not have their
Figure 14. Schematic representation of the PrSUIT protocol developed in this work.
thermodynamic stability modulated through interaction with the ligand (see “no interaction” SPROX curves in Figure 15A). However, proteins that do have their thermodynamic stability modulated through interaction with the ligand will have a denaturant dependence to their oxidation in the first round of oxidation that is different in the presence and absence of ligand (see “interaction” SPROX curves in Figure 15A). The $^{18}$O enrichment observed for methionine-containing peptides derived from regions of protein structure that experience different degrees of global and/or subglobal protection upon ligand binding will be more (or less) enriched with $^{18}$O at one or more denaturant concentrations (see Figure 15A and B).

4.3.2 PrSUIT Analysis of a Model Protein Mixture

The protocol outlined in Figure 14 was used to characterize the binding of CsA to the individual proteins in a model protein mixture, which was mainly composed of seven yeast proteins including: Cpr1, Cna1, Cnb1, Pyk1, Gdh2, Tma108, and Gsy2. The mixture also contained a number of additional yeast proteins that were co-purified in the IMAC step used to isolate the above seven proteins from yeast cell lysates in which they were each overexpressed. The primary goal of this work was to determine if the known direct and indirect interactions of CsA with Cpr1 and Cna1, respectively, could be detected using the stable isotope labeling protocol outlined here.

Initially, a pooled sample containing equal aliquots of the 10 combined sample pairs (see Figure 14) was subjected to an LC-MS/MS analysis to sequence the peptides and identify the proteins present in the sample. A total of 120 peptides from 49 different proteins were identified using Spectrum Mill to search the mass spectral data obtained in
Figure 15. Theoretical structure of PrSUIT data and its relationship to SPROX data. A. Hypothetical PrSUIT data (i.e., $^{18}$O versus [Den] plots) expected for proteins that do not (Case I) and do (Cases II or III) have their thermodynamic stability altered in the presence of the ligand. Cases II and III correspond to situations in which the thermodynamic stability of the identified protein is stabilized and destabilized, respectively, as a result of ligand binding. B. Schematic representation of the molecular $^{16}$O and $^{18}$O oxidized species generated in the 1st and 2nd oxidation reactions in the PrSUIT protocol at three different points in a protein’s SPROX curves generated in the absence, (-), and in the presence, (+), of a binding ligand.
five LC-MS/MS runs. These Spectrum Mill identifications included 29 methionine-containing peptides from a total of 21 different proteins and 13 methionine-containing peptides from six of the seven primary proteins in the mixture. While Cnb1 was successfully identified as a component of the mixture, its identification was not made with a methionine-containing peptide. An additional seven methionine-containing peptides from the seven primary proteins in the mixture were also identified based on their expected mass and isotope envelope using Agilent's Bioconfirm software. The sequences of these additional seven methionine-containing peptides were subsequently confirmed in LC-MS/MS experiments specifically targeting these peptides. Ultimately, 36 methionine-containing peptides from 21 different proteins were analyzed in the PrSUIT analysis described here (see Table 5).

The 10 paired and combined samples were each subjected to an LC-MS analysis, and the LC-MS data (i.e., the relative intensities of the isotopologues detected in the spectra obtained from the methionine-containing peptides) was used to determine the %[^18]O incorporation in each methionine-containing peptide at each denaturant concentration (see Figure 16). The resulting %[^18]O incorporation values were then normalized to correct for systematic errors that could result from either differential protein recoveries and/or differences in the actual H[^16]O[^2] and H[^18]O[^2] concentrations. Such values for 27 of the 36 methionine-containing peptides identified in the model mixture could be readily determined. The normalized %[^18]O labeling values for 9 of the methionine-containing peptides identified in the model mixture were compromised

91
because the peptides had low intensity ion signals and/or ion signals that overlapped with those of other co-eluting peptides.

PrSUIT plots (i.e., Normalized $^{18}\text{O}$ Labeling versus [GdmCl] plots) were constructed for 27 methionine-containing peptides detected in the model mixture analysis (see Table 5). These 27 methionine-containing peptides included 20 peptides from six of the seven target proteins in this study as well as seven peptides from seven other random proteins that were apparently co-purified with the seven target proteins. The seven peptides from these seven random proteins in the mixture served as negative controls. Shown in Figure 17 are PrSUIT plots for several methionine-containing peptides, including one with a relatively constant $^{18}\text{O}$ labeling at the different denaturant concentrations (Figure 17A) and two with $^{18}\text{O}$ labeling that is significantly low (see below) at several denaturant concentrations (Figure 17B and C).

Shown in Figure 18 is the distribution of normalized $^{18}\text{O}$ labeling values obtained at all the denaturant concentrations for all of the 27 methionine containing peptides in this work. Also shown in Figure 18 is the distribution of normalized $^{18}\text{O}$ labeling values obtained for just the negative controls. Both distributions are similar, although the distribution observed for the negative controls is not as smooth due to the relatively small number of points. The range of normalized $^{18}\text{O}$ labeling values observed in these experiments was largely defined by the random error associated with the weighted mass average determinations for each peptide. The largest source of error in the weighted mass average determinations came from the variability associated with the relative ion signal intensities of the different isotopologues of a given peptide. It is
this error that largely defined the 6% standard deviation observed for all the normalized %\(^{18}\)O labeling values in Figure 18.

The distribution of all the normalized %\(^{18}\)O labeling values for all the peptides in Figure 18 is also centered at the average value, 39%, which is close to the 45% value expected for methionine-containing peptides derived from proteins that do not interact with CsA. The %\(^{18}\)O incorporation into methionine-containing peptides from proteins that do not interact with CsA was expected to be 45% if the H\(_2\)^\(^{18}\)O\(_2\) reagent was indeed 90% enriched with \(^{18}\)O as reported by the supplier, and if the H\(_2\)^\(^{18}\)O\(_2\) and H\(_2\)^\(^{16}\)O\(_2\) reagent concentrations in the first round oxidation reactions were indeed identical. The 6% discrepancy between the average %\(^{18}\)O value observed in this experiment, 39%, and the expected value, 45%, is most likely due to the H\(_2\)^\(^{18}\)O\(_2\) and H\(_2\)^\(^{16}\)O\(_2\) reagent concentrations not being exactly identical during the first round of oxidation. These results suggest that the actual hydrogen peroxide concentration in the H\(_2\)^\(^{18}\)O\(_2\) stock solution may have been slightly lower than that in the H\(_2\)^\(^{16}\)O\(_2\) stock solution. Hydrogen peroxide solutions can decompose over time to water and oxygen, thus it is possible that the actual hydrogen peroxide concentrations in the H\(_2\)^\(^{18}\)O\(_2\) and H\(_2\)^\(^{16}\)O\(_2\) stock solutions may have been differentially lower than that reported by the manufacturer. Any oxidization of the protein material in the samples prior to PrSUIT analysis (i.e., oxidation products generated during the protein purification steps) could also have contributed to the systematic reduction in \(^{18}\)O levels observed in the PrSUIT analyses.

The distribution of all the normalized %\(^{18}\)O labeling values at all the denaturant concentrations (see Figure 18) was used to help identify peptides with significantly
Table 5. List of identified methionine-containing peptides

<table>
<thead>
<tr>
<th>Protein</th>
<th>A.A.</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cna1</td>
<td>355-368</td>
<td>AAVLKYEENVMNIR</td>
</tr>
<tr>
<td></td>
<td>184-198</td>
<td>HLTYFTFKNEMLHKK</td>
</tr>
<tr>
<td></td>
<td>332-354</td>
<td>VGTGPSLITMFASPYLDTYHKN</td>
</tr>
<tr>
<td></td>
<td>471-486</td>
<td>MFSVLREESEKVEYLK(^a)</td>
</tr>
<tr>
<td></td>
<td>326-354</td>
<td>MYKNKVTGFPSTMFASPYLDTYHKN(^a)</td>
</tr>
<tr>
<td></td>
<td>329-354</td>
<td>NNKVTGFPSTMFASPYLDTYHKN(^a)</td>
</tr>
<tr>
<td></td>
<td>93-109</td>
<td>ILNMSTVALSKEPNLLK(^a)</td>
</tr>
<tr>
<td>Pyk1</td>
<td>313-337</td>
<td>AEVSDVGNAILDGADCVMLSGETAK</td>
</tr>
<tr>
<td></td>
<td>201-225</td>
<td>FGVKNGVHMVFASFIKTANDVLTIR</td>
</tr>
<tr>
<td></td>
<td>338-369</td>
<td>GNYPINAVTTMAETAVIAEQAILYPNYDDMR</td>
</tr>
<tr>
<td></td>
<td>241-264</td>
<td>IENQQGVNNFDEILKVTGDMVAR</td>
</tr>
<tr>
<td></td>
<td>205-216</td>
<td>NGVHMVFASFIR</td>
</tr>
<tr>
<td></td>
<td>287-312</td>
<td>SNLAGKPVICATQMLESMYTPRPR</td>
</tr>
<tr>
<td></td>
<td>92-119</td>
<td>TGTNTNDVDPYIPPNHEMIHTITDDKYAK</td>
</tr>
<tr>
<td>Gdh2</td>
<td>9-38</td>
<td>GALNSLNDPIASLSSMSDYHVFDGFPGK</td>
</tr>
<tr>
<td></td>
<td>920-943</td>
<td>GGVTSSSMEVLASLALNDNFVHK</td>
</tr>
<tr>
<td></td>
<td>120-144</td>
<td>IITNDNHAIFMESNTGVISIDSQKQ</td>
</tr>
<tr>
<td></td>
<td>710-729</td>
<td>SPSSLGIPHDIEYGMSTSLGVR</td>
</tr>
<tr>
<td></td>
<td>212-250</td>
<td>LTFVYESVYMPAGDPAGDISSQDLDLKDIESIDKTMYK</td>
</tr>
<tr>
<td>Tma108</td>
<td>702-711</td>
<td>VMSQVLFNLK(^b)</td>
</tr>
<tr>
<td></td>
<td>515-548</td>
<td>FINQLSTEEKDSLQEDVPYQVPLFGVLPDGFMDKT</td>
</tr>
<tr>
<td></td>
<td>426-431</td>
<td>GIMLR</td>
</tr>
<tr>
<td></td>
<td>212-231</td>
<td>TPLMTTSTVGFSIGDLEFK</td>
</tr>
<tr>
<td>Rsc2</td>
<td>463-486</td>
<td>LFKNNEVMKGTKQYRDHVLVSNLGVK(^c)</td>
</tr>
<tr>
<td>Cpr1</td>
<td>54-74</td>
<td>VIPDFMLQGGDFTAGNGTGGK</td>
</tr>
<tr>
<td>Clu1</td>
<td>360-383</td>
<td>NFNDEFQAIKDLTSTLQDRIEMER</td>
</tr>
<tr>
<td>Cog3</td>
<td>142-176</td>
<td>LSNLTESIKALHYFEVULDPMRRRNHATSPAIVK</td>
</tr>
<tr>
<td>Gsy2</td>
<td>474-497</td>
<td>MIHPEFLNANNPILGLDYDEFVR</td>
</tr>
<tr>
<td>Nip100</td>
<td>651-675</td>
<td>LNEENIRLKEVLQKENMLETEK(^c)</td>
</tr>
<tr>
<td>Pnc1</td>
<td>3-24</td>
<td>TLIVDMQNDISPGLSTVPK</td>
</tr>
<tr>
<td>Myo3</td>
<td>912-928</td>
<td>IMIKVGPTIEYHKQPNK(^c)</td>
</tr>
<tr>
<td>Rpl9A</td>
<td>90-110</td>
<td>MRYYAHIPVINIVIVEKDGA(^b)</td>
</tr>
<tr>
<td>Rif2</td>
<td>116-158</td>
<td>VEHHQAYGIDRAVSETLSLVINVVIEMNDYLMKEGIQSSK(^c)</td>
</tr>
<tr>
<td>Bem2</td>
<td>1159-1178</td>
<td>SFTTTTTVLENMAKRYVGAK(^c)</td>
</tr>
<tr>
<td>Npl3</td>
<td>127-155</td>
<td>LFVRPPELVDQESNELNIEIFPGPFGPMEKV</td>
</tr>
<tr>
<td>Met1</td>
<td>128-173</td>
<td>IPINTFHKPESTFNMIPTWDPKGSLQISVTTNNGYLANRIK(^c)</td>
</tr>
<tr>
<td>Tef2</td>
<td>265-288</td>
<td>VETGVIKPGMVVTFAPAGVTEVK(^b)</td>
</tr>
<tr>
<td>Ty1A</td>
<td>331-354</td>
<td>HLNMVTAVELFDHIAYEEQQGSR</td>
</tr>
<tr>
<td>Ugpl</td>
<td>331-346</td>
<td>LIESNLEMEIIPNQK</td>
</tr>
<tr>
<td>Tip1</td>
<td>356-370</td>
<td>GVEYFEVITFEMGQK</td>
</tr>
</tbody>
</table>

\(^a\)Peptides were only identified in experiment with CnA alone.
\(^b\)No PrSUIT plot due to interference from other peptide.
\(^c\)No PrSUIT plot due to low signal intensity.
Figure 16. Relative ion signal intensities observed (black bars) and theoretically calculated (grey bars) for the isotopologues of the methionine-containing peptide, Cpr1(54-74) of sequence VIPDFMLQGDFTAGNGTGGK, that were detected in the PrSUIT analysis of the model protein mixture. The data obtained at two different denaturant concentrations, 3.5 and 2.0 M GdmCl, are shown in A and B, respectively. In each case M represents the monoisotopic mass of the peptide, 2096.9837 Da. The theoretical distributions of the relative ion signal intensities for the isotopologues in A and B are those expected for the oxidized Cpr1(54-74) peptide with 41 and 11% $^{18}$O labeling, respectively, in the PrSUIT protocol.
altered $^{18}\text{O}$ labeling. Methionine-containing peptides with normalized $^{18}\text{O}$ labeling greater than 51% or less than 27% were deemed significantly altered based on the observed distribution of the measured values (Figure 18), which revealed 96% of the measured values were within the 27-51% range. The center of this range, 39%, was the average normalized $^{18}\text{O}$ labeling value determined for all the methionine-containing peptides at all the denaturant concentrations. The range was also equivalent to the average value ±2 standard deviations. It was also reasoned that tight binding interactions would produce such altered normalized $^{18}\text{O}$ labeling values that were consistently greater than 51% or consistently less than 27% at two or more consecutive denaturant concentrations. This selection strategy is analogous to that which have previously been described for selecting tight binding interactions in SPROX analyses using an isobaric mass tagging strategy [74].

A total of 24 of the 27 methionine-containing peptides analyzed here had normalized $^{18}\text{O}$ labeling values between 27% and 51% at all the denaturant concentrations and were clearly not “hits.” One peptide had a normalized $^{18}\text{O}$ labeling value outside the 27% and 51% at only one denaturant concentration and was also not categorized as a “hit.” Two methionine-containing peptides, including one derived from Cpr1 and one derived from Cna1 were identified as “hits” as they had normalized $^{18}\text{O}$ labeling values that were less than 27% at two or more consecutive denaturant concentrations. The methionine-containing peptide from Cpr1, Cpr1(54-74) of sequence VIPDFMLQGGDFTAGNGTGGK, had normalized $^{18}\text{O}$ labeling values less than 27% at five denaturant concentrations between 1 M and 3 M GdmCl (see Figure 17B),
indicating an increase in thermodynamic stability of the Cpr1 protein in the presence of CsA. The Cna1 peptide, Cna1(332-354) of sequence VTGFPSTMFSAPNYLDTYHNK had normalized %\(^{18}\)O labeling values less than 27% at two consecutive denaturant concentrations between 2 M to 3 M (Figure 17C), indicating an increase in thermodynamic stability of the Cna1 protein in the presence of CsA.

### 4.3.3 PrSUIT Analyses of Purified Cpr1 and Purified Cna1

As part of this work, IMAC-purified samples of overexpressed Cpr1 and Cna1 were each subjected to a PrSUIT experiment using the CsA ligand. The single methionine-containing peptide of Cpr1 (i.e., Cpr1(54-74)) that was detected in the PrSUIT analysis of Cpr1 in the model protein mixture was also the single methionine-containing Cpr1 peptide detected in the PrSUIT analysis of the purified protein. The PrSUIT plot obtained for the Cpr1(54-74) peptide in the experiment on the purified Cpr1 protein (see Figure 17B, open symbols) had the same overall structure as the PrSUIT plot obtained for the peptide in the experiment performed on the model protein mixture (Figure 17B, closed symbols), although the exact %\(^{18}\)O labeling values recorded in the two experiments were not identical. Both plots show similarly reduced \(^{18}\)O labeling at multiple denaturant concentrations between 1 M and 3 M GdmCl (see Figure 17B), indicating an increase in thermodynamic stability of the Cpr1 protein in both PrSUIT experiments, a result that is consistent with Cpr1 being a direct protein target of CsA. The discrepancies between the %\(^{18}\)O labeling values recorded in the two experiments can
Figure 17. Representative PrSUIT results obtained in the CsA binding analysis using the model protein mixture (closed symbols) and the purified proteins (open symbols). A. PrSUIT plot obtained for a methionine-containing peptide, Gdh2(920-943) of sequence GGVTSSSMELASLALNDNDVHK, derived from a protein with no CsA-induced interactions. B. and C. PrSUIT plots obtained for two methionine-containing peptides, Cpr1 (54-74) of sequence VIPDFMQLGGDFTAGNGTGK and Cna1(332-354) of sequence VTGFPSLITMFSAPNYLDTYHK, respectively, each derived from proteins with known CsA-induced interactions. The CsA binding interaction with Cna1 is also known to require Cpr1. Note that the purified protein plots were constructed with % $^{18}$O labeling values that were not normalized.
Figure 18. Distribution of all the normalized $^{18}$O labeling values at all the denaturant concentrations (black bars) obtained on the 27 methionine-containing peptides analyzed in the PrSUIT experiment performed on the model protein mixture, as well as the contribution of normalized $^{18}$O values from just the negative control peptides (white bars).
be explained by small differences in the experimental parameters (e.g., mismatched 
$H_2^{18}O_2$ and $H_2^{16}O_2$ reagent concentrations) and/or potential differences in the biophysical 
properties of the Cpr1 folding reaction in the presence of Cna1 and Cnb1 (see Discussion).

Seven unique methionine-containing peptides from Cna1 were identified in the 
PrSUIT experiment performed on the purified Cna1 construct (see Table 5). The seven 
methionine-containing peptides included one Cna1 peptide that was identified as a “hit” 
in the PrSUIT analysis of the model protein mixture, two Cna1 peptides that were detected in the PrSUIT analysis of the model mixture but not identified as “hits,” and four 
new methionine-containing Cna1 peptides that were not detected in the PrSUIT analysis 
of the model protein mixture. The PrSUIT plots obtained for all seven of the methionine-
containing Cna1 peptides identified in the experiment on the purified Cna1 construct 
showed relatively constant $^{18}O$ incorporation at all denaturant concentrations. Of 
particular significance is that the PrSUIT plot for the Cna1(332-354) peptide, which had 
displayed reduced $^{18}O$ incorporation at multiple denaturant concentrations between 2 M 
and 3 M GdmCl (see Figure 17C) when Cna1 was analyzed in the model protein mixture, 
did not yield such reduced $^{18}O$ incorporation when Cna1 was analyzed separately from 
the mixture (see Figure 17C). This change in the PrSUIT behavior of the Cna1(332-354) 
peptide is consistent with Cna1 being an indirect protein target of CsA.

4.3.4 Binding Affinity Measurements

The PrSUIT data collected on the “hit” peptides identified in a PrSUIT 
experiment was used to generate SPROX curves (see Section 4.2.5) for each peptide hit
in the presence and absence of ligand. The resulting SPROX curves were used to quantify the affinity of the detected protein-ligand interaction. Shown in Figure 19 are the SPROX curves generated from the PrSUIT data collected here on the Cna1(332-354) and Cpr1(54-74) peptides derived from the protein mixture (see Figure 17B and C). The $\Delta G_f$ values extracted from the SPROX curves generated for the Cpr1(54-74) peptide with and without the CsA ligand (-12.0 ± 2.6 and -5.5 ± 2.3 kcal/mol, respectively) and the $\Delta G_f$ values extracted from the SPROX curves generated for the Cna1(332-354) peptide with and without the CsA ligand (-7.1 ± 2.1 and -4.7 ± 2.2 kcal/mol, respectively) were used to calculate $K_d$ values of 6 and 120 nM for the Cpr1-CsA and Cpr1-CsA-Cna1 complexes, respectively.

4.4 Discussion

The primary goal of this work was to determine if the stable isotope labeling protocol outlined here could detect and quantify the direct binding of CsA to Cpr1 and the indirect binding interaction of CsA with calcineurin that involves the Cpr1-CsA complex. The binding affinity of CsA for human cyclophilin A has been previously measured using more conventional techniques, and dissociation constants in the 30-200 nM range have been reported [73, 84-87]. In earlier SPROX experiments on cyclophilin A, including work on the intact human protein using a MALDI readout [47] and on the yeast protein using a quantitative proteomics readout involving an isobaric tagging strategy [74], the $K_d$ values determined for the CsA-cyclophilin A complex mass ranged
Figure 19. SPROX curves generated from PrSUIT data collected in this work on Cpr1 and Cna1 in the model mixture analysis. A. SPROX curves generated for the Cpr1(54-74) peptide of sequence VIPDFMLQGGDFTAGNGTGGK using the PrSUIT data shown in Figure 17B. B. SPROX curves generated for the Cna1(332-354) peptide of sequence VTGFPSTLMFSAPNYLDTYHNK using the PrSUIT data shown in Figure 17C.
from 26-100 nM. The \( K_d \) value determined here from the PrSUIT data, 6 nM, is close to this range (i.e., within 5-fold of the lower end).

The interaction of calcineurin with the CsA-cyclophilin A complex has been established in enzymatic [126, 127] and X-ray crystallographic studies [128], and the interaction has been shown in these studies to require both cyclophilin and CsA (i.e., neither cyclophilin or CsA interact with calcineurin in the absence of the other). The PrSUIT results on the yeast proteins in this study are also consistent with this observation as the Cna1(332-354) peptide was not identified as a “hit” in the PrSUIT experiment on the isolated Cna1 protein. It was, however, identified as a “hit” in the protein mixture analysis, in which Cna1 was analyzed in the presence of Cpr1. A \( K_d \) value has not been previously reported in the literature for the calcineurin binding interaction with the CsA-cyclophilin A complex. However, the human cyclophilin A-CsA complex has been shown to inhibit the phosphatase activity of calcineurin with a \( K_i \) value of 32 nM [126], which is within 5-fold of the \( K_d \) value of 120 nM that was determined here.

Five of the proteins in the model mixture were identified to interact (either directly and/or indirectly) with CsA in earlier work in which the binding assay was performed in the context of all the endogenous proteins in a yeast cell lysate [74]. With the exception of Cpr1, peptide hits from these proteins were not identified in the PrSUIT experiment described here. These results suggest that CsA does not directly interact with these proteins. Identification of these proteins as hits in earlier analyses of these proteins in the context of the yeast cell lysate suggests that there are other yeast proteins that mediate CsA’s interaction with these proteins.
The PrSUIT strategy described here is fundamentally related to SPROX. Thus, as is the case for SPROX analyses, a requirement for PrSUIT analyses is that the protein(s) under study must have methionine residues in the primary amino acid sequence that are globally protected in the protein’s folded three-dimensional structure. Such globally protected methionine residues are methionine residues that can only react with the hydrogen peroxide reagent when the protein globally unfolds. The successful analyses of such proteins using the LC-MS-based proteomics readout described here further requires that tryptic peptides containing these methionine residues also be detected in the mass spectrometry readout. The results obtained in this work and in an earlier SPROX study [74], indicate that approximately 20-25% of the peptides identified in the LC-MS-based readout contained at least one methionine residue and these peptides mapped to between 30 and 45% of the identified proteins.

Of particular significance in both SPROX and PrSUIT analyses of protein-ligand binding is that the methionine residues need not be derived from regions of a protein’s three-dimensional structure that are at or near the binding site. They can be derived from any globally protected region of the folding domain(s) in the protein that are involved in binding. This is substantiated by the PrSUIT results, which identified both the Cna1(332-354) and Cpr1(54-74) peptides as “hits.” These yeast Cna1 and Cpr1 peptide sequences were aligned in human homologues of Cna1 and Cpr1, on which there is X-ray crystallographic data for the CsA-Cpr1 and CsA-Cpr1-Cna1 complexes, and it was determined that the Cna1(332-352) peptide is at the Cpr1-Cna1 binding interface, but that the Cpr1(54-74) peptide was not near the Cpr1-CsA binding interface.
As with SPROX, there are several caveats for the use of PrSUIT to detect and quantify protein-ligand binding interactions. The determination of $K_d$ values requires an accurate evaluation of the free ligand concentration (see equation 4-3). The successful detection of a protein-ligand binding interaction using either SPROX or PrSUIT also requires a sufficiently high free ligand concentration to produce a measurable shift (i.e., $>0.5\ M\ GdmCl$) in the protein’s SPROX curve. The minimum free ligand concentration needed to produce a $>0.5\ M$ shift will not be the same for all protein-ligand systems. Complexes with larger $K_d$ values (i.e., weaker binding interactions) and larger $m$-values (i.e., more cooperative folding reactions) will require higher free ligand concentrations for their detection than those with smaller $K_d$ values and smaller $m$-values. The results with the Cpr1 and Cna1 protein systems suggest that nanomolar $K_d$ values should be readily measured using free ligand concentrations in the 10 to 400 $\mu$M range for proteins with $m$-values in the 2-4 kcal mol$^{-1}$ M$^{-1}$, respectively.

In SPROX and PrSUIT analyses of protein-ligand binding, it is also important that the oxidation products of H$_2$O$_2$-mediated oxidation reaction do not alter the ligand binding affinities of the protein(s) under study. The reasonably good agreement between the PrSUIT derived binding affinities determined in this work and those previously reported for the CsA-Cpr1 and CsA-Cpr1-Cna1 complexes suggest that the oxidation products did not complicate the analyses described here. It is also important that the protein-ligand binding affinity not be compromised by the presence of the chemical denaturant.
The basic structure of PrSUIT data is expected to be similar to what is summarized in Figure 15A. However, the specific $\%^{18}$O labeling values observed in a PrSUIT experiment will depend on several experimental parameters and on the biophysical properties of the protein-ligand interaction under study. For example, the baseline $\%^{18}$O labeling values observed in a PrSUIT analysis will be maximized when the $^{18}$O enrichment of the $\text{H}_2^{18}\text{O}_2$ reagent is high, the $\text{H}_2^{18}\text{O}_2$ and $\text{H}_2^{16}\text{O}_2$ reagent concentrations are exactly matched, and “air” oxidation of the sample prior to analysis is minimized. The magnitude of the “dip” in $\%^{18}$O labeling values expected for binding proteins will be greatest when the baseline $\%^{18}$O labeling values are maximized. The exact magnitude of the “dip” in $\%^{18}$O labeling values expected for binding proteins will also vary from protein complex to protein complex. The magnitude of the “dip” is largely dependant on the amplitudes of the SPROX curves expected for the protein and the protein-ligand complex. The *Fraction Oxidized* values in the post-transition baselines of SPROX curves (see Figure 15) are typically 1. However, fraction oxidized values in the pre-transition baselines of SPROX curves can vary depending on how well a specific methionine residue is protected from oxidation in the SPROX experiment. Methionine residues that are buried in the hydrophobic core of a protein (or protein-ligand complex) and that are in conformationally restricted regions of the protein's folded three-dimensional structure will be more highly protected from oxidation in the SPROX experiment and will yield SPROX curves with low pre-transition baselines and high amplitudes. However, globally protected methionine residues in more dynamic (i.e.,
conformationally flexible) regions of protein structure can yield SPROX curves with larger pre-transition baseline values and small amplitudes.

In the PrSUIT experiment, it is important that the same protein concentration be used in the plus and minus ligand samples and that there is a consistent level of protein recovery in the PrSUIT sample workup (particularly in the protein precipitation and re-dissolution steps). Differences in protein concentrations will skew the $\%^{18}$O values observed in the baselines of the PrSUIT plots. It is relatively straightforward to ensure that the same protein concentration is used for the plus and minus ligand samples by employing the same protein stock solution to prepare both sets of samples. However, it can be more difficult to ensure a consistent level of protein recovery in the PrSUIT sample workup. For example, the protein precipitation and re-dissolution steps can yield different amount of protein from the chemical denaturant-containing buffers. If the protein recovery is not consistent in the samples that are paired prior to the LC-MS/MS analysis, then the resulting $\%^{18}$O values will be skewed. In ligand binding analyses involving protein mixtures where few protein-ligand binding interactions are expected (i.e., the majority of methionine-containing peptides are expected to exhibit non-binding behavior shown in Figure 15A), it is possible to correct for such sample-to-sample variation in protein recoveries using the normalization procedure described above.

The PrSUIT strategy developed here for the detection and quantitation of protein ligand binding interactions by SPROX has several advantages over the isobaric mass tagging strategy we previously reported. The labeling scheme in PrSUIT generates only one form (i.e., the sulfoxide form) of all the methionine-containing peptides in a sample,
and it enables quantitative data to be extracted from the mass spectral data. This is in contrast to the isobaric mass tagging strategy we previously demonstrated with SPROX, which generates two forms (i.e., oxidized and non-oxidized) of each methionine-containing peptide in the samples, and requires product ion mass spectra for quantitation. The reduction in sample complexity that comes with generation of only the oxidized form of methionine-containing peptides in PrSUIT may facilitate the analysis of more complex protein mixtures. The ability to extract quantitative data from mass spectral data may also facilitate more complete coverage of the methionine-containing peptides generated in the PrSUIT analysis of proteins in complex protein mixtures.
5. Protocol for SPROX Analysis of Protein-Ligand Binding on the Proteomic Scale

5.1 Introduction

An advantage of SPROX is the ability to make quantitative measurements of all classes of ligands on a global scale, unlike the yeast two-hybrid assay or tandem affinity purification (TAP) studies. Because the SPROX reaction specifically targets methionine residues, which are the only amino acids useful for providing thermodynamic information within the limits of SPROX, a selective strategy for the isolation of Met peptides would be extremely beneficial and would expand the scope of this platform.

Methionine residues naturally occur in the primary sequences of proteins with a frequency of only 2.37% [129]. However, they exist in 96.9% of known protein sequences [129]. As a result, proteomic strategies isolating methionine-containing peptides (Met peptides) would maximize proteome coverage while minimizing the number of peptides identified. Previous approaches for the selection of Met peptides have included separations-based [130-132] and solid phase-based strategies [133-136]. The solid phase-based approach is based on the selective covalent reaction between immobilized bromoacetyl groups with the thioether groups from methionine side chains at pH 2-3, with the subsequent release and complete regeneration of the Met peptide under reducing conditions at pH 9 [133]. The solid phase-based approach has previously been demonstrated for its selectivity of Met peptides from an E. coli lysate [134].

This chapter describes the incorporation of a solid phase-based Met peptide isolation strategy with the existing SPROX platform for thermodynamic analysis on a
proteomic scale. Reported here is the first application and characterization of this solid phase-based approach combined with the SPROX technique to isolate isobaric mass tag-labeled Met peptides from yeast cell lysates in order to globally profile protein interactions with three classes of ligands, including small molecules (resveratrol and manassantin A), the cofactor nicotinamide adenine dinucleotide (NAD$^+$), and other proteins.

5.2 Experimental

5.2.1 Yeast Growth

*S. cerevisiae* strain Y258 (*MATa pep4-3, his4-580, ura 3-53, leu2-3,112*) was grown with YPD (1% yeast extract, 2% peptone, and 2% dextrose) and harvested when the OD$_{600}$ reached approximately 0.8-1.0. For purposes of this study, this yeast was referred to as the “endogenous” yeast. Additionally, Y258 strains overexpressing YDL215C, YPL240C, YAL038W, and YCR012W under control of the GAL1 promoter were grown as described in sections 3.2.3 and 4.2.1, and were used in the NAD$^+$ binding experiment and the protein-protein interaction study.

5.2.2 NAD$^+$ Cofactor Binding Study

Yeast pellets overexpressing YDL215C were lysed in 500 µL of 20 mM phosphate, pH 7.4 containing 1 mM AEBSF, 50 µM bestatin, 15 µM E64, 20 µM leupeptin, and 10 µM peptstatin A (Pierce). Cell lysis was accomplished by vortexing the pellets with 425-600 µm acid-washed glass beads (Sigma-Aldrich) for 20 s ten times, with 1 min intervals on ice in between. The samples were then centrifuged at 14,000 x g
for 5 min to pellet the insoluble material, and the supernatant lysate was saved for SPROX analysis. Lysates ranged from 9 to 19 mg/mL. The lysate was divided into two 225 µL aliquot. A total of 25 µL of 50 mM β-nicotinamide adenine dinucleotide (NAD+) (Sigma) was added to one aliquot, and 25 µL 20 mM phosphate was added to the other. After incubation with the NAD+ for 2 h, 20 µL of the lysate sample was added to 75 µL SPROX buffer composed of 20 mM phosphate, with concentrations of urea ranging between 0 and 8 M, and equilibrated 30 min. To initiate oxidation, 5 µL 9.8 M H2O2 (Sigma-Aldrich) was added, and the sample was oxidized 6 min before quenching with 950 µL 300 mM free L-methionine (Sigma-Aldrich). There were a total of 8 SPROX samples for each with and without NAD+ lysate sample, and the specific final urea concentrations in the SPROX reactions were 0, 1.0, 2.0, 2.7, 3.3, 4.0, 4.9, and 6.0 M.

5.2.3 Resveratrol Binding Study

Endogenous yeast pellets were lysed, and the lysates were 7-10 mg/mL. Lysate was divided into two 150 µL aliquots, and 16.67 µL 10 mM resveratrol (Sigma-Aldrich) in DMSO was added to one while 16.67 µL DMSO was added to the other. The lysates were allowed to equilibrate in the presence or absence of resveratrol for 30 min before adding 20 µL of the lysate samples to 25 µL SPROX buffer. Further equilibration time in the SPROX buffers was allowed to establish protein folding equilibria. Oxidation was initiated by adding 5 µL 9.8 M H2O2, and the oxidation reactions proceeded for 3 min before quenching with 1 mL 300 mM L-methionine. Final [GdmCl] in the SPROX reactions were 0.5, 1.0, 1.3, 1.7, 2.0, 2.5, 3.0, and 3.5 M.
5.2.4 Manassantin A Binding Study

Endogenous yeast pellets were lysed, and the lysate was 8.4 mg/mL. The lysate was divided into two 170 µL aliquots, and 18.8 µL of 2 mM manassantin A in DMSO was added to one while DMSO was added to the other. Manassantin A was a gift from Professor Jiyong Hong at Duke University. The SPROX conditions were identical to those described in Section 5.2.3, except that the final [GdmCl] were 0, 0.5, 1.0, 1.3, 1.7, 2.0, 2.5, and 3.0 M.

5.2.5 Protein-Protein Interactions Studies

Yeast overexpressing pyruvate kinase (Pyk1, YAL038W) and 3-phosphoglycerate kinase (Pgk1, YCR012W) were grown according to Section 5.2.1. Overexpressed proteins were purified with Talon resin as described in Section 4.2.1. The purified Pyk1 was 1.3 mg/mL (approximately 15 µM), and Pgk1 was 3.6 mg/mL (approximately 50 µM), as determined from Bradford assays [125].

Endogenous yeast pellets were lysed, and the lysate was 13 mg/mL. The lysate was divided into 160 µL aliquots, and either 40 µL of one of the purified proteins or of 20 mM phosphate (control) was added to the lysate. The SPROX conditions were identical to those described in Section 5.2.3. The same control sample was used for both protein-protein interaction experiments.

5.2.6 General Proteomics Procedure

To remove the denaturant and methionine, which would interfere with digestion and iTRAQ labeling steps, the lysate protein samples were precipitated with the addition of 200 µL ice-cold 1 g/mL TCA. The protein samples were then incubated on ice
overnight. The protein samples were pelleted by centrifuging at 8,000 x g for 30 min at 4°C. The supernatants were subsequently removed, and the protein pellets were washed with 300 µL ice-cold ethanol. After a brief centrifuging step, the ethanol was removed, and the washing step was repeated 2 additional times. The remainder of the ethanol was evaporated using a SpeedVac concentrator. The protein pellets were then redissolved using 30 µL 0.5 M triethylammonium bicarbonate (Sigma), pH 8.5, and 1.5 µL 2% (w/w) SDS (Sigma), along with a series of vortexing and sonicating steps. After dissolution, 5 µL of each sample was removed for a Bradford assay to determine how much protein was recovered following the TCA precipitation. The remainder of the sample was treated with 2.5 µL 50 mM tris(2-carboxyethyl)phosphine (Thermo) and was heated for 1 h at 60°C to reduce the disulfide bonds. Following reduction, 1.25 µL 200 mM S-methyl methanethiosulfonate (Sigma) in isopropanol was added to alkylate the free cysteines. After 10 min at room temperature, proteomics grade trypsin (Sigma) was added at an approximate trypsin:protein ratio of 1:20-1:50, and the protein samples were digested at 37°C overnight.

5.2.7 iTRAQ Labeling

Digestion efficiency was confirmed by removing 1 µL of the samples, using C18 ZipTips for sample cleanup, eluting the samples onto a stainless steel MALDI target with 2 µL saturated α-cyano-4-hydroxycinnamic acid (Sigma) in 50:50:0.1 ACN/H₂O/TFA, and observing the presence of peptides on the Bruker MALDI-TOF/TOF. The peptide samples were then labeled with iTRAQ 8-plex isobaric mass tags (Applied Biosystems). Each set of samples (i.e., with and without ligand) was labeled with 0.5 units of the
iTRAQ tags dissolved in 50 µL isopropanol. After addition of the iTRAQ tags, the pH of the samples was tested to confirm it was above pH 8. The iTRAQ labeling reaction proceeded for 2 h at room temperature, and the samples were then stored at -20°C until further cleanup prior to mass spectrometry analysis. Samples that were not enhanced for Met peptides were cleaned up with C18 resin following iTRAQ labeling prior to MS analysis.

5.2.8 Met Peptide Isolation

Approximately 25-30 µg of each iTRAQ-labeled sample was used for the methionine enhancement protocol. This amount was removed from each iTRAQ-labeled sample and combined with the other 7 labeled samples within each set of samples. The pooled samples containing all 8 iTRAQ tags were reduced in volume to approximately 50-75 µL using a SpeedVac concentrator. Approximately 25-30 µL glacial acetic acid was added to lower the pH of the sample to between 2-3.

Pi³™ Methionine Reagent (The Nest Group, Southboro, MA) was prepared by adding 200 µL methanol and vortexing for 15 minutes. The methanol was removed, the samples were added to the prepared Pi³™ Methionine Reagent, and the samples and reagent were mixed vigorously for 90 min to allow covalent binding between the immobilized bromoacetyl group and the unoxidized methionine side chains. Following the capture, 100 µL 0.2 BME in 25% acetic acid was added, and the samples were vortexed for 30 min. The liquid was removed from the resin and discarded, and the resin was washed three times each with 0.2 M BME in 25% acetic acid, 70% ACN/0.1%TFA, and then water. Following the washes, 86 µL 1 M ammonium bicarbonate (pH 9) and 14
µL BME was added to release the methionine residues from the resin, and the samples were vortexed for 2 h. The supernatant containing the methionine peptides was removed and saved, and 50 µL 20% ACN/0.1% TFA was added to the resin. The samples were vortexed for 15 min, and the supernatant was removed and added to the solution containing the methionine peptides. The resin was washed twice with 400 µL 2% TFA, and each wash was saved and added to the solution containing the methionine peptides. The peptide samples were cleaned up and desalted by adding to 50 µg C18 resin (The Nest Group), washing with 0.1% TFA, and eluting with 60-70% ACN/0.1% TFA. The ACN was evaporated using a SpeedVac concentrator, and 0.1% TFA was added to bring the total volume to about 150 µL.

5.2.9 Mass Spectrometry Analysis

Samples were analyzed using an Agilent 6520 Q-TOF mass spectrometer system equipped with a Chip Cube Interface. The HPLC Chip used in this experiment contained a 150 mm x 75 µm column with Zorbax 300SB-C18 5 µm packing. The tryptic peptides were eluted using a linear gradient consisting of an increase from 5% to 15% Buffer B in 2.5 min, 15% to 45% Buffer B in 78 min, and then 45% to 100% Buffer B over 10 min. Buffer A was 0.1% FA in water and Buffer B was 0.1% FA in ACN. The flow rate was 0.4 µL/min. The capillary voltage was 1850 V. The drying gas (N₂) was 350°C at a flow rate of 6 L/min. The skimmer and fragmentor were set to 65 V and 175 V, respectively. The inclusion window width for precursor ions was 4 m/z units. The scan rate was 3 scans per second in the mass spectra and 2 scans per second in the product ion mass spectra, and there were 4 precursors selected for fragmentation per cycle.
Multiple LC-MS runs were performed for each set of samples. The exact numbers of runs for each experiment were as follows. In the NAD$^+$ binding experiment, 3 LC-MS runs were performed for each Met peptide enhanced and non-enhanced samples for with and without ligand. In the resveratrol binding experiment, 4 non-enhanced runs and 15 Met peptide enhanced runs were performed for samples in the presence of resveratrol, and 3 non-enhanced runs and 11 Met peptide enhanced runs were performed for samples in the absence of resveratrol. For the Manassantin A and PPI experiments, 1 non-enhanced and 3 Met peptide enhanced runs were performed for with and without ligand.

5.2.10 iTRAQ Data Analysis

Peptides were identified using Agilent’s Spectrum Mill MS Proteomics Workbench software, Rev A03.03.084 SR4. Search parameters included alkylation of cysteines with MMTS as a fixed modification, both the N-terminus and lysine residues labeled with 8plex iTRAQ reagent as a fixed modification, deamidation of glutamines and asparagines as a variable modification, and oxidation of methionines as a variable modification. The precursor and product ion mass tolerances were set to 20 ppm, the protein cleavage chemistry was set for trypsin with 3 maximum missed cleavages, and the maximum ambiguous precursor charge was set to 7. The peptide fragment products and peptide precursor masses were searched against the NCBI nr database for S. cerevisiae. Searches performed on the Met peptide enhanced samples were searched against a database of tryptic Met peptides. Searches were also performed against a decoy
database using the same search parameters to determine false positive rates for the searches.

The Spectrum Mill output containing peptide identifications also contained the extracted iTRAQ intensities, which were exported into a Microsoft Excel format. For each peptide, the iTRAQ intensities were averaged, and the intensity at each denaturant concentration was then divided by that average to give iTRAQ values centered around 1. All of these intensities from peptides that did not contain methionines were then averaged at each denaturant concentration to give a set of normalization values. The peptides used for the set of normalization values did not contain any zero intensity at any denaturant concentration, had a total iTRAQ intensity sum of at least 1000 across all denaturant concentrations, had a Spectrum Mill identification score of at least 8, and did not contain any phenylalanine residues. The iTRAQ intensities of the Met peptides were then divided by this set of normalization values at each corresponding denaturant concentration.

Following normalization of iTRAQ intensity values, the normalized iTRAQ intensity values from Met peptides that appeared in both the with and without ligand samples were combined and compared to determine if any type of interaction was induced as a result of the ligand. A program that was developed in-house was used to match the iTRAQ intensity values for the Met-containing peptides. The two inputs required for this program, one for each with and without ligand, contained the peptide sequence, protein name, identification score, iTRAQ intensity sum, and each of the normalized iTRAQ intensity values. The program took every peptide that appeared more
than once within the same input and averaged the iTRAQ intensity values from those peptides to result in just one set of iTRAQ values for each peptide. Then, if a peptide was present in both with and without ligand inputs, the program matched the two sets of intensity values and output as a text file the list of Met peptides, along with the scores and normalized iTRAQ intensity values, for with and without ligand. The list of matched peptides was then manually investigated to identify curves with transition midpoints that appeared to shift as a result of added ligand. The criteria for midpoint analysis involved choosing an iTRAQ intensity threshold that defined the general transition from pre- to post-transition based on the distributions of the iTRAQ intensities at low and high denaturant. Curves were analyzed for SPROX curve midpoints, and curves that shifted by 2 or more denaturant data points in the presence of ligand were considered potential hits. Curves that had more than one outlier (i.e., a high intensity in the post-transition region or a low intensity in the pre-transition for unoxidized Met peptides) were disregarded.

5.2.11 SPROX on Purified eIF4A

Yeast overexpressing eIF4A (Tif2) was grown and purified using Talon Metal Affinity Resin as described in Section 4.2.1. SPROX was performed by adding 5 µL 9.8 M H$_2$O$_2$ to 45 µL purified eIF4A containing GdmCl (final [GdmCl] after addition of H$_2$O$_2$ was 0.5, 1.0, 1.3, 1.7, 1.9, 2.3, 2.5, 2.8, 3.0, and 3.5 M) in the presence (400 µM) and absence of resveratrol. Oxidation proceeded for 3 min before quenching with 1 mL 300 mM L-methionine, and a TCA precipitation was performed to remove the denaturant. The protein was dissolved in 50 mM ammonium bicarbonate (pH 8.5) with 0.1%
RapiGest and subsequently reduced, alkylated, and digested with trypsin. Peptides were analyzed using both MALDI-TOF/TOF and nanoLC-Q-TOF.

5.3 Results and Discussion

5.3.1 General Strategy

A platform for the thermodynamic analysis of protein-ligand binding on a proteomic scale [74] was improved with the adoption of a Met peptide enhancement strategy, which is portrayed in Figure 20. Previously, a two-dimensional MudPIT strategy [137] with TMT isobaric mass tags [138] was used. The MudPIT strategy allowed the identification and analysis of hundreds of proteins, but 80% of the peptides identified did not contain any methionines. Additionally, each MudPIT run takes approximately 1 day to complete. In an effort to improve the efficiency of this platform, a commercially available Met peptide isolation strategy was employed to enhance the numbers of identified Met peptides. Here, isobaric tags for relative and absolute quantitation (iTRAQ) [139] were used with one-dimensional nanoLC runs (see Figure 20). Following SPROX, the protein samples were subjected to standard proteomic procedures that included a TCA precipitation with subsequent tryptic digestion. After digestion, samples at each denaturant concentration were labeled with a different iTRAQ 8-plex reporter tag. After pooling the iTRAQ-labeled samples together, the samples were subjected to a Met peptide enhancement strategy that uses commercially available immobilized bromoacetyl groups that covalently bind to the thioether group of unoxidized methionine residues at low pH. After washing non-Met peptides away, the
Figure 20. Schematic representation of the protocol utilized in this study. SPROX was performed on yeast lysates, samples were labeled with iTRAQ, and Met peptides were isolated using immobilized bromoacetyl functional groups, enhancing the frequency of identified Met peptides in subsequent LC-MS/MS analyses.
Met peptides were then released under reducing conditions and high pH and analyzed by LC-MS/MS. This strategy was applied to detect protein targets of NAD$^+$, resveratrol, and manassantin A (see Figure 21), as well as protein-protein interactions (PPIs) involved with Pyk1 and Pgk1.

### 5.3.2 Selection Efficiency

Between 60-80% of the identified peptides from all the experiments in the Met peptide enhanced samples were Met peptides, which is a 3-4 times higher frequency than the 13-19% rate of Met peptides found in the non-enhanced samples (see Table 6). This is comparable to the efficiency of Met peptide isolation that has previously been reported [136]. The non-Met peptides found in the enhanced samples were due to non-specific binding to the resin, as the covalent binding to the bromoacetyl functional group is very selective for Met peptides at pH 2-3 [133].

The numbers of peptides and proteins identified with 99% confidence from each experiment are shown in Table 7. The total number of Met peptides identified from the Met peptide enhanced runs was greater than the non-enhanced runs. Even taking into account that more runs were generally performed with the Met peptide enhanced samples, the numbers of Met peptides identified in a direct run-to-run comparison was at least two times higher, as can be observed from the NAD$^+$ binding experiment where equal numbers of Met peptide enhanced and non-enhanced runs were performed. The numbers of peptides and proteins that were actually assayed for binding is also shown in Table 7. These were the numbers of peptides that were identified in both with and without ligand samples. Only peptides above a total iTRAQ intensity sum threshold
Figure 21. Three of the ligands investigated in this study. A. NAD$^+$  B. Resveratrol  C. Manassantin A
### Table 6. Met peptide selection efficiency

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Fraction of Met peptides after Met peptide enhancement&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Fraction of Met peptides before Met peptide enhancement&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAD&lt;sup&gt;+&lt;/sup&gt;</td>
<td>86%</td>
<td>18%</td>
</tr>
<tr>
<td>Resveratrol</td>
<td>59%</td>
<td>13%</td>
</tr>
<tr>
<td>Manassantin A</td>
<td>63%</td>
<td>13%</td>
</tr>
<tr>
<td>Pgk1</td>
<td>59%</td>
<td>17%</td>
</tr>
<tr>
<td>Pyk1</td>
<td>66%</td>
<td>19%</td>
</tr>
</tbody>
</table>

<sup>a</sup>Fraction of methionine-containing peptides identified from all LC-MS runs (i.e., +/-ligand)

### Table 7. Assay statistics

<table>
<thead>
<tr>
<th>Experiment</th>
<th># Met peptides (proteins) in enhanced samples&lt;sup&gt;a&lt;/sup&gt;</th>
<th># Met peptides (proteins) in non-enhanced samples&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Number of peptides (proteins) assayed for binding&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAD&lt;sup&gt;+&lt;/sup&gt;</td>
<td>214 (124)</td>
<td>103 (70)</td>
<td>238 (126)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Resveratrol</td>
<td>331 (184)</td>
<td>67 (42)</td>
<td>410 (243)&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Manassantin A</td>
<td>451 (261)</td>
<td>51 (37)</td>
<td>178 (100)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Pgk1</td>
<td>108 (70)</td>
<td>39 (24)</td>
<td>189 (96)&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Pyk1</td>
<td>144 (89)</td>
<td>45 (26)</td>
<td>191 (94)&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>Number of peptides identified in all LC-MS runs with ≥ 99% confidence

<sup>b</sup>Number reflects total number of oxidized and unoxidized Met peptides that appear in both +/- ligand with Spectrum Mill identification scores ≥ 3.0

<sup>c</sup>Total iTRAQ intensity sum cutoff of 1000 used

<sup>d</sup>Total iTRAQ intensity sum cutoff of 875 used
made it through to the matching between with and without ligand, which varied between 875 and 1000, depending on the experiment. The purpose of the iTRAQ intensity threshold was to only include higher quality iTRAQ data. For each experiment, all peptides above a Spectrum Mill identification score of 3.0 were assayed for binding to respective ligands. The score is based on the mass accuracy of the precursor and product ions, in addition to fragmentation efficiency. A higher score demonstrates higher confidence in the identification.

One advantage of the SPROX analysis is that lower identification scores could be used when assaying for binding hits. This is because the only Met peptides of interest were the peptides whose SPROX midpoint shifted in the presence of ligand. If a certain hit had two low identification scores (both with and without ligand), then further investigation could be performed to verify its identity. However, based on the shape of the curves and because of the Met enhancement protocol, the chance that the peptide was a Met peptide was nearly 100%, while only 1 in 5 peptides in non-enhanced samples are Met peptides. Additionally, the retention times for with and without ligand could be checked to confirm similar retention times. So the fact that the hypothetical low score peptide hit was identified as a Met peptide in both with and without ligand samples with the same retention time added more confidence to the identification.

The effectiveness of the Met peptide enhancement was investigated by performing equal numbers of LC-MS runs of the enhanced and non-enhanced samples in the NAD$^+$ binding experiment. Shown in Figure 22 are the numbers of peptides from each set of samples and the number of peptides that overlap between the Met peptide
enhanced and non-enhanced samples. As can be clearly seen, the Met peptide enhancement essentially doubled the number of Met peptides identified with 99% confidence when compared to the non-enhanced samples. The number of identified Met peptides that were identified in both runs was only 34% of the Met peptide enhanced samples but 70% of the non-enhanced samples, which means that 2/3 of the Met peptides identified from the enhanced samples were not identified in the non-enhanced samples. For this reason, all following experiments utilized many more LC-MS runs from the Met peptide enhanced samples to more efficiently target Met peptides.

In the resveratrol binding and PPI studies, the 118 iTRAQ reporter ion intensity was discarded because these experiments suffered from an atmospheric contaminant at m/z 118.1221, which overlapped with the iTRAQ 118 reporter ion. As a result, the intensity value from this tag was not used in the normalizations or the curve analyses for these two experiments.

The distributions of the iTRAQ intensities at high and low denaturants for the unoxidized Met peptides are shown in Figure 23. Because the SPROX experiment probed the stability of proteins using the disappearance of unoxidized methionines with increasing denaturant for the majority of the peptides that were assayed (approximately 80%), the SPROX curves generally decreased in intensity from low to high denaturant. The other 20% of the peptides were oxidized Met peptides that predominantly came from the non-enhanced samples. As expected, most of the normalized iTRAQ intensities at low denaturant were distributed from 1.4 to 1.8, while the higher denaturant intensities were centered around 0.4 to 0.8. The only exception was in the NAD$^+$ binding
Figure 22. A. Numbers of peptides identified with $\geq 99\%$ confidence in the NAD$^+$ binding assay. B. Number of proteins from the peptides in (A).
Figure 23. Distributions of iTRAQ intensities at low (dashed line) and high (solid line) denaturant for (A) NAD$^+$ binding, (B) manassantin A binding, (C) resveratrol, (D) and protein-protein interaction experiments.
experiment (Figure 23A) where the magnitude between the peaks of the distributions was not quite as great. It is unclear why the magnitude was not as large; however, in all cases, the distributions intersect close to a normalized iTRAQ intensity of 1.0, which was used as the intensity threshold for pre- and post-transitions.

5.3.3 NAD$^+$ Binding Hits

The iTRAQ intensities were manually investigated for ligand binding to NAD$^+$ using a midpoint analysis of the SPROX transitions. All peptides that contained a total iTRAQ intensity sum of 1000 or greater in both sets of data (i.e., with and without NAD$^+$) were matched against the other set for identical peptides. No score filter was applied in this step because peptides could be filtered out according to scores after identifying the peptides that appeared in both with and without ligand samples. If a peptide with a high score in one set was matched to the same peptide in the corresponding set of data with a low score, other determining factors such as retention time and product ion spectra could be used to establish whether they were in fact the same peptide. After matching peptides from with and without NAD$^+$, a total of 238 peptides from 126 proteins were assayed (see Table 7). The criteria for selecting hits involved determination of the transition in each curve with and without NAD$^+$. Transition midpoints that shifted by 1-2 M urea (2 denaturant data points) as a result of added NAD$^+$ were deemed significant. The percentage of “perfect” SPROX curves, i.e., curves that contained no outliers as described in Section 5.2.10, was 28%. Shown in Figure 24A are representative SPROX iTRAQ intensities for a peptide that does not display ligand binding, as the midpoints of the SPROX curve are both found at
approximately 2.3 M urea. Figure 24B shows an example of a hit peptide, which belongs to NAD\(^+\)-dependent glutamate dehydrogenase, because the midpoint shifted from approximately 2 M to 3M with the addition of NAD\(^+\). In all, 10 peptides from 9 proteins were selected as potential hits with regards to NAD\(^+\) binding (see Table 8), and 5 of the 9 proteins are known NAD\(^+\) binders.

The five hits that have known binding interactions with NAD\(^+\) were glyceraldehyde-3-phosphate dehydrogenase (GAPDH), glutamate dehydrogenase, inosine monophosphate dehydrogenase, 2-oxoglutarate dehydrogenase, and alcohol dehydrogenase. Additionally, one hit was an NADP\(^+\)-dependent enzyme (isocitrate dehydrogenase). Porcine NADP\(^+\)-dependent isocitrate dehydrogenase has previously been shown to bind with cofactor analogues that do not contain a 2’ phosphate; however, the catalytic efficiency was greatly reduced [140]. Also, porcine NADP\(^+\)-dependent isocitrate dehydrogenase was specifically shown to bind NAD\(^+\), although the catalytic efficiency was reduced to 2.4%-12% depending on pH [141]. Additionally, NAD\(^+\) was used to study the stereochemistry of the nucleotide binding site in *E. coli* NADP\(^+\)-dependent isocitrate dehydrogenase [142]. This suggests that NADP\(^+\)-dependent isocitrate dehydrogenase has an affinity for NAD\(^+\); however, it is evident that the enzyme catalysis is greatly reduced using NAD\(^+\) as a cofactor. Inosine monophosphate dehydrogenase and alcohol dehydrogenase had a midpoint shift to the left towards lower denaturant concentration with added NAD\(^+\), suggesting destabilization. However, these proteins are multi-domain proteins, and dehydrogenases are known to undergo large conformational changes upon NAD\(^+\) binding [143, 144]. Likely, the domains these
peptides originated from were not located near the NAD\(^+\) binding site and were destabilized. No crystal structure exists for inosine monophosphate dehydrogenase, but the peptide from alcohol dehydrogenase is located in a loop region distant from the NAD\(^+\) binding site (PDB 2HCY). The remaining 4 hits have no known interactions with NAD\(^+\). Because the cofactor NAD\(^+\) and NAD\(^+\) binding sites within proteins are well characterized, this suggests that the false positive rate for this experiment was approximately 40%.

An additional survey of the data revealed one additional NAD\(^+\)-dependent enzyme that was not selected as a hit based on the criteria used from the midpoint analysis. This protein, mitochondrial aldehyde dehydrogenase, did not have a significant midpoint shift in the presence of NAD\(^+\). Six other proteins known to bind NAD\(^+\) had one or more curves that were uninterpretable, and so it was impossible to determine anything regarding thermodynamic stability changes as a result of NAD\(^+\) for those proteins. The presence of one protein that distinctly displayed non-binding behavior suggested that the false negative rate was around 14%. In addition to the possibility that this hit was just missed, two additional reasons it was not selected as a hit include the possibility that this protein’s affinity for NAD\(^+\) was weaker than the detection limit of the assay, or the Met peptide came from a region of the protein that was not affected by NAD\(^+\) binding. Assuming an \(m\)-value of 1 kcal mol\(^{-1}\) M\(^{-1}\), one independent binding site, and a midpoint shift of 1.5 M urea, the lowest affinity \(K_d\) that would have been expected to be detected was approximately 80 \(\mu\)M. The false positive and false negative rates attained with
Figure 24. A. Representative SPROX curve from the peptide IADTVGKDVVVQLYQEQLSAQGMPMIK from glucokinase-1. B. One of the selected peptides that displayed NAD$^+$ binding, GDIESISDKMYK from NAD$^+$-dependent glutamate dehydrogenase. The first iTRAQ intensity (0 M GdmCl) in the curve in the presence of NAD$^+$ is an example of an overlooked outlier intensity.
Table 8. List of peptides selected as hits from NAD$^+$ binding SPROX analysis

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Protein</th>
<th>Midpoint shift (M urea)$^a$</th>
<th>Known interaction?$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>VINDAFGIEEGLM(ox)TTVHSLTATQK</td>
<td>glyceraldehyde-3-phosphate dehydrogenase NAD$^+$-dependent glutamate dehydrogenase</td>
<td>1.0</td>
<td>yes</td>
</tr>
<tr>
<td>GDIESISDKTMYK</td>
<td></td>
<td>1.0</td>
<td>yes</td>
</tr>
<tr>
<td>GGNIPMPGWMDFPTGK</td>
<td>hexokinase 2</td>
<td>-1.2</td>
<td>no</td>
</tr>
<tr>
<td>LLIVDDNGNLVSMLSR</td>
<td>inosine monophosphate dehydrogenase</td>
<td>-1.5</td>
<td>yes</td>
</tr>
<tr>
<td>GVMNAVNNVNVNVIAAAFVK</td>
<td>enolase 2</td>
<td>-1.7</td>
<td>no</td>
</tr>
<tr>
<td>NIPMMPAGEPVLEAIFEVDANGILK</td>
<td>heat shock protein SSB1</td>
<td>1.3</td>
<td>no</td>
</tr>
<tr>
<td>TFSPQEISAMVLTK</td>
<td>heat shock protein SSB1</td>
<td>-1.3</td>
<td>no</td>
</tr>
<tr>
<td>LIDDMVAQMIK</td>
<td>NADP$^+$-specific isocitrate dehydrogenase</td>
<td>1.1</td>
<td>no$^c$</td>
</tr>
<tr>
<td>SVELGVEDIVLMGMAHR</td>
<td>2-oxoglutarate dehydrogenase</td>
<td>1.3</td>
<td>yes</td>
</tr>
<tr>
<td>LPLVGGHEAGVVVMGGENVK</td>
<td>alcohol dehydrogenase 1</td>
<td>-1.5</td>
<td>yes</td>
</tr>
</tbody>
</table>

$^a$Negative values mean that the midpoint was shifted towards lower [urea] in the presence of NAD$^+$
$^b$Refers to known instances of NAD$^+$ binding in yeast proteins
$^c$Has been shown in other species to bind NAD$^+$ (e.g., Ref. 141)
SPROX were similar in magnitude to other global scale protein-protein interaction studies that utilize yeast two-hybrid screens or TAP-MS, where the false positive rates are either unknown [8, 9, 25] or range from 25-40% [27, 31, 145], and false negative rates are around 20% [26].

5.3.4 Resveratrol Binding Hits

Resveratrol is a stilbenoid found in red grapes that has been associated with lifespan extension and other health benefits (see Figure 21). Previously, Lomenick et al. [146] identified eukaryotic initiation factor 4A (eIF4A) as a protein target of resveratrol (3,5,4’-trihydroxy-trans-stilbene) using a new technique termed drug affinity responsive target stability (DARTS), a protease protection approach for the identification of protein targets of drugs. A previous study by Smith et al. [147] also demonstrated that deletion of the open reading frame (ORF) for eIF4A extended the lifespan of yeast. However, separate studies have identified the NAD\(^+\)-dependent deacetylase Sir2 as the target for the lifespan extending effects of resveratrol [148, 149], but these results remain controversial [150, 151]. Other targets have been postulated, but a recent review suggested that the direct targets of resveratrol still remain uncertain [152]. In an effort to shed light on resveratrol’s mechanism of action, SPROX was performed in yeast lysate in the presence and absence of resveratrol.

The data was investigated identically as that described in the Section 5.3.3. In all, 410 peptides from 243 proteins were assayed for binding to resveratrol (see Table 7). An increased number of peptides and proteins assayed was due in part to an increased number of LC-MS runs. In total, 45% of the curves were considered perfect curves. This
increase in high-quality curves may have also been due to an increase in the number of LC-MS runs performed. The more a certain peptide was sequenced, the more the iTRAQ reporter ion intensities could be averaged in the construction of the SPROX curves. This reduced the random “noise” of the reporter ion intensities and resulted in clearer SPROX curves for peptides that were sequenced more often.

Because eIF4A was a protein of interest based on prior studies [146, 147], the behavior of this protein was first investigated. Four Met peptides from eIF4A were identified and assayed. Shown in Figure 25 are the SPROX curves from these peptides. All the midpoints were around 1.5 M GdmCl, and no significant change in the transition midpoint was observed for any of these peptides. The helicase eIF4A is a 394-residue dumbbell-shaped protein with two distinct domains separated by an 11-residue linker region [153]. Three of the peptides (eIF4A(24-35) FDDMELDENLLR, (52-72) AIMPIIEGHDVLAQAQSGTGK, and (90-100) APQALMLAPTR) came from the N-terminal domain, while eIF4A(357-373) KGVAINFVTNEDVGAMR is located near the C-terminus in the C-terminal domain. Because all of the SPROX midpoints were nearly the same from the iTRAQ data, this suggests that the stability of the two domains of eIF4A were approximately equal (see Figure 25).

A separate SPROX experiment performed on purified eIF4A confirmed that eIF4A does not directly bind to resveratrol (see Figure 26). Both a MALDI-TOF/TOF and ESI-Q-TOF were used to collect this data, and SPROX curves looked identical between the two different MS readouts. The SPROX curves for with and without resveratrol did not appear significantly different within experimental error, although the
curves from the N-terminal peptides in the absence of resveratrol consistently had slightly lower pre-transitions compared to those same peptides in the presence of resveratrol. This had the effect of shifting the \( C_{SPROX}^{1/2} \) value to the left in the presence of resveratrol. Two additional Met peptides from the N-terminal domain were also identified and exhibited similar behavior (data not shown). However, this effect is most likely due to experimental uncertainty. If the shifted \( C_{SPROX}^{1/2} \) values were a genuine experimental artifact, this would suggest that resveratrol weakly bound the unfolded N-terminal domain. It should be pointed out that the experiments on the purified eIF4A still contained a C-terminal 19 kDa purification tag, which could possibly affect the stability of the protein or a potential binding interaction. However, taken together with the SPROX/iTRAQ study, these results suggest that eIF4A is not a direct target of resveratrol. Using an \( m \)-value ranging from 2.5-5 kcal mol\(^{-1} \) M\(^{-1} \) based on the data from the purified eIF4A SPROX experiment and the size of the domains [56], a \( K_d \) for eIF4A-resveratrol of 10-50 \( \mu \)M or weaker would go undetected in the iTRAQ experiment. It should be noted that the direct binding of resveratrol to eIF4A using DARTS was not conclusively established. Although eIF4A-initiated translation did appear to be inhibited in the presence of resveratrol, the authors themselves pointed out that the inhibition effect could be the result of resveratrol binding to ribosomal proteins [146]. Based on the SPROX results obtained here, however, there is no evidence for direct binding of resveratrol to eIF4A, which suggests that the previously observed eIF4A-initiated translation inhibition was the result of another protein target of resveratrol. Additional proteins of interest from previous studies were proteins belonging to the sirtuin family.
Neither Sir2 nor any other sirtuins made it through the iTRAQ intensity filters for the resveratrol binding assay (i.e., they had low iTRAQ intensities), but Sir1 and Sir3 were initially identified in product ion spectra from the Spectrum Mill output.

Based on the SPROX midpoint analysis, however, four potential protein targets of resveratrol were identified (see Table 9 and Figure 27). These included elongation factor 3A, ribosomal proteins S9-A and L8-A, and cytosolic aldehyde dehydrogenase, and shown in Figure 27 are the SPROX curves from those proteins. Interestingly, many of these proteins are associated with eIF4A, either directly or indirectly. Elongation factor 3A interacts with eIF4A [9], and cytosolic aldehyde dehydrogenase has been shown to interact with elongation factor 2 [154], which interacts with eIF4A [8]. However, even more significantly, a previous study also reported on the interaction between cytosolic aldehyde dehydrogenase and resveratrol through activity studies [155], which proposed that resveratrol binds to the aldehyde substrate binding pocket of aldehyde dehydrogenase [155]. It was shown that resveratrol increased the activity of aldehyde dehydrogenase at high concentrations of substrate but inhibited it at low concentrations of substrate. From the iTRAQ data, the $K_d$ for the interaction between aldehyde dehydrogenase and resveratrol was approximately 10 µM using an estimated $m$-value of 2 kcal mol$^{-1}$ M$^{-1}$. Two ribosomal proteins, 40S ribosomal protein S9-A and 60S ribosomal protein L8-A, were also selected as hits, both of which indirectly associate with eIF4A. Many ribosomal proteins were previously identified as additional targets of resveratrol-mediated effects using DARTS [146], but the two ribosomal proteins identified in this study were not the same as had previously been identified. However, if resveratrol
Figure 25. Methionine-containing peptides from eIF4A in the SPROX iTRAQ resveratrol-binding experiment. Arrows denote transition midpoints. A. (52-72) AIMPIIEGHDVLAQAQSGTGK  B. (90-100) APQALMLAPTR  C. (24-35) FDDMELDENLLR  D. (357-373) KGVAINFVTNEDVGAMR. All peptides display similar midpoint transitions with and without resveratrol.
Figure 26. SPROX curves from purified eIF4A in presence (○) and absence (●) of resveratrol. A-D correspond to the same peptides found in Figure 25A-D. The N-terminal peptides found in A-C exhibited a slight change in SPROX behavior in the presence of resveratrol.
Figure 27. Potential hits of resveratrol. A. Cytosolic aldehyde dehydrogenase (QQFDTIMNYIDIGK) B. Elongation factor 3A (MPELIPVLSETMWDTK) C. 60S ribosomal protein L8-A (MGVPIAYIVK) D. 40S ribosomal protein S9-A (QIVNIPSFMVR).
<table>
<thead>
<tr>
<th>Peptide</th>
<th>Protein</th>
<th>Midpoint shift (M GdmCl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QQFDTIMNYIDIGK</td>
<td>cytosolic aldehyde dehydrogenase</td>
<td>1.1</td>
</tr>
<tr>
<td>MPELIPVLSETMWDTK</td>
<td>elongation factor 3A</td>
<td>0.7</td>
</tr>
<tr>
<td>MGVPYAIVK</td>
<td>60S ribosomal protein L8-A</td>
<td>1.1</td>
</tr>
<tr>
<td>QIVNIPSFMVR</td>
<td>40S ribosomal protein S9-A</td>
<td>0.7</td>
</tr>
</tbody>
</table>
inhibits the protein translation machinery as demonstrated previously [156], it is not surprising that the stability of additional ribosomal proteins and elongation factors may be affected by resveratrol. These results add to the existing evidence that resveratrol interacts with proteins involved with the translation machinery.

5.3.5 Manassantin A Binding Hits

Manassantin A is a compound found in herbal medicines that has been shown to be an inhibitor of the activity of HIF-1 [157], TNF-α [158], NF-κB [159], and sterol O-acyltransferase (ACAT) [160]. Additionally, it has also shown anti-cancer properties [161, 162]. However, specific protein targets of manassantin A have not been identified. The structure of manassantin A is shown in Figure 21. Here, a S. cerevisiae lysate was assayed for binding to manassantin A to identify yeast protein targets of this small molecule drug.

A total of 178 Met peptides from 100 different proteins were assayed for binding. Forty-four of the 178 Met peptides were the oxidized and unoxidized pairs of 22 Met peptides. Out of the 356 curves from with and without manassantin A, 34% were considered “perfect”, and only 11% were considered uninterpretable.

After manual investigation and midpoint analysis of the data, no protein targets were identified from this data. However, only 100 proteins in total were assayed from the entire yeast lysate. While this is more than would have been assayed if no Met peptide enhancement protocol were used, this is still a relatively small fraction of the yeast proteome. It is hypothesized that the protein target of manassantin A is a
mitochondrial protein\textsuperscript{1}, so isolating the mitochondrial fraction and performing SPROX on it would possibly increase the likelihood that a protein target could be found.

5.3.6 Protein-Protein Interactions Hits

Methods for the detection of PPIs on a global scale have been reported [9, 25, 154, 163], but a technique for quantitating PPIs on a global scale is lacking. SPROX is unique in its ability to measure protein-ligand interactions with many classes of ligands, including other proteins. To demonstrate its ability in measuring PPIs in complex mixtures, two model systems were used, which included Pgk1 and Pyk1. Pgk1 and Pyk1 are common enzymes involved in glycolysis. These two proteins were overexpressed and purified from yeast, and then spiked in excess into yeast lysates. SPROX was then performed on these systems to identify proteins that interact with Pgk1 and Pyk1 through changes in their thermodynamic stability.

5.3.6.1 Pgk1 Hits

A total of 189 Met peptides that came from 96 different proteins were assayed for binding to Pgk1 (see Table 7). Out of the 378 curves from the 189 Met peptides, 16\% curves were deemed uninterpretable because of the random scattering of the iTRAQ intensities, and 34\% were considered “perfect”. A total of 9 proteins from the 96 proteins assayed showed evidence for interaction with Pgk1 (see Table 10 and Figure 28). One of the protein hits was Pgk1, as this protein interacts with itself [164]. Two of the 9 proteins, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and pyruvate

\textsuperscript{1} Communication with Professor Jiyong Hong
decarboxylase (Pdc1), have previously been reported in the literature to interact with Pgk1 specifically in yeast from tandem affinity purification (TAP) experiments [8, 9, 154], and these interactions have been characterized in other species as well. Three different Met peptides from GAPDH indicated a stability increase in the presence of Pgk1. One peptide from pyruvate decarboxylase (Pdc1) showed a clear shift in the midpoint transition (0.7 M), while 2 peptides displayed what appeared to be a small shift (0.4 M). Additionally 2 peptides from Pdc1 appeared to show a SPROX transition in the absence of added Pgk1 and then no transition with added Pgk1. The six other proteins that were detected to interact with Pgk1 are shown in Table 10. These include elongation factors 1α and 3A, thiol-specific peroxiredoxin, heat shock protein (Hsp) SSA2, ubiquitin and adenylate kinase 1. Relative affinities for Pgk1 from low to high ranged from peroxiredoxin (0.5 M shift, $K_d \sim 900$ nM) to translational elongation factor EF-1α (1.0 M shift, $K_d \sim 130$ nM).

5.3.6.2 Pyk1 Hits

A total of 191 Met peptides that came from 94 different proteins were assayed for binding to Pyk1. Out of the 382 curves from the 191 Met peptides, 11% of the curves were deemed uninterpretable because of the random scattering of the iTRAQ intensities, and 45% were considered “perfect”. A total of 7 proteins from the 94 proteins assayed showed evidence for interaction with Pyk1 (see Table 10), including Pyk1 itself. Two of the proteins, GAPDH and alcohol dehydrogenase, are enzymes also involved with glycolysis, so it isn’t surprising that Pyk1 would interact with them on some level. In fact, pyruvate kinase and glyceraldehyde-3-phosphate dehydrogenase from Mycoplasma
Table 10. Proteins identified as targets of the proteins studied in this work

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Protein</th>
<th>Ligand</th>
<th>Midpoint Shift (M)</th>
<th>Previously Reported Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>MVLGPPGAGK</td>
<td>Adenylate kinase</td>
<td>Pgk1</td>
<td>0.7</td>
<td>no</td>
</tr>
<tr>
<td>MKETAESYLGAK</td>
<td>Hsp SSA2</td>
<td>Pgk1</td>
<td>0.7</td>
<td>no</td>
</tr>
<tr>
<td>MVAEVDMK</td>
<td>Elongation factor 3A</td>
<td>Pgk1</td>
<td>0.7</td>
<td>no</td>
</tr>
<tr>
<td>LTGMAFR</td>
<td>GAPDH</td>
<td>Pgk1</td>
<td>0.7</td>
<td>yes b</td>
</tr>
<tr>
<td>VINDAFGIEEGLM(ox)TTVHSLTATQK</td>
<td>GAPDH</td>
<td>Pgk1</td>
<td>0.7</td>
<td>yes b</td>
</tr>
<tr>
<td>VVITAPSSTAPMFVMGVNEEK</td>
<td>GAPDH</td>
<td>Pgk1</td>
<td>1.0</td>
<td>yes b</td>
</tr>
<tr>
<td>IYVEGMR</td>
<td>Pyruvate decarboxylase</td>
<td>Pgk1</td>
<td>0.7</td>
<td>yes b</td>
</tr>
<tr>
<td>VADKIQLIDNLDKVDIIIIGGGM(ox)A FTFKK</td>
<td>Pgk1</td>
<td>Pgk1</td>
<td>0.8</td>
<td>yes c</td>
</tr>
<tr>
<td>MPQTVEWSK</td>
<td>Thiol-specific peroxiredoxin</td>
<td>Pgk1</td>
<td>0.5</td>
<td>no</td>
</tr>
<tr>
<td>VETGVIKGMVTVFAPAGVTEVK</td>
<td>Participation factor EF-1a</td>
<td>Pgk1</td>
<td>1.0</td>
<td>no</td>
</tr>
<tr>
<td>MQIFVK</td>
<td>Ubiquitin</td>
<td>Pgk1</td>
<td>0.7</td>
<td>yes d</td>
</tr>
<tr>
<td>NMEIWK</td>
<td>Acetohydroxyacid reductoisomerase</td>
<td>Pyk1</td>
<td>0.7</td>
<td>no</td>
</tr>
<tr>
<td>VTDGVMVAR</td>
<td>Alcohol</td>
<td>Pyk1</td>
<td>0.5</td>
<td>yes</td>
</tr>
<tr>
<td>ANGTTLVLYMPAGAK</td>
<td>Dehydrogenase 1</td>
<td>Pyk1</td>
<td>0.7</td>
<td>no</td>
</tr>
<tr>
<td>MVAEVDMK</td>
<td>Elongation factor 3A</td>
<td>Pyk1</td>
<td>1.0</td>
<td>no</td>
</tr>
<tr>
<td>AAATAAMTK</td>
<td>Elongation factor 3A</td>
<td>Pyk1</td>
<td>0.7</td>
<td>no</td>
</tr>
<tr>
<td>LTGMAFR</td>
<td>GAPDH</td>
<td>Pyk1</td>
<td>0.5</td>
<td>no</td>
</tr>
<tr>
<td>TWETYEMR</td>
<td>40S ribosomal protein S20</td>
<td>Pyk1</td>
<td>0.7</td>
<td>no</td>
</tr>
<tr>
<td>MGVPYAIVK</td>
<td>60S ribosomal protein L8-A</td>
<td>Pyk1</td>
<td>0.7</td>
<td>no</td>
</tr>
</tbody>
</table>

a Previously reported interaction between yeast proteins
b Ref. [8]
c Ref. [164]
d Ref. [165, 166]
*pneumoniae* were previously shown to form a protein complex using TAP-MS [167]. The other four interacting proteins were acetohydroxyacid reductoisomerase, elongation factor 3A, 60S ribosomal protein L8-A, and 40S ribosomal protein S20. Most of the Pyk1 interactors displayed relatively the same affinity for Pyk1, with a shift of approximately 0.7 M (corresponding to a $K_d$ of roughly 100 nM). Exceptions were Pyk1, which showed the least affinity for itself ($K_d \sim 2 \, \mu M$), and elongation factor 3A, which displayed the tightest binding to Pyk1 ($K_d \sim 30 \, nM$).

**5.3.6.3 Analysis of Protein Hits**

Pgk1 is a highly abundant glycolytic enzyme and has previously exhibited more than just its glycolytic functions within the cell. Prior studies have shown that mammalian PGK and lactate dehydrogenase are present in the nuclei and exert non-glycolytic functions [168, 169]. PGK has been identified as a subunit of primer recognition proteins [169, 170], it was found to stimulate mRNA transcription at the elongation step [171], and it was discovered that PGK bound mRNA for post-transcriptional regulation [172]. Additionally, PGK and translation elongation factor 1a from *Nepenthes paniculata* were both recruited to bind the 3′-UTR of Bamboo mosaic virus RNA [173]. In this light then, the interaction of Pgk1 and elongation factors 1a and 3A is not unexpected. The SPROX results add to the evidence that PGK may have the ability to regulate other cellular processes in addition to glycolysis. In fact, many other glycolytic enzymes have various non-glycolytic functions as well [174]. GAPDH, for example, is another common glycolytic enzyme involved with cellular processes such as transcription [175] and apoptosis [176, 177].
Figure 28. Representative SPROX curves from detected PPIs. A. Pyruvate decarboxylase with Pgk1. B. Glyceraldehyde-3-phosphate dehydrogenase with Pgk1. C. Acetohydroxyacid reductoisomerase with Pyk1. D. A peptide from Pyk1 in the presence of excess Pyk1.
Pyk1 functions as a homotetramer, and so the presence of excess Pyk1 should have increased the stability of Pyk1 [72]. Indeed, one peptide from Pyk1 (VTDGVMVAR) showed an increased transition midpoint by 0.5 M GdmCl. However, it should be noted that both identification scores for this peptide were lower than the 99% confidence limit, and were closer to an 84% confidence level. Again, it should be emphasized that these lower identification scores have much more confidence in the SPROX binding assays based on the similar retention times, precursor $m/z$ and charge state, the shape of the curves, and the fact that it was independently identified by Spectrum Mill in both cases. Because of this, the peptide from Pyk1 was considered a confident hit.

A comparison of all the proteins that were assayed in the PPI study to the Saccharomyces Genome Database [178] allowed the identification of proteins that might have also showed up as hits based on affinity-MS experiments. For example, there were three additional proteins that have been shown to physically interact with Pgk1 that were also assayed, yet they were not identified as hits in the SPROX assay. Alcohol dehydrogenase and fructose 1,6-bisphosphate aldolase had multiple Met peptides identified and assayed, but none of them showed evidence of interaction with Pgk1. The third protein, enolase II, had 8 different Met peptides assayed (4 unique peptides with both oxidized and unoxidized forms), but unfortunately most of the peptide curves from the sample with excess Pgk1 were uninterpretable. As the curves in the absence of Pgk1 all showed cleared transitions, it is unclear why the curves in the presence of Pgk1 would have curves with indiscernible midpoints. One peptide from enolase II did have two
high-quality curves from with and without excess Pgk1, but there was not a significant midpoint shift to consider it a hit.

One caveat in this PPI study, like other common types of interaction studies, is that these detected interactions were made in vitro and may not actually reflect the associations that normally occur in the cell in vivo where proteins are much more compartmentalized.

5.4 Conclusions

The global protein profiling SPROX platform was improved with the incorporation of a Met peptide enhancement strategy. The enhancement strategy increased the numbers of Met peptides that were identified and assayed by at least two-fold, and this assay was used for the detection of protein-ligand interactions to five different ligands, including NAD$^+$, manassantin A, resveratrol, Pgk1, and Pyk1. The increase of Met peptides therefore expands the scope of the SPROX platform.

A few model systems were used to characterize the SPROX platform, which included the cofactor NAD$^+$ and the proteins used in the PPI study. Multiple known NAD$^+$-binding proteins were identified. A false positive and false negative rate from the NAD$^+$ binding study were assessed to be approximately 40% and 14%, respectively. Additionally, many known PPIs were detected to the proteins Pgk1 and Pyk1. Common targets of the two proteins were glycolytic enzymes and elongation factors, which are proteins that would be expected to interact with Pgk1 and Pyk1 based on known biological functions.
The investigation of binding to manassantin A and resveratrol, biologically active compounds with less well-understood targets, was also performed. No targets to manassantin A were identified, and 4 targets to resveratrol were identified. One of the resveratrol hits was cytosolic aldehyde dehydrogenase, which has previously been identified as a direct target of resveratrol. The other three proteins were all involved in the protein translation machinery, which has also been implicated as a target of resveratrol. Moreover, as part of the translation machinery, they were also associated, directly or indirectly, with eIF4A, which has previously been identified as a protein of interest. Additionally, the SPROX/iTRAQ platform did not detect any interaction between eIF4A and resveratrol, and at the most eIF4A had a low affinity for resveratrol (i.e., $K_{d} < 10^{-50}$ µM).

The Met peptide enhancement strategy increased the number of peptides assayed in a SPROX-based platform for the proteomic analysis of ligand binding. However, this is still a small fraction of the total proteome. Further advantages of the Met peptide enhancement strategy should be seen with the incorporation of two-dimensional separation (e.g., MudPIT or an additional isoelectric focusing step prior to reversed phase chromatography) or SPROX on more targeted systems (e.g., mitochondrial fractions or ribosomal fractions). Overall, this should be seen as a useful discovery tool. Additional steps to validate hits from these types of experiments are also needed, and work is currently in progress to accomplish this.
6. Conclusions and Future Directions

This dissertation project was largely focused on developing methodologies for measuring protein-ligand interactions in two types of multiplexed assays. The first involved the analysis of multiple ligands binding to a target protein, and the second involved the analysis of multiple proteins binding to a specific ligand of interest. In the case of the first assay, this work demonstrated the ability to screen moderately sized (1,000-10,000 member) chemical libraries at a rate of 6 s/compound. This rate most likely represents the upper limit to the throughput of this SUPREX-based assay, which is comparable to many HTS strategies. An advantage of the single-point SUPREX assay is that hits were determined based on stability increases due to ligand binding and not changes in enzyme activity, making this a general screening strategy for most proteins of interest. This assay should have broad application for the identification of potential therapeutic agents for a number of proteins.

The second type of multiplexed analysis, involving one ligand and many potential protein targets, was described in terms of three studies. The first, involving purification of a protein of interest following SPROX, would likely be used in instances where only the thermodynamic analysis of one (or a few) protein(s) of interest is desired. However, for drug mode-of-action studies, where the analysis of as many proteins as possible is advantageous, PrSUIT or SPROX combined with isobaric mass tags and Met peptide enhancement would be ideal. These two approaches enabled the analysis of dozens to hundreds of proteins for the evaluation of ligand binding. However, this likely does not represent the upper limit to the peptide and protein coverage possible in the analysis of
complex biological mixtures. This work primarily involved the use of a one-dimensional reversed phase chromatographic separation for peptide identification. However, an additional order of magnitude of the numbers of proteins assayed will likely be obtained through the use of additional fractionation and/or additional chemical modification strategies. Two dimensional chromatographic approaches (e.g., reversed phase combined with strong-cation exchange [137] or isoelectric focusing [179, 180]) routinely allows for the identification of thousands of peptides and proteins in proteomics. Additionally, the incorporation of additional chemical modification strategies, such as lysine amidination [181], would expand the scope of this assay even further by increasing the numbers of peptides that would be useful for thermodynamic measurements. Through a combination of these methods, the thermodynamic analysis of 1,000 proteins or more for binding to ligands of interest should be possible.
References


Biography

Patrick DeArmond was born June 15, 1983 in Decatur, Indiana. He graduated summa cum laude from Indiana Wesleyan University in Marion, Indiana in April 2005 with his Bachelor of Science degree, where he studied in chemistry. While at Indiana Wesleyan University, he performed undergraduate research under the supervision of Professor John R. Lakanen. Patrick matriculated into the graduate school at Duke University in the Chemistry Department in August 2006 where he performed research under the supervision of Professor Michael C. Fitzgerald. While in graduate school, he was awarded the William Krigbaum/Hobbs Chemistry Fellowship, and he contributed to designing and teaching a chemistry seminar course entitled “A Scientist’s Food for Thought: Following Food from the Field to the Table and Beyond”. He received multiple conference travel fellowships from Duke University to present his research at the American Society for Mass Spectrometry conferences in 2008, 2009, and 2010, as well as the Human Proteome Organization conference in 2011. He was a member of the American Society for Mass Spectrometry and the American Association for the Advancement of Science. Patrick graduated from Duke University in May 2011 with his degree of Doctor of Philosophy.

Publications:


