Development and Application of Covalent-Labeling Strategies for the Large-Scale Thermodynamic Analysis of Protein Folding and Ligand Binding

by

Yingrong Xu

Department of Chemistry
Duke University

Date:_____

Approved:

_____
Prof. Michael C. Fitzgerald, Supervisor

_____
Prof. Terrence G. Oas

_____
Prof. Katherine J. Franz

_____
Prof. Qiu Wang

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

2016

ABSTRACT

Development and Application of Covalent-Labeling Strategies for the Large-Scale Thermodynamic Analysis of Protein Folding and Ligand Binding

by

Yingrong Xu

Department of Chemistry
Duke University

Date:_____
Approved:

_____
Prof. Michael C. Fitzgerald, Supervisor

_____
Prof. Terrence G. Oas

_____
Prof. Katherine J. Franz

_____
Prof. Qiu Wang

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

2016

Abstract

Thermodynamic stability measurements on proteins and protein-ligand complexes can offer insights not only into the fundamental properties of protein folding reactions and protein functions, but also into the development of protein-directed therapeutic agents to combat disease. Conventional calorimetric or spectroscopic approaches for measuring protein stability typically require large amounts of purified protein. This requirement has precluded their use in proteomic applications. Stability of Proteins from Rates of Oxidation (SPROX) is a recently developed mass spectrometry-based approach for proteome-wide thermodynamic stability analysis. Since the proteomic coverage of SPROX is fundamentally limited by the detection of methionine-containing peptides, the use of tryptophan-containing peptides was investigated in this dissertation. A new SPROX-like protocol was developed that measured protein folding free energies using the denaturant dependence of the rate at which globally protected tryptophan and methionine residues are modified with dimethyl (2-hydroxyl-5-nitrobenzyl) sulfonium bromide and hydrogen peroxide, respectively. This so-called Hybrid protocol was applied to proteins in yeast and MCF-7 cell lysates and achieved a ~50% increase in proteomic coverage compared to probing only methionine-containing peptides. Subsequently, the Hybrid protocol was successfully utilized to identify and quantify both known and novel protein-ligand interactions in cell lysates. The ligands under study included the well-known Hsp90 inhibitor geldanamycin and the less well-understood omeprazole sulfide that inhibits liver-stage malaria. In addition to protein-

small molecule interactions, protein-protein interactions involving Puf6 were investigated using the SPROX technique in comparative thermodynamic analyses performed on wild-type and Puf6-deletion yeast strains. A total of 39 proteins were detected as Puf6 targets and 36 of these targets were previously unknown to interact with Puf6. Finally, to facilitate the SPROX/Hybrid data analysis process and minimize human errors, a Bayesian algorithm was developed for transition midpoint assignment. In summary, the work in this dissertation expanded the scope of SPROX and evaluated the use of SPROX/Hybrid protocols for characterizing protein-ligand interactions in complex biological mixtures.

Dedication

I would like to dedicate this dissertation to my parents for always standing by my side, through the highs and lows of my life, no matter we are close in distance or separated by the Pacific Ocean.

Contents

List of Tables

List of Figures

List of Abbreviations

[GdmCl]	guanidine chloride concentration
$C_{1/2}$	denaturant concentration at the transition midpoint of a denaturation curve
cyt c	cytochrome c
DMSO	dimethyl sulfoxide
FA	formic acid
H/D exchange	hydrogen/deuterium exchange
HNSB	dimethyl (2-hydroxy-5-nitrobenzyl) sulfonium bromide
Hsp90	heat shock protein 90
iTRAQ	isobaric tags for relative and absolute quatitation
K_d	dissociation constant
L/H ratio	ratio of the light- and heavy-labeled peptide intensities
LC	liquid chromatography
m-value	$\delta\Delta G_f/\delta$[GdmCl] or $\delta\Delta G_f/\delta$[urea]
MALDI	matrix-assisted laser desorption/ionization
Met	methionine
MMTS	methyl methanethiolsulfonate

MS	mass spectrometry
OD_{600}	optical density at 600 nM
PP	pulse proteolysis
Puf6	pumilio/fem-3 mRNA-binding factor 6
Q-TOF	quadrupole time-of-flight
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SILAC	stable isotope labeling by amino acids in cell culture
SPROX	stability of proteins from rates of oxidation
SUPREX	stability of unpurified proteins from rates of H/D exchange
TCA	trichloroacetic acid
TCEP	tris (2-carboxyethyl) phosphine
TEAB	tetraethylammonium bromide
TFA	trifluoroacetic acid
Trp	tryptophan
ΔG_f	Gibbs free energy of protein folding in the absence of denaturant

Acknowledgements

First of all, I would like to thank my research advisor, Professor Michael C. Fitzgerald, for his guidance in my PhD research. With his patience, resourcefulness and optimism, Professor Fitzgerald is one of the greatest academic advisors I have ever had. I would also like to thank Professor Franz, Professor Oas and Professor Wang for being my committee members and for their support in my thesis research.

My 5-year stay at the Fitzgerald lab cannot be as pleasant and satisfying without the company of my lab mates. I want to thank Dr. Erin Strickland for being my mentor when I first entered the lab and for her contribution in the development of the Hybrid protocol described in Chapter 2. I'm also extremely grateful for the help of Dr. Michelle Ariel Geer Wallace with the geldanamycin binding experiments in Chapter 3 and the assistance from Dr. Jagat Adhikari with the SILAC-PP experiments on omeprazole sulfide binding in Chapter 4. I also want to thank all of my current group members, Lorrain Jin, Julia Roberts, Ryenne Ogburn, Fang Liu, He Meng, Arthur Ma and Jaret Karnuta for creating such an enjoyable and supportive working environment. I would like to give special thanks to the IT analyst William Day for the maintenance of our lab computers and the departmental instrument operations director George Dubay for his help with my mass spectrometry sample analysis.

I want to acknowledge my friends Ying Shi, Nuoya Yang, Jiachen Ma and Jin Liang, for bringing me so much joy and encouragement. Last but not the least, I would

like to thank my parents for their incredible support and understanding over the last 5 years, without which I can never achieve all of this.

1. Introduction

1.1 Significance of protein folding and stability measurements

Probing the thermodynamic stability of proteins and protein-ligand complexes can offer crucial insights not only into the fundamental properties of protein folding reactions and protein functions, but also into the development of protein-directed therapeutic agents to combat disease. A widely used measure of protein thermodynamic stability is the folding free energy ΔG_f, which is the free energy difference between the native three-dimensional structure proteins adopt in solution and the random coil-like structures they adopt under denaturing conditions. Protein stability changes can occur when mutations, post-translational modifications, or binding interactions take place. This dissertation lays emphasis on protein stability changes associated with ligand binding including those involving small molecule drugs and other proteins.

Investigating protein-drug interactions helps elucidate drug mode-of-action and facilitates drug development. For example, studies on interactions between platinum-based drugs and serum proteins provided evidence of drug transport and anticancer mechanism, as well as inspired the design of novel metallodrugs for cancer chemotherapy(*1*). Besides protein-drug interactions, knowledge on protein-protein interactions can shed light on numerous cellular processes and help characterize disease states. For example, understanding of the protein-protein interactions involved in insulin signaling network offered knowledge of how the pathways were regulated in type II diabetes(*2*).

1.2 Conventional approaches for protein folding and thermodynamics measurements

For many decades, mainly calorimetric(*3-5*) and spectroscopic(*6-9*) methods have been utilized to make quantitative measurements of protein stability and to characterize the thermodynamic properties of protein-ligand binding. As a powerful calorimetric technique, differential scanning calorimetry (DSC) tracks protein heat capacity changes during thermal denaturation with increasing temperature. However, a typical DSC experiment requires 1~2 mL purified protein solution with concentration of at least 1 mg/mL.(*6*) Such high concentration and large amounts of purified protein material precludes its application in proteome-wide studies, which is the focus of this dissertation. Furthermore, potential issues such as aggregation and self-association caused by the high concentration impairs the accuracy of the measurement. Fluorescence emission, circular dichroism (CD) and nuclear magnetic resonance (NMR) are among the most frequently utilized spectroscopic methods. These optical techniques are sensitive to changes in the protein folding/unfolding equilibrium in solution with alterations in environmental factors such as chemical denaturant concentration, temperature and pH. The intrinsic fluorescence of aromatic residues such as tryptophan and tyrosine is prone to changes in local geometry around the chromophores; the far UV (180-260 nm wavelengths) CD spectrum characterizes secondary structures of proteins; NMR (commonly coupled with hydrogen exchange) provides detailed knowledge of protein conformational properties by monitoring shifts in backbone and side-chain

resonances. Similar to calorimetric methods, the above spectroscopic approaches also require considerable amounts of purified proteins. Furthermore, the fluorescence technique often requires mutagenesis to incorporate tryptophan or tyrosine into native structure, perturbing stability of the native protein.

To enable stability measurements in complex protein mixtures, a mass spectrometry-based method termed "Stability of Unpurified Proteins from Rates of amide H/D Exchange" (SUPREX) was developed(*10-13*). In SUPREX, protein sample is diluted into a series of deuterated H/D exchange buffers containing increasing concentration of chemical denaturant. Each sample is then allowed to undergo H/D exchange for a specific amount of time before MALDI-TOF analysis. Thermodynamic parameters including ΔG_f and m-values can be obtained from the denaturant dependence of protein mass increase. SUPREX can be used for as little as ~10 picomoles of material(*14*). Unfortunately, the conditions of amide H/D exchange are not compatible with commonly used fractionation techniques such as column chromatography and gel electrophoresis. This ultimately limits SUPREX's application to complex biological mixtures.

In summary, the above approaches are not generally amenable to proteome-wide analyses. To maximize the amount of protein thermodynamic stability information and achieve a comprehensive understanding of protein-ligand binding interactions (including both direct and indirect interactions), the need arises for high-throughput, large-scale approaches amenable to unpurified samples.

1.3 Energetics-based proteome-wide approaches

1.3.1 Experimental framework

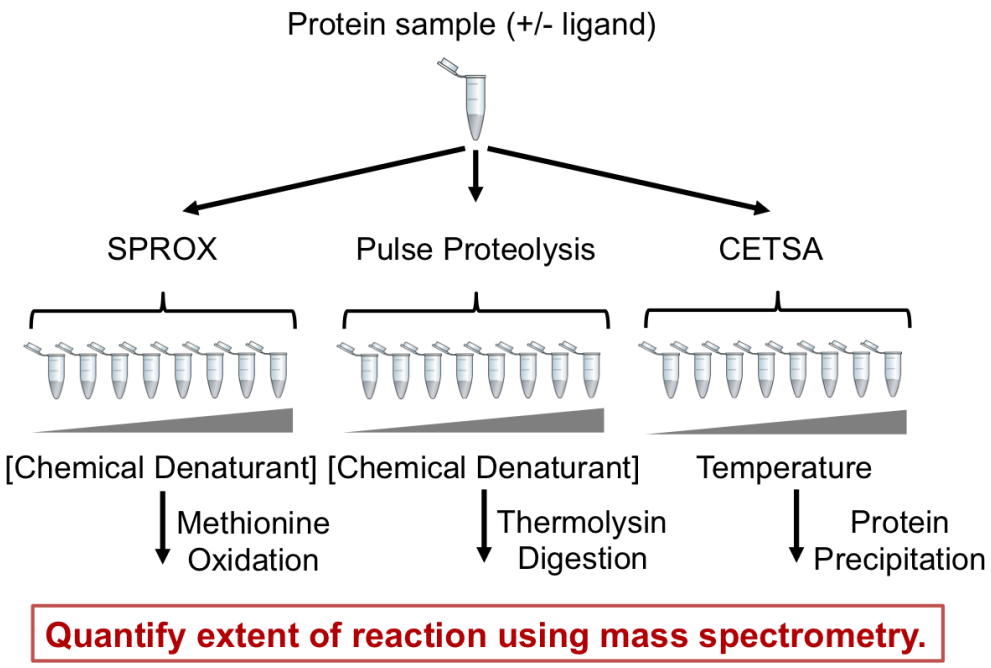

Figure 1: Schematic representation of the energetics-based proteome-wide approaches.

Recently several energetics-based techniques have been developed to probe protein stability and characterize protein-ligand binding interactions on proteomic scale(*15-18*). The basic experimental workflow used for these energetics-based techniques is shown in Figure 1. These techniques generally utilize the chemical denaturant or temperature dependence of a chemical or enzymatic reaction to evaluate the thermodynamic stability of proteins. Increasing the denaturant concentration or temperature unfolds proteins and exposes globally protected reaction sites. The extent of reaction at these sites as a function of the chemical denaturant concentration or the temperature is then quantified in the presence and absence of a ligand. Any significant

difference in the chemical or thermal denaturation behavior upon ligand binding will be interpreted as ligand-induced stabilization or destabilization and under ideal circumstances the corresponding binding affinity K_d value can be determined. In non-ideal circumstances the differential chemical (or thermal) denaturation behavior can be used in a qualitative manner to detect ligand binding. To better accommodate the requirement for high-throughput and proteome-wide analysis, mass spectrometry is frequently chosen as readout due to its unparalleled identification and quantification power.

Summarized in this introductory chapter are the three recently established, energetics-based techniques that have been developed for the large-scale analysis of protein folding and ligand binding. The techniques covered in this chapter include Stability of Proteins from Rates of Oxidation (SPROX), pulse proteolysis and Cellular Thermal Shift Assay (CETSA).

1.3.2 SPROX

1.3.2.1 Methodology overview

As a mass spectrometry- and covalent modification-based method, SPROX utilizes the denaturant dependence of hydrogen peroxide (H_2O_2)-mediated oxidation of globally protected methionine residues to evaluate the solution-phase thermodynamic properties of proteins and protein-ligand complexes. The inherent chemical stability of the oxidized protein products enables the use of standard protein and peptide fractionation techniques, providing compatibility with bottom-up proteomics analysis

platform. In a SPROX experiment, a protein sample is diluted into buffers containing increasing concentration of chemical denaturant. A set amount of H_2O_2 is added to each sample and initiates the oxidation reaction. The reaction conditions are tuned so that the pseudo-first order reaction of the unprotected methionine residues is allowed to proceed for ~3-5 half-lives before quenched with excess amount of methionine or catalase. For intact protein study, the samples are analyzed by MALDI-TOF; while for proteome-wide study, the samples are subjected to quantitative, bottom-up shotgun proteomics sample preparation and analysis with LC-MS/MS. An increase in oxidized protein/methionine peptide signal or a decrease in wild-type methionine peptide signal across the denaturant concentrations can be observed due to the fact that more globally protected methionine residues are solvent exposed with increasing denaturant. Protein stability information such as ΔG_f, m-value and transition midpoint ($C_{1/2}$) can be extracted from the chemical denaturation curves. The oxidation state of the methionine residues can report on the thermodynamic properties of the protein from which they are derived. In ligand binding studies, a significant increase (or decrease) in $C_{1/2}$ upon ligand binding indicates ligand-induced stabilization (or destabilization). Given the free ligand concentration and m-value, a K_d for the protein-ligand complex can be calculated in the stabilization case(*19-21*).

1.3.2.2 Intact protein level SPROX

The SPROX technique was initially demonstrated in 2008(*22*). As part of this initial study, four model proteins: ubiquitin, ribonuclease A (RNase A), bovine carbonic

anhydrase II (BCA II) and cyclophilin A (CypA) were investigated(*22*). The guanidine chloride (GdmCl)-induced equilibrium unfolding behavior of the purified proteins was studied by SPROX and analyzed with MALDI-TOF. Meaningful ΔG_f and m-values were only determined when the protein unfolding reaction is well modeled by a two-state model. The above thermodynamic parameters were evaluated for known two-state folders ubiquitin and RNase A using SPROX and were in good agreement with those obtained by CD spectroscopy. For non-two-state folders BCA II and CypA, although no meaningful ΔG_f and m-value were assessed, K_d values (nanomolar to low micromolar range) for BCA II-4-carboxybenzenesulfonamide and CypA-CsA complexes were determined and were consistent with literature.(*22*)

1.3.2.3 SPROX using isobaric mass tagging strategy

To facilitate SPROX for proteome-wide analysis, after oxidation reaction and tryptic digestion the peptides are labeled with isobaric mass tags (e.g., TMTsixplex™ or iTRAQ® Reagents-8plex) and subjected to shotgun LC-MS/MS analysis.(*15*) The multiplex capability of the isobaric mass tags decreases SPROX analysis time by combining samples from different denaturant concentrations into a single LC-MS/MS run. The reporter ion intensities of the wild-type and oxidized methionine peptides are plotted as a function of denaturant concentration to generate denaturation curves. $C_{1/2}$ values can then be evaluated and compared between lysate with and without the incubation of ligand. The peptides with a significant $C_{1/2}$ difference will be interpreted as hits and potentially interact with the ligand. Since effective denaturation curves can

only be obtained from methionine peptides, the proteomic coverage of SPROX depends on the number of methionine peptides/proteins identified. The frequency of methionine residues in naturally occurring protein sequences is estimated to be ~2.5%(*23*, *24*) and is relatively low. To increase SPROX coverage, a commercially available resin was used which chemo-selectively isolated wild-type methionine peptides and the percentage of methionine peptides identified in the shotgun proteomics experiment was raised from ~20% to ~70%.(*20*)

In the first reported proteome-wide experiment, SPROX was applied to study protein targets of the immunosuppressive drug CsA in yeast cell lysate using TMTsixplex™ and Multidimentional Protein Identification Technology (MudPIT)(*19*, *25*). In this study, 327 proteins were assayed in this study (without the methionine enrichment strategy) and 10 proteins were interpreted as hits that potentially bound CsA. 2 of the protein hits, CypA and UDP-glucose-4-epimerase were previously reported in literature to interact with CsA. CypA directly binds CsA and was calculated to have a K_d of 26-100 nM in this SPROX study; while UDP-glucose-4-epimerase was involved in an indirect binding event with CsA. The other 8 protein hits were not known to interact with CsA. This study successfully demonstrated SPROX's ability to detect both on- and off-target effects of protein-drug interactions.

SPROX with iTRAQ 8-plex quantitation was used in a study to identify protein targets of the well-known and ubiquitous enzyme cofactor, β-nicotinamide adenine dinucleotide (NAD^+) in yeast cell lysate.(*20*) In this study, 122 proteins were assayed and

11 proteins were interpreted as NAD^+ binding hits. Notably 7 out of the 11 protein hits were dehydrogenases and 5 of the 7 dehydrogenases were known to bind NAD^+. K_d value of 6 μM was calculated for alcohol dehydrogenase and was consistent with literature(*26*). A total of 2 of the non-dehydrogenase hits were previously reported to interact indirectly with NAD^+.

Protein targets of another ubiquitous and well-studied enzyme cofactor adenosine triphosphate (ATP) were investigated by iTRAQ-SPROX experiments in yeast cell lysate.(*21*) The non-hydrolyzable ATP analog adenylyl imidodiphosphate (AMP-PNP) was used as ligand in the study. A total of 373 proteins were assayed and 28 proteins were detected as ATP-interacting hits. 14 of the 28 protein hits were annotated as ATP-binding in the SGD database. The newly identified hits were largely from protein classes of hydrolases, oxidoreductases and nucleic acid-binding proteins, unlike the previously annotated protein hits, which were mostly kinases, ligases and chaperones.

Resveratrol is a biologically active ligand with various health benefits and its yeast protein targets were studied with iTRAQ-SPROX in the lysate.(*20*) Unlike NAD^+ and ATP, protein targets of resveratrol were less well-understood. A total of 243 proteins were assayed and 7 proteins were determined as hits. Cytosolic aldehyde dehydrogenase was one of the protein hits and was previously reported to interact directly with resveratrol. A total of 4 of the other protein hits were associated with protein translation machinery, which was implicated in resveratrol's mode-of-action.

Further experiments are required to distinguish between true protein targets from false positives among the 6 newly identified hits in the future.

The false-positive rate of protein target discovery using the iTRAQ-SPROX technique was also evaluated by performing a control experiment where two identical sets of iTRAQ-SPROX samples were prepared without incubation of ligand.(*27*) For Q-TOF and Orbitrap mass spectrometer systems the false-positive rates were calculated to be 1.2-2.2% and <0.8% respectively. The false-positive rate was found to be largely associated with the random quantitation errors of the isobaric mass tags.

1.3.2.4 SILAC-SPROX

The SPROX methodology was also coupled with a Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC) approach to identify the protein targets of ligands.(*28*) The SILAC-SPROX samples can be prepared both in solution and in gel. For the solution phase protocol, cell pellets with light or heavy amino acids (lysine or arginine or both) are lysed and the ligand is spiked into either the light or the heavy lysate. The (-) and (+) ligand lysate stock solutions are then diluted into a series of buffers containing increasing concentrations of chemical denaturant. Each sample is allowed to react with a fixed amount of H_2O_2 for 3-24 min before quenched with excess methionine. The (-) and (+) ligand samples with the same denaturant concentration are combined and precipitated with TCA. After bottom-up proteomics sample preparation, the combined samples are subjected to LC-MS/MS analysis and the corresponding L/H (or H/L) ratios can be obtained for each denaturant concentration. Peptides with

significantly altered L/H ratios at two or more consecutive denaturant concentrations are considered as hits. The gel phase protocol is the same as the solution phase one until TCA precipitation. For the gel phase protocol, after precipitation, cyanogen bromide (CnBr) is added and selectively cleaves after the wild-type methionine residues. The samples are fractionated by SDS-PAGE and analyzed by LC-MS/MS after in-gel digestion. Since more methionine residues become solvent exposed and oxidized as denaturant concentration increases, the appearance of intact protein signal and the disappearance of fragment signal can be observed. The most important advantage of the gel phase protocol is that all peptides instead of just methionine peptides can be used as effective probes and report on the folding/unfolding thermodynamics of the proteins they derive from. The criteria for hit determination are similar to the solution phase protocol.

In a proof-of-principle study, the SILAC-SPROX protocol was applied to detect the direct binding of CsA to CypA and identify protein targets of ATP both in yeast cell lysate.(*28*) CypA was successfully detected as a hit with CsA-induced stabilization. For the ATP-binding study, 526 proteins were assayed from both solution and gel phase experiments, 139 proteins were interpreted as hits, of which 37 proteins were previously known to bind ATP. The false-positive rate of the methodology was calculated to be 0.4-3.5%. A good way to confirm hits and eliminate false-positives is to corroborate with iTRAQ-SPROX experiment.

Recently, the solution phase SILAC-SPROX was utilized to achieve large-scale characterization of breast cancer disease states.(*29*) Similar to the ligand-binding experiment, the disease state study starts with cell pellets from two biological states, one with heavy amino acid labels and the other with light. Unlike the traditional gene/protein expression level study, the SILAC-SPROX approach provides novel perspective into the molecular basis of breast cancer by performing a global thermodynamic stability profile. Different representative stages of breast cancer were investigated in this study by using well-studied cell line models MCF-10A, MCF-7 and MDA-MB-231. For each comparison (MCF-7 vs. MCF-10A and MCF-7 vs. MDA-MB-231) ~800 proteins were assayed and ~180 proteins were determined to have significantly altered stability. Interestingly, ~45% of these hit proteins did not have altered expression levels. The discovery of novel protein signatures for breast cancer highlights the complementary ability of protein stability measurements to traditional expression level research.

1.3.2.5 Other SPROX-related techniques

Apart from isobaric mass tagging and SILAC quantitation, $H_2{}^{16}O_2$ and $H_2{}^{18}O_2$ labeling strategy was reported to be used with SPROX.(*30*) In this protocol termed Probing with SPROX Using Isotope Tags (PrSUIT), protein samples with and without ligand incubation were diluted into buffers containing increasing chemical denaturant, and were reacted twice with $H_2{}^{16}O_2$ and $H_2{}^{18}O_2$ (the first time at intact protein level, the second time at peptide level after trypsin digestion). The (-) and (+) ligand samples at the

same denaturant concentration were pooled together and analyzed by LC-MS/MS. For methionine peptides, the denaturant dependence of $^{18}O/^{16}O$ ratio was used to characterize ligand-induced stability changes. In a proof-of-principle experiment, PrSUIT successfully detected both the direct interaction of CypA and the indirect interaction of calcineurin with CsA in a multi-component protein mixture.

One requirement of SPROX is that the protein should contain at least one globally protected methionine residue to provide any thermodynamic information. To overcome this limitation, the potential of other amino acids to be covalently labeled and interfaced with shotgun, quantitative, bottom-up proteomics was explored. A SPROX-like strategy involving s-methyl thioacetimidate (SMTA) modification of lysine residues was developed and applied to both purified proteins and yeast cell lysate.(*31*) For intact protein study using MALDI-TOF readout, the SMTA protocol successfully evaluated thermodynamic properties of purified proteins including ubiquitin, BCA II, RNase A, 4-oxalocrotonate tautomerase and lysozyme. K_d value of 27 mM was calculated for the lysozyme-N-acetyl-D-glucosamine complex was calculated using the SMTA approach and was in reasonable agreement with literature. The approach was also amenable to proteome-wide analysis using LC-MS/MS readout and applied for protein stability measurements in yeast cell lysate. The SMTA protocol was found to be complementary with SPROX in that ~80% of the assayed lysine peptides in the protocol did not contain methionine.

Instead of using chemical denaturation, SPROX technique coupled with thermal denaturation was developed.(*32*) In a test experiment, thermal SPROX was applied to determine the protein target of CsA in a 4-protein model mixture consisting of ubiquitin, RNase A, BCA II and CypA. The known direct binding interaction between CypA and CsA was detected using both MALDI-TOF and LC-MS-based readouts. However, protein aggregation in complicated biological mixtures was found to be problematic for the proteome-wide application of thermal SPROX(*32*).

1.3.3 Pulse Proteolysis

1.3.3.1 Basic pulse proteolysis protocol

In the basic pulse proteolysis experiment,(*33*) protein sample is incubated with buffers containing increasing concentration of urea and allowed to reach equilibrium. Thermolysin is added into each sample and the pulse of proteolysis is quenched with EDTA after 1 min. Thermolysin is a robust enzyme and retains its protease activity even at high concentration of urea; thermolysin has broad specificity and hydrolyzes peptide bonds at hydrophobic residues from the unfolded region of the protein. The protein samples at each urea concentration are then fractionated by SDS-PAGE to separate the undigested intact protein from the digested fragments. Finally the protein bands are stained and quantified with a gel-scanner. Since more regions of the protein become unfolded with the increase in denaturant, a decrease in intact protein band intensity can be observed across urea concentrations. The fraction of folded proteins (f_{fold}) is calculated based on band intensities and plotted against urea concentration to generate

denaturation curves, where transition midpoint C_m can be determined. At C_m, half of the protein is unfolded. The global protein stability in water $\Delta G_{unf}°$ is evaluated by multiplying C_m with an estimated m-value based on protein size.

As a proof-of-principle study, basic pulse proteolysis was utilized to assess global stability of E. coli ribonuclease HI (RNase H*) and its variants I53A, I53D.(*33*) The $\Delta G_{unf}°$ values obtained were consistent with those measured by CD spectroscopy. Notably, similar C_m value of RNase H* was acquired when the protocol was applied to crude lysate of *E. coli* cells overexpressing RNase H*. Similar to SPROX, the protocol was also applied to monitor protein-ligand binding.(*33*) The stability difference between maltose-binding protein (MBP) and MBP-maltose complex was measured, providing K_d value of 0.15 μM, which was in agreement with literature.

The basic pulse proteolysis protocol has been applied to evaluate the thermodynamic stability of wild-type and mutants of cystathionine β-synthase (CBS)(*34, 35*), *Mycobacterium tuberculosis* recA intein(*36*), mouse H-ferritin(*37*), *E. Coli* thioredoxin(*38*), of single-domain knotted proteins YibK and YbeA(*39*), of the double mutant of *E. Coli* dihydrofolate reductase (DHFR)(*40, 41*), of *E. Coli* glyceraldehyde-3-phosphate dehydrogenase (GAPDH), of purified SH3 domains(*42*), and of mutants of MBP(*43*). The protocol was also used to assess ligand-induced protein stability changes, such as the stabilization of CBS by S-adenosylmethionine (AdoMet)(*44*), the binding of AdoMet to YibK and YbeA(*39*) and the interaction of lipophilic drugs with intestinal fatty acid binding protein (I-FABP)(*45*). Furthermore, the basic pulse proteolysis

protocol extended its application to characterize protein-protein interactions. In one study, pulse proteolysis was used to evaluate binding affinities between the multistate folding protein β-lactamase TEM-1 and various β-lactamase inhibitor protein (BLIP) mutants(*46*).

1.3.3.2 Pulse and Western

Although the basic protocol is tolerant to impurities in protein samples and can be applied in context of cell lysate, the protein of interest is required to be highly abundant for SDS-PAGE analysis. To overcome this limitation and enable the analysis of low abundant proteins in complicated mixture, the strategy termed "Pulse and Western" was developed, combining pulse proteolysis with quantitative Western blotting(*47*). The target protein is quantified with the specific antibody by Western blotting after SDS-PAGE readout. In a test experiment, Pulse and Western was applied to assess the thermodynamic stability of low-concentration H-ras in both *E. Coli* and Jurkat T cell lysates. The measured C_m and $\Delta G_{unf}°$ values were in good agreement with those obtained from purified H-ras and by CD spectroscopy(*47*).

Pulse and Western was utilized for protein thermodynamic stability measurements in tissue extracts. The approach was used to assess the stability of arginine kinase(*48*) and glutamate dehydrogenase (GDH)(*49*) in crayfish tail muscle tissue extracts. In another study, Pulse and Western was applied to measure the protein stability of MAPK-activated protein kinase-2 (MAPKAPK2) in squirrel skeletal muscle tissue extracts at euthermic and hibernating conditions(*50*).

1.3.3.3 Pulse proteolysis on proteomic scale

To increase experimental throughput and allow proteomic scale discovery of protein-ligand interactions, a modified pulse proteolysis strategy was established using 2D gel electrophoresis as the fractionation technique(*16*). Protein sample with and without ligand is incubated in a urea-containing buffer. Instead of using a series of urea concentrations, only one concentration is selected here. The (-) and (+) samples are then exposed to thermolysin for 1 min before quenched with excess EDTA and the remaining intact proteins are fractionated by 2D gel electrophoresis. Gel spots with different protein intensities are digested and the corresponding proteins are identified by MALDI-TOF/TOF. In the proof-of-principle study, the new strategy was applied to identify protein targets of ATP in *E. Coli* cell lysate using the nonhydrolyzable ATP analogue adenosine 5'-[γ-thio]triphosphate (ATPγS)(*16*). 12 gel spots were chosen for the downstream digestion and mass spectrometry analysis. 10 proteins were identified from the 12 gel spots and 6 out of the 10 proteins were previously known to bind ATP with K_d values ranging from 0.025 μM to 600 μM. 3 out of the other 4 proteins were also validated to bind ATP. Several caveats should be considered for future applications: the poor resolution and reproducibility of 2D gel electrophoresis can cause difficulties in result interpretation.

To avoid the above complications from the 2D gel electrophoresis approach, an alternative pulse proteolysis method was investigated using anion exchange chromatography fractionation with 1D SDS-PAGE readout(*51*). In this approach, cell

lysate is fractionated by anion exchange chromatography. Each fraction is then diluted into a series of urea buffers with or without the incubation of ligand. The pulsed proteolysis reaction by thermolysin is performed and the samples are separated with SDS-PAGE. The urea dependence of f_{fold} can be quantified by gel intensities and selected proteins can be identified using mass spectrometry after in-gel digestion. This approach was again applied to detect ATP-binding proteins in *E. Coli* lysate using ATPγS and identified 30 protein hits in total. 9 of the 30 proteins were not known previously to bind ATP and one of the 9 novel hits was confirmed to be stabilized strongly by ATP(*51*). Although reproducibility is improved compared to the 2D gel electrophoresis approach, this new method requires the tedious handling of a large number of SDS-PAGE gels. Additionally, the fact that pulse proteolysis is performed after cell lysate fractionation may preclude the detection of some indirect interactions.

1.3.3.4 SILAC-Pulse Proteolysis (SILAC-PP)

Recently, pulse proteolysis was combined with SILAC quantitation to enable the unbiased detection and quantitation of protein-ligand interactions in complicated biological mixtures(*52*). In SILAC-PP, cell pellets with heavy and light amino acids (lysine or arginine or both) are lysed and incubated with or without ligand. The (-) and (+) stock solutions are then diluted into a series of urea buffers and exposed to a 1 min pulse of thermolysin. The (-) and (+) samples with the same urea concentration are combined and subjected to SDS-PAGE fractionation. The gel lanes are digested and analyzed with LC-MS/MS. The urea dependence of L/H (or H/L) ratios can be obtained

and hits are identified with significantly altered L/H ratios for at least two consecutive urea concentrations.

Protein targets of two test ligands CsA and ATP were investigated using SILAC-PP in yeast cell lysate(*52*). The tight binding of CsA to CypA was successfully detected and 33 proteins were interpreted as hits in the ATP-binding experiment using AMP-PNP as ligand. 21 of the 33 hits were previously known to bind ATP. The method's false-positive rate was evaluated by control experiments to be in the range of 2.1% to 3.6%.

1.3.4 CETSA

CETSA is a method to probe protein-ligand interactions in cell lysate or living cells based on the biophysical principle that target proteins complexed with ligand are more resistant to thermal denaturation. In the first report of CETSA(*17*), the (-) and (+) ligand cell lysates were divided into multiple aliquots and heated individually at different temperatures for 3 min. After cooling to room temperature, ultracentrifugation was performed to separate the supernatants from the precipitated proteins. The supernatants were then analyzed by SDS-PAGE followed by Western blot quantitation. The disappearance of the intact protein signal with increasing temperature was monitored and the melting temperature T_m where half of the protein had precipitated was obtained as an indication of protein thermal stability. In the case of a direct binding event, T_m may shift to higher temperatures whereas during an indirect binding event, T_m may shift to lower temperatures. As a proof-of-principle study, CETSA was used to

identify protein targets of antifolate cancer drugs methotrexate and raltitrexed in both intact K562 cells and cell lysates(*17*). For intact cell experiments, living cells were exposed to drugs before lysis while for cell lysate experiments drugs were directly incubated with lysates. The known targets DHFR and thymidylate synthase (TS) were successfully identified with a large thermal shift. In the same study, target engagement of other small molecule inhibitors including TNP-470, Vermurafenib, Olaparib and Suramin was also investigated using CETSA in both cell lines and organ tissues. An isothermal dose-response fingerprint ($ITDRF_{CETSA}$) experiment is usually performed to study the drug concentration effects on target engagement. For $ITDRF_{CETSA}$, lysate aliquots are exposed to different concentrations of drug at a fixed temperature. However, due to the irreversibility of thermal denaturation, $ITDRF_{CETSA}$ can not yield absolute binding constants.

To improve sample throughput, a screen-compatible format of CETSA was developed using antibody pair selection for quantitation (AlphaScreen technology)(*53*). This screen format involves the use of microtiter plate-compatible system and automated liquid handling equipment. Screen format CETSA was used for the identification of intracellular p38α stabilizers among a small test compound library containing 352 different small molecules(*53*). To achieve proteomic scale drug target discovery, CETSA was combined with quantitative mass spectrometry (TMT10 quantitation)(*18, 54*). This approach, also termed "Thermal Proteome Profiling (TPP)", was used to generate quantitative thermal stability data for more than 7000 proteins in

human cells(*54*). TPP was applied to study target engagement of two well-known promiscuous kinase inhibitors staurosporine and GSK3182571 in K562 cell lysate. For staurosporine, 51 of the 175 protein kinases assayed showed significant T_m shift and passed all hit criteria(*54*).

1.4 Focus of dissertation

The focus of this dissertation is the development and application of energetics-based proteome-wide approaches for measurement of protein stability and characterization of protein-ligand interactions. SPROX has its unique advantages compared with pulse proteolysis and CETSA in that it evaluates K_d values and provides domain-level thermodynamic stability information. However, unlike pulse proteolysis and CETSA, SPROX requires globally protected methionine residues to report thermodynamic properties of the corresponding proteins. Due to signal suppression and low ionization efficiency, it becomes a challenge to detect methionine residues in shotgun proteomics experiments using data-dependent acquisition. To expand the scope of SPROX, this thesis work explores the use of tryptophan as an alternative stability probe. Chapter 2 will describe the development of the SPROX-like protocol involving covalent modification of tryptophan. The new protocol was evaluated on selected model proteins and yeast cell lysate. A ~50% increase in proteomic coverage was successfully achieved using the Hybrid protocol, which combined tryptophan modification and methionine oxidation in one experiment(*55*). Chapter 3 and Chapter 4 will describe the application of SPROX/Hybrid protocols in the detection of known and novel protein-

small molecule binding interactions, including the well-studied Hsp90 inhibitor geldanamycin and the less well-understood omeprazole sulfide with antimalarial activity. Chapter 5 will describe the use of the SPROX technique in characterization of protein-protein interactions involving Puf6. Chapter 6 will demonstrate the application of a Bayesian algorithm for SPROX/Hyrbid data analysis.

2. The development of a SPROX-like protocol using tryptophan modification

2.1 Introduction

This chapter comes largely from the research paper titled "Thermodynamic Analysis of Protein Folding and Stability using a Tryptophan Modification Protocol"(*55*). The SPROX methodology is especially attractive for such large-scale analyses because it can be interfaced with bottom-up, shotgun proteomics approaches using mass spectrometry(*15, 19, 20*). One requirement in SPROX is that the target protein contain at least one globally protected methionine residue in its three-dimensional structure. The frequency of Met residues in protein sequences is relatively low (e.g., ~2.5%)(*23*). However, nearly all protein sequences contain at least one Met apart from the initiator Met(*56*). Thus, SPROX is not fundamentally limited by the low frequency of Met residues in protein. However, there is a practical limitation associated with the use of SPROX in large-scale analyses that is defined by the number of methionine-containing peptides that can be detected in the bottom-up shotgun proteomics experiment. One approach for expanding the protein coverage in proteome-wide SPROX experiments is to exploit chemical modification reactions that target other amino acid residues. To this end, SPROX-like protocols have been developed that exploit the reaction of lysine residues with SMTA (s-methyl thioacetimidate)(*31*) and the H/D exchange properties of histidine side chains(*57*). While these protocols have expanded the application of SPROX to some proteins, they have not been useful for

those that do not have histidine and lysine residues at buried positions in their folded three-dimensional structure.

With the goal of further increasing the peptide and protein coverage in SPROX, we report here a SPROX-like protocol using dimethyl (2-hydroxy-5-nitrobenzyl) sulfonium bromide (HNSB) to label tryptophan (Trp) residues in proteins. Trp residues are attractive probes in the SPROX experiment because ~90% of the Trp residues in proteins are located in globally protected regions of protein structure(*23*). The frequency of Trp residues in protein sequences, ~1.5%(*23*), is slightly lower than that of Met residues. However, analogous to Met probes in SPROX, one Trp probe can report on the stability and ligand binding properties of the entire domain to which it maps. HNSB and a closely related compound, 2-hydroxy-5-nitrobenzyl bromide (HNB), have been previously shown to selectively modify Trp residues of proteins in both acidic and neutral solutions(*58, 59*). Over the years HNB and HNSB have been used to probe the solvent accessibility of Trp residues in protein structures(*60-63*), to quantify the number of Trp residues in proteins(*64*), and to evaluate structure-function relationships in proteins(*58, 59, 65-67*). Here we utilize the chemical denaturant dependence of the rate at which Trp residues in proteins are modified with HNSB to generate thermodynamic information about protein folding and ligand binding interactions. We also demonstrate that the HNSB protocol developed here can be incorporated into the SPROX technique to increase the protein and peptide coverage in proteomic applications of thermodynamic stability measurements.

2.2 Experimental

2.2.1 Materials

The following materials were from Sigma-Aldrich (St. Louis, MO): luteinizing hormone releasing hormone (LH-RH; 97 wt %), substance P acetate salt hydrate (Sub P; ≥95 wt %), insulin oxidized B chain (ProteoMass, MALDI-MS standard), cytochrome c (cyt c) from equine heart (≥95 wt %, mainly in oxidized form), lysozyme from chicken egg white (≥90 wt %), carbonic anhydrase II (BCA II) from bovine erythrocytes (93 wt %), ribonuclease A (RNase A) from bovine pancreas (≥60 wt %), ubiquitin from bovine erythrocytes (≥98 wt %), trypsin from porcine pancreas (proteomics grade, dimethylated), aldolase from rabbit muscle (≥80 wt %), trypsin inhibitor from soybean, L-tryptophan, L-methionine, Dimethyl (2-hydroxy-5-nitrobenzyl) sulfonium bromide (HNSB; ≥95 wt %), brinzolamide, guanidine hydrochloride (GdmCl), tris (2-carboxyethyl) phosphine hydrochloride (TCEP · HCl), S-methyl methanesulfonate (MMTS), triethylammonium bicarbonate buffer (1.0 M, pH 8.5) (TEAB), sodium dodecyl sulfate (SDS), hydrogen peroxide (H_2O_2; 30% w/w), ammonium acetate, 1,1,3,3-tetramethoxypropane (TMP), pyrrolidine, α-Cyano-4-hydroxycinnamic acid (4-HCCA), sinapic acid (SA), trichloroacetic acid (TCA), trifluoroacetic acid (TFA) and urea. UltraLink hydrazide resin, Coomassie Plus Bradford Assay reagent and formic acid (FA) were purchased from Thermo Scientific (Rockford, IL). Methanol and isopropyl alcohol were from VWR International (West Chester, PA). MacroSpin columns (Silica C18) and Pi^{3TM} Methionine Selective Resin were from the Nest Group (Southborough, MA). The

iTRAQ 8-plex Reagents were from AB Sciex (Foster City, CA). The acetonitrile (ACN) and water used for HPLC were from Honeywell Burdick & Jackson (Muskegon, MI). The C18 pipette tips, or Zip-tips™, were from Millipore (Billerica, MA). Ethanol (completely denatured) was from Avantor Performance Materials, Inc. (Phillipsburg, NJ). Sodium phosphate dibasic was from EM Science (Gibbstown, NJ). Sodium phosphate monobasic was from Mallinckrodt Backer, Inc. (Paris, KY). N,N',N''-triacetylchitotriose ($(NAG)_3$) was from Toronto Research Chemicals Inc. (Toronto, Ontario, Canada).

2.2.2 HNSB reaction rate determination

Timecourse experiments were conducted both on model peptide LH-RH and model protein cyt c. An aliquot of peptide/protein stock solution was added to a series of pH 7.4, 20 mM phosphate buffers with a certain GdmCl concentration. LH-RH stock (3.3 mM) was made in pH 7.4 20 mM phosphate buffer. The HNSB reaction was initiated by adding an aqueous solution of HNSB (32.3 mM for LH-RH, 46.2 mM for cyt c). In the 20 μL reaction mixture, the final peptide/protein concentration was 33 μM and HNSB concentration was 1.6 mM for LH-RH, 2.3 mM for cyt c. At specific time-points (0 s, 2 s, 5 s, 10 s, 20 s, 30 s, 60 s, 120 s, 300 s and 600 s), the reaction was quenched with 40 μL 49 mM L-Tryptophan aqueous solution. The 0 s time point was performed by spiking in HNSB solution after addition of L-Tryptophan. Timecourse experiments for LH-RH were carried out in the presence of different [GdmCl] that included: 0 M, 0.2 M, 1 M, 2 M, 3 M, 4 M and 7 M. For cyt c, the buffer used for each time point was pH 7.4 20 mM phosphate with 8 M GdmCl. Measurements of pH were made using a Mettler Toledo

(Columbus, OH) SevenEasy pH meter. GdmCl concentrations were determined using a Bausch & Lomb (Rochester, NY) refractometer. The relationship between the refractive index and [GdmCl] used in this work is the same as reported in reference (*68*). The protein samples in the quenched reaction mixtures were subjected to a desalting procedure using Zip-tips.

2.2.3 Intact protein analysis

Stock solutions containing 1 mM of each purified protein sample (lysozyme, cyt c and BCA II) were prepared in 20 mM phosphate buffer (pH 7.4). The zinc-bound BCA II solution, which was used in the brinzolamide binding study, was made with 300 μM BCA II and 450 μM Zn^{2+}, and it was equilibrated at room temperature for 4 h before use. Aliquots of each protein stock solution were diluted into a series of 20 mM phosphate buffers (pH 7.4) containing [GdmCl] ranging from 0.25 to 8 M. Proteins were equilibrated in the GdmCl-containing buffers for 30 min before aliquots of an aqueous HNSB stock solution were added to each sample to initiate the modification reaction. The final reaction volume was 20 μL. In each case the final protein concentration was ~30 μM and the final HNSB concentration was 2-3 mM. After 1 min, the modification reaction in each GdmCl buffer was quenched with 40 μL of a 49 mM solution of L-tryptophan. The protein samples in the quenched reaction mixtures were each subjected to a concentration and desalting step using C18 Zip-tips™ according to the manufacturers protocol. Ultimately, 10 MALDI-TOF mass spectra were acquired on each

sample and the data were used to generate ΔG_f- and *m*-values ($\delta\Delta G_f/\delta$[GdmCl]) in an analogous fashion to that which we have previously described in SPROX(*22*).

2.2.4 MALDI-TOF analysis

MALDI mass spectra were acquired on an Ultraflex II TOF/TOF mass spectrometer (Bruker Daltonics, Billerica, MA) equipped with a Nd:YAG laser. A total of 100 laser shots were used to acquire the data in each mass spectrum. The following instrument parameters were used: an ion source 1 voltage of 25.00 kV, an ion source 2 voltage of 23.00 kV for model protein analysis and 23.65 kV for model peptide analysis, a lens voltage of 6.75 kV for proteins and 6.25 kV for peptides, pulsed ion extraction of 250 ns for proteins and 0 ns for peptides. Saturated MALDI matrix solutions were prepared in water/ACN solutions (30/70 v/v) containing 0.1% TFA. 4-HCCA was the matrix used for peptides and SA was the matrix for proteins. Insulin oxidized B chain was the internal mass calibrant for the model peptide studies, trypsin inhibitor was the internal mass calibrant for lysozyme and cyt c, aldolase was the internal mass calibrant for BCA II.

2.2.5 Yeast cell lysate preparation

The yeast cell lysate sample was derived from two yeast GAL1 overexpression strains: YLR294C and YGR030C from Y258 host library. Both of the overexpression strains were purchased from Open Biosystems. A yeast colony from each strain was incubated in 50 mL YPD medium at 30 °C overnight to reach an OD_{600} of ~1.6. 20 mL of the overnight culture was then inoculated with 1 L YPD media to have an OD_{600} of ~0.3.

This cultured was incubated at 30 °C until the OD_{600} of the solution was between 1.2-2.0. Yeast pellets were achieved by centrifuging 250 mL fractions of the final solution and the pellets were stored at -20 °C.

Yeast pellets were lysed with 500 μL pH 7.4 20 mM phosphate buffer and 25 μL 20× protease inhibitor cocktail (containing 20 mM AEBSF, 0.2 mM pepstatin A, 0.4 mM leupeptin, 0.3 mM E-64 and 1 mM bestatin). Cell lysis was accomplished using glass beads (0.5 mm) at 4 °C with 20 s of disruption 20 times with 1 min intervals on ice in between. Lysate was centrifuged at 4 °C for 10 min and the supernatant was saved for subsequent analysis. The protein concentration in the lysate determined by Bradford assay was ~14 mg/mL.

2.2.6 Quantitative bottom-up proteomics analysis

Aliquots (20-μL each) of a yeast cell lysate sample were combined with 27.6 μL volumes of a series of 20 mM phosphate buffers (pH 7.4) containing increasing concentrations of GdmCl. The final eight denaturant concentrations used in Experiments 1, 2, and 3 are summarized in Table 1. The eight denaturant-containing protein samples were equilibrated at room temperature for 30 min. In the HNSB protocol, the modification reactions were initiated by adding 2.5 μL of 46.2 mM solution of HNSB to each sample, and they were quenched after 1 min by adding 1 mL of 4.9 mM solution of L-tryptophan to each sample. In the hybrid protocol, 5 μL of 30% (w/w) (9.8 M) H_2O_2 was initially added to each sample. After 2 min, 2.5 μL of a 46.2 mM solution of HNSB was added to each sample. After 3 min, a 1 mL solution containing L-methionine (0.5 M)

and L-tryptophan (4.9 mM) was added to each sample. The proteins in each sample were precipitated by combining 200 μL of TCA (1g/mL) with each sample, incubating the samples on ice overnight, and centrifuging the samples at 14,000 x g for 20 min. The resulting protein pellets were washed three times with 300 ◎L of ice-cold ethanol.

The protein pellets derived from each denaturant containing reaction buffer were combined with 30 μL 0.5 M of TEAB buffer (pH 8.5) and 1.5 μL of a 2% (w/v) SDS solution. The samples were sonicated, vortexed and heated at 60 ℃ to facilitate their dissolution. Disulfide bond reduction was carried out with addition of 3 μL 50 mM TCEP·HCl to each sample and incubated for 1-hour at 60 ℃. A 1.6 μL aliquot of 200 mM MMTS was added to each sample before a 10 min incubation at room temperature. A 1 μL aliquot of trypsin stock solution (1 mg/mL) was added to each sample and the samples were incubated at 37 ℃ overnight. A total of 0.5 unit of iTRAQ reagent was dissolved in 50 μL isopropyl alcohol and was added to each of the 8 samples

The combined samples were either subject to a desalting protocol using C18 reversed-phase resin, a methionine-containing peptide enrichment strategy, or a tryptophan-containing peptide enrichment strategy prior to LC-MS/MS analysis. The C18 desalting and methionine-enrichment protocols were performed as previously described(*15, 20*). The tryptophan-enrichment was performed as described below. In Experiment 1 the C18 purified sample (i.e., the non-enhanced sample) was subject to two replicate LC-MS/MS analyses. In Experiments 2 and 3 the C18 purified samples (i.e., the

nonenhanced samples), the tryptophan-enriched samples, and the methionine-enriched samples were each subject to a single LC-MS/MS analysis.

The tryptophan enrichment protocol used here was adapted from reference (*69*). A total of ~240 µl of the combined iTRAQ-labelled sample was evaporated to dryness, and 100 µL of 80% TFA in water (v/v) was added to each sample along with 1.8 µL TMP. The samples were vortexed for 1-hour at room temperature, before 2 mL of deionized water was added to each sample. A 400 µL aliquot of a hydrazide resin suspension in a 0.02% sodium azide solution (50% w/v) was taken and washed 2 × 200 µL pH 5 100 mM ammonium acetate buffer. The resin was added to the sample and the mixture was vortexed overnight to capture Trp-containing peptides. After capture, the resin was washed 2 × 400 µL methanol, 2 × 400 µL 8 M urea and once with 400 µL pH 5 100 mM ammonium acetate buffer for 15 min. Trp-containing peptides were eluted by vortexing the resin twice with 400 µL 500 mM pyrrolidine in water for 1 h, and then twice with 200 µL 500 mM pyrrolidine in water/methanol (v/v 1:1) for 30 min. The eluted fractions were pooled and concentrated to dryness.

Table 1: Summary of the iTRAQ labeling scheme used in the 3 proteome-wide experiments. Shown are the GdmCl concentrations (M) of the samples corresponding to each of the 8 iTRAQ tags.

	iTRAQ reporter ion m/z							
	113	114	115	116	117	118	119	121
Experiment 1	0	0.6	1.3	1.8	2.3	3.0	3.5	3.9
Experiment 2	1.2	1.3	1.7	2.0	0.6	2.5	3.5	3.1
Experiment 3	1.0	1.4	1.7	2.0	0.5	2.6	3.6	3.2

2.2.7 LC-MS/MS analysis

The LC-MS/MS analyses were performed at the Proteomics Resource at the Fred Hutchinson Cancer Research Center on an Orbitrap Elite ETD mass spectrometer equipped with an EASY-nLC system (ThermoFisher Scientific Inc., Waltham, MA). The trapping column was a 100 ◎m × 2 cm Integrafrit column (New Objective) packed with 200 Å Magic C18 AQ 5 ◎m material (Michrom), and the column was a 75 ◎m × 25 cm PicoFrit column (New Objective) packed with 100 Å Magic C18 AQ 5 ◎m material (Michrom). The flow rate was 400 nl/min. The LC gradient was 5 to 7% solvent B over 2 min, 7 to 35% solvent B over 90 min, 35 to 50% solvent B over 1 min, isocratic at 50% solvent B for 9 min, increased from 50 to 95% B over 1 min, and finally isocratic at 95% solvent B for the last 8 min. Product ion scans (resolution 15,000) were collected for the top 10 most intense peaks in a given precursor scan (resolution 60,000) with an intensity threshold of 5,000, and dynamic exclusion of a given m/z ratio was set at 1 scan in a 0.75 min window. The precursor isolation width was 1.2 m/z. The scan range for the precursor scan was 400-1,800 m/z and 100-2,000 m/z for the product ion scan. Collision induced dissociation was achieved using HCD with normalized collision energy of 40% and an HCD activation time of 0.1 ms.

The LC-MS/MS data were searched using Proteome Discoverer with fixed modifications of MMTS on cysteine and iTRAQ® 8-Plex on N-terminus and lysine residues. Proteome Discoverer was set with a variable modification of oxidation on methionine residues. In Experiments 2 and 3 a variable modification of +151 on

tryptophan residues was also employed. In all cases only peptides identified with high and medium confidence (i.e., false discovery rates (FDR) <1% and 5%, respectively) were used in the data analysis.

2.3 Data analysis

2.3.1 HNSB reaction rate determination

The extent of HNSB modification was determined by taking the weighted average mass of the singly charged protein ion signals corresponding to modified and wild-type species that were observed in the MALDI-TOF mass spectrum. The $Mass_{wt,av}$ values derived from the 10 MALDI-TOF mass spectra collected on each sample were averaged, and a $\Delta Mass_{wt,av}$ value was determined for each sample by subtracting the mass of wild type protein from $Mass_{wt,av}$. Ultimately, plots of $\Delta Mass_{wt,av}$ versus reaction time were constructed. Pseudo-first-order reaction rate constants were obtained by fitting the $\Delta Mass_{wt,av}$ versus reaction time plots to Equation 1, the 3-parameter exponential rise to max equation in SigmaPlot 2001 (San Jose, CA).

$$\Delta Mass_{wt,av}=\Delta Mass_{wt,av,0}+A(1-e^{-kt}) \qquad \text{Equation 1}$$

In Equation 1, $\Delta Mass_{wt,av,0}$ is the $\Delta Mass_{wt,av}$ at 0 s, A is the amplitude of the curve and k is the pseudo-first-order rate constant of modification at the specified [HNSB]. Ultimately, k was divided by [HNSB] to generate the second-order rate constant for the modification of unprotected Trp.

2.3.2 Intact protein analysis

The calculated $\Delta Mass_{wt,av}$ values were plotted against the GdmCl concentration. In order to extract ΔG_f and m-values, the $\Delta Mass_{wt,av}$ versus [GdmCl] curves were fit to Equation 2.

$$\Delta Mass_{wt,av}=\Delta M_{\infty}+(\Delta M_0-\Delta M_{\infty})e^{-\frac{k_{mod}t}{1+K_{fold}}} \qquad \text{Equation 2}$$

In Equation 2, which is analogous to equations we have previously described for the analysis of SUPREX(*10*), SPROX(*22*), and SMTA-labeling(*31*) data. $K_{fold}=e^{-(\Delta G_f+m[GdmCl])/RT}$, ΔM_0 is the $\Delta Mass_{wt,av}$ before global modification, ΔM_{∞} is the $\Delta Mass_{wt,av}$ after global modification, t is the reaction time in seconds, k_{mod} is the pseudo-first-order rate constant for the modification of an unprotected Trp residue, R is the gas constant, T is the temperature in Kelvin, and [GdmCl] is the GdmCl concentration. During the fitting, ΔM_0, ΔM_{∞}, ΔG_f and m were allowed to float and k_{mod} was assigned a value based on the exact HNSB concentration used in the model protein study and the determined second-order rate constant for the modification of an unprotected Trp residue.

2.3.3 Proteome-wide experiments

The iTRAQ reporter ion intensities measured in the LC-MS/MS analyses were normalized using similar strategy as reported(*15*). Briefly, the 8 iTRAQ reporter ion intensities in each product ion mass spectra were averaged and the raw intensity of each reporter ion in the product ion mass spectra was divided by the average value. This

process is referred to as N1 normalization. Then the N1-normalized values for all the non-Trp-containing peptides (in the HNSB protocol) or for all the non-Trp and non-Met-containing peptides (in the Hybrid protocol) were averaged within each reporter ion, to generate so-called N2-normalization factors for the Trp- and Met-containing peptides in each experiment. Summarized in Table 2 are the N2-normalization factors generated for each reporter ion (and corresponding denaturant concentration) in Experiments 1, 2, and 3. The above normalization procedures were only carried out for control samples (samples without Trp or Met enrichment). Ultimately, the N2-normalization factors were used in a second so-called N2 normalization, where the N1-normalized values for each Trp (or Met)-containing peptide were divided by the corresponding N2-normalization factors to obtain the final normalized reporter ion intensity in this work. If multiple product ion spectra were obtained for a given peptide (e.g., it was identified multiple times in a single LC-MS/MS run) the normalized reporter ion intensity values were averaged using Runcompare (an AWK script developed in-house). In the end, a single data set was obtained.

The distribution of N2-normalized reporter ion intensity values for the tag corresponding to the high [GdmCl] (7.4 M) and the low [GdmCl] (0 M) were plotted for the Trp- and Met-containing peptides in each experiment (Figure 2). The point at which the two distributions crossed best separated the pre- and post-transition baselines of the denaturation curves. The midpoint of the curve was the [GdmCl] corresponding to the reporter ion intensity value. This value was 1.1 for Trp-containing peptides and 1.0 for

Met-containing peptides (Figure 2). Typically, the [GdmCl] values flanking the transition were averaged and the average value was taken as the transition midpoint. If there was a normalized reporter ion intensity of 1.1 ± 0.1 for Trp-containing peptides (or 1.0 ± 0.1 for Met-containing peptides) at the transition, then that [GdmCl] was assigned as midpoint. If multiple normalized reporter ion intensity values of 1.1 ± 0.1 for Trp-containing peptides (or 1.0 ± 0.1 for Met-containing peptides) existed at the transition, the average of all these [GdmCl] values was assigned as midpoint. Chemical denaturation data sets in which more than one normalized reporter ion intensity was not >1.1 or <1.1 for Trp-containing peptides (or >1.0 or <1.0 for Met-containing peptides) in the pre- and post-transition baselines, respectively, were classified as "poor quality" and not assigned a midpoint. When only a single outlying value existed, the value was removed from the data set and the remaining 7 values were used for denaturation curve construction and transition midpoint assignment. In cases where the outlying value was at or near the transition, two alternate midpoints could be assigned. In some cases the reporter ion intensities were all within 10% of 1.1 for Trp-containing peptide (or 1.0 for Met-containing peptides). These cases are indicated with an "NT" for "no transition."

The $C_{1/2}$ values extracted from the chemical denaturation data sets were used to calculate ΔG_f values. This was accomplished by using the equation below (for derivation see reference (*70*)) and an m-value of 2 kcal/(mol M).

$$\Delta G_f = -RT \ln\left(\frac{k_{mod}t}{0.693} - 1\right) - mC_{1/2} \qquad \text{Equation 3}$$

In the Equation 3, k_{mod} is the pseudo-first-order rate constant for the HNSB modification or H_2O_2 oxidation (0.09 s^{-1} and 0.013 s^{-1} respectively), t is the corresponding reaction time (60 s and 180 s respectively).

To assess the accuracy and precision of the HNSB and Hybrid protocols, $C_{1/2}$ values were compared for the same matched peptides from different experiments. Absolute $C_{1/2}$ differences were calculated and peptides with more than one isobaric mass tag differences (≥ 0.6 M in case of the two technical replicates from Experiment 1, ≥ 0.5 M in all other cases, since the [GdmCl] spacing between tags is different in Experiment 1 is different than that in Experiment 2 and 3, see Table 1) were considered "hit peptides". Data quality of the "hit peptides" was further evaluated and two features were inspected: retention time and isolation interference. "Hit peptides" with very different retention times or with high isolation interference (typically above 40%) were removed from the "hit list". After the above data curation process, only peptides remaining on the "hit list" were used to calculate false discovery rate.

Table 2: N2-normalization factors generated for each reporter ion in all proteome-wide experiments.

Experiment	Statistics	iTRAQ reporter ion m/z							
		113	114	115	116	117	118	119	121
1	Average	0.99	0.90	0.78	1.13	1.07	0.83	1.20	1.09
	S.D.[1]	0.17	0.16	0.14	0.13	0.12	0.14	0.15	0.14
2	Average	0.97	0.84	0.88	0.96	1.00	1.12	1.07	1.16
	S.D.	0.17	0.14	0.16	0.15	0.20	0.16	0.15	0.15
3	Average	1.04	1.11	0.89	1.03	0.70	1.10	1.06	1.08
	S.D.	0.20	0.20	0.20	0.18	0.22	0.17	0.18	0.17

[1]Standard deviation.

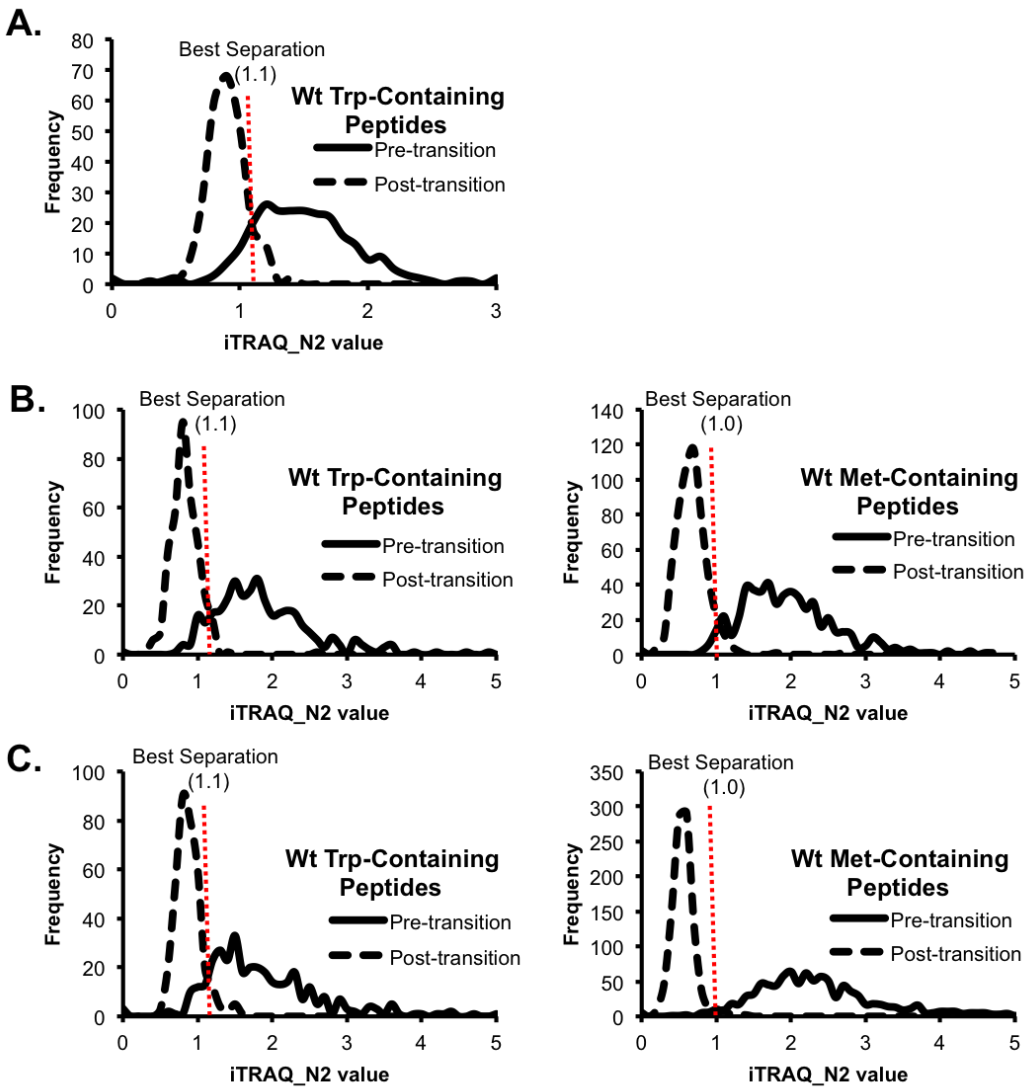

Figure 2: Distribution of the N2-normalized iTRAQ reporter ion intensity ratios observed for the wild-type Trp- and Met-containing peptide probes in (A) Experiment 1 (B) Experiment 2, and (C) Experiment 3. Shown in each case is the data for the reporter ions at the highest (pre-transition) and lowest (post-transition) denaturant concentration used in each case. The vertical dotted lines indicate the normalized reporter ion ratio that best separates the pre- and post-transition baselines (i.e., the normalized iTRAQ reporter ion intensity ratio expected at the $C_{1/2}$ value of the SPROX data recorded for the wild-type (non modified) Trp- and Met-containing peptides.

2.4 Results

2.4.1 Protocol development

The protocol developed here (Figure 3) involves reacting equal aliquots of a protein sample with HNSB in the presence of increasing concentrations of the chemical denaturant, GdmCl. The HNSB reaction conditions (i.e., HNSB concentration and reaction time) are identical for the protein samples in each denaturant-containing buffer. The optimum reaction conditions were found to be 2.3 mM HNSB with a 1 min reaction time. Under these conditions, the pseudo-first order reaction of HNSB with an unprotected tryptophan residue (Figure 4) proceeds for eight half-lives. The optimum reaction conditions were established based on the results of kinetic studies on three model systems, including two chemically denatured proteins (cyt c and lysozyme) and an unstructured peptide (LH-RH) of sequence EHWSYGLRPG (Figure 5). An average second-order rate constant was determined from the data collected on the model systems (Table 3), and this average value of 39 $M^{-1}s^{-1}$ was used to help establish the optimum HNSB concentration and reaction time for the protocol.

In theory, HNSB reaction times greater than 1 min can be used in the described protocol (see Equation 2 and 3). In practice, we found that after several minutes the reactivity of HNSB is significantly (>25%) diminished (see Figure 6). Presumably, this is because the reagent is hydrolyzed. Under the optimized reaction conditions unprotected Trp residues were singly modified and there was minimal loss (<20%) in HNSB reactivity over the course of the reaction. Cross-reactivity of the HNSB with other amino

acid other than Trp was also not observed under the above conditions (Figure 7). Free cysteine residues have been reported to react with HNSB, though their reactivity is less than half that of Trp(*59*). However, we note that the extent of cysteine modification under the above HNSB reaction conditions was also found to be minimal as judged by the results of our bottom-up proteomics analyses described below.

It is important in the protocol that the molar ratio of HNSB to Trp residues be less than 70:1. Under such conditions the Trp residues in the model proteins and peptides studied here were singly modified with HNSB (n=1 in Figure 4). At molar ratios >70:1 multiple modifications (n>1 in Figure 4) were observed on a single Trp residue (Figure 8). Keeping the number of modifications per Trp residue to 1 decreases the likelihood that the modification will perturb the folding and unfolding equilibrium being measured.

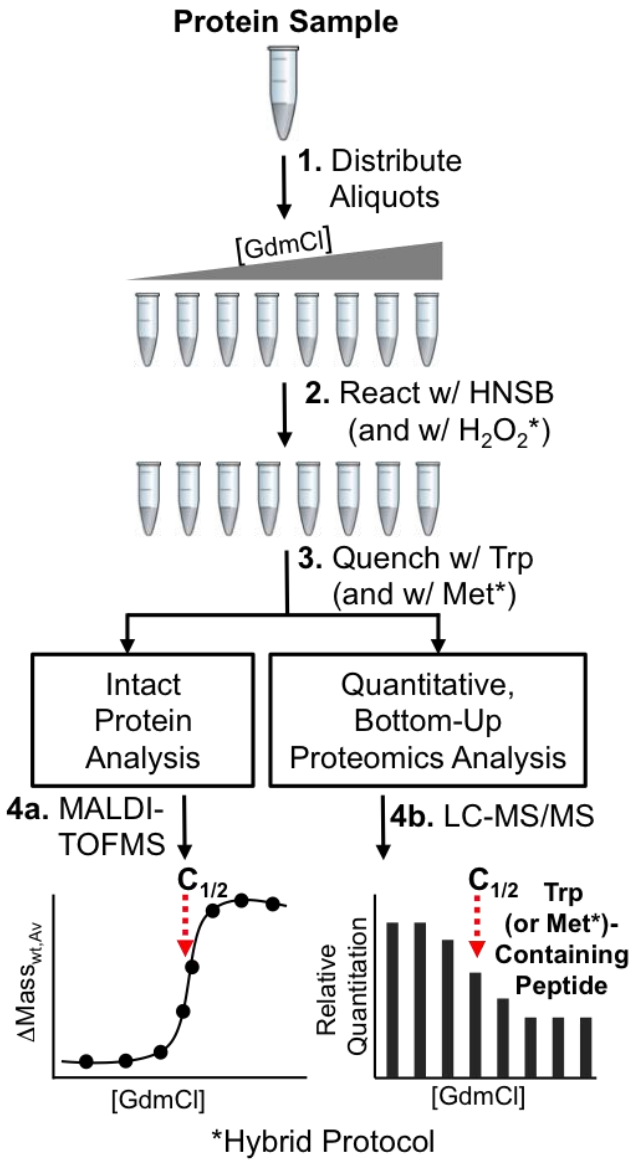

Figure 3: Schematic representation of the HNSB protocol developed in this work.

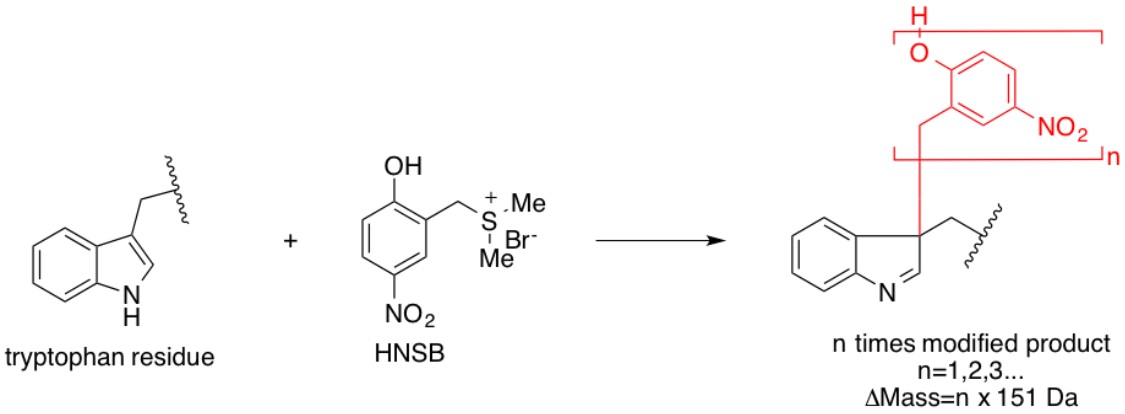

Figure 4: Schematic representation of HNSB modification reaction involving the Trp side chain.

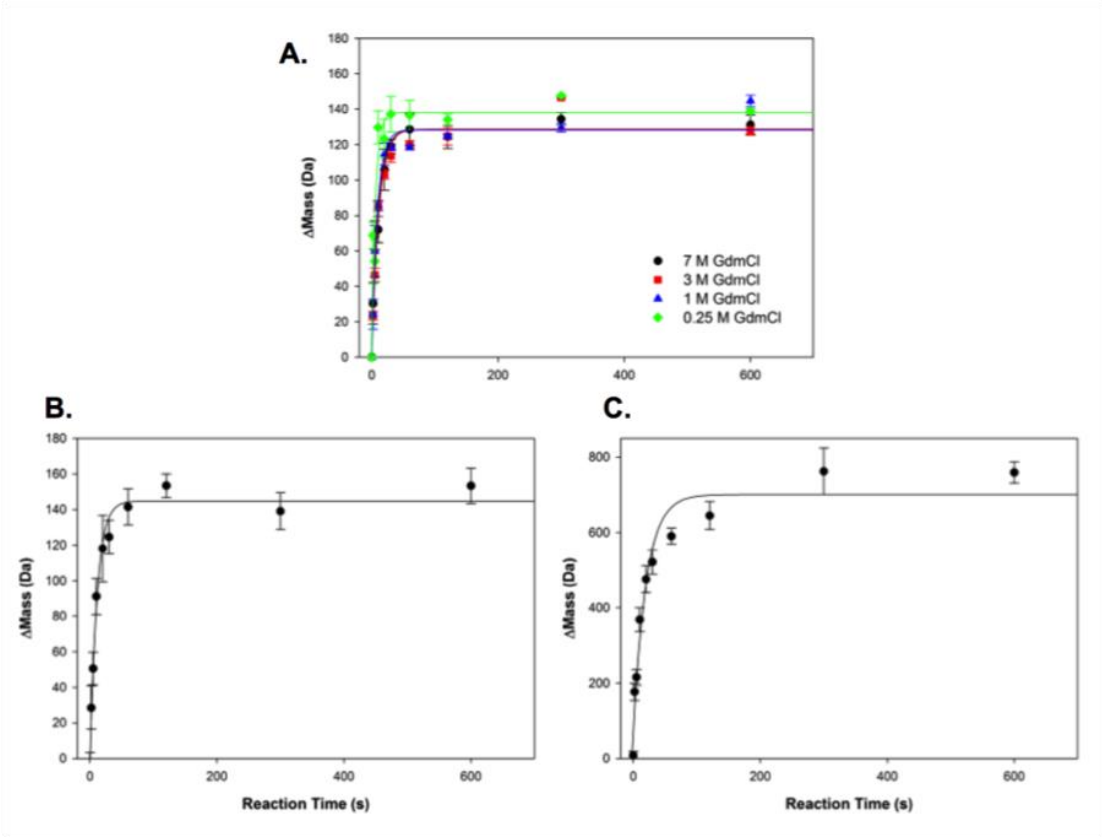

Figure 5: Time course of HNSB labeling of (A) LH-RH at different concentrations of GdmCl ranging from 0.25 M to 7 M, (B) cyt c in 8 M GdmCl, (C) lysozyme in 8 M GdmCl. The HNSB concentration in each case was 2.3 mM and in large excess over the peptide and protein concentrations. The lines represent the best fit of the data to a three-parameter single-exponential rise to max equation using SigmaPlot. The error bars represent ±1 standard deviation of the values obtained from 10 replicate MALDI mass spectra. The pseudo-first order reaction rate constants and corresponding second-order rate constants extracted from the data are summarized in Table 3).

Table 3: Pseudo-first order rate constant of HNSB reaction with model peptide and proteins.

Peptide/Protein	[GdmCl] (M)	k_{HNSB} (s^{-1})[1]
LH-RH	0	0.26±0.02[2]
	0.25	0.28±0.06[2]
	1	0.16
	2	0.14±0.01[3]
	3	0.13
	4	0.07±0.02[2]
	7	0.13
Cyt c	8	0.08±0.02[4]
Lysozyme	8	0.033±0.006[3]

[1]All converted to the condition with [HNSB]=2.33 mM. [2]Results after 2 replicate measurements. [3]Results after 3 replicate measurements. [4]Results after 8 replicate measurements.

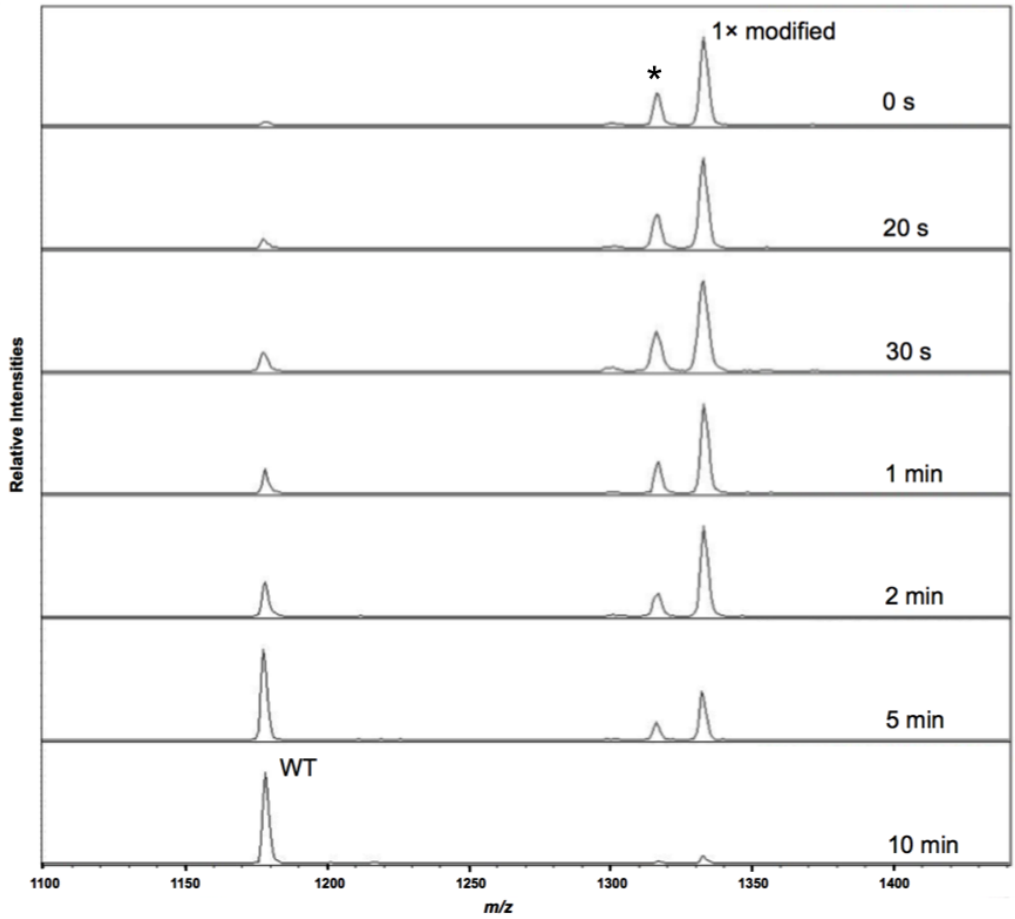

Figure 6: Impact of HNSB hydrolysis on modification reaction. LH-RH was reacted with HNSB ([HNSB]:[LH-RH] = 70:1) for 1 min in each case. However, the HNSB was pre-incubated in the reaction buffer (20 mM phosphate buffer ,pH 7.4, containing with 7 M GdmCl) for 0 s, 20 s, 30 s, 1 min, 2 min, 5 min, and 10 mins prior to reaction with LH-RH. The ion signals for the wild-type and HNSB modified peptide species are indicated as "WT" and "1 x modified," respectively.

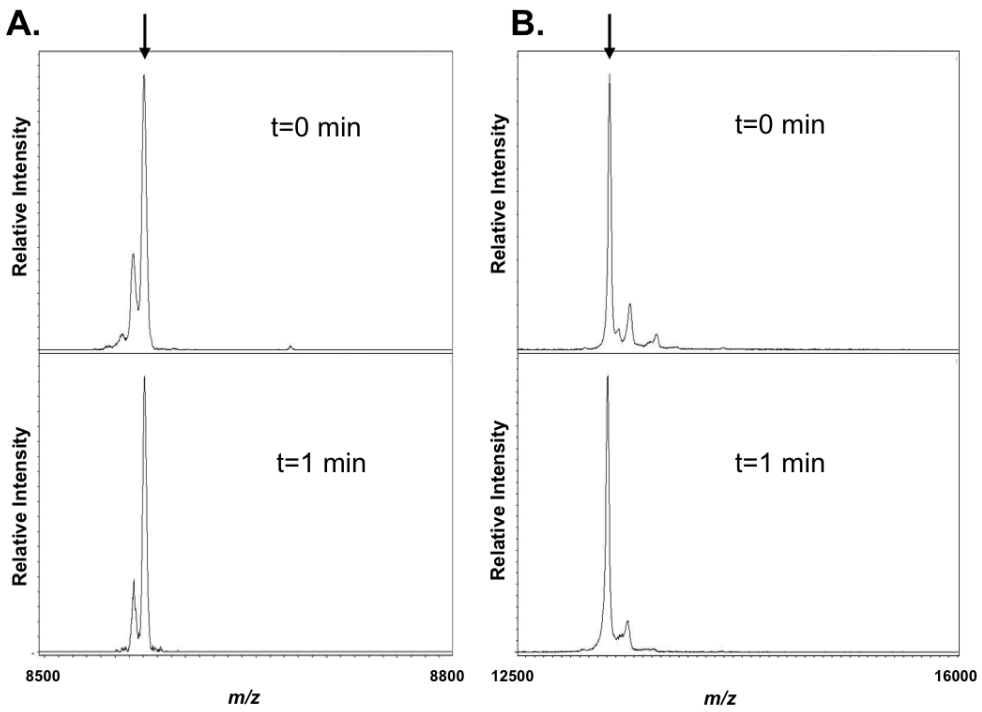

Figure 7: MALDI-TOF mass spectra obtained after reaction of (A) ubiquitin and (B) RNase A with 2.3 mM HNSB in 7 M GdmCl reaction buffers. Both proteins contain no Trp. RNase A has 8 Cys, all in disulfide bonds, while ubiquitin has no Cys. The arrows denote the MALDI ion signal expected for the wild-type protein.

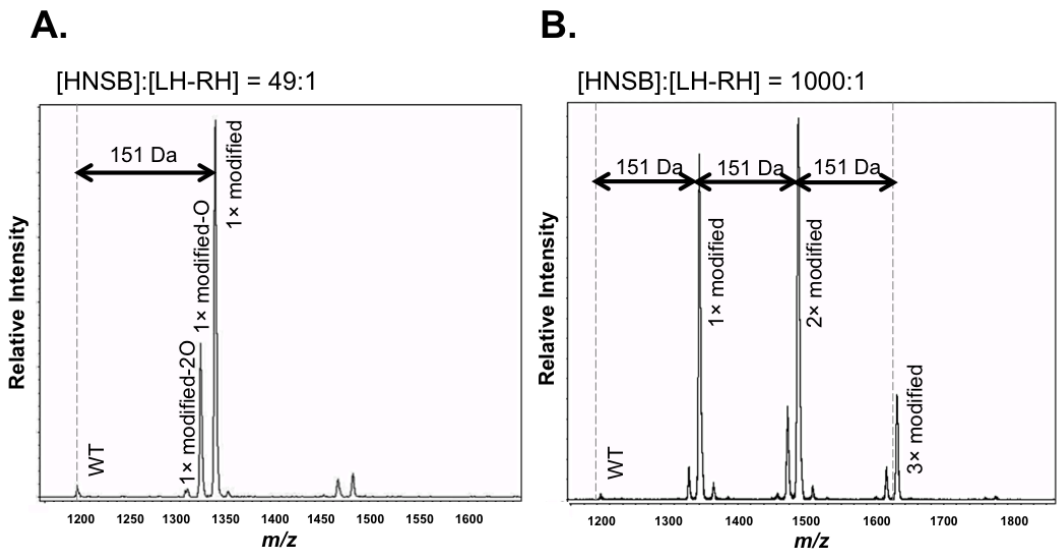

Figure 8: MALDI-TOF mass spectra of HNSB modified LH-RH in the presence of different relative concentrations of HNSB. Shown in (A) and (B) are the reaction products detected using [HNSB] : [LH-RH] rations of 49:1 and 1000:1 with a 5 min reaction time. WT: wild type; -O and -2O represent the successive loss of two oxygen atoms consistent with known MALDI artifacts observed with nitro aromatic compounds(71).

2.4.2 Intact protein analysis

Three purified proteins (lysozyme, cyt c, and BCA II), were each analyzed using the HNSB protocol illustrated in Figure 3. Plots of $\Delta Mass_{wt,av}$ to [GdmCl] were generated (Figure 9) and thermodynamic parameters (Table 4) were determined for each protein. The $\Delta Mass_{wt,av}$ values in the post-transition baselines of the curves generated for cyt c and BCA II (~150 and 1000 Da, respectively) were consistent with that expected for complete modification of the Trp residues found in these two proteins (1 and 7 for cyt c and BCA II, respectively). The $\Delta Mass_{wt,av}$ values in the pre-transition baselines of the cyt c and BCA II curves (~0 in each case) were also consistent with all of the Trp residues in these two proteins being positioned in buried regions of protein structure. The

$\Delta Mass_{wt,av}$ values in the pre-transition baseline of the lysozyme curve suggest that 1 Trp is solvent exposed and modified even at low [GdmCl] concentrations. This is consistent with the crystal structure of lysozyme (PDB 1HEL)(*72*) and surface accessibility calculations(*62*), which suggest that one of the Trp residues, W62, is indeed solvent exposed. The ~750 Da $\Delta Mass_{wt,av}$ values in the post-transition baseline of the lysozyme curve (Figure 9B) also suggest that only 5 out of 6 Trp residues in lysozyme were modified. In theory, all 6 Trp residues should be modified at high GdmCl concentrations, since lysozyme looses its higher order structure in concentrated GdmCl solutions(*73*). One reason for the incomplete reaction observed with lysozyme may be related to the presence of a Trp-Trp sequence in the protein. After one of the Trp residues is modified, steric hindrance from bulky (2-hydroxy-5-nitrobenzyl) group may keep adjacent Trp residue from being modified.

Lysozyme was analyzed in the presence of $(NAG)_3$ (Figure 9B), which is known to bind lysozyme with measured K_d values ranging from 5 to 19 ◎M(*74*). BCA II was analyzed in the presence of brinzolamide (Figure 9C), which is known to bind human carbonic anhydrase II (hCA II) with a measured K_d value of 3 nM(*75*). Ligand-induced protein stabilizations were detected in both cases. Only the lysozyme-$(NAG)_3$ analysis generated a K_d value that was in reasonable agreement with that previously reported in the literature (Table 4). The inability of the HNSB protocol to generate an accurate K_d value for this complex is most likely related to the non-two-state folding behavior of this protein (see section 2.5).

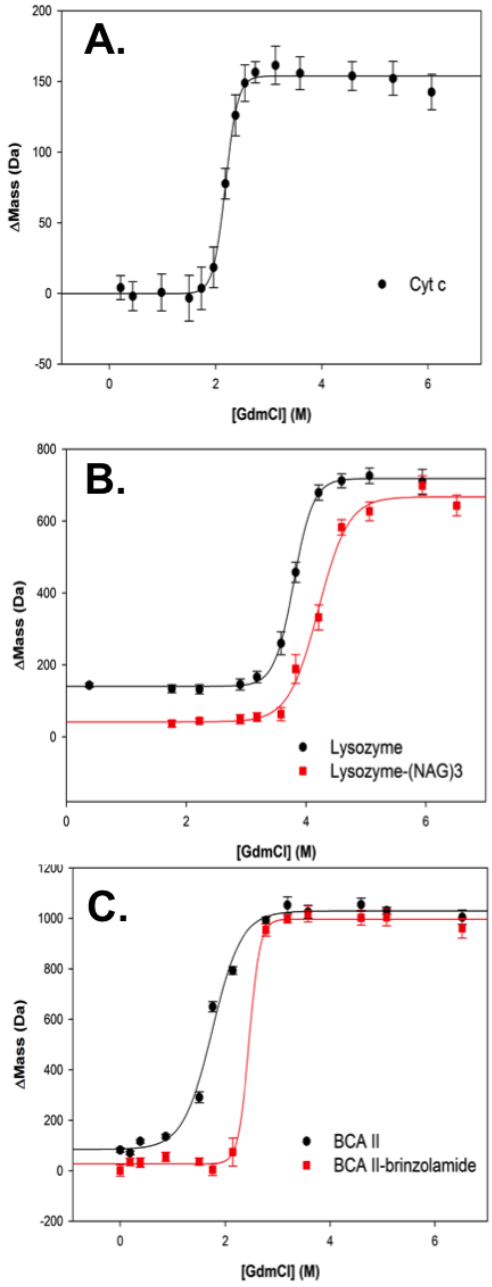

Figure 9: Data obtained using the HNSB protocol on the intact protein level. Denaturation curves obtained in the analysis of cyt c, lysozyme, and BCA II are shown in (A), (B), and (C), respectively. Also shown in (B) and (C) are the denaturation curves obtained for $(NAG)_3$-bound lysozyme and brinzolamide-bound BCA II, respectively.

Table 4: Thermodynamic parameters of model proteins determined using HNSB protocol and comparison with literature values.

Model protein	Protocol	m (kcal mol^{-1} M^{-1})	ΔG_f (kcal mol^{-1})	K_d
lysozyme	HNSB	2.6 ± 0.6[1]	-11.1 ± 1.6[1]	
	Optical Spectroscopy	2.33[2]	-8.9[3]	
lysozyme +$(NAG)_3$	HNSB	2.4 ± 0.6[4]	-10.4 ± 2.3[4]	33 μM (5-19 μM)[5]
cyt c	HNSB	3.5 ± 0.4[6]	-9.9 ± 0.9[6]	
	Fluorescence	3.42[7]	-9.6[7]	
BCA II	HNSB	1.5 ± 0.3[6]	-4.0 ± 0.5[6]	
	SUPREX	6.1 ± 0.4[8]	-16.1 ± 0.6[8]	
	SPROX	3.7 ± 0.7[9]	-8.2 ± 1.2[9]	
BCA II + brinzolamide	HNSB	4.4 ± 0.6[6]	-12.4 ± 1.5[6]	0.06 nM (3 nM)[10]

[1]Error is the standard deviation from 7 replicate measurements. [2]Value from reference (*76*). [3]Value from reference (*73*). [4]Error is the standard deviation from 4 replicate measurements. [5]Value from reference (*74*). [6]Fitting errors are reported. [7]Value from reference (*77*). [8]Value from reference (*13*). [9]Value from reference (*22*). [10]Value from reference (*75*).

2.4.3 Bottom-up proteomics analysis

The HNSB-modification protocol was combined with a bottom-up, shotgun proteomics strategy to characterize the protein folding behavior of unpurified proteins in a yeast cell lysate. As part of the quantitative, bottom-up proteomics analysis the proteins in each denaturant-containing buffer were precipitated with TCA after the HNSB reaction was quenched, reduced with TCEP, alkylated with MMTS, digested with trypsin, and labeled with a set of iTRAQ 8-plex reagents. Ultimately, an LC-MS/MS analysis of the iTRAQ-labeled samples was used to evaluate the denaturant dependence

of the HNSB-modification reaction. This was accomplished using the reporter ions of the iTRAQ 8-plex to quantify the relative amounts of wild-type (unmodified) Trp-containing peptides as a function of chemical denaturant concentration (Figure 10A). As proteins in the yeast cell lysate sample are unfolded in the presence of increasing concentrations of chemical denaturant, buried Trp residues become exposed, and are subject to modification with HNSB. Thus, the relative amount of a specific wild-type (unmodified) Trp-containing peptide derived from a globally protected region of protein structure will decrease as a function of the chemical denaturant concentration in the HNSB-modification reaction at the protein level. The iTRAQ data acquisition and analysis protocols used here are directly analogous to those we have previously described for SPROX(*15, 20*).

Summarized in Table 5 is the proteomic coverage obtained using the HNSB-modification protocol to characterize the chemical-denaturant induced equilibrium unfolding behavior of the proteins in a yeast cell lysate (Experiment 1). A total of 182 of the 250 unique Trp-containing peptides detected in Experiment 1 yielded high quality data with well-defined pre- and/or post-transition baselines (Figure 10A). Approximately 57% of these 182 Trp-containing peptides had $C_{1/2}$ values between 0.8 M and 1.3 M GdmCl. A total of 23 of the 250 unique Trp-containing peptides detected in Experiment 1 showed no transition, indicating that they were from either solvent exposed or very stable ($C_{1/2}$ > 4 M) regions of protein structure. A total of 41 of the 250

unique Trp-containing peptides detected in the LC-MS/MS analyses from Experiment 1 yielded poor quality data (Figure 10B), which precluded $C_{1/2}$ value assignments.

The compatibility of the HNSB labeling protocol with SPROX (see Figure 11) prompted us to investigate the utility of a hybrid protocol for the proteome-wide analysis of protein folding and stability. In the hybrid protocol (Figure 3), the protein samples in the chemical-denaturant-containing buffers were simultaneously labeled with both hydrogen peroxide and HNSB. After the labeling reactions were quenched, the samples were prepared for a bottom-up proteomics analysis using isobaric mass tags, as described above. In the hybrid protocol both the Met- and Trp-containing peptides detected in the LC-MS/MS readout can be used to report on the proteins to which they map. In order to maximize coverage of Met- and Trp-containing peptides in the LC-MS/MS readout a portion of the iTRAQ-labeled peptide mixture can be subject to a Met-containing peptide enrichment protocol and another portion of the iTRAQ-labeled peptide mixture can be subject to a Trp-containing peptide enrichment protocol.

Shown in Table 5 is a summary of the proteomics results obtained in two replicate analyses of the proteins in a yeast cell lysate using the hybrid protocol and both the Met- and Trp-containing peptide enrichment strategies (Experiments 2 and 3). The efficiency of the Met-containing peptide enrichment was comparable to that which we have previously reported(*20*). The Trp-containing peptide enrichment strategy increased the fraction of Trp-containing peptides detected in the LC-MS/MS readouts from 3-7% to 20-23%. More importantly, the hybrid protocol enabled the analysis of 750 unique

proteins with 1851 unique peptide probes (see Figure 12). The peptide and protein coverage achieved in these two biological replicates of the hybrid protocol was 50% and 25% (respectively) greater than that which would have been achieved using the SPROX protocol alone (Figure 12).

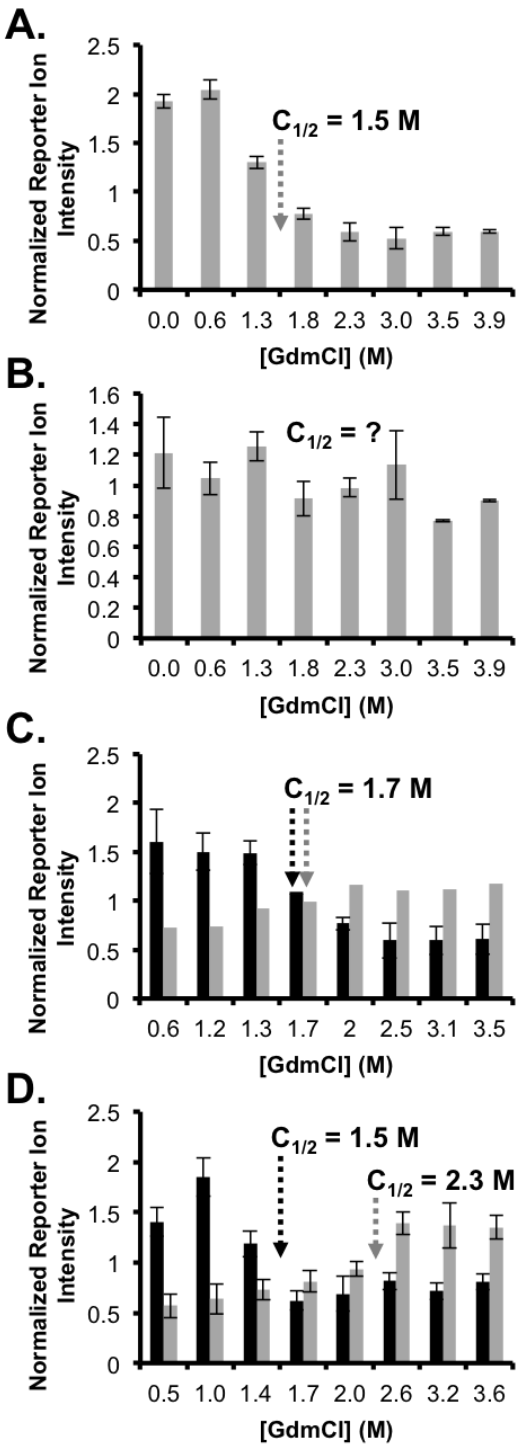

Figure 10: Representative data obtained using the HNSB protocol for proteome-wide analyses. (A) An example of high quality chemical denaturation data recorded in Experiment 1 for a wild-type Trp-containing peptide

(AIEMEGLTWGAHQFIPIGFGIK) from YAL003W, (B) An example of poor quality data obtained in Experiment 1 for a wild-type Trp-containing peptide (EKEVDQVIVVTVDNPFANQAWA) from YLR109W. (C) Chemical denaturation data recorded in Experiment 2 for YDR385W using wild-type (dark shaded bars) and HNSB-modified (light shaded bars) peptides of the same amino acid sequence (GTVAFGSGLHGWAFTIR). The data are an example in which the $C_{1/2}$ values of the wild-type and modified peptide probes are consistent indicating the HNSB modification does not alter the stability of the protein. (D) Chemical denaturation data recorded in Experiment 2 for YEL021W using wild-type (dark shaded bars) and HNSB-modified (light shaded bars) peptides of the same amino acid sequence (IAEWADITNAHGVVGPGIVSGLK). The data are an example in which the $C_{1/2}$ values of the wild-type and modified peptide probes are not consistent indicating the HNSB modification alters the stability of the protein. In cases where multiple product ion mass spectra were collected the averaged data are shown with error bars representing +/- one standard deviation. The arrows point to the denaturant concentration at the $C_{1/2}$ value extracted from each data set.

Table 5: Proteomic coverage obtained in bottom-up proteomics analyses.

Sample	Total[1]	Trp(wt)[2]	Met(wt)[3]
Experiment 1 (HNSB protocol)			
Nonenhanced run 1	2839(763)	200(137)	419(235)
Nonenhanced run 2	2822(754)	199(137)	399(236)
Nonenhanced Runs 1 and 2 Combined	3244(866)	250(161)	484(271)
Experiment 2 (Hybrid protocol)			
Nonenhanced	3366(768)	107(76)	116(84)
Trp-enhanced	1916(568)	384(226)	NA[4]
Met-enhanced	1941(546)	NA[4]	539(309)
Nonenhanced+enhanced	NA[4]	439(244)	575(333)
Experiment 3 (Hybrid protocol replicate)			
Nonenhanced	2591(764)	115(90)	106(87)
Trp-enhanced	1719(554)	393(213)	NA[4]
Met-enhanced	1479(610)	NA[4]	1040(501)
Nonenhanced+enhanced	NA[4]	436(234)	1091(535)

[1]Total number of unique peptide (proteins) identified. [2]Total number of unique wild-type Trp-containing peptides (proteins) identified. [3]Total number of unique wild-type Met-containing peptides (proteins) identified. [4]Not applicable.

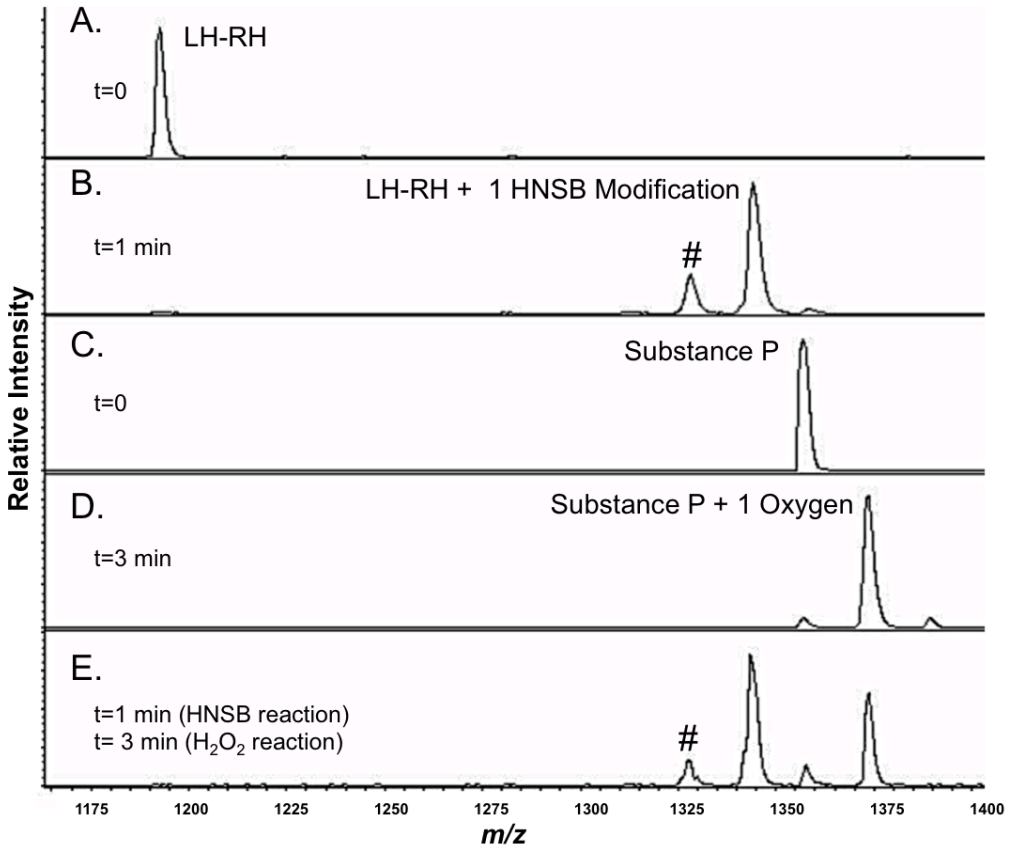

Figure 11: MALDI-TOF mass spectra obtained on model peptides subjected to the HNSB and H_2O_2 labeling conditions in the HNSB, SPROX, and hydrid protocols. The LH-RH (EHWSYGLRPG) and Substance P (RPKPQQFGLM) peptides contain 1 trp and 1 met residue, respectively. (A) and (B) Mass spectra obtained before and after (respectively) reaction of LH-RH with HNSB in the HNSB protocol. (C) and (D) Mass spectra obtained before and after (respectively) reaction of substance P with hydrogen peroxide in the SPROX protocol. (E) The simultaneous reaction of LH-RH and Substance P with HNSB and H_2O_2 in the hybrid protocol. In each case the peptide concentration was ~16 mM, and the H_2O_2 and HNSB concentrations were 1 M and 1.6 mM, respectively. The ion signal indicated with a "#" is a MALDI-TOF artifact consistent with that previously reported for nitro aromatic compounds(71). The other ion signals are from the [M+H]+ species of the indicated analyte.

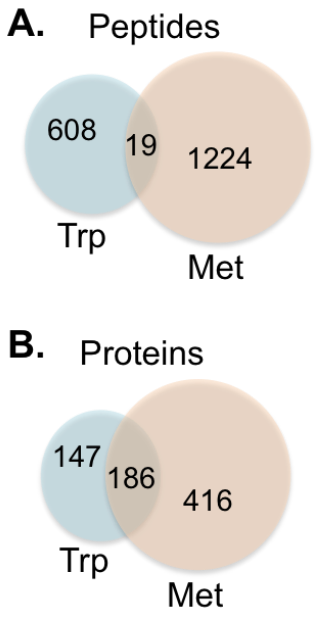

Figure 12: Venn diagrams showing the peptide (A) and protein (B) coverage obtained using the Met- and Trp-containing peptide probes detected in Experiments 2 and 3 using the hybrid protocol.

2.5 Discussion

2.5.1 Evaluation of thermodynamic parameters

The HNSB-based covalent labeling protocol developed here is closely related to SPROX. As with SPROX there are several caveats associated with using the described approach to evaluate the ΔG_f and *m*-values (i.e., $\partial\Delta G_f/\partial$[Denaturant]) of protein folding reactions.(22) These caveats include: (1) that the GdmCl-induced equilibrium folding/unfolding reaction of the protein be well-modeled by a reversible, two-state process, (i.e., significant populations of partially folded intermediate states cannot exist); (2) that the refolding rate of the protein be greater than the modification rate of an unprotected Trp; (3) that the protein of interest contain at least one globally protected Trp residue; (4) that the folding behavior of the HNSB modified protein not deviate

significantly from that of the wild-type protein; and (5) that the wild-type protein and HNSB-modified protein reaction products have similar ionization efficiencies in MALDI. The last caveat specifically applies to intact protein analyses using the MALDI-TOF read-out. The two-state folding caveat is important in any experiment to evaluate ΔG_f and *m*-values from chemical denaturation data(*78*). The need to have protein folding kinetics that are faster than the labeling reaction kinetics is important in the derivation of the equations used to analyze our HNSB-labeling data (i.e., Equation 2 and 3).

An important feature of the HNSB modification approach described here is that a single globally protected tryptophan residue in a protein can report on the thermodynamic stability and ligand binding properties of the entire protein, provided the protein is a two-state folding protein. For example, the chemical denaturant dependence of the HNSB labeling reaction involving the single tryptophan residue in the two-state folding protein, cyt c(*79, 80*), was used here to accurately evaluate the ΔG_f and *m*-values of this two-state folding protein (Table 4). In the case of non-two-state folding proteins, the HNSB labeling protocol does not yield meaningful ΔG_f and *m*-values for a protein's global unfolding reaction. However, in such cases the labeling protocol can still be used for the qualitative analysis of ligand binding by measuring the change in $C_{1/2}$ values upon ligand binding. For example, the binding of brinzolamide to the non-two-state folding protein, BCA II(*81*), resulted in a $C_{1/2}$ value shift of 0.5 M (Figure 9C).

In the case of large multi-domain proteins that do not unfold in a highly concerted manner, the LC-MS/MS readout in the quantitative, bottom-up proteomics analysis can be used to evaluate the folding and ligand binding behavior of individual domains using the HNSB or hybrid protocols (Figure 3). Theoretically, tryptic peptide probes derived from the same protein folding domain should yield the same ΔG_f values. This was observed, for example, in the multi-domain subunit A of the V1 peripheral membrane domain of V-ATPase, in which the hedgehog/intein domain yielded a ΔG_f of -4.2±0.2 kcal/mol using 2 unique Met-containing peptide probes and a value of -4.5 kcal/mol using a Trp-containing probe. Interestingly, Met- and Trp-containing peptide probes from another region of this protein (domain not assigned) also yielded consistent ΔG_f values (-3.2±0.5 kcal/mol and -3.5±0.5 kcal/mol, respectively). However, this other region of the protein was less stable than the hedgehog/intein domain.

The internal consistency of our Met- and Trp-containing peptide probe data was further assessed using the 186 proteins identified with both Met- and Trp-containing peptide probes in Experiments 2 and 3 (Figure 12B). The Met- and Trp-containing peptide probes were mapped to specific domains using the SCOP (Structural Classification of Proteins) database (1.75 release, June 2009)(*82*), and the 134 protein domains that were successfully analyzed with unique Met- and Trp-containing peptides are summarized. Among the 134 domains for which ΔG_f values were calculated using both Met- and Trp-containing peptide probes, 111 (or ~83%) of the domains had ΔG_f values that were consistent (i.e., within 30%) between the Met- and Trp-containing

probes. This redundancy in the hybrid protocol increases the confidence of $C_{1/2}$ value assignments and ΔG_f value determinations.

2.5.2 Accuracy and precision

A potential complication associated with using the HNSB and hybrid protocols described here to evaluate the thermodynamic stability of proteins and protein domains is that the HNSB modification of Trp residues and/or the oxidation of Met residues can perturb the folding/unfolding equilibrium. We note that such a complication was not observed in our use of the HNSB protocol to analyze cyt c and lysozyme using the intact protein readout. We also note that the above complication is minimized when wild-type Trp- and Met-containing peptide probes are used in the quantitative bottom-up proteomics readout, as such probes must come from protein molecules that were never modified. In theory, wild-type Trp- and Met-containing peptide probes and their corresponding modified peptide probes should yield similar $C_{1/2}$ values (Figure 10C). However, if the modification perturbs the folding/unfolding equilibrium then the $C_{1/2}$ value of the modified peptide probe will be altered (Figure 10D).

It is possible that modification of a Trp- or Met-residue in one region of the protein could disrupt the folding behavior of another region of the protein. In such cases, even the $C_{1/2}$ and ΔG_f values obtained using wild-type Trp- and Met-peptide probes from the affected region will be compromised. In order to evaluate the potential of such structural perturbations to impact $C_{1/2}$ and ΔG_f value determinations using the HNSB and hybrid protocols, we compared chemical denaturation data obtained for the same

peptide probe in the HNSB, SPROX, and hybrid protocols. Comparison of data obtained on wild-type Trp-peptide probes in the hybrid and HNSB protocols enabled an evaluation of the effects of Met oxidation. Similarly, the effects of Trp-modification could be assessed by comparing data obtained on wild-type Met-peptides in the hybrid and SPROX protocols. The SPROX data set used here was from the yeast cell lysate analysis performed in reference(*27*).

A total of 108 Met-containing peptide probes in Experiment 3, which employed the hybrid protocol, overlapped with those previously reported in the SPROX experiment, and 89% of these Met-peptide probes showed consistent $C_{1/2}$ values (i.e., midpoints that shifted as a result of no more than one isobaric mass tag difference). A total of 96 Trp-peptide probes assayed in the Experiment 1, which employed the HNSB protocol, overlapped with those in Experiment 3, which employed the hybrid protocol, and 95% percent of these Trp-peptide probes showed consistent $C_{1/2}$ values. These results suggest that the accuracy of $C_{1/2}$ and ΔG_f values was not compromised by the Met- and Trp-modification reactions in the large majority of cases.

A comparison of the $C_{1/2}$ values obtained on the same Trp- and Met-peptide probes in Experiments 2 and 3 provides a measure of the biological variability in the described protocol. The biological variability appears to be in the range of 3-8% (i.e., this is the fraction of $C_{1/2}$ values that differed by >0.5 M and were deemed inconsistent). A comparison of the $C_{1/2}$ values generated for the 128 Trp-containing peptide probes assayed in both of the technical replicates performed in Experiment 1, revealed that only

2% of the $C_{1/2}$ values were significantly shifted (i.e., different by >0.6 M). This technical reproducibility is similar to that which we have previously reported for the SPROX protocol. The fraction of Met-peptide probes that showed inconsistencies in the above comparison (11%) is slightly higher than that expected from the biological variability in the thermodynamic measurements performed here (3-8%). These results suggest that the HNSB modification of Trp residues does perturb the equilibrium unfolding properties of proteins in a relatively small number of cases.

Side reactions of H_2O_2 and HNSB with residues other than Met and Trp (respectively) can potentially compromise the accuracy of ΔG_f value determinations. However, such side reactions only become an issue if they produce significant amounts of a protein product that has a different stability from the wild-type protein. The H_2O_2 and HNSB reaction conditions in the described protocols are optimized to specifically target Met and Trp residues. Our ability to detect many Cys residues that are only modified with MMTS in our Trp and Met-probes suggests that such side reactions involving HNSB and H_2O_2 (which are probably most likely to occur at Cys residues) are not significant. This, combined with the internal consistency of our results generated using Met- and Trp-containing probes in the HNSB, SPROX, and hybrid protocols, suggests that such side reactions do not significantly interfere with the experiment.

2.6 Conclusions

An HNSB modification protocol is described that enables the thermodynamic properties of protein folding and ligand binding interactions to be evaluated. The HNSB

protocol can generate reasonably accurate thermodynamic measurements (ΔG_f, and K_d values) for two-state folding protein and protein-ligand systems. The technique is also amenable to proteome-wide analyses using quantitative, bottom-up, shotgun proteomics strategies. Moreover, the HNSB protocol can be performed simultaneously with the SPROX protocol, and use of the two protocols together significantly increases the peptide and protein coverage in proteome wide analyses. The increased proteomic coverage in the hybrid protocol will greatly benefit future applications of the methodology such as the high-throughput identification of protein targets in drug mode-of-action studies.

3. Measurement of geldanamycin binding affinity with Hsp90 in cell lysates using the Hybrid protocol

3.1 Introduction

Hsp90 is an important chaperone protein that is involved in cellular signaling. Hsp90 binds to substrate proteins at a late stage of folding when proteins are nearly folded into their native conformations(*83, 84*), and many of its substrate proteins are involved in signal transduction. Hsp90 operates under stress conditions(*84*), and its function is regulated by many different protein-protein interactions with co-chaperones including Hsp70, Hsp40, Hop/Sti1, Aha1, p50/Cdc37, Pih1, Tah1 and p23/Sba1(*84-88*). Hsp90 inhibitors are attractive drugs for cancer therapy because Hsp90 is involved in maturation of oncogenic signaling proteins and has been shown to selectively target cancer cells over normal cells(*89-93*).

Geldanamycin (Figure 13) is one of several natural product Hsp90 inhibitors that exhibit anticancer activity by disrupting the association of Hsp90 with client proteins(*94-100*). Geldanamycin binds to the N-terminal ATP-binding domain of Hsp90, inhibits the ATPase activity of Hsp90, and prevents binding of a co-chaperone, p23/Sba1, to Hsp90(*101-104*). The effects of geldanamycin and a number of geldanamycin derivatives against cancer and disease states have been well-studied over the years(*91, 99, 105*). One interesting observation is that the binding affinity (i.e., K_d value) of geldanamycin and its derivatives for purified Hsp90 is on the order of 1 μM, whereas the IC_{50} is in low nanomolar range(*106-109*).

A number of different hypotheses have been proposed to explain the discrepancy between the binding affinity of geldanamycin and its IC_{50} value. One hypothesis for the low IC_{50} value is that other protein-protein interactions in the cell actually modulate the binding affinity, and the K_d value determined to date have necessarily been determined on highly purified sample preparations of Hsp90. Interestingly, studies have shown that the binding affinity for the Hsp90-geldanamycin interaction in tumor cells is lower than in normal cells(*89, 107, 108, 110*). In one study, the binding affinity of a geldanamycin derivative 17-AAG to Hsp90 was investigated in various human cell lysates(*89*). The IC_{50} value of ~40 nM in MCF-7 cells was nearly 20-fold lower than that measured in non-tumorigenic cells such as Primary Renal Proximal Tubule Epithelial Cells (RPTEC). The increasing binding affinity observed in the MCF-7 cells was attributed to the elevated levels of Hsp90 co-chaperones, p23/Sba1 and Hop, in the MCF-7 cells(*89*). Although it is noteworthy that the above study determined IC50 values based on pull-down experiment using biotinylated-geldanamycin probe followed by Western blot quantitation(*89*) and did not establish a direct measurement of the binding affinity. In fact the only true thermodynamic measurements of geldanamycin binding to Hsp90 have been made on purified Hsp90.

Here we report on the use of the Hybrid protocol(*55*) described in Chapter 2, to measure the binding affinity of geldanamycin to Hsp90 in the context of two different cell lysates. The yeast (normal cell line) and MCF-7 cell (tumor cell line) lysates used in this study each contains Hsp90, but a larger fraction of Hsp90 is complexed with co-

chaperones (e.g., p23/Sba1 and Hop) in the tumor cells than in the normal yeast cells. The Hybrid protocol utilized in this work enables the quantitative thermodynamic analysis of protein folding and ligand binding reactions in the context of whole cell lysates, and the technique has been established in a series of different model systems(*19-21*). The results of this work suggest that the binding affinity of geldanamycin for Hsp90 is not modulated by endogenous protein-protein interactions, at least under the conditions of these experiments, which include the use of a 30 min protein-drug equilibrium time in the lysates, prior to SPROX analysis.

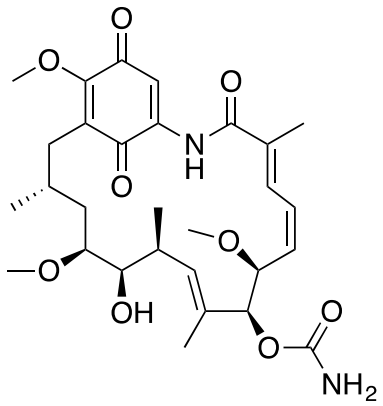

Figure 13: chemical structure of geldanamycin, known Hsp90 inhibitor.

3.2 Experimental

3.2.1 Materials

The Saccharomyces cerevisiae open reading frame strain Y258 was purchased from Open Biosystems (Waltham, MA). The MCF-7 cell line was purchased from Duke Cell Culture Facility (Durham, NC). The following reagents were from Sigma-Aldrich (St. Louis, MO): trypsin from porcine pancreas (proteomics grade, dimethylated), L-

methionine, L-tryptophan, HNSB (≥95 wt%), GdmCl, MMTS, TEAB buffer (1.0 M, pH 8.5), SDS, H_2O_2 (30%, w/w), ammonium acetate, TMP, pyrrolidine, TCA, fetal bovine serum (FBS) and TFA. The following materials were from Thermo Scientific (Waltham, MA): TCEP•HCl, UltraLink hydrazide resin and Coomassie Plus Bradford Assay reagent. The following cell culture reagents were from Gibco (Grand Island, NY): DMEM (containing 4.5 g/L D-Glucose, L-Glutamine, 110 mg/L sodium pyruvate), MEM NEAA (100×), Antibiotic-Antimycotic (Anti-Anti, 100×) and insulin (human recombinant zinc). HyClone™ HyQTase™ was from GE Healthcare Life Sciences (Logan, UT). Methanol and isopropyl alcohol were from VWR International (West Chester, PA). MacroSpin columns (Silica C18) and Pi³™ methionine reagent kits were from the Nest Group (Southborough, MA). Ethanol (completely denatured) was from Avantor Performance Materials, Inc. (Phillipsburg, NJ). Sodium phosphate dibasic was from EM Science (Gibbstown, NJ). Sodium phosphate monobasic was from Mallinckrodt Backer, Inc. (Paris, KY). Geldanamycin (≥98 wt%) was from Chem-Impex International, Inc. (Wood Dale, IL). Dimethyl sulfoxide (DMSO) was from Acros (NJ). Zirconia/silica beads (1 mm) were from Biospec Products (Bartlesville, OK).

3.2.2 Cell culture and cell lysate preparations

Yeast cell lysates were prepared as described in Chapter 2. MCF-7 cells were grown in DMEM media supplemented with NEAA, Anti-Anti, 10% FBS and 10 μg/ml insulin. The cells were washed twice with PBS before harvesting with HyQTase in a 15 mL tube. The harvested cells were immediately spun down at 800 rpm for 6 min. The

pelleted cells were then stored at -20°C for future use. Cell pellets were lysed with pH 7.4 20mM phosphate buffer containing protease inhibitor cocktail which included 1 mM AEBSF, 10 μM pepstatin A, 20 μM leupeptin, 15 μM E-64 and 50 μM bestatin. Cell lysis was accomplished with zirconia/silica beads (1 mm) and 20 s of disruption using Disruptor Genie (Scientific Industries, Bohemia, NY) 10 times with 1 min intervals on ice in between. The suspension was centrifuged at 15,000g for 20 min at 4°C and the supernatant was used in the following experiments as cell lysate.

3.2.3 Hybrid protocol

Both the yeast and MCF-7 lysates were subjected to the same Hybrid protocol as described in Chapter 2. For each analysis, the lysate (15-17 mg/mL determined by Bradford assay) was divided into two 180 μL portions, with 20 μL DMSO added to the (-) ligand group, 20 μL 500 μM geldanamycin stock solution (in DMSO) to the (+) ligand group. The lysate was equilibrated with the drug for 30 min before 20 μL aliquots of the (-)/(+) lysate were combined with 25 μL volumes of a series of pH 7.4 20 mM phosphate buffers containing increasing concentrations of GdmCl. The final eight GdmCl concentrations used in Yeast-Hybrid and MCF-7-Hybrid experiments are summarized in Table 6. The 16 denaturant-containing protein samples were equilibrated at room temperature for 30 min before reaction was initiated by adding 5 μL H_2O_2 (30%, w/w) to each sample. After 2 min, 2.5 μL of a 46.2 mM solution of HNSB was spiked into each sample and the reaction was quenched by 1 mL quenching cocktail containing 0.5 M L-methionine and 4.9 mM L-tryptophan at 3 min. The proteins in each sample was

precipitated by addition of 200 μL of 1g/mL TCA and incubated on ice at 4°C overnight. The samples were then centrifuged at 8,000g, 4°C for 30 min. The resulting protein pellets were washed 3 times with 300 μL pre-chilled ethanol and dried in a Savant™ SpeedVac™ concentrator (Thermo Scientific, Waltham, MA) for 5 min.

To redissolve protein pellets, 30 μL 0.5 M of TEAB buffer (pH 8.5) and 1.5 μL of a 2% (w/v) SDS solution were added to each sample. The samples were sonicated, vertexed, and heated at 60°C to facilitate their dissolution. Disulfide bond reduction was carried out with addition of 3.5 μL 50 mM TCEP•HCl and incubation at 60°C for 1 h. A 1.8 μL aliquot of 200 mM MMTS was added to each sample before a 10 min incubation at room temperature. The proteins in each sample were digested with 2 μL trypsin (1 mg/mL in 1mM HCl) at 37°C overnight.

Each of the (-) and (+) peptide samples was labeled with 0.5 unit of an iTRAQ-8plex tag dissolved in 50 μL isopropanol and the labeling scheme can be found in Table 6. The labeling reactions were allowed to proceed at room temperature for 2 h. Within each set, (-) or (+), 10 μL or 30 μL of each sample was combined. The combined samples, either 80 μL or 240 μL in volume, were either subjected to a desalting protocol using C18 MacroSpin columns (80 μL sample), a methionine-containing peptide enrichment strategy (240 μL sample), or a tryptophan-containing peptide enrichment strategy (240 μL sample) prior to LC-MS/MS analysis. The above 3 fractionation protocols were performed as described in Chapter 2. For the MCF-7-Hybrid experiment, the C18 purified samples (i.e., the non-enhanced samples) were each subjected to a single LC-

MS/MS analysis while the methionine/tryptophan-enriched samples were each subjected to two replicate LC-MS/MS analyses. For the Yeast-Hybrid experiment, all samples were each subjected to a single LC-MS/MS analysis.

LC-MS/MS analyses were performed at the Proteomics Resource at the Fred Hutchinson Cancer Research Center on an Orbitrap Elite ETD mass spectrometer equipped with an EASY-nLC system (ThermoFisher Scientific Inc., Waltham, MA). The instrument parameters were the same as described in Chapter 2. The data were searched using Proteome Discoverer with fixed modifications of MMTS on cysteine and iTRAQ® 8-Plex on N-terminus and lysine residues. Proteome Discoverer was set with a variable modification of oxidation on methionine residues. In all cases only peptides identified with high and medium confidence (i.e., false discovery rates (FDR) <1% and 5%, respectively) were used in the data analysis.

Table 6: Summary of the iTRAQ labeling scheme used in the 2 Hybrid experiments. Shown are the GdmCl concentrations (M) of the samples corresponding to each of the 8 iTRAQ tags.

Experiment	iTRAQ reporter ion m/z							
	113	114	115	116	117	118	119	121
Yeast-Hybrid	0	0.5	1.0	1.2	1.6	1.9	2.4	2.8
MCF-7-Hybrid	0	0.5	1.0	1.3	1.5	1.7	2.0	2.5

3.3 Data analysis

3.3.1 Hybrid protocol data analysis

The iTRAQ reporter ion intensities measured in the LC-MS/MS analyses were normalized using similar strategy as reported(*15*). Only peptides with raw iTRAQ

intensities that summed to greater than 1000 were included in the analysis. Briefly, the 8 iTRAQ reporter ion intensities in each product ion mass spectra were averaged and the raw intensity of each reporter ion in the product ion mass spectra was divided by the average value. This process is referred to as N1 normalization. Then the N1-normalized values for all the non-Trp and non-Met-containing peptides were averaged within each reporter ion, to generate so-called N2-normalization factors for the Trp- and Met-containing peptides in each experiment. Summarized in Table 7 are the N2-normalization factors generated for each reporter ion (and corresponding denaturant concentration) in both experiments. The above normalization procedures were only carried out for non-enhanced samples (samples without Trp or Met enrichment). Ultimately, the N2-normalization factors were used in a second so-called N2 normalization, where the N1-normalized values for each Trp (or Met)-containing peptide were divided by the corresponding N2-normalization factors to obtain the final normalized reporter ion intensity in this work. If multiple product ion spectra were obtained for a given peptide (e.g., it was identified multiple times in a single LC-MS/MS run) the normalized reporter ion intensity values were averaged using Runcompare (an AWK script developed in-house). Only product ion spectra with isolation interferences smaller than 30% were used for averaging. In the end, for either methionine or tryptophan probes, 3 different data sets containing peptides with their averaged reporter ion intensities were obtained: one data set containing peptides identified in the (-) group,

one containing peptides identified in the (+) group, and another containing the peptides identified in both (-) and (+) groups.

For each of the matched peptides, the averaged reporter ion intensities of the (-) group were subtracted from those of the (+) group, obtaining differences for all 8 iTRAQ tags. The range of differences encompassing ~70% of the data was determined and then the differences were filtered to identify peptides with at least 2 consecutive differences in the outlying 30% of the data. Such peptides were pasted into a new sheet for midpoint assignment. For the purpose of hit identification, the midpoint assigned here was not represented by the actual [GdmCl] (M), instead was represented by the corresponding iTRAQ tag number. For example, the lowest [GdmCl] was assigned tag 1, the second lowest tag 2, and so on, the highest [GdmCl] tag 8.

The distribution of N2-normalized reporter ion intensity values for the tag corresponding to the highest [GdmCl] and the lowest [GdmCl] were plotted for the Trp- and Met-containing peptides in each experiment (Figure 14). The point at which the two distributions crossed best separated the pre- and post-transition baselines of the denaturation curves. The midpoint of the curve was the [GdmCl] corresponding to the reporter ion intensity value. This value was 1.0 for both Trp- and Met-containing peptides (Figure 14). Typically, the tags flanking the transition were averaged and the average value was taken as the transition midpoint. If there was a normalized reporter ion intensity of 1.0 ± 0.1 at the transition, then that tag was assigned as midpoint. If multiple normalized reporter ion intensity values of 1.0 ± 0.1 existed at the transition, the

average of all these tags was assigned as midpoint. Chemical denaturation data sets in which more than one normalized reporter ion intensity was not >1.0 or <1.0 in the pre- and post-transition baselines, respectively, were classified as "poor quality" and not assigned a midpoint. When only a single outlying value existed, the value was removed from the data set and the remaining 7 values were used for denaturation curve construction and transition midpoint assignment. In cases where the outlying value was at or near the transition, two alternate midpoints could be assigned, and were classified as "ambiguous". In some cases the reporter ion intensities were all within 10% of 1.0 and these cases are indicated with an "NT" for "no transition."

Peptides were selected with midpoint shift ≥ 1.5 iTRAQ tag to either a higher or lower [GdmCl]. Among these peptides, hit peptides were identified with at least two consecutive intensity differences at or between the [GdmCl] range defined by the (-) and (+) midpoints.

Table 7: N2-normalization factors generated for each reporter ion in both Hybrid experiments.

Experiment	Group	Statistics	iTRAQ reporter ion m/z							
			113	114	115	116	117	118	119	121
Yeast-Hybrid	(-)	Average	0.94	0.81	0.94	1.05	1.12	0.95	1.15	1.03
		S.D.	0.19	0.20	0.13	0.20	0.17	0.14	0.16	0.24
	(+)	Average	1.12	0.91	0.94	0.99	1.10	0.87	0.96	1.10
		S.D.	0.32	0.19	0.15	0.18	0.18	0.15	0.17	0.21
MCF-7-Hybrid	(-)	Average	0.95	0.93	1.04	1.12	0.98	0.75	1.14	1.07
		S.D.	0.16	0.19	0.14	0.20	0.16	0.20	0.15	0.18
	(+)	Average	0.57	1.02	1.10	1.13	1.03	0.48	1.30	1.36
		S.D.	0.30	0.16	0.16	0.17	0.17	0.28	0.19	0.22

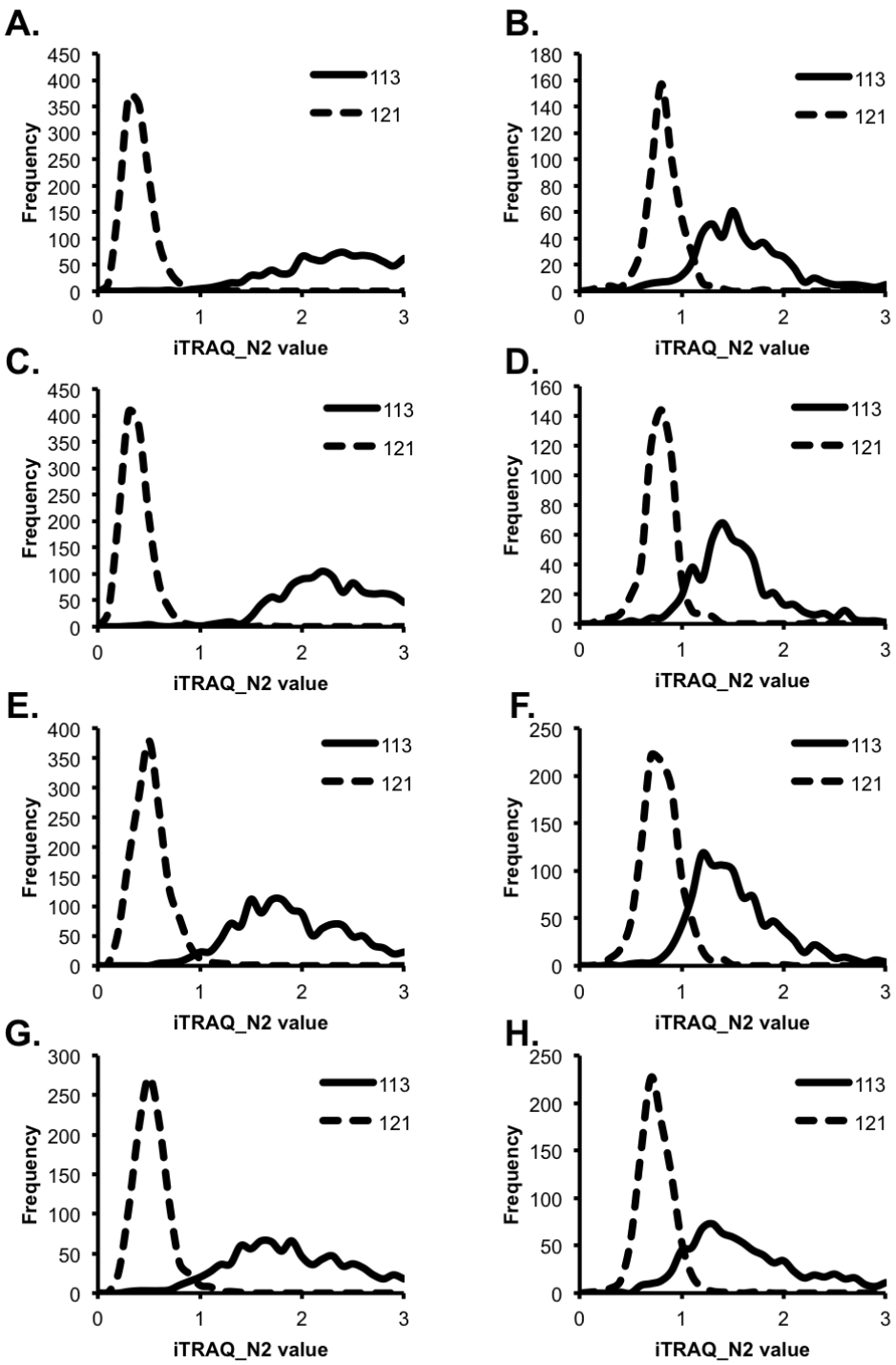

Figure 14: Distribution of the N2-normalized iTRAQ reporter ion intensity ratios observed for the wild-type Trp- and Met-containing peptide probes in (A)-(D) Yeast-Hybrid and (E)-(H) MCF-7-Hybrid. Distribution of the wild-type Met peptides

is shown in (A), (C), (E), (G), while the distribution of the wild-type Trp peptides is shown in (B), (D), (F), (H). Shown in each case is the data for the reporter ions at the lowest (iTRAQ 113 m/z reporter ion, pre-transition) and highest (iTRAQ 121 m/z reporter ion, post-transition) denaturant concentration used in each case.

3.3.2 Binding affinity calculation

To calculate K_d value, midpoint differences of the hit peptides above were converted from tag to the corresponding [GdmCl] (M). Since the final geldanamycin concentration 20 μM in large excess over the concentration of any individual protein in the lysates, Equation 4 and Equation 5 were used to determine binding free energy $\Delta\Delta G_f$ and K_d.

$$\Delta\Delta G_f = -m \times \Delta C_{1/2} \qquad \text{Equation 4}$$

$$K_d = \frac{[L]}{e^{-\Delta\Delta G_f/RT} - 1} \qquad \text{Equation 5}$$

In Equation 4, $\Delta C_{1/2}$ is the midpoint difference in [GdmCl] (M). A m-value of 2.6 kcal $mol^{-1}M^{-1}$ was used, estimated based on the average protein domain size of 100 amino acids and the average contribution of 0.026 kcal $mol^{-1}M^{-1}$ of GdmCl to the m-value of the protein(*111*). In Equation 5, [L] is the free ligand concentration. Due to the large excess of geldanamycin, 20 μM was directly used for [L]. R is the universal gas constant, T is temperature in Kelvin. For the calculation, R was 0.001987 kcal $mol^{-1}K^{-1}$, and T was 298.15 K.

3.4 Results and discussion

3.4.1 Peptide and protein coverage

The proteins in S. cerevisiae Y258 and MCF-7 cell lysates were probed for geldanamycin binding using the Hybrid protocol, which involves the chemical modification of methionine and tryptophan residues in proteins with H_2O_2 and HNSB, respectively. The experimental workflow used in the Hybrid protocol is illustrated in Figure 15. The yeast and MCF-7 cell lysates were each analyzed using the Hybrid protocol both in the presence and absence of geldanamycin. The Hybrid protocol samples were analyzed using Orbitrap Elite mass spectrometer and the resulting proteomic coverage is summarized in Table 8.

Greater proteomic coverage was observed using the Hybrid protocol compared to using the methionine oxidation strategy alone. Based on Table 8 as well as the number of overlapping peptides between the Met and Trp probes (data not shown), for Yeast-Hybrid, the total number of peptide probes was increased by ~37%, leading to a ~17% increase in protein coverage compared to using the Met probes alone; for MCF-7-Hybrid, the total number of peptide probes was increased by ~66%, leading to a ~58% increase in protein coverage. If the enriched samples from Yeast-Hybrid experiment were each subjected to 2 replicates of LC-MS/MS analysis instead of only 1, the proteomic coverage increase would have been greater. Apart from the total peptide and protein coverage, the advantage of using the Hybrid protocol can also be demonstrated with Hsp90 peptides (Table 9). Additional coverage was obtained by using the Trp probe: for yeast

Hsp90, the Hybrid protocol successfully identified 1 Trp peptide from its C-terminal domain; for human Hsp90, the protocol identified 1 Trp peptide from its N-terminal domain and 2 from its large middle domain.

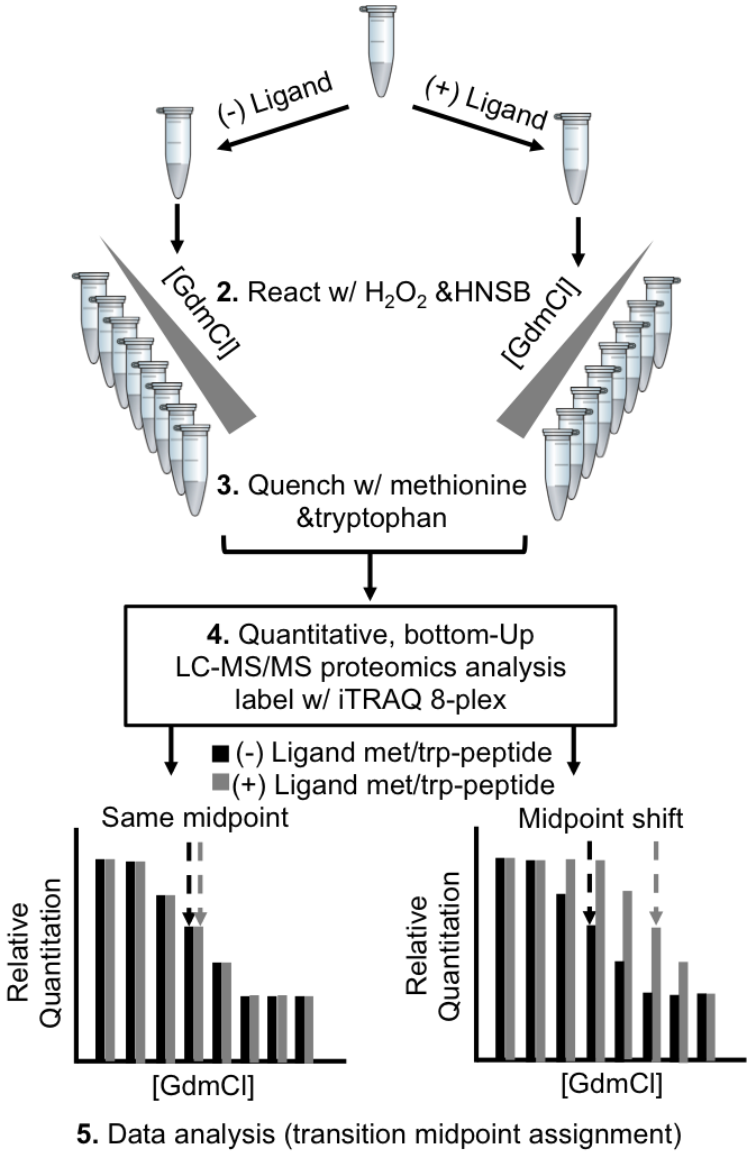

Figure 15: Schematic representation of the experimental workflow for the Hybrid protocol.

Table 8: Protein and peptide coverage achieved in the 2 geldanamycin Hybrid experiments.

Experiment	Met-Assayed[1]	Trp-Assayed
Yeast-Hybrid	1202 (474)[2]	484 (264)
MCF-7-Hybrid	1104 (546)	761 (470)

[1]The number of assayed methionine/tryptophan peptides (proteins), which are peptides identified in both (-) and (+) groups. [2]The number of unique peptides (proteins).

Table 9: Assayed Hsp90 peptides in the 2 experiments and their corresponding protein domains.

Experiment	Peptide sequence	Protein domain[1]
Yeast-Hybrid	DSGIGMTK	N-terminal
	EMLQQNK	Large middle
	MPEHQK	Small middle
	TGQFGWSANMER	C-terminal
	DSSMSSYMSSK	
MCF-7-Hybrid	**TLTIVDTGIGM(OX)TK**[2]	N-terminal
	HNDDEQYAWESSAGGSFTVR	
	EMLQQSK	Large middle
	SLTNDWEDHLAVK	
	TKPIWTR	
	HGLEVIYM(OX)IEPIDEYCVQQLK	Small middle
	YYTSASGDEMVSLK	
	IMKDILEK	C-terminal
	DNSTM(OX)GYM(OX)AAK	

[1]Domain information for yeast Hsp90 was obtained from reference (*101*) and the domain information for human Hsp90 was obtained from reference (*112*). [2]This bolded peptide is identified as a hit in MCF-7-Hybrid experiment.

3.4.2 Evaluation of Hsp90-geldanamycin binding affinity

A number of methionine- and tryptophan-containing peptides from Hsp90, the known direct target of geldanamycin, were assayed in experiments described here. These peptides came from different Hsp90 domains (Table 9). One advantage of the Hybrid protocol is that one peptide probe can report on the binding properties of the domain to which it maps. As expected the hit peptide detected in the human lysate

experiment (TLTIVDTGIGM(OX)TK (M98)) was from the N-terminal ATP binding domain of Hsp90 (Table 9), which is also known to be where geldanamycin binds(*101, 103, 104*). Interestingly, the yeast peptide DSGIGMTK was not a hit in Yeast-Hybrid due to insignificant midpoint shift (Figure 16A) but it is noteworthy that this peptide was identified as a hit in a previous iTRAQ-SPROX experiment(*113*), which only utilized the methionine oxidation reaction. One explanation is that the HNSB modification of the Trp residues might perturb the three-dimensional structure of Hsp90 in a way that precludes binding(*55*). A comparison of the yeast and human Hsp90 protein sequences reveals that all the Trp residues are conserved between the yeast and human Hsp90 except for W357 in yeast Hsp90(*114*). It is possible that the HNSB modification of W357 abrogated binding of geldanamycin to the N-terminal domain of Hsp90 via an allosteric effect.

The chemical denaturant data for the hit methionine-containing peptide can be used to calculate a K_d value for the Hsp90-geldanamycin complex. The observed $\Delta C_{1/2}$ value of 0.6 M (Figure 16B) was used to calculate a K_d value of 1.2 μM for the Hsp90-geldanamycin complex in yeast lysate, which is identical to the previously established K_d value of 1.2 μM measured by isothermal titration calorimetry with purified protein(*104*). Similarly, the observed $\Delta C_{1/2}$ value of 0.7 M (Figure 16C) was used to calculate a K_d value of 1.0 μM for the Hsp90-geldanamycin complex in MCF-7 cell lysate. This is in agreement with the range of K_d values from 0.08 μM to 0.60 μM determined for purified human Hsp90 with similar protein-drug equilibrium times of 1-1.5 h(*89,*

115). The K_d value of 1.0 μM is consistent with the upper limit of the range of previously measured values.

The goal of this study is to compare the binding affinity of Hsp90-geldanamycin complex in the two different cell lysates studied here. The K_d values of the Hsp90-geldanamycin are comparable in yeast lysate and in MCF-7 cell lysate using protein-drug equilibrium time of 30 min, which seems to suggest that there is no significant binding affinity difference between normal and tumor lysates. The results presented here suggest that the endogenous protein-protein interactions of Hsp90 with its co-chaperones do not affect Hsp90-geldanamycin binding affinity.

The results of other studies have suggested that the Hsp90-geldanamycin complex requires an extended incubation time to reach equilibrium(*115*). It has been suggested that geldanamycin binds to Hsp90 slowly and requires >10 h to reach equilibrium(*115*). In order to evaluate the binding affinity at long equilibration times, the Hybrid protocol could be performed using a long >10 h protein-drug equilibration time. The results of a recent iTRAQ-SPROX analysis using a 24-hour protein-drug equilibration time suggest that this slow-tight binding property of human Hsp90 may exist (Figure 16D). Using an extended denaturant concentration range (i.e., 0.5-4.1 M), the slow, tight binding event was successfully detected with a K_d value of 30 nM (Figure 16D), which was ~30 times tighter than the K_d determined for the 30 min incubation time. This K_d value was roughly consistent with the previously reported 9 nM for the human Hsp90-geldanamycin complex after 24-hour equilibrium(*115*).

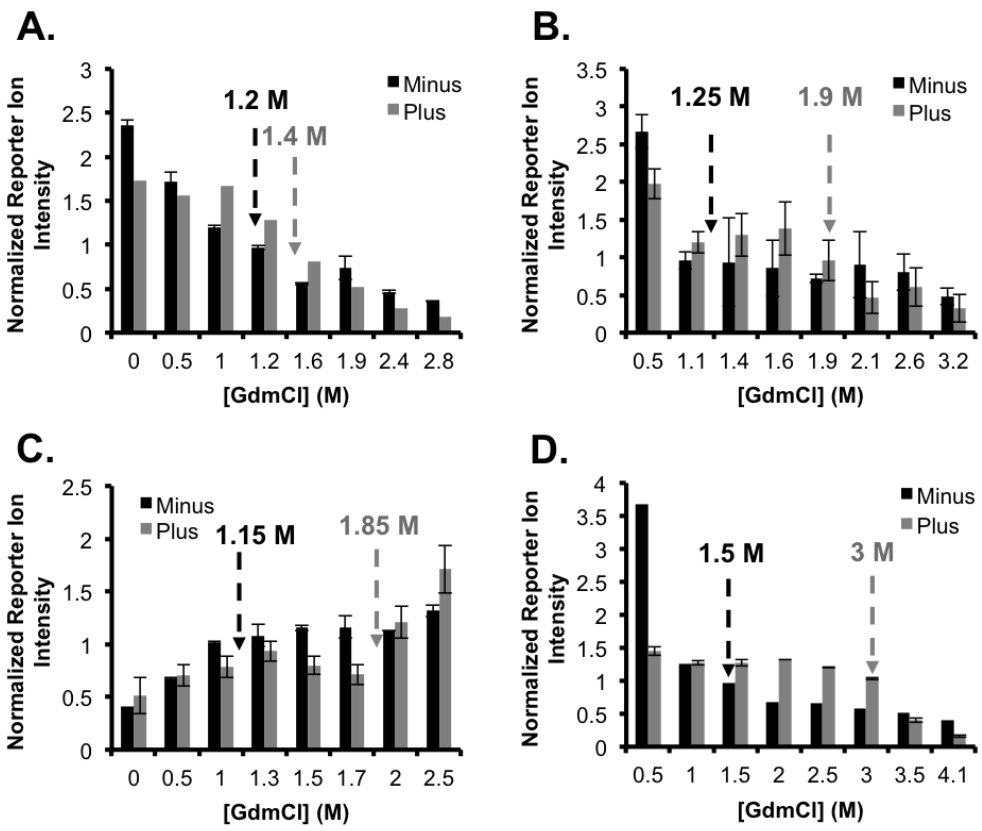

Figure 16: Chemical denaturant data recorded for yeast Hsp90 peptide (A) (B) DSGIGMTK and human Hsp90 (C) TLTIVDTGIGM(OX)TK (D) TDTGEPMGR. (A) is derived from Yeast-Hybrid experiment, (B) is from previous iTRAQ-SPROX experiments(113), (C) is from MCF-7-experiment and (D) is from a recent iTRAQ-SPROX experiment on MCF-7 lysate. (A)-(C) used 30 min equilibrium time while (D) used an extended incubation time of 24 h. In cases where multiple product ion mass spectra were collected the averaged data are shown with error bars representing ±1 standard deviation. The arrows point to the $C_{1/2}$ values extracted from each data set.

3.5 Conclusion

The Hybrid protocol was applied to characterize the direct binding of geldanamycin to Hsp90 in both yeast and MCF-7 cell lysate using 30 min protein-drug equilibrium time. Proteomic coverage was increased by 20%-60% compared to using the methionine peptide probes alone. A methionine-containing peptide located in the N-

terminal ATP-binding domain of Hsp90 was successfully detected as hit in MCF-7 cell lysate. Such hit behavior was not observed in the Yeast-Hybrid experiment possibly due to the structural perturbation by HNSB modification. However, the yeast peptide was identified as hit in a previous iTRAQ-SPROX experiment(*113*) with K_d value of 1.2 μM, comparable to the K_d value of 0.97 μM calculated from MCF-7-Hybrid experiment. As a result, for 30 min incubation time, there is no significant difference in binding affinity for the Hsp90-geldanamycin complex between the normal and tumor cell lysates, which seems to suggest the endogenous protein-protein interactions with Hsp90 co-chaperones (e.g., p23/Sba1 and Hop) have little effect on Hsp90-geldanamycin binding affinity. Nonetheless, since one previous study(*115*) argues that 30 min incubation is too short to reach equilibrium, a recent iTRAQ-SPROX experiment with extended [GdmCl] range and 24-hour incubation time confirmed the slow, tight binding event between Hsp90 and geldanamycin.

4. Identification of protein targets of omeprazole sulfide in yeast lysate using SPROX-related techniques

4.1 Introduction

Malaria continues to be a major health burden to the world, especially to pregnant women and children under 5 years old in endemic areas of Africa. Although unprecedented efforts in development of therapeutic agents and implementation of preventive measures decreased the malaria death rate by 60% globally between 2000 and 2015, the World Health Organization revealed that there were still about 438,000 malaria deaths worldwide(*116*). Human malaria infection is a two-stage process, including a one-time asymptomatic liver stage followed by a cyclic symptomatic blood stage. *Plasmodium* sporozoites are first deposited under the host skin with the bite of an infected female *Anopheles* mosquito and then migrate to the liver where each sporozoite develops into tens of thousands of merozoites within hepatocytes(*117, 118*). The merozoites are finally released from the hepatocytes to invade red blood cells, which marks the onset of the disease(*118*). To date, most anti-malaria drug discovery efforts and the majority of existing therapies target the blood stage(*119-123*).

Recently, several liver-stage screens have been developed(*124-127*) to identify therapeutic agents that can target this initial stage of infection. To date, 2 compounds from the imidazolopiperazine family, GNF179 and KAF156 have been identified to show some promise in fighting liver-stage malaria and the latter is currently in Phase II trials(*128, 129*). Inhibitors for liver-stage malaria are of high therapeutic value due to

their potentials for causal prophylaxis and possibly eradication(*123, 126*). In an effort to discover more liver-stage anti-malaria agents, a high-throughput liver-stage malaria screen based on a luciferase-expressing sporozoite strain of *Plasmodium berghei* ANKA was used to identify liver-stage inhibitors among a library of 5,375 known bioactive compounds(*126*). This screen identified 37 liver-stage inhibitors, only 20% of which were known anti-parasite agents, and 47% of which possessed other therapeutic activities(*126*). Among the hits were 3 ulcer drugs including omeprazole (Prilosec, Figure 17), esomeprazole (Nexium) and tenatoprazole. All 3 of ulcer drugs were determined to have low micromolar liver-stage IC_{50} values(*126*). Omeprazole in its active sulfenamide form suppresses gastric acid secretion by inhibiting the proton pump(*130-132*). Notably, the omeprazole derivative, omeprazole sulfide (Figure 17), although inactive against proton pumps, inhibited liver-stage malaria with IC_{50} value of 0.6 μM (Derbyshire and co-workers, unpulished data).

In order to better understand the liver stage inhibition activity of omeprazole sulfide, here both the SPROX and Hybrid protocols were utilized to identify novel protein targets of this compound. As energetics-based approaches amenable to proteome-wide analysis of protein stability and ligand binding(*15, 19*), the SPROX-related techniques were chosen in this study due to advantages over traditional approaches. Particularly, the Hybrid protocol is a SPROX-like technique with improved proteomic coverage by exploiting both methionine and tryptophan probes(*55*). A total of one SPROX and 4 biological replicates of the Hybrid experiment were performed to

identify potential protein targets of omeprazole sulfide in the context of yeast lysate. To corroborate the hit results, two biological replicates of the SILAC-PP experiment were conducted according to reference (*52*).

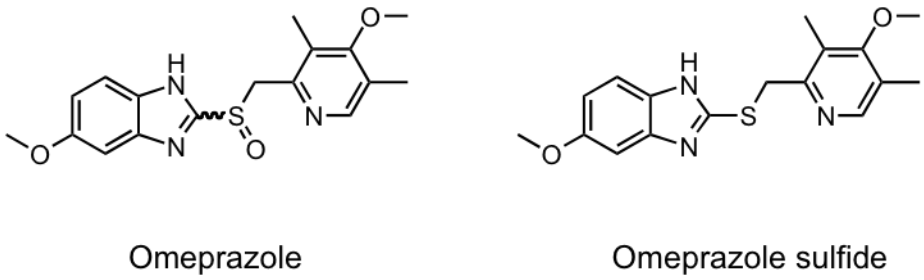

Figure 17: chemical structures of omeprazole and omeprazole sulfide.

4.2 Experimental

4.2.1 Materials

The *Saccharomyces cerevisiae* open reading frame strain Y258 was purchased from Open Biosystems (Waltham, MA). The following reagents were from Sigma-Aldrich (St. Louis, MO): trypsin from porcine pancreas (proteomics grade, dimethylated), thermolysin from *Bacillus thermoproteolyticus rokko*, L-methionine, L-tryptophan, HNSB (≥95 wt%), GdmCl, MMTS, TEAB buffer (1.0 M, pH 8.5), SDS, H_2O_2 (30%, w/w), ammonium acetate, TMP, pyrrolidine, TCA, urea and TFA. The following materials were from Thermo Scientific (Waltham, MA): TCEP•HCl, UltraLink hydrazide resin and Coomassie Plus Bradford Assay reagent. Methanol and isopropyl alcohol were from VWR International (West Chester, PA). MacroSpin columns (Silica C18) and Pi^{3TM} methionine reagent kits were from the Nest Group (Southborough, MA). Ethanol (completely denatured) was from Avantor Performance Materials, Inc. (Phillipsburg,

NJ). Sodium phosphate dibasic was from EM Science (Gibbstown, NJ). Sodium phosphate monobasic was from Mallinckrodt Backer, Inc. (Paris, KY). Dimethyl sulfoxide (DMSO) was from Acros (NJ). Tris-HCl buffer (1 M, pH 8) and SDS-sample buffer (6×, nonreducing) were from Boston BioProducts (Ashland, MA). Omeprazole sulfide is a gift from Professor Emily R. Derbyshire at Duke University.

4.2.2 Hybrid protocol and iTRAQ-SPROX protocols

A total of 4 biological replicates of Hybrid protocol and 1 replicate of SPROX were performed. The Hybrid and the iTRAQ-SPROX experiments were conducted as described in Chapter 3 and reference (*15*), respectively. Yeast cell lysates were prepared as described in Chapter 2, with total protein concentration in the range of 12-15 mg/mL determined by Bradford assay. Briefly, the lysate was divided into two 180 μL portions, with 20 μL DMSO added to the (-) ligand group, 20 μL 10 mM omeprazole sulfide stock solution (in DMSO) to the (+) ligand group. The lysate was equilibrated with the ligand for 1.5 h before 20 μL aliquots of the (-)/(+) lysate were combined with 25 μL volumes of a series of pH 7.4 20 mM phosphate buffers containing increasing concentrations of GdmCl. The final eight GdmCl concentrations used in Hybrid and SPROX experiments are summarized in Table 10. The final ligand concentration was 500 μM. The 16 denaturant-containing protein samples were equilibrated at room temperature for 2 h before reaction was initiated by adding 5 μL H_2O_2 (30%, w/w) to each sample. For the Hybrid protocol, after 2 min, 2.5 μL of a 46.2 mM solution of HNSB was spiked into each sample and the reaction was quenched by 1 mL quenching cocktail containing 0.5 M L-

methionine and 4.9 mM L-tryptophan at 3 min. For the SPROX protocol, only H_2O_2 was added and the oxidation reaction was quenched by 1 mL 0.5 M L-methionine at 3 min. The proteins in each sample was precipitated by addition of 200 μL of 1g/mL TCA and incubated on ice at 4°C overnight. The samples were then centrifuged at 8,000g, 4°C for 30 min. The resulting protein pellets were washed 3 times with 300 μL pre-chilled ethanol and dried in a Savant™ SpeedVac™ concentrator (Thermo Scientific, Waltham, MA) for 5 min.

To redissolve protein pellets, 30 μL 0.5 M of TEAB buffer (pH 8.5) and 1.5 μL of a 2% (w/v) SDS solution were added to each sample. The samples were sonicated, vortexed, and heated at 60°C to facilitate their dissolution. Disulfide bond reduction was carried out with addition of 3.5 μL 50 mM TCEP•HCl and incubation at 60°C for 1 h. A 1.8 μL aliquot of 200 mM MMTS was added to each sample before a 10 min incubation at room temperature. The proteins in each sample were digested with 2 μL trypsin (1 mg/mL in 1mM HCl) at 37°C overnight.

Each of the (-) and (+) peptide samples was labeled with 0.5 unit of an iTRAQ-8plex tag dissolved in 50 μL isopropanol and the labeling scheme can be found in Table 10. The labeling reactions were allowed to proceed at room temperature for 2 h. Within each set, (-) or (+), 10 μL or 30 μL of each sample was combined. The combined samples, either 80 μL or 240 μL in volume, were either subjected to a desalting protocol using C18 MacroSpin columns (80 μL sample), a methionine-containing peptide enrichment strategy (240 μL sample), or a tryptophan-containing peptide enrichment strategy (240

μL sample) prior to LC-MS/MS analysis. Only methionine enrichment was carried out for the SPROX protocol. The above 3 fractionation protocols were performed as described in Chapter 2. For SPROX and Hybrid Exp 1, the samples were each subjected to a single LC-MS/MS analysis. For Hybrid Exp 2-4, the C18 purified samples (i.e., the non-enhanced samples) were each subjected to a single LC-MS/MS analysis while the methionine/tryptophan-enriched samples were each subjected to two replicate LC-MS/MS analyses.

Table 10: Summary of the iTRAQ labeling scheme used in the SPROX experiment and 4 Hybrid experiments. Shown are the GdmCl concentrations (M) of the samples corresponding to each of the 8 iTRAQ tags.

Experiment	iTRAQ reporter ion m/z							
	113	114	115	116	117	118	119	121
SPROX	1	1.3	1.7	2	0.5	2.5	3.5	3
Hybrid	0.5	1	1.2	1.6	1.9	2.4	2.9	3.3

4.2.3 SILAC-PP

SILAC-PP experiments were conducted as described in reference (*52*). A total of 2 biological replicates of SILAC-PP experiment were performed, referred to as PP-Exp1 and PP-Exp2. The yeast deletion strain BY4739 (Open Biosystems, Lafayette, CO) was grown on a petri dish containing synthetic complete (SC) media with Difco™-agar and L-lysine at 30 °C for 3 days. Per liter media for the plate consisted of 1.7 g of yeast nitrogen base, 5 g of ammonium sulfate, 20 g of Difco™-agar, 2 g of glucose, 1.92 g of synthetic drop out mix without lysine and 30 mg of L-lysine. Per liter SC media for the subsequent culture contained 1.7 g of yeast nitrogen base, 5 g of ammonium sulfate, 20 g

of glucose, 1.92 g of synthetic drop out mix without lysine. One colony from the plate was inoculated into 10 mL of SC media containing 30 mg/L light L-lysine (light SC media) and incubated at 30 °C overnight. A 10 μL aliquot of the cell culture was transferred into 100 mL of light SC media, while another 10 μL aliquot was transferred into 100 mL of SC media containing 30 mg/L heavy lysine ($^{13}C_6$, $^{15}N_2$) (Cambridge Isotope Laboratories, MA) (heavy SC media). The above two cell cultures were incubated at 30 °C for 24 h before 100 μL light and heavy cell cultures were transferred into 100 mL corresponding light and heavy SC media. The above two cell cultures were incubated at 30 °C for 24 h and this step was repeated 3 times. Finally, when the measured OD_{600} was 6-8, 100 μL light and heavy cell cultures were transferred into 1 L corresponding light and heavy SC media before incubation at 30 °C overnight. When the OD_{600} of each culture reached ~1.8, the cells were pelleted at 1,000 g at 4 °C for 10 min, washed with D.I. and frozen at -20 °C.

For each experiment, one light and one heavy yeast pellet was each suspended in 400 μL lysis buffer containing 20 mM Tris-HCl buffer (pH 8.0) with 50 mM NaCl, 10 mM $CaCl_2$ and 1×house-made protease inhibitor cocktail (same composition as described in Chapter 2). Light and heavy cell lysates were prepared as described in Chapter 2 and the concentrations of the two cell lysates were normalized to be the same using lysis buffer. The lysate concentration was 4.5 mg/mL for PP-Exp1 and 7 mg/mL for PP-Exp2, determined by Bradford assay. The heavy lysate was incubated with omeprazole sulfide (final concentration 1 mM) while the light lysate was incubated with equal volume of

DMSO at room temperature for 1 h. 25 μL aliquots of (-) or (+) ligand lysate were equilibrated with 75 μL urea buffers in 20 mM Tris-HCl (pH 8.0) overnight. The final [urea] in each reaction buffer was 0, 0.9, 2, 2.6, 2.9, 3.5, 3.9, 4.5, 4.9, 5.3, 6 and 6.9 M and the final ligand concentration was 750 μM.

The pulse proteolysis reaction was initiated by the addition of 9 mg/mL thermolysin to each sample, with protein : protease mass ratio of 10:1. The reaction was allowed to proceed for 1 min before quenching with 15 μL of 0.5 M EDTA. Afterwards, 100 μL of (+) and 100 μL (-) samples were combined from the same [urea]. To each combined sample, 57 μL gel loading buffer (made by mixing 600 μL 6× non-reducing SDS sample buffer with 225 μL BME) was added. The samples were heated at 95 °C for 5 min, cooled to room temperature and then run on a 4-20% midsize polyacrylamide gel (BioRad, Hercules, CA) at 120 V for ~1 h. The gels were incubated in fixing solution containing 25% isopropanol and 10% acetic acid for ~20 min. The gels were then stained overnight with a staining solution that contained 0.01% R-250 and 10% acetic acid. Finally, the gels were destained with 10% acetic acid overnight and different molecular weight regions of each gel were excised and digested using the protocol described in reference (*133*). The gel bands excised in PP-Exp1 and PP-Exp2 were 20-25 kDa, 25-34 kDa, 37-50 kDa, 50-75 kDa and 75-100 kDa.

4.2.4 LC-MS/MS analyses

For SPROX and Hybrid Exp1, LC-MS/MS analyses were performed at the Proteomics Resource at the Fred Hutchinson Cancer Research Center on an Orbitrap

Elite ETD mass spectrometer equipped with an EASY-nLC system (ThermoFisher Scientific Inc., Waltham, MA). The instrument parameters were the same as described in Chapter 2. The data were searched using Proteome Discoverer with fixed modifications of MMTS on cysteine and iTRAQ® 8-Plex on N-terminus and lysine residues. Proteome Discoverer was set with a variable modification of oxidation on methionine residues. In all cases only peptides identified with high and medium confidence (i.e., false discovery rates (FDR) <1% and 5%, respectively) were used in the data analysis.

For Hybrid Exp 2-4, LC-MS/MS analyses were performed at the Duke Proteomics Facility using a Thermo Scientific Q-Exactive Plus high-resolution mass spectrometer with a nanoAcuity UPLC system (Waters Corp) and a nanoelectrospray ionization source. Approximately 4 μL of sample was injected and trapped on a Symmetry C18 180 μm × 20 mm trapping column for 3 min at 5 μL/min (99.9/0.1 v/v water/acetonitrile 0.1% formic acid), followed by the analytical separation on a 75 μm × 250 mm column packed with 1.7 μm Acquity HSST3 C18 stationary phase (Waters Corp). Peptides were separated using a gradient of 3 to 30% acetonitrile with 0.1% formic acid over 90 min at a flow rate of 0.4 microliters/minute (μL/min) with a column temperature of 55 °C. Data collection on the QExactive Plus mass spectrometer was performed in a data-dependent acquisition (DDA) mode of acquisition with a resolution (r) of 70,000 (at m/z 200) for full MS scan from m/z 375 – 1600 with a target AGC value of 1×10^6 ions in profile mode, followed by 20 MS/MS scans at r=17,500 (at m/z 200) in centroid mode, using an AGC target value of 1×10^5 ions, a max fill time of 60 ms, and normalized collision energy of 30

V. A 30 s dynamic exclusion was employed to decrease MS/MS oversampling. The total analysis cycle time for the sample injection was approximately 125 minutes. The resulting data were searched on Spectrum Mill Workbench Software B.04 Rev. B against SwissProt *Saccharomyces cerevisiae* database with trypsin digestion including tryptic peptides with up to 3 missed cleavages. Fixed modifications of MMTS on cysteine and iTRAQ® 8-Plex on N-terminus and lysine residues were included. Oxidation of methionine residues and deamidation of asparagine residues were included as variable modifications. The precursor ion mass tolerance was set to 15 ppm and the product ion mass tolerance was set to 0.02 Da. Only peptides identified with global FDR < 1% were used in data analysis.

For PP-Exp1 and PP-Exp2, samples from gel bands 20-25 kDa, 25-34 kDa, 37-50 kDa, 50-75 kDa were analyzed on Agilent 6520 Q-TOF mass spectrometer equipped with a Chip Cube interface. The ProtID-Chip43(II) HPLC chip with a 40 nL enrichment column and a 75 μm × 43 mm analytical column packed with Zorbax 300SB-C18 5 μm packing material. The peptides were eluted with a linear gradient from 3 to 5% solvent B (ACN with 0.1% FA) for 2 min, 15 to 60% solvent B for 18 min, 60 to 90% solvent B for 3 min, isocratic 100% solvent B for 2 min and isocratic 5% solvent B for 3 min. The flow rate was 0.4 μL/min. Drying gas was set at 4 L/min at 350 °C. The skimmer and fragmentor were set at 65 and 175 V, respectively. The collision energy was determined by the linear equation with slope of 3.5 V/100 Da and offset of -4.8 V. The inclusion window width for precursor ions was 4 m/z. The scan rate was 3 scans/s for precursor

mass spectra and 2 scans/s for product ion spectra. Data-dependent acquisition was performed with the top 4 precursor ions selected for fragmentation during each cycle. All spectra were acquired in profile mode.

For PP-Exp1 and PP-Exp2, samples from gel band 75-100 kDa were analyzed on a Thermo Scientific Q-Exactive Plus high-resolution mass spectrometer with a nanoAcuity UPLC system (Waters Corp) and a nanoelectrospray ionization source. First, the sample was trapped on a Symmetry C18 180 μm × 20 mm trapping column for 3 min at 5 μL/min (99.9/0.1 v/v water/acetonitrile 0.1% formic acid), followed by the analytical separation on a 75 μm × 250 mm column packed with 1.7 μm Acquity BEH130 C18 stationary phase (Waters Corp). Peptides were separated using a gradient of 5 to 40% acetonitrile with 0.1% formic acid over 60 min at a flow rate of 0.4 microliters/minute (μL/min) with a column temperature of 55 °C. Data collection on the QExactive Plus mass spectrometer was performed in a data-dependent acquisition (DDA) mode of acquisition with a resolution (r) of 70,000 (at m/z 200) for full MS scan from m/z 375 – 1600 with a target AGC value of 1×10^6 ions in profile mode, followed by 10 MS/MS scans at r=17,500 (at m/z 200) in centroid mode, using an AGC target value of 5×10^4 ions. A 20 s dynamic exclusion was employed to decrease MS/MS oversampling. The total analysis cycle time for the sample injection was approximately 95 minutes.

The SILAC-PP data were searched on Spectrum Mill Workbench Software B.04 Rev. B against SwissProt *Saccharomyces cerevisiae* database with trypsin digestion including tryptic peptides with up to 3 missed cleavages. Fixed modifications of SILAC

0-8 Da on lysine residues and carbamidomethylation on cysteine residues were included. Methionine oxidation and deamidation of asparagine residues were included as variable modifications. The precursor ion mass tolerance was set to 20 ppm for the Q-TOF and 5 ppm for the Q-Exactive Plus. The product ion mass tolerance was set to 50 ppm for the Q-TOF and 0.02 Da for the Q-Exactive Plus.

4.3 Data analysis

The Hybrid and iTRAQ-SPROX data analysis process was as described in Chapter 3. Only peptides with raw iTRAQ intensities that were summed to greater than 1000 were included in analysis. In SPROX and Hybrid Exp1, product ion spectra with isolation interference smaller than 30% were used for averaging; while in Hybrid Exp2-4, product ion spectra with precursor isolation purity percent larger than 70% were used for averaging. Summarized in Table 11 are the N2-normalization factors generated for each reporter ion. For the (+) ligand group in Hybrid Exp4, the relative standard deviation of the 117 tag was ~80%. To prevent this tag from distorting the data of a given peptide, the (+)-Exp4 data were re-analyzed discarding the 117 tag. Instead, the N2-normalized value of the 117 tag was re-constructed by averaging the N2-normaliazed values of the 115 and 118 tags.

SILAC-PP data analysis was performed as described in reference (*52*). Briefly, a median L/H ratio for each experiment was determined from the L/H ratios for all the peptides identified. The median L/H ratios were calculated separately for Q-TOF and Q-Exactive Plus data. For PP-Exp1, the median L/H ratio was 0.8 for both Q-TOF and Q-

Exactive Plus data; for PP-Exp2, the median L/H ratio was 1.1 for Q-TOF data and 1.2 for Q-Exactive Plus data. The L/H ratios of all peptides from a particular protein at a given [urea] in each gel band were averaged to generate a single L/H ratio. Finally, the protein-average L/H ratios were used to generate SILAC-PP datasets. Only proteins identified in at least 4 urea concentrations were assayed for potential changes in stability changes. Hit proteins were identified with >1.7 fold deviations from the median L/H ratio at 2 or more consecutive urea concentrations. Hits were ultimately confirmed by visual inspection of the data.

Table 11: N2-normalization factors generated for each reporter ion in the SPROX experiment and Hybrid Exp1-4.

Experiment	Statistics	iTRAQ reporter ion m/z							
		113	114	115	116	117	118	119	121
(-)-SPROX	Average	0.86	1.02	0.65	0.96	1.08	1.15	1.39	0.90
	S.D.	0.20	0.24	0.19	0.21	0.23	0.28	0.35	0.22
(+)-SPROX	Average	1.08	0.70	0.34	1.18	1.11	1.33	1.08	1.18
	S.D.	0.20	0.18	0.18	0.21	0.27	0.25	0.19	0.29
(-)-Exp1	Average	0.91	1.02	0.93	0.98	0.96	1.06	1.03	1.12
	S.D.	0.19	0.17	0.18	0.19	0.21	0.17	0.17	0.19
(+)-Exp1	Average	1.06	0.88	0.97	0.78	0.95	1.03	0.52	1.80
	S.D.	0.22	0.17	0.21	0.18	0.20	0.31	0.14	0.42
(-)-Exp2	Average	0.93	1.06	1.08	1.06	1.02	1.00	0.89	0.96
	S.D.	0.13	0.12	0.11	0.10	0.08	0.09	0.21	0.10
(+)-Exp2	Average	0.96	1.12	1.07	1.09	1.11	0.85	0.86	0.95
	S.D.	0.12	0.12	0.10	0.13	0.09	0.12	0.19	0.09
(-)-Exp3	Average	0.62	1.34	1.10	0.78	1.51	0.86	1.04	0.75
	S.D.	0.27	0.34	0.31	0.30	0.39	0.33	0.26	0.32
(+)-Exp3	Average	0.63	0.79	1.10	1.13	1.21	1.21	0.93	0.99
	S.D.	0.39	0.49	0.18	0.19	0.20	0.19	0.17	0.15
(-)-Exp4	Average	0.90	1.10	0.81	1.10	1.24	1.12	0.91	0.81
	S.D.	0.12	0.13	0.15	0.10	0.21	0.12	0.14	0.13
(+)-Exp4	Average	1.12	1.21	0.91	1.14	0.78	1.11	0.82	0.91
	S.D.	0.21	0.20	0.14	0.15	0.61	0.14	0.14	0.14
(+)-Exp4	Average	1.08	1.17	0.88	1.10	NA[2]	1.08	0.79	0.88
7 tags[1]	S.D.	0.18	0.16	0.12	0.12	NA	0.10	0.13	0.13

[1]Data analysis discarding tag 117. [2]Not applicable.

4.4 Results and discussion

4.4.1 Proteomic coverage

The SPROX and Hybrid protocols were used to identify protein targets of omeprazole sulfide in a yeast cell lysate. Multiple biological replicates were performed to eliminate false positives. Summarized in Table 12 are the proteomic coverages obtained in the different experiments performed here. By exploiting both methionine

and tryptophan probes, the hybrid protocol achieved a 30-50% increase in proteomic coverage compared to using only methionine probes. In total, 899 unique proteins were assayed for omeprazole binding interactions using 3365 peptide probes and 88 unique peptides which mapped to 70 unique proteins were identified as hits. Further analysis of the behavior for the 70 proteins (e.g., consistency across all 5 experiments) was then performed to reach a final hit list containing 20 proteins (see Table 13).

Table 12: Peptide (protein) coverage achieved in all experiments and the corresponding number of hits.

Experiment	Assayed[1] M	Hits M	Assayed W	Hits W
SPROX	1044 (443)[2]	45 (39)	NA[3]	NA
Hybrid-Exp1	448 (254)	8 (8)	634 (328)	5 (5)
Hybrid-Exp2	1228 (517)	9 (8)	769 (371)	2 (2)
Hybrid-Exp3	883 (369)	13 (11)	456 (223)	0 (0)
Hybrid-Exp4	1053 (443)	7 (7)	691 (339)	7 (6)

[1]Peptides identified in both (-) and (+) ligand groups. [2]Number of unique peptides (number of corresponding proteins). [3]Not applicable.

Table 13: Final list of the 20 hit peptides (proteins) and their corresponding transition midpoints.

Accession number	Peptide sequence	Experiment[1]	$C_{1/2}$-(-) (M)	$C_{1/2}$-(+) (M)
P24031	QLQMLAR	Exp1[2]	1.2	1.8
P22147[3]	LIEAMQPYLR	Exp2	1.9	1.4
P52488	SHIFNIPMK	Exp2	1.9	1.4
P32471	VSLDDLQQSIEEDEDHVQSTDIAAM(OX)QK	Exp4	2.6	1.2
P60010	APEALFHPSVLGLESAGIDQTTYNSIM(OX)K	Exp4	2.6	1.9
P32775	FWNPENPYK	Exp1	1.2	2.2
Q00711	GSDWLGDQDSIHYM(OX)TR	Exp1	1.4	2.2
P04840	LGWSLSFDA	Exp1	1.8	2.4
P05749	NQLRDWIR	Exp1	1.9	3.1
P34730	MKGDYHR	SPROX	1.6	2.2
P33298	ILIELLTQMDGFDQSTNVK	SPROX	1.6	2.2
P32472	GWDQGVAGMCVGEK	SPROX	1.1	1.6
P40069	ADDTMNAK	SPROX	2.4	1.8
P53312	GLSPQMAK	SPROX	1.2	1.8
P40414	EADLNSEQMGR	SPROX	1.1	1.8
P06169	M(OX)SANISETTAMITDIATAPAEIDR	SPROX	2.2	1.1
P37291	IGAPAMTTR	SPROX	1.1	1.8
Q08601	AMQWLVK	SPROX	1.8	1.2
P06634	HIVEDCDMTPVGER	SPROX	1.2	1.8
P07262	ICYAFMR	SPROX	1.9	1.1

[1]The experiment where the peptide was identified as hit. [2]Hybrid-Exp1. [3]The bolded proteins were also assayed in the SILAC-PP experiments.

4.4.2 Global hit analysis

Based on the results summarized in Table 12, the peptide hit rate for each experiment was in the range of 1-4%, which is on the same order as the false-positive rate of hit discovery, previously determined for the SPROX and Hybrid approaches(*27, 55*). The hit rate of the SPROX experiment, ~4%, was slightly higher than that of the

Hybrid experiments, ~1-2%, most likely due to the aberrantly low N2-normalization factor of tag 115 in the (+) ligand group (see Table 11).

A bioinformatics analysis was performed on the 899 total assayed proteins and on the 70 protein hits using PANTHER(*134, 135*). Proteins were classified into different protein classes or molecular functions and the percentage of proteins in each class was calculated (Figure 18). One goal of the bioinformatics analysis was to determine if the hits were enriched in a specific protein class or molecular function. However, no significant enrichments were observed (Figure 18).

Performing multiple biological replicates helps eliminate false positives. The original list of 88 hit peptides from 70 proteins was inspected across all experiments and any peptide with inconsistent behavior (hit in one experiment and not a hit in another) was then removed from the hit list. Unfortunately, no peptide/protein was identified as hit in more than one experiment. In other words, the 20 proteins in the final hit list are those that were only assayed in one experiment (Table 13).

Consequently, the following evidence (1) experimental hit rate is close to the protocol false-positive rate; (2) no enrichment of protein classes or molecular function was observed among hits; (3) no peptide/protein was identified as hit in 2 or more experiments suggests that no strong protein hits were identified in the protein target discovery experiments performed here using the Hybrid and iTRAQ-SPROX protocols.

Protein Class	Hits	All
calcium-binding protein	1.1	1.3
chaperone	3.4	3.9
cytoskeletal protein	2.3	2.6
hydrolase	11.5	12.5
isomerase	3.4	3.4
kinase	1.1	2.2
ligase	9.2	5.9
lyase	4.6	4.5
nucleic acid binding	20.7	20.2
oxidoreductase	8	10.5
phosphatase	1.1	1.7
protease	3.4	2.6
receptor	2.3	1.2
signaling molecule	1.1	0.6
transcription factor	1.1	2.1
transfer/carrier protein	2.3	1.4
transferase	16.1	12.2
transporter	6.9	4.3

Molecular Function	Hits	All
binding	20.7	25.6
catalytic activity	52.4	47.7
nucleic acid binding transcription factor activity	1.2	1.9
receptor activity	2.4	1.2
structural molecule activity	7.3	10.3
translation regulator activity	8.5	4.7
transporter activity	7.3	5.2

Figure 18: Bioinformatics analysis of the 899 total assayed proteins ("All") and the 70 total hit proteins ("Hits") from the original hit list. The numbers in the boxes represent the percentage of proteins belonging to each protein class or with a particular molecular function. The classification was performed using PANTHER(134, 135) version 10.0.

4.4.3 Hit validation by SILAC-PP

Pulse proteolysis is another energetics-based approach that can be used to evaluate protein thermodynamic stability and protein-ligand interactions(*33*, *52*). The pulse proteolysis technique is particularly well-suited for the validation of SPROX results(*52*). As part of this work, SILAC-PP was used to validate the 20 hits in Table 13. A total of 2 biological replicates of SILAC-PP experiment were conducted and the protein average L/H ratios were plotted against [urea] to generate PP datasets (Figure 19). If the L/H ratio is not significantly altered in at least 2 consecutive urea concentrations, there is no interaction detected between the protein and the ligand. The

SILAC-PP data obtained on P05694, a non-hit protein in all the Hybrid and iTRAQ-SPROX experiments, is consistent with non-hit behavior (Figure 19A). A total of 5 of the 20 hit proteins in Table 13 were assayed in SILAC-PP experiments. None of the 5 proteins showed hit behavior (Figure 19B). Therefore, it is likely that these 5 proteins were false positives in the Hybrid and iTRAQ-SPROX experiments. The remaining 15 protein hits identified in the Hybrid and iTRAQ-SPROX experiments require further experiments to determine if they are false positives or not.

Unfortunately, the Hybrid and iTRAQ-SPROX techniques failed to conclusively identify any protein targets of omeprazole sulfide. The primary reason for this failure is most likely related to the limited proteomic coverage of the SPROX-like techniques. Although the use of the Hybrid protocol increases proteomic coverage that can be achieved, a total of only 899 proteins were assayed in this study. This is a relatively small fraction, ~14%, of the yeast proteome(*136*). Efforts to increase the proteomic coverage obtained in the Hybrid and iTRAQ-SPROX protocols (e.g., the use of fractionation techniques such as isoelectric focusing and 2D-HPLC, or simply increasing the number of LC-MS/MS analyses) could potentially result in the discovery of novel protein targets of omeprazole sulfide.

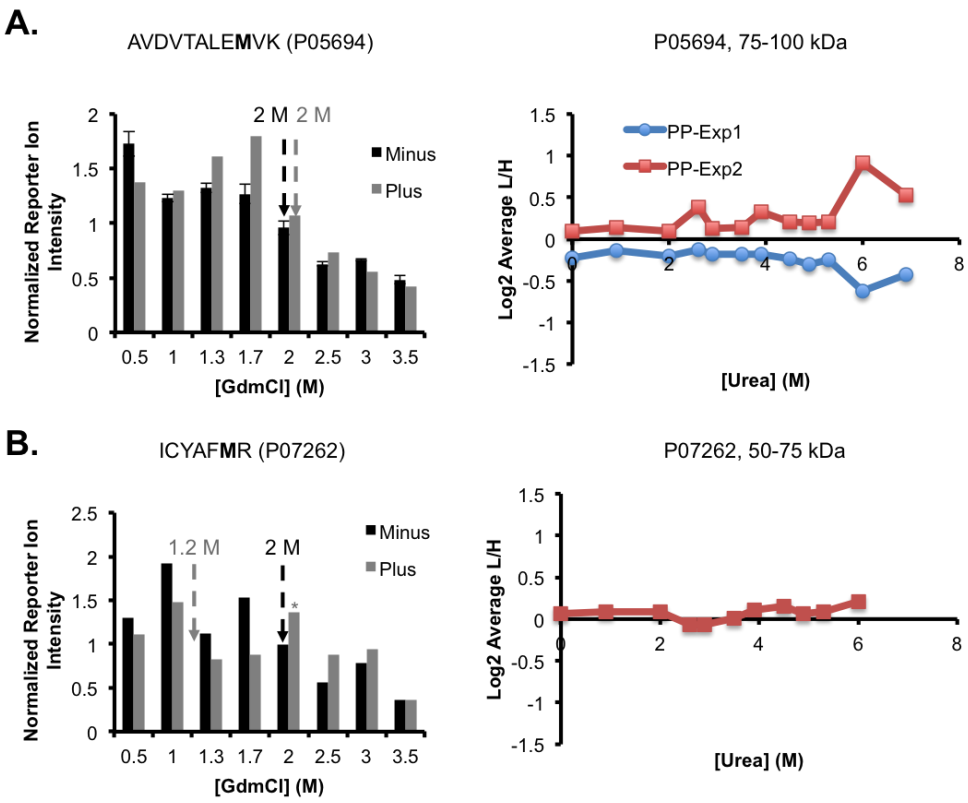

Figure 19: Hit validation using SILAC-PP. (A) SPROX data of the peptide AVDVTALEMVK from the non-hit protein P05694 and SILAC-PP data of the same protein from the 75-100 kDa gel band. (B) SPROX data of the hit peptide ICYAFMR from protein P07262 and SILAC-PP data (PP-Exp2) of the same protein from the 50-75 kDa gel band. In cases where multiple product ion mass spectra were collected the averaged data are shown with error bars representing ±1 standard deviation. The arrows point to the $C_{1/2}$ values extracted from each dataset. The asterisk (*) sign denotes data outlier. Intact protein molecular weight is 85.9 kDa for P05694 and 49.6 kDa for P07262.

4.5 Conclusion

Compared to blood-stage malaria, liver-stage malaria is less well understood and of higher therapeutic value in causal prophylaxis and eradication. The omeprazole derivative, omeprazole sulfide, was determined to have low micromolar liver-stage IC_{50} value. To elucidate the liver-stage activity of omeprazole sulfide, SPROX-related

techniques were applied to identify its protein targets using yeast as model organism. A total of 1 SPROX experiment and 4 biological replicates of Hybrid experiments were performed in the context of yeast lysate, assaying 899 proteins overall. With no protein hits appearing in more than 1 experiment, 20 one-time protein hits were identified with no contradictory data from other 4 experiments. To eliminate false positives, 2 biological replicates of SILAC-PP experiments were conducted. A total of 5 one-time hits were considered to be false positives with no hit behavior detected in SILAC-PP experiments. In conclusion, no strong protein hits were determined in this study, probably due to the limited proteomic coverage. Future experiments with improved proteomic coverage or directly studying *Plasmodium* parasite proteins can possibly identify more promising protein targets.

5. Identification of Puf6 protein-protein interactions in a yeast cell lysate using the SPROX technique

5.1 Introduction

Messenger RNA (mRNA) localization helps achieve spatial regulation of protein translation, which is important for the functional and structural asymmetries in cells. During events like cell division, embryonic development, cell migration and synaptic plasticity, mRNA localization is coupled with translational repression to spatially and temporally control protein production(*137, 138*). In budding yeast, localization of ASH1 mRNA at the bud tip of daughter cells during anaphase inhibits HO endonuclease expression and prevents mating type switching(*139-141*). The RNA-binding protein (RBP) She2 binds co-transcriptionally to the nascent ASH1 mRNA(*142, 143*) and recruits it to the type V myosin motor myosin 4 (Myo4) and the RBP, She3(*144-146*). The functional locasome complex consisting of She2, She3 and Myo4 then transports the ASH1 mRNA along the actin cytoskeleton to the bud tip(*140, 141, 146*). In the nucleus, Puf6, Khd1, Loc1, She2 together with ASH1 mRNA participate in the assembly of a pre-messenger ribonucleoprotein particle (mRNP)(*147-151*). Particularly, ASH1 mRNA translation is repressed by Puf6 and Khd1 during transport(*150, 151*) and released by the phosphorylation of the 2 factors at the bud tip(*148, 149*).

Puf6 is a non-canonical member of the pumilio/ferminization of XX and XO animals (fem)-3 mRNA binding factor (PUF) family. Puf6 and its human ortholog Puf-A are localized in nucleolus while classical PUF proteins such as human PUM1 are

localized in cytoplasm(*152*). In terms of protein structure, canonical PUF proteins possess a characteristic crescent shape consisting of 8 α-helical PUM repeats(*153-157*), whereas Puf6 and Puf-A each contains 11 PUM repeats(*152*). PUF proteins repress translation by binding to the 3'untranslated region (3'UTR) of the mRNA(*150, 158*) and canonical PUF proteins recognize RNA bases with conserved side chains in the 8 PUM repeats(*153*). Interestingly, unlike canonical PUF proteins that are sequence-specific when binding to single-stranded RNAs *in vitro*(*152, 153, 159-161*), Puf6 binds RNA and DNA promiscuously without sequence specificity *in vitro*(*152*).

The specificity of Puf6 to ASH1 mRNA *in vivo* suggests that other proteins might interact with Puf6 to enhance its specificity. A recent study revealed that Puf6 is recruited onto the 3'UTR of ASH1 mRNA by forming a ternary complex with She2 and Loc1(*162*). GST pull-down experiments *in vitro* demonstrated that in the ternary complex, Loc1 acted as a bridge by binding directly with Puf6 and She2(*162*).

Here iTRAQ-SPROX is utilized to identify additional protein-protein interactions involving Puf6. Most traditional techniques to detect protein-protein interactions including yeast two-hybrid assay(*163-165*), affinity mass spectrometry(*166-168*), protein chips(*169*) and genetic approaches(*170, 171*) suffer from long analysis time, large sample quantity and inability to quantitatively evaluate binding affinity. SPROX as a mass spectrometry-based proteome-wide approach has advantages over traditional techniques in that it provides quantitative measurement of protein binding affinities and it is able to detect both direct and indirect protein-protein interactions(*19, 22*). As part of

this work the iTRAQ-SPROX technique is used to profile the thermodynamic stability of proteins in 2 different yeast cell lysates, one derived from a wild-type yeast strain and the other derived from a Puf6 deletion strain. Ultimately, the stability profiles generated on the 2 yeast strains were compared, and the proteins that were differentially stabilized in the 2 strains were identified. These differentially stabilized protein hits were taken to be those that have either direct or indirect binding interactions with Puf6.

5.2 Experimental

5.2.1 Materials

The following reagents were from Sigma-Aldrich (St. Louis, MO): trypsin from porcine pancreas (proteomics grade, dimethylated), L-methionine, MMTS, TEAB buffer (1.0 M, pH 8.5), SDS, H_2O_2 (30%, w/w), TCA, urea and TFA. The following materials were from Thermo Scientific (Waltham, MA): TCEP•HCl and Coomassie Plus Bradford Assay reagent. Methanol and isopropyl alcohol were from VWR International (West Chester, PA). MacroSpin columns (Silica C18) and Pi^{3TM} methionine reagent kits were from the Nest Group (Southborough, MA). Ethanol (completely denatured) was from Avantor Performance Materials, Inc. (Phillipsburg, NJ). Sodium phosphate dibasic was from EM Science (Gibbstown, NJ). Sodium phosphate monobasic was from Mallinckrodt Backer, Inc. (Paris, KY).

Both the wild-type and Puf6 deletion yeast cell pellets were derived from strain YPH499 and provided by Chen Qiu from National Institute of Environmental Health Sciences (Research Triangle Park, NC). Yeast lysates were prepared as described in

Chapter 2, with a normalized total protein concentration of 13 mg/mL, as determined by Bradford assay.

5.2.2 iTRAQ-SPROX

A total of 4 iTRAQ-SPROX experiments were conducted including 2 on the wild-type yeast lysate (WT1, WT2) and 2 on the Puf6 deletion strain (Delta1, Delta2). The SPROX experiments were performed as described in reference (*15*). Briefly, 20 μL aliquots of lysate were combined with 75 μL volumes of a series of pH 7.4 20 mM phosphate buffers containing increasing concentrations of urea. Since RNA (DNA) was likely to be involved in Puf6 interactions, urea was used instead of GdmCl to minimize the potential disruption of electrostatic interactions(*20, 172*). The 8 denaturant-containing protein samples were equilibrated at room temperature for 2 h before the oxidation reaction in SPROX was initiated by adding 5 μL H_2O_2 (30%, w/w) to each sample. The final urea concentrations in the reaction buffers were 0, 1, 2.1, 2.7, 3.3, 4, 4.9 and 7.5 M. The oxidation reactions were quenched with 1 mL 0.5 M L-methionine after 3 min. The proteins in each sample was precipitated by addition of 200 μL of 1 g/mL TCA and incubated on ice at 4°C overnight. The samples were centrifuged at 8,000g, 4°C for 30 min. The resulting protein pellets were washed 3 times with 300 μL pre-chilled ethanol and dried in a Savant™ SpeedVac™ concentrator (Thermo Scientific, Waltham, MA) for 5 min.

To redissolve protein pellets, 30 μL 0.5 M of TEAB buffer (pH 8.5) and 1.5 μL of a 2% (w/v) SDS solution were added to each sample. The samples were sonicated,

vortexed, and heated at 60°C to facilitate protein dissolution. Disulfide bond reduction was carried out with addition of 3.5 μL 50 mM TCEP•HCl and incubation at 60°C for 1 h. A 1.8 μL aliquot of 200 mM MMTS was added to each sample before a 10 min incubation at room temperature. The proteins in each sample were digested with 2 μL trypsin (1 mg/mL in 1mM HCl) at 37°C overnight.

Each of the peptide samples was labeled with 0.5 unit of an iTRAQ-8plex tag dissolved in 50 μL isopropanol. The samples derived from the 0, 1, 2.1, 2.7, 3.3, 4, 4.9 and 7.5 M SPROX buffers were labeled with the 113, 114, 115, 116, 117, 118, 119 and 121 iTRAQ tags. The labeling reactions were allowed to proceed at room temperature for 2 h. Within the 8 samples, 10 μL or 30 μL of each sample was combined. The combined samples, either 80 μL or 240 μL in volume, were either subjected to a desalting protocol using C18 MacroSpin columns (80 μL sample) or a methionine-containing peptide enrichment strategy (240 μL sample) prior to LC-MS/MS analysis. The above fractionation protocols were performed as described in Chapter 2. In the case of WT1, Delta1 and Delta2 SPROX experiments, the non-enhanced samples were each subjected to a single LC-MS/MS analysis while the methionine-enriched samples were each subjected to two replicate LC-MS/MS analyses. In the WT2 SPROX experiment, both the non-enhanced and the methionine-enriched samples were each subjected to a single LC-MS/MS analysis.

5.2.3 LC-MS/MS analyses

Samples from the WT1, Delta1 and Delta2 SPROX experiments were analyzed on a Thermo Scientific Q-Exactive Plus high-resolution mass spectrometer with a nanoAcuity UPLC system (Waters Corp) and a nanoelectrospray ionization source. Approximately 4 μL of sample was injected and trapped on a Symmetry C18 180 μm × 20 mm trapping column for 3 min at 5 μL/min (99.9/0.1 v/v water/acetonitrile 0.1% formic acid), followed by the analytical separation on a 75 μm × 250 mm column packed with 1.7 μm Acquity HSST3 C18 stationary phase (Waters Corp). Peptides were separated using a gradient of 3 to 30% acetonitrile with 0.1% formic acid over 90 min at a flow rate of 0.4 microliters/minute (μL/min) with a column temperature of 55 °C. Data collection on the Q-Exactive Plus mass spectrometer was performed in a data-dependent acquisition (DDA) mode of acquisition with a resolution (r) of 70,000 (at m/z 200) for full MS scan from m/z 375 – 1600 with a target AGC value of 1×10^6 ions in profile mode, followed by 20 MS/MS scans at r=17,500 (at m/z 200) in centroid mode, using an AGC target value of 1×10^5 ions, a max fill time of 60 ms, and normalized collision energy of 30 V. A 30 s dynamic exclusion was employed to decrease MS/MS oversampling. The total analysis cycle time for the sample injection was approximately 125 minutes. The resulting data were searched on Spectrum Mill Workbench Software B.04 Rev. B against SwissProt *Saccharomyces cerevisiae* database with trypsin digestion including tryptic peptides with up to 3 missed cleavages. Fixed modifications of MMTS on cysteine and iTRAQ® 8-Plex on N-terminus and lysine residues were included. Oxidation of

methionine residues and deamidation of asparagine residues were included as variable modifications. The precursor ion mass tolerance was set to 15 ppm and the product ion mass tolerance was set to 0.02 Da. Only peptides identified with global FDR < 1% were used in data analysis.

Samples from the WT2 SPROX experiment were analyzed with an Easy-nLC 1000 (Thermo Scientific) coupled to an Orbitrap Fusion mass spectrometer (Thermo Scientific). The LC system configured in a vented format consisted of a fused-silica nanospray needle (PicoTip™ emitter, 75 μm inner diameter, New Objective) packed in-house with Magic C18 AQ 100Å reverse-phase media (Michrom Bioresources Inc., 25 cm) and a trap (IntegraFrit™ Capillary, 100 μm inner diameter, New Objective) containing Magic C18 AQ 200Å (2 cm). 1.8 μg of peptides was loaded onto the column from each sample and separated using a two-mobile-phase system consisting of 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B). The chromatographic separation was achieved over a 106-min gradient from 3% to 50% B (3-6% B for 2 min, 6–33% B for 90 min, 30–50% B for 7 min, and 50% B for 5 min) at a flow rate of 300 nL/min. The mass spectrometer was operated in a data-dependent MS/MS mode over the m/z range of 350–1500. The mass resolution was set to 120,000. The cycle time was set to 3 seconds, and the most abundant ions from the precursor scan were selected for MS/MS analysis using 40% normalized HCD collision energy and analyzed with an Orbitrap with the resolution set to 15000. Selected ions were dynamically excluded for 30 seconds. The resulting data were searched on Proteome Discoverer 1.4.1.14 against the

Saccharomyces Genome Database with trypsin digestion including tryptic peptides with up to 2 missed cleavages. Fixed modifications of MMTS on cysteine and iTRAQ® 8-Plex on N-terminus and lysine residues were included. Oxidation of methionine residues was included as variable modifications. The precursor ion mass tolerance was set to 10 ppm and the product ion mass tolerance was set to 0.6 Da. Only peptides identified with high (global FDR < 1%) and medium (global FDR < 5%) scores were used in data analysis.

5.2.4 SPROX data analysis

The iTRAQ-SPROX data analysis process was as described in Chapter 3. Only peptides with raw iTRAQ intensities that were summed to greater than 1000 were included in analysis. For WT1, Delta1 and Delta2, product ion spectra with precursor isolation purity percent larger than 70% were used for analysis; while for WT2, product ion spectra with isolation interference smaller than 30% were used for analysis. Summarized in Table 14 are the N2-normalization factors generated for each reporter ion.

Table 14: N2-normalization factors generated for each reporter ion in the 4 SPROX experiments.

Experiment	Statistics	iTRAQ reporter ion m/z							
		113	114	115	116	117	118	119	121
WT1	Average	1.06	1.10	1.06	1.03	1.01	0.86	1.09	0.80
	S.D.	0.19	0.15	0.14	0.16	0.14	0.14	0.25	0.18
WT2	Average	0.77	0.76	1.14	0.90	1.07	1.24	1.01	1.10
	S.D.	0.48	0.39	0.51	0.33	0.38	0.43	0.40	0.64
Delta1	Average	1.18	1.40	0.97	0.86	0.73	0.86	1.00	1.00
	S.D.	0.19	0.28	0.14	0.14	0.14	0.16	0.28	0.21
Delta2	Average	1.38	1.16	0.98	0.84	0.90	1.04	0.96	0.74
	S.D.	0.37	0.28	0.20	0.20	0.24	0.30	0.22	0.24

5.3 Results and Discussion

5.3.1 Proteomic coverage

A total of 4 different comparative analyses were conducted to identify Puf6 protein-protein interaction. The different comparisons, the number of assayed peptides (proteins) in each comparison and the number of peptide and protein hits identified in each comparison are summarized in Table 15. A total of 492 unique proteins were assayed in the 4 comparisons. The hit rate for each comparison is ~5%, which is higher than the 1-2% false-positive rate of protein-target discovery that has been previously established for the iTRAQ-SPROX technique(*27*). The proteomic coverages in WT2 vs. Delta1 and WT2 vs. Delta2 comparisons are smaller than those in comparisons involving WT1. This is most likely due to the differences in LC-MS/MS acquisition methods between the Q-Exactive Plus and Orbitrap Fusion (see Experimental). Among the 492 assayed proteins, 16 proteins were previously reported to either directly bind or co-exist in stable complexes with Puf6(*173*). These physical interactions described in literature were detected by traditional techniques, mainly affinity capture MS(*173*). Out of the 16 proteins, 3 proteins were also hits in the iTRAQ-SPROX experiments (Table 16).

Table 15: Peptide (protein) coverage achieved in all 4 comparisons and the corresponding number of hits.

Comparison	Assayed[1]	Hits
WT1 vs. Delta1	703 (329)	36 (27)
WT1 vs. Delta2	668 (318)	38 (31)
WT2 vs. Delta1	260 (133)	13 (11)
WT2 vs. Delta2	264 (140)	19 (16)

[1]Peptides (proteins) identified in both groups of a given comparison.

5.3.2 Global hit analysis

Summarized in Table 16 are the hit peptides (proteins) identified with significant Puf6-induced thermodynamic stability changes. All of the hit peptide and proteins in Table 16 yielded consistent results in all of the comparative analysis. That is, if a protein hit was assayed in multiple comparisons, it was consistently identified as a hit. The hit peptides can be grouped into 3 classes (see Table 16). Included in class I are the strongest hits (identified as hits in 3 or 4 comparisons). The hits in class II include those that were consistently identified as hits in 2 comparisons, and class III includes peptides that were only identified as hits in one comparison. Out of the 492 total assayed proteins, a total of 46 peptides (39 proteins) were determined as hits, among which 2 peptides belong to class I, 11 peptides belong to class II and 33 belong to class III. Similar to protein-small molecule experiments by iTRAQ-SPROX (as described in Chapter 4), here the Puf6 deletion strain was used as the (-) ligand group with no detectable level of Puf6, while the wild-type strain was used as the (+) ligand group with endogenous Puf6 concentration of ~1 μM(*174*).

A bioinformatics analysis was performed using PANTHER(*134, 135*) to categorize the 492 total assayed proteins, the 39 hit proteins identified here, and the 95 proteins In the SGD that were previously annotated to have physical interactions with Puf6. Proteins were categorized based on their molecular functions and the percentage of proteins corresponding to each class was calculated (Figure 20). Notably, proteins with "translation regulator activity" and "catalytic activity" were enriched in the hits

(11% and 61%, respectively) as compared to all the assayed proteins (4% and 47%, respectively). Puf6 is known to regulate ASH1 mRNA translation(*150*) and it is not surprising that hit proteins with translation regulator activity might facilitate the process by interacting with Puf6. Interestingly, the proteins previously annotated in the SGD to have physical interactions with Puf6 showed a different distribution of molecular function (Figure 20), with "binding" and "structural molecule activity" being the 2 dominant classes instead of "binding" and "catalytic activity". The different enrichment pattern could be an indication that iTRAQ-SPROX provides complementary information to traditional techniques such as affinity capture MS.

Table 16: Hit peptides (proteins) and their corresponding transition midpoints (urea concentrations).

Accession number	Peptide Sequence	WT 1 (M)	WT 2 (M)	Delta 1 (M)	Delta 2 (M)	Hit class[1]
P32471	AIEM(OX)EGLTWGAHQFIPIGFGIK	2.1	2.4	3.6	4	II
Q12377	EFIPHSTEYMMQFAK	2.4	ND	4	3.6	III
P07264	AGMIKPDETTFQYTK	2.4	2.4	4.4	PQ[2]	II
P50861	AGIDEAHSMHNHGEDWGAAAVEMAVK	2.7	2.7	4.4	3.6	I
P26784[3]	GMVSHK	2.1	ND	4.4	PQ	III
P04046	YMDEDVHR	2.1	ND	3.3	ND	III
P03965	TIMINYNPETVSTDFDEVDR	2.4	ND	PQ	1	III
P53852	FPMHVTAVQNAIESITK	2.1	ND	3.6	3.6	II
P38791	ATDLIEAMK	2.7	ND	1.6	1.6	II
P32324	QATGGQAFPQM(OX)VFDHWSTLGSDPLDPTSK	3.3	ND	4.4	PQ	III
P00925	SIVPSGASTGVHEALEMRDEDKSK	3	NA	2.1	PQ	III
P00360	VINDAFGIEEGLM(OX)TTVHSMTATQK	2.7	3	6.2	ND	II
P00358	KVVITAPSSTAPMFVM(OX)GVNEEK	2.4	ND	6.2	ND	III
	VVITAPSSTAPM(OX)FVMGVNEEK	2.7	3.3	4.4	PQ	III
P00359	VINDAFGIEEGLM(OX)TTVHSLTATQK	3	3	6.2	6.2	I

	IALSRPNVEVVALNDPFITNDYAAYM(OX)FK	3.6	3.6	6.2	6.2	II
P32589	ENEMLAQDK	2.4	ND	0.5	0.5	II
P12398	VQGSVIGIDLGTTNSAVAIMEGK	2.7	ND	1.6	1.6	II
	DAGLSTSDISEVLLVGGMSR	2.7	ND	1.6	1.6	III
	MKETAEAYLGKPVK	3.3	ND	ND	1.6	III
P26637	EDYYTGTMIYGPYKGEK	2.4	ND	0.5	ND	III
P11632	ALSAYMFFANENR	2.4	ND	3.3	PQ	III
P15624	SLVAMGTHDLDSIEGPFHYR	3	ND	2.1	2.1	II
Q12447	IPHEYITIESMGK	2.4	ND	1	ND	III
P52893	GGYMELTGFSHEMR	2.4	ND	4	ND	III
Q12118	AMANKDYELAINK	2.4	ND	0.5	ND	III
P15019	FRFDLNEDAMATEK	2.7	ND	1.6	ND	III
P40070	GQQMQIELK	2.1	ND	4	ND	III
P05753	AILMQR	2.1	ND	PQ	3.6	III
	TDTTYPAGFMDVITLDATNENFR	2.7	ND	1.6	1	III
P06169	NIVEFHSDHM(OX)K	3.3	4	ND	2.1	III
P06787	LTDAEVDDMLR	2.4	ND	PQ	3.3	III
P07258	ITLADHDPVPYVNPMK	2.4	ND	ND	1	III
P12695	TVIENPLEMLL	2.1	2.4	ND	3.6	II
P16521	MTPSGHNWVSGQGAGPR	2.1	2.4	PQ	3.6	II
	AAATAAMTK	ND	2.4	0.5	1	III
P25491	VGIVPGEVIAPGMR	2.4	2.7	PQ	3.6	III
P33298	MSLAPEADLDSLIIR	2.4	ND	PQ	3.3	III
P38861	ISNTVQFMDPTTLQTADLSPSVYWR	2.7	ND	PQ	1.6	III
Q04344	LHDIPDEFLTDAMPIAK	3	PQ	PQ	4.4	III
Q08647	GMIIGNYNFSDASLNLGDLK	2.1	ND	ND	3.6	III
Q12389	ELAMQTFNVFK	2.4	ND	ND	3.3	III
P15108	MPEHQK	ND	2.4	ND	3.6	III
P47096	GGFTVMIVGGPNER	2.1	2.4	4	3.3	III
P17967	NMAPEYVK	ND	2.4	ND	1	III

[1]A total of 3 hit classes are defined here. Class I: hits identified in 3 or more comparisons; class II: hits identified in 2 comparisons; class III: hits identified in 1 comparison. [2]PQ: poor quality, ND: not detected. [3]Bolded protein were also identified to have physical interaction with Puf6. P26784: ref (*175*); P38861: ref (*176*); Q12389: ref (*175*).

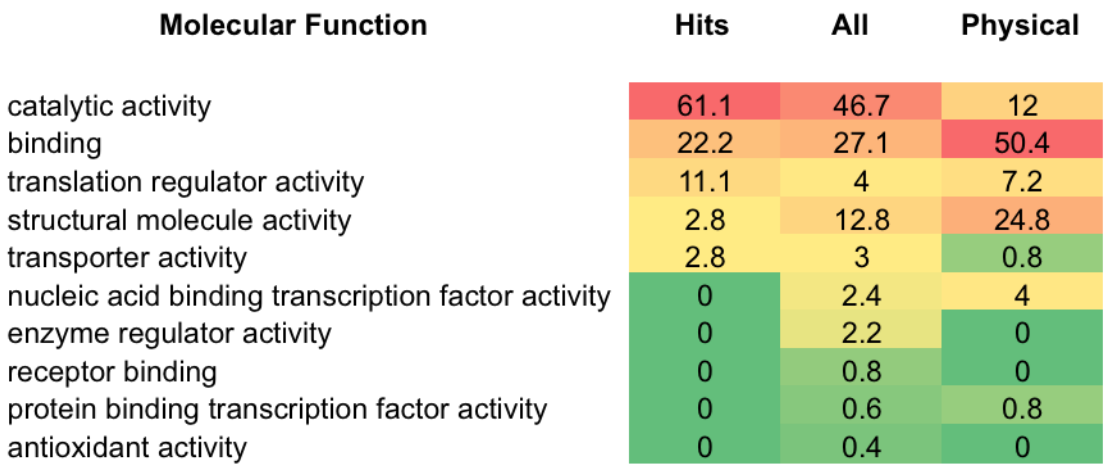

Figure 20: Bioinformatics analysis of the 492 total assayed proteins ("All"), the 39 hit proteins from Table 16 ("Hits") and the 95 proteins annotated to have physical interactions with Puf6 by SGD database(173) ("Physical"). The numbers in the boxes represent the percentage of proteins with a particular molecular function. The classification was performed using PANTHER(134, 135) version 10.0.

5.3.3 Previously annotated hits

A recent study demonstrated that Puf6, She2 and Loc1 form a ternary complex during ASH1 mRNA localization(*162*). Fun12 is also known to interact with Puf6 and the Fun12-Puf6 interaction represses ASH1 mRNA translation(*148*). Unfortunately, She2, Loc1 and Fun12 were not assayed in this study. Among the 39 hit proteins identified here (Table 16), 3 proteins were previously annotated in the SGD to directly interact with Puf6. Puf6 is known to participate in ribosomal large subunit biogenesis(*177*) and the 3 hit proteins, 60S ribosomal protein L16-A (P26784), 60S ribosomal export protein NMD3 (P38861) and ATP-dependent RNA helicase DBP10 (Q12389) are either a component of or closely related to the ribosomal large subunit(*178, 179*). Shown in

Figure 21 are the iTRAQ-SPROX data of the above 3 hit proteins. Of the 3 hit peptides (proteins), 2 peptides display destabilization behavior (P26784 and Q12389), which possibly results from indirect protein-protein interactions and/or allosteric effects of ligand binding. In addition to these 3 proteins, 13 other proteins previously annotated to be physical interactors of Puf6 in SGD were also assayed in this study. The 13 proteins did not show hit behavior in the iTRAQ-SPROX experiments. There are 2 possible explanations for this: 1) the protein domains to which the assayed methionine-containing peptides mapped do not participate in binding; or 2) these physical interactors identified in the affinity capture MS results could be false positives. For the remaining hits in Table 16, further experiments such as affinity MS, immunoprecipitation or SILAC-PP are required to validate the hit behavior. Furthermore, since some of the protein-protein interactions involving Puf6 are RNA-dependent. For example, the Puf6-Fun12 interaction is known to be RNA-dependent(*148*). To capture these RNA-dependent interactions, the iTRAQ-SPROX experiments may need to be performed with the addition of ASH1 mRNA.

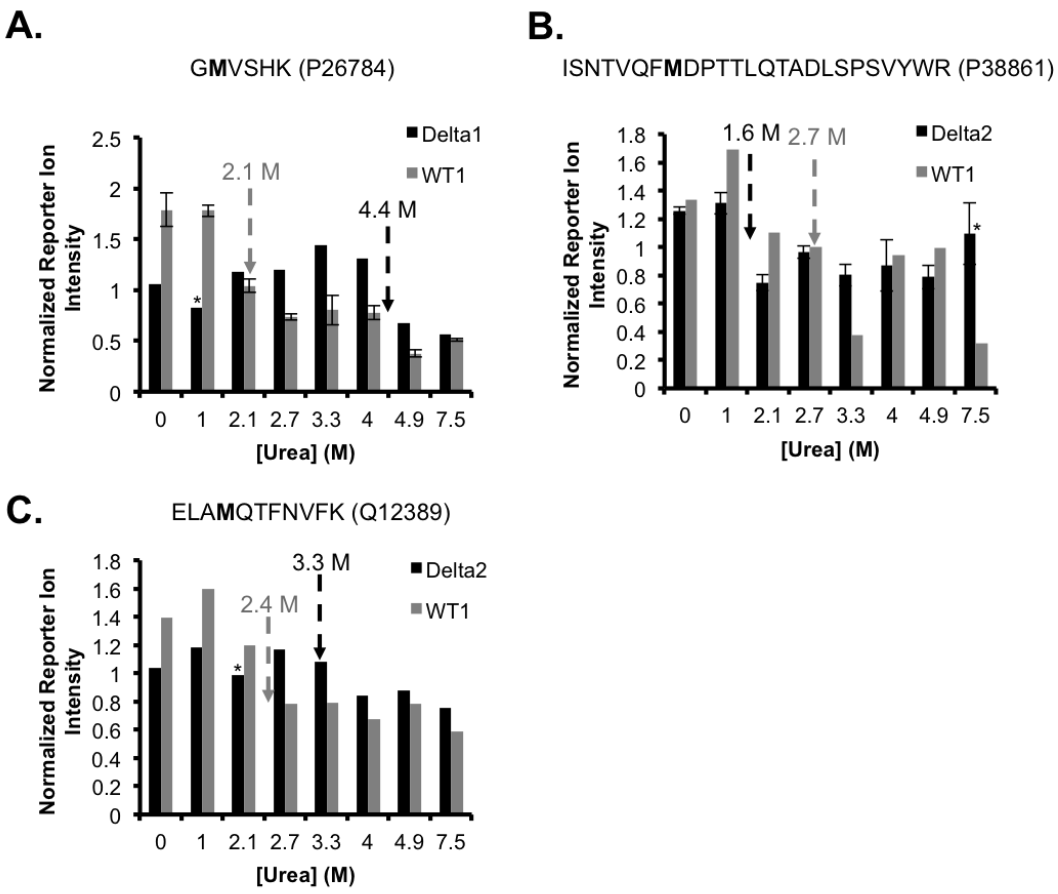

Figure 21: SPROX data of the 3 hit proteins that are also annotated to have physical interaction with Puf6 in SGD(173) (bolded proteins in Table 16). The arrows point to the $C_{1/2}$ values extracted from each dataset. The asterisk (*) sign denotes data outlier. In cases where multiple product ion spectra were collected the averaged data are shown with error bars representing ±1 standard deviation.

5.4 Conclusions

Puf6 is a non-canonical member of the PUF protein family and represses yeast ASH1 mRNA translation during localization. *In vivo*, Puf6 binds specifically to the 3'UTR of ASH1 mRNA for translation repression. However, no sequence-dependence was observed for Puf6-RNA interactions *in vitro*. It is hypothesized that other proteins interact with Puf6 to facilitate its recognition of ASH1 mRNA. To identify protein

interaction partners of Puf6, iTRAQ-SPROX experiments were conducted on both the wild-type and Puf6-deletion yeast lysates in this study. Among the 492 total assayed proteins, 39 proteins were identified as hits due to significant differences in protein thermodynamic stability between the wild-type and Puf6-deletion strains. These proteins possibly interact with Puf6 directly or indirectly to facilitate the RNA localization process. Out of the 39 proteins, 3 protein hits were ribosomal large subunit-associated proteins and were annotated to have physical interactions with Puf6 in SGD. Future experiments such as affinity MS, immunoprecipitation or SILAC-PP are required to validate the remaining 36 protein hits.

6. SPROX/Hybrid data analysis using a Bayesian approach

6.1 Introduction

In a typical proteome-wide SPROX/Hybrid experiment, 1000-2000 methionine/tryptophan-containing peptides from 500-1000 proteins are typically identified in both the minus and plus ligand samples (see references (*19-21*) and Chapter 3, 4) and subjected to downstream data analysis. Among these assayed peptides, ~30% are identified with at least 2 consecutive significant differences in N2-normalized iTRAQ reporter ion intensities (see Chapter 3 for detailed data analysis process). Typically these peptides with significant N2-normalized iTRAQ reporter ion intensity differences are assigned transition midpoints. Currently, SPROX/Hybrid transition midpoints are assigned by applying a set of "rules" after a visual inspection of the resulting iTRAQ intensities for each peptide(*15*). The visual inspection strategy has two major drawbacks: (1) since the process is not automated, it can be time-consuming and involve human errors, especially for large datasets; (2) instead of continuous concentration values, the transition midpoints assigned via visual inspection are discrete values corresponding to the specific chemical denaturant concentrations at or midway between the 8 chemical denaturant concentrations used in the experiment. To automate the assignment of transition midpoints and to facilitate the assignment of more specific and continuous midpoint values, a Bayesian nonlinear regression algorithm was developed in this work.

The number of data points generated in SPROX/Hybrid denaturation curves (typically 8) is relatively small compared to the number of parameters typically used to fit the data (4 parameters in the sigmoid model). This, combined with the frequent occurence of data outliers, makes it especially challenging to directly utilize nonlinear regression methods to fit the chemical denaturation curves generated in SPROX/Hybrid experiments. Bayesian approaches have the advantage of using prior knowledge of parameters and can be especially useful for obtaining reasonable parameter estimations when data points are sparse (*180*). To make the Bayesian algorithm developed here more robust to outliers, the student's t-distribution instead of normal distribution, was used for the mean function. Here a Bayesian algorithm for fitting SPROX/Hybrid data is developed and used in combination with a difference analysis to evaluate the false discovery rate of protein target discovery using the SPROX/Hybrid approach. The algorithm is also used to analyze the data generated in the omeprazole sulfide-binding experiments described in Chapter 4. The results obtained are compared to those obtained by visual inspection.

6.2 Methods

6.2.1 Bayesian algorithm for midpoint assignment

The mean function used in the Bayesian algorithm is the 4-parameter sigmoidal equation as follows:

$$y=A+\frac{B-A}{1+e^{-\frac{x-C_{1/2}}{b}}} \qquad \text{Equation 6}$$

In the context of a SPROX/Hybrid experiment, y denotes N2-normalized iTRAQ intensities, x denotes denaturant concentration (M), A is the pre-transition baseline, B is the post-transition baseline, $C_{1/2}$ represents transition midpoint and b represents the steepness of the transition.

The R-based Bayesian algorithm is included in the Appendix. Briefly, a student's t-distribution with 8 degrees of freedom was used for the mean function. A hierarchical model was used for parameter estimation and moderately informative priors were applied for the hyper-parameters. The prior knowledge of the 4 parameters (A, B, $C_{1/2}$ and b) was based on the parameter estimation by non-Bayesian regression method (the NLS function in R). Gibbs sampling was performed using the rjags package in R (need some more software info here) and 10000 iterations with 100 burn-ins were used for each denaturation curve. The average $C_{1/2}$ value and its standard deviation were calculated for the Gibbs samples. A relative standard deviation (RSD) was determined and used as the criteria for the goodness-of-fit evaluation. Only peptide data with RSD less than 30% were considered good quality data and used for hit identification.

6.2.2 Hit identification using the Bayesian algorithm

The experimental data used in this study were data from Hybrid-Experiments 1-4 described in Chapter 4. Initially, the control data generated in the 4 experiments were used to establish the false-positive rate of protein target discovery using Bayesian algorithm. In this so-called, Super Control experiment, the denaturation data generated

for the controls (i.e., the (-) ligand samples) in Hybrid-Experiments 1 and 2 were compared to the denaturation data generated for the controls (i.e., the (-) ligand samples) in Hybrid-Experiments 3 and 4. For simplicity, only denaturation data from the wild-type methionine-containing peptides were analyzed here. Apart from midpoint assignment, the data analysis process was the same as previously described in Chapter 4. A peptide was identified as a hit if it met all 3 criteria: (1) the absolute $C_{1/2}$ difference between the (-) and (+) ligand groups was larger than or equal to 0.5 M; (2) the RSD value for the fit was smaller than 30% for both groups; (3) at least 2 consecutive differences fell within the denaturant concentration range defined by the $C_{1/2}$ values of the (-) and (+) ligand group. Step (3) is the difference analysis process.

6.3 Results and Discussion

6.3.1 Benchmarking analysis of purified proteins

Initially, the Bayesian algorithm was used to extract transition midpoints from purified protein data. The purified protein data include denaturation curves of cyt c and lysozyme generated by HNSB protocol (Chapter 2). To better mimic the proteome-wide data which relied on iTRAQ-8plex quantitation, only 8 data points were chosen for each protein: 3 data points in the pre-transition region, 2 in the transition region and 3 in the post-transition region. Shown in Figure 22 are the denaturation data and the resulting sigmoidal curves constructed with parameters estimated by the Bayesian algorithm. The estimated $C_{1/2}$ values for cyt c and lysozyme are 1.95 ± 0.02 M and 3.75 ± 0.02 M, respectively, which are consistent with $C_{1/2}$ values, 1.94 and 3.75 M, estimated by non-

Bayesian approaches (the 4-parameter sigmoidal equation in SigmaPlot and the NLS function in R). The above benchmarking analysis suggests that the Bayesian algorithm is a valid approach for assigning transition midpoints to SPROX/Hybrid data sets consisting of 8 data points.

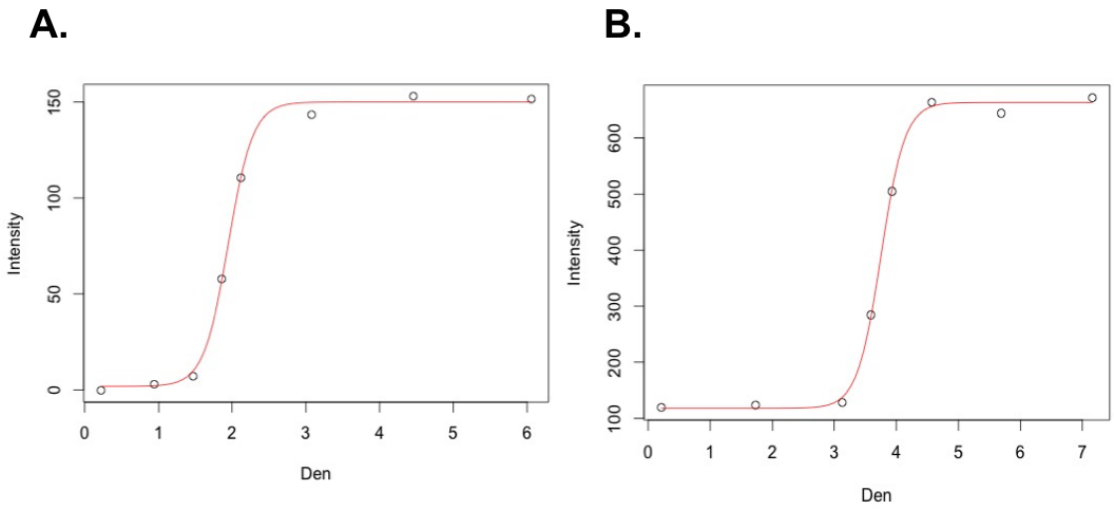

Figure 22: Denaturation data for (A) cyt c and (B) lysozyme generated by intact-protein level HNSB protocol (Chapter 2). The red curves represent the sigmoidal curves constructed with parameters estimated by the Bayesian algorithm.

6.3.2 Proteome-wide Hybrid data analysis

The Bayesian algorithm was next applied to iTRAQ-SPROX/Hybrid data generated in proteome-wide experiments. As part of the work in this Chapter, the Bayesian algorithm was used to assign midpoints to the chemical denaturation data generated from the (-) ligand samples from Hybrid-Exp1 and Hybrid-Exp2. Denaturation data from a total of 1045 wild-type methionine-containing peptides were used for regression. The goodness of fit was evaluated by the RSD values of the $C_{1/2}$ estimation. Only peptide denaturation curves with RSD less than 30% were treated as

good quality data and the resulting $C_{1/2}$ values were then compared with those obtained by visual inspection. Out of the 1045 denaturation curves, 497 or 48% of the data were considered good quality. The rest of the chemical denaturation data sets were poorly fit by the Bayesian algorithm. The poor fits were due to several reasons: (1) the denaturation curve did not have a clear transition (e.g., the amplitude of the transition was low; (2) the data was noisy; (3) the data did not have the expected sigmoidal structure .

The Bayesian algorithm was able to rescue denaturation data with outliers, which could possibly lead to ambiguous midpoint assignment by visual inspection. Figure 23A highlights a situation where there was a data outlier in the transition region and the $C_{1/2}$ value could be either 1.4 M or 2.15 M by visual inspection. However, although slightly affected by the outlier, the Bayesian algorithm still provided a $C_{1/2}$ estimation of 1.58 M with RSD of 29%. In contrast to the visual inspection strategy, the Bayesian algorithm is sensitive to the overall shape of the denaturation curve instead of only the transition region. As shown in Figure 23B, the relatively high first data point defined a high pre-transition baseline, which led to a steep transition and a relatively small $C_{1/2}$ value of 0.78 M. Since the visual inspection strategy only takes into account where N2-normalized iTRAQ intensities transition from the pre- to the post-transition region (i.e., cross 1), the first data point had no effect on the final $C_{1/2}$ assignment and the $C_{1/2}$ was assigned to be 1.4 M (Figure 23B).

$C_{1/2}$ values were compared between the Bayesian algorithm and the visual inspection strategy for the 497 good quality denaturation curves. The distribution of the $C_{1/2}$ differences between the two transition midpoint strategies is illustrated in Figure 23C. On average, the $C_{1/2}$ value assigned by the Bayesian algorithm was 0.28 M smaller than that assigned by visual inspection. This trend is likely due to the sensitivity of the Bayesian algorithm to the relatively high first data point. Overall the midpoint assignments using these two strategies were generally similar, as ~90% of all $C_{1/2}$ differences fell within -0.5 to 0.5.

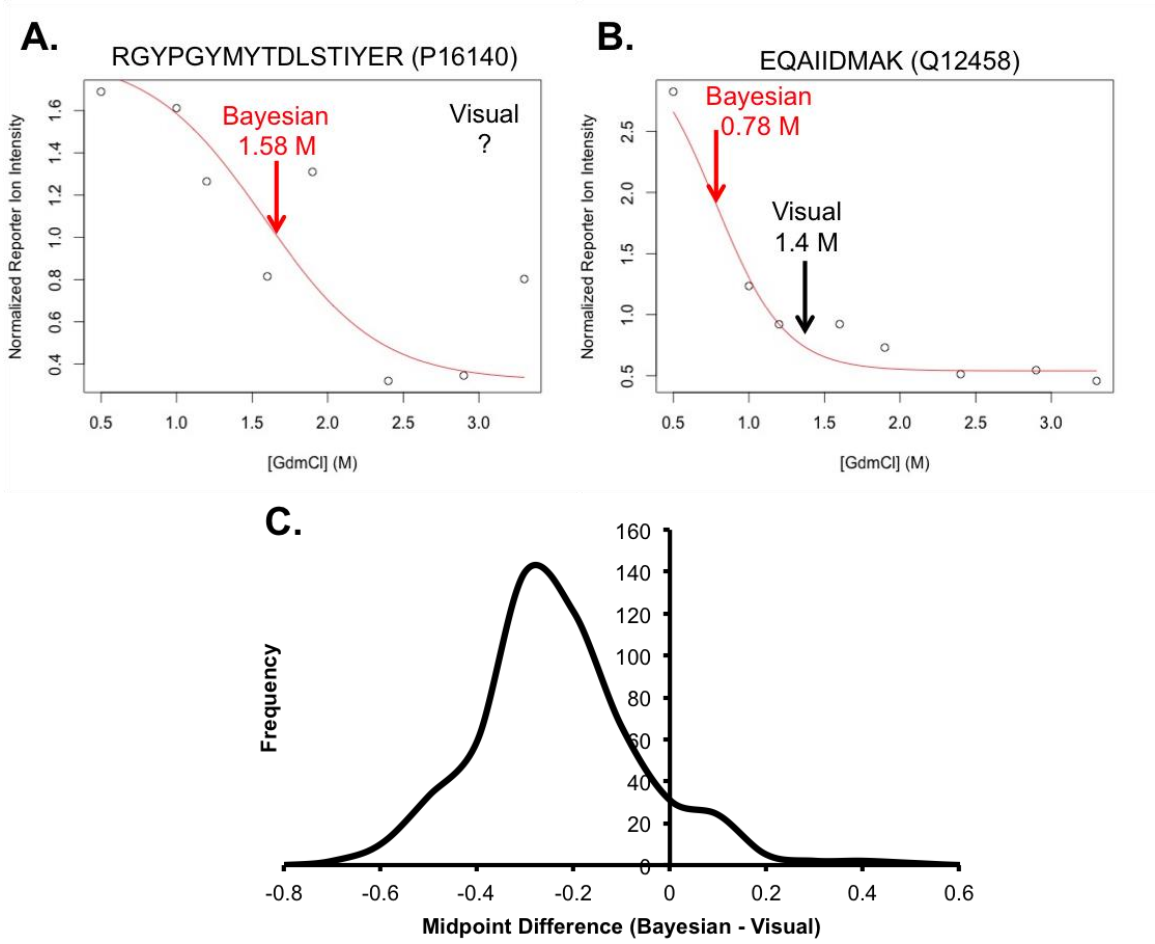

Figure 23: Application of the Bayesian approach to proteome-wide Hybrid data analysis. The dataset used here is the combined dataset of the (-) ligand groups from Hybrid-Exp1 and Hybrid-Exp2. (A) Example denaturation curve with a clearly assigned $C_{1/2}$ value by the Bayesian algorithm but an ambiguous $C_{1/2}$ assignment by visual inspection. (B) Example denaturation curve with a relatively high first data points, leading to a smaller $C_{1/2}$ value by the Bayesian algorithm than that by visual inspection. (C) Distribution of the $C_{1/2}$ differences between the Bayesian algorithm and visual inspection for the 497 good quality denaturation curves.

6.3.3 Application to the omeprazole sulfide-binding study

The Bayesian algorithm was also used to analyze the ligand binding data generated in the 4 Hybrid experiments described in Chapter 4. For this analysis, only the denaturation data for wild-type methionine-containing peptides were analyzed.

Summarized in Table 17 are the number of hit peptides determined by the Bayesian algorithm and the corresponding number of hits determined by visual inspection. In the Super Control experiment, the combined denaturation data from the (-) ligand groups of Hybrid-Exp1 and Hybrid-Exp2 were compared with the combined denaturation data from the (-) ligand groups of Hybrid-Exp3 and Hybrid-Exp4. In theory, hits from the Super Control experiment should be false positives. Not surprisingly, the peptide hit rate determined by the Bayesian approach was small ~0.5%. This is slightly smaller than the 1-3%, false-positive rate previously reported for SPROX using the visual inspection strategy (*27*). When applied to Hybrid-Exp1-4, the Bayesian approach generated a total of 3 peptide hits, 2 of which were also identified as hits by visual inspection (Table 17).

There are 3 possible scenarios that can occur when comparing between the hit lists generated by the Bayesian algorithm to that generated by visual inspection: (1) a peptide can be identified as a hit using both the Bayesian algorithm and visual inspection approach (Figure 24A); (2) a peptide can be identified as a hit by the Bayesian algorithm but not be a hit by visual inspection (Figure 24B); or (3) a peptide can be identified as a hit by visual inspection but not be a hit using the Bayesian algorithm (Figure 24C). For the peptide example in Figure 24A, the $C_{1/2}$ shift determined by the Bayesian algorithm is 1.34 M→2.01 M, which is consistent with the $C_{1/2}$ shift determined by visual inspection 1.4 M→1.9 M. For the peptide example in Figure 24B, the $C_{1/2}$ shift determined by the Bayesian algorithm is 0.92 M→1.45 M, while $C_{1/2}$ determined by visual inspection is 1.4 M and shows no shift. This discrepancy between the Bayesian

and visual inspection results highlighted in Figure 24B is likely due to the sensitivity of the Bayesian algorithm to the relatively high first data point, as discussed in section 6.3.2. For the peptide in example highlighted in Figure 24C, the $C_{1/2}$ shift determined by visual inspection is 1.9 M→1.1 M, while $C_{1/2}$ determined by the Bayesian algorithm is ~1.7 M and shows no shift. In this case the data in red color is of poor quality. The fifth data point needs to be overlooked for the visual inspection and the RSD value for the Bayesian estimation is ~40%, which is higher than the 30% criteria for hit identification.

Table 17: The number of assayed wild-type methionine-containing peptides (proteins) for each experiment and the corresponding number of hits determined by both Bayesian algorithm and visual inspection.

Experiment	Assayed[1]	Visual hits	Bayesian hits	Overlapping hits[2]
Super Control	1045 (543)	16 (15)	5 (5)	2 (2)
Hybrid-Exp1	393 (242)	8 (8)	1 (1)	1 (1)
Hybrid-Exp2	1003 (507)	8 (8)	0 (0)	0 (0)
Hybrid-Exp3	684 (356)	13 (11)	2 (2)	1 (1)
Hybrid-Exp4	835 (427)	5 (5)	0 (0)	0 (0)

[1]Wild-type methionine-containing peptides identified in both groups. [2]Common peptide (protein) hits between visual and Bayesian.

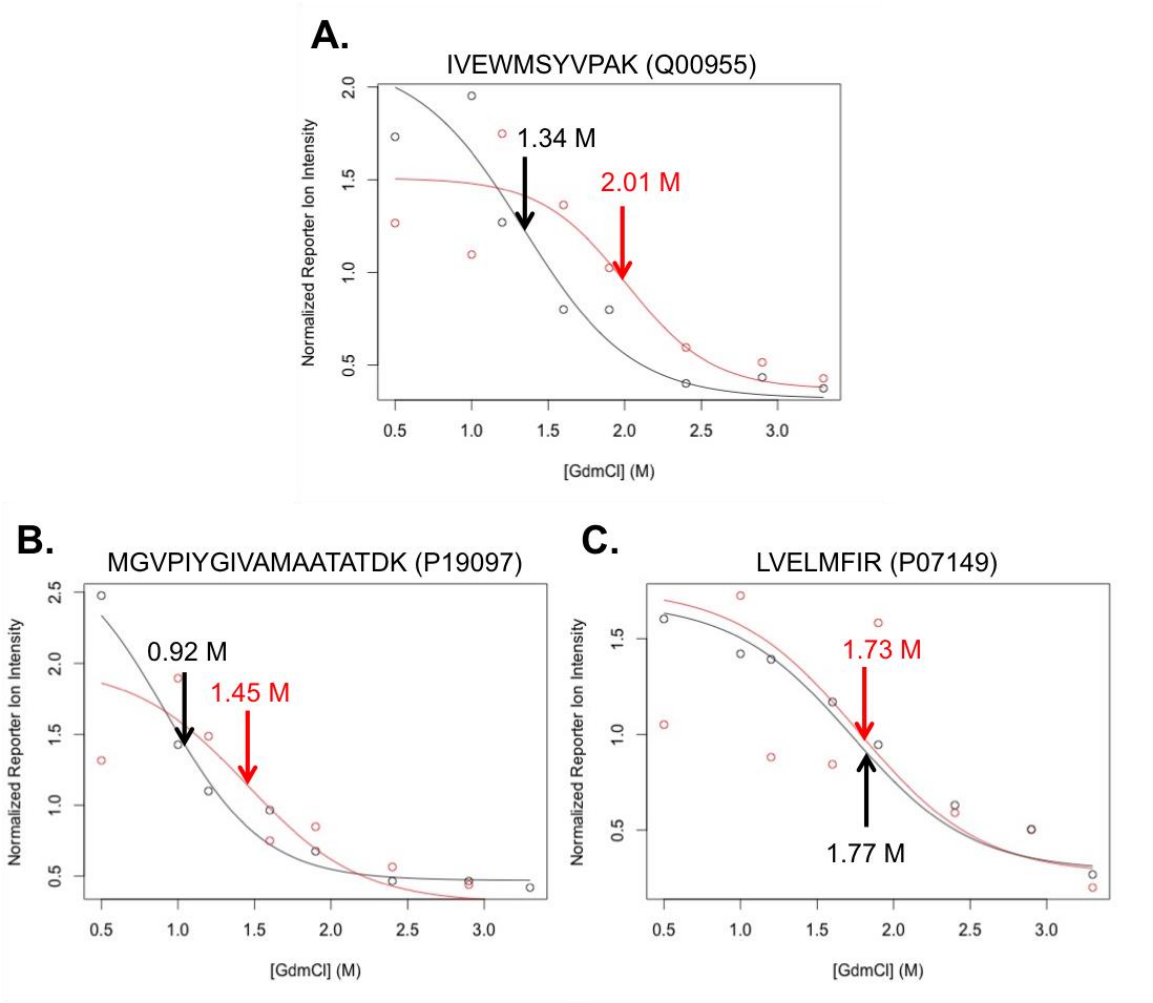

Figure 24: Application of the Bayesian algorithm in the omeprazole sulfide binding study. (A) Example peptide identified as a hit in Hybrid-Exp3 by both the Bayesian algorithm and visual inspection. (B) Example peptide identified in the Super Control experiment as a hit by the Bayesian algorithm but not a hit by visual inspection. (C) Example peptide identified in the Super Control experiment as a hit by visual inspection but not a hit by the Bayesian algorithm. The sigmoidal curves were constructed from the parameters estimated by the Bayesian algorithm. In (A), the black color denotes the (-) ligand group, while the red color denotes the (+) ligand group. In (B) (C), the black color denotes the combined data from Hybrid-Exp1,2, while the red color denotes the combined data from Hybrid-Exp3,4. The $C_{1/2}$ values estimated by the Bayesian algorithm are marked with arrows.

6.4 Conclusions

To automate the midpoint assignment process in the SPROX/Hybrid experiment and minimize human errors, a Bayesian algorithm was developed using student's t-distribution for the mean function. The algorithm, was successfully benchmarked using purified protein data and generated $C_{1/2}$ values consistent with those obtained by non-Bayesian approaches. Combined with the difference analysis process previously used for SPROX/Hybrid data analysis, the Bayesian algorithm was tested on proteome-wide Hybrid data and applied to identify hit peptides in the omeprazole sulfide-binding experiment described in Chapter 4. Overall the Bayesian algorithm provided similar $C_{1/2}$ assignments to visual inspection. Compared to visual inspection, the Bayesian algorithm was more sensitive to the data structure as modeled by the sigmoidal equation and was more tolerable to data outliers in the transition region. To further test its potential in SPROX/Hybrid data analysis, the Bayesian algorithm will be applied to more ligand-binding datasets.

7. Conclusions and future directions

Recent years have witnessed the development of several mass spectrometry-based approaches including SPROX, pulse proteolysis and CETSA, for the large-scale analysis of protein stability and protein-ligand binding interactions. Among these approaches, SPROX is a unique technique because it provides thermodynamic measurement of protein folding free energies that map to specific regions of the protein structure. Unlike pulse proteolysis and CETSA, one requirement in SPROX is that the protein under study should be identified with at least one globally protected methionine residue. To expand the scope of SPROX, the use of tryptophan peptides as stability probes was investigated in Chapter 2 of this dissertation and a SPROX-like protocol (termed "Hybrid protocol") was developed with ~50% increase in proteomic coverage. Chapter 3 describes the application of the Hybrid protocol in a proof-of-principle ligand-binding study. The interaction between Hsp90 and the well-studied Hsp90 inhibitor geldanamycin was successfully identified and quantified in the context of MCF-7 cell lysate. A K_d value consistent with literature values obtained on purified Hsp90 was determined for the complex and the Hybrid protocol also showed a protein domain-level specificity. Briefly, peptides were detected in all 3 major domains of Hsp90 by the Hybrid protocol and only the peptide from the N-terminal ligand-binding domain was identified as hit.

In addition to the well-known ligand geldanamycin, the small molecule omeprazole sulfide with less well-understood protein targets was studied by the Hybrid

protocol and the result was demonstrated in Chapter 4. The peptide hit rate for each replicate was 1-4%, which was on the same order as the false-positive rate of hit discovery previously determined for the SPROX technique. The 20 protein hits identified in all 4 biological replicates were likely to be false-positives since none of the proteins were identified as hits in more than one replicate. The main reason for the lack of convincing hits was the limited proteomic coverage. Although the Hybrid protocol was used, a total of 899 proteins were identified, which only mapped to ~14% of the yeast proteome.

Apart from protein-small molecule interactions, protein-protein interactions involving Puf6 were investigated by the iTRAQ-SPROX technique and described in Chapter 5. Comparative thermodynamic analyses were performed on wild-type and Puf6-deletion yeast strains. A total of 492 proteins were assayed and 39 proteins were determined to be targets of Puf6. Out of the 39 protein hits, 36 were previously unknown to interact with Puf6. Interestingly, bioinformatics analysis revealed that proteins with translation regulator activity were enriched in the hits as compared to all assayed proteins. This is not surprising since Puf6 was known to regulate ASH1 mRNA regulation and these newly identified protein targets likely facilitated the process by interacting with Puf6. Before this study, limited success had been achieved by applying the iTRAQ-SPROX technique to study protein-protein interactions. The primary reason for the success here was that instead of using wild-type cell lysates (-)/(+) spiked-in protein of interest, wild-type vs. deletion strain lysates were used. The deletion lysate

was a better (-) ligand sample since the endogenous protein level in the wild-type lysate could possibly interfere with hit detection.

A typical SPROX/Hybrid experiment generates 1000-2000 denaturation curves based on iTRAQ-8plex quantitation. During the data analysis process, ~30% of the denaturation data were subjected to transition midpoint assignment by visual inspection. To facilitate the midpoint assignment process and minimize human errors, a robust Bayesian algorithm was developed which estimated midpoint values using nonlinear regression. The algorithm and its application to the Hybrid datasets were described in Chapter 6. Around 40% of the SPROX/Hybrid data generated good-quality fitting results and overall the midpoints assigned by the algorithm were consistent with those assigned by visual inspection. Compared with visual inspection, the Bayesian algorithm was more tolerable to outliers in the transition region. When combined with difference analysis, the Bayesian approach provided a lower false-positive rate in ligand-binding studies.

Despite efforts in increasing proteomic coverage of the SPROX technique, the scope of the SPROX-like protocols is still fundamentally limited by the detection of selected amino acid residues. To further increase the proteomic coverage, more LC-MS/MS analyses or fractionation techniques such as SDS-PAGE, isoelectric focusing and 2D-HPLC can be used. Future research can also explore the use of SPROX technique for probing membrane proteins.

As discovery-based techniques, the SPROX/Hybrid protocols face the challenge of hit validation. It was demonstrated in Chapter 4 that performing biological replicates and corroborating with other energetics-based approaches such as pulse proteolysis could help eliminate false positives. Additionally, direct binding assays and pull-down assays can validate direct binding events. Furthermore, in order to relate the changes in protein stability with biological functions and processes, enzyme activity assays can be performed for selected hit proteins.

According to the fitting results by the Bayesian algorithm, ~60% of the data were considered poor quality. There are several explanations: (1) ratio compression of the iTRAQ quantitation leads to smaller transition amplitudes; (2) the data structure deviates from the sigmoidal model; (3) the data are noisy. To minimize co-elution and ratio compression, MS3 quantitation can be used in the future to generate better quality data with less interference. The MS3 quantitation is likely to increase the transition amplitude of the denaturation curve and lead to more accurate estimation of transition midpoint.

Appendix: R-based Bayesian algorithm for midpoint assignment

```
library("rjags")
setwd("/Users/Input file directory")
set.seed(42)
AllData=read.csv("Input file.csv") ##Put in file name.
Den=c(0.5,1,1.2,1.6,1.9,2.4,2.9,3.3) ##Enter denaturant concentration values.
iters=10000 ##Number of Gibbs samples.
Total=nrow(AllData)
Summary=matrix(NA,nrow=Total,ncol=8)
colnames(Summary)=c("C1/2","sd-C1/2","b","sd-b","Intensity0","sd-
Intensity0","Post_Intensity","sd-Post_Intensity")

for(j in 1:Total){
 STD=sd(as.numeric(AllData[j,5:12]))
 if (STD<=0.15){
  Summary[j,]=rep(NA,times=8)
 }
 if (STD>0.15){
  SelectedData=AllData[j,5:12]
  Average=as.matrix(cbind(Den,t(SelectedData)))
  N=nrow(Average)
  Intensity=as.numeric(Average[,2])
  Intensity.MAX=sort(Intensity)[N]
  Intensity.MAX2=sort(Intensity)[N-1]
  Intensity.MIN=min(Intensity)
  if(max(Intensity[3:N])==Intensity.MAX){
   upperbound=Intensity.MAX2
  }else{upperbound=Intensity.MAX}
  if(Intensity.MIN>=0.4){
   lowerbound=Intensity.MIN
  }else{lowerbound=0}
  jags=NULL
  Samples=NULL
  s.model.wt="
  model{
  for (i in 1:N){
  Intensity[i]~dt(mu[i],prec.Intensity,8)
  mu[i]<-Intensity0.Fac*upperbound+(Post_Intensity.Fac-
Intensity0.Fac)*upperbound/(1+exp(-(Den[i]-Midpoint)/b)) ##Mean function.
```

```
  }
  Logvar.Intensity~dunif(-2,6)
  prec.Intensity<-1/exp(Logvar.Intensity)

  Midpoint~dt(Midpoint.mean,prec.Midpoint,8) T(0,3.5)
  Midpoint.mean~dnorm(1.4,25/16) T(0,3.5)
  Logvar.Midpoint~dunif(-2,3)
  prec.Midpoint<-1/exp(Logvar.Midpoint)

  b~dt(b.mean,prec.b,8) T(0,1)
  b.mean~dnorm(0.1,1000) T(0,1)
  Logvar.b~dunif(-6,-1)
  prec.b<-1/exp(Logvar.b)

  Intensity0.Fac<-1.4*Intensity0.Fac.pre
  Intensity0.Fac.pre~dbeta(Intensity0.a,Intensity0.b) T(0,0.99)
  Intensity0.aOverb~dgamma(2,0.3)
  Intensity0.b~dgamma(2,0.1)
  Intensity0.a<-Intensity0.aOverb*Intensity0.b

  Post_Intensity.Fac<-0.7*Post_Intensity.Fac.pre
  Post_Intensity.Fac.pre~dbeta(Post_Intensity.a,Post_Intensity.b)
T(lowerbound/upperbound,0.99)
  Post_Intensity.aOverb~dgamma(2,7)
  Post_Intensity.b~dgamma(2,0.1)
  Post_Intensity.a<-Post_Intensity.aOverb*Post_Intensity.b
  }"

  writeLines(s.model.wt,con="s-model-wt.bug")
  jags=jags.model("s-model-
wt.bug",data=list(Intensity=Intensity,Den=Den,N=N,lowerbound=lowerbound,upperbou
nd=upperbound),n.chains=4,n.adapt=100)

Samples=jags.samples(jags,c("Intensity0.Fac","Post_Intensity.Fac","Midpoint","b"),iters)
  Summary[j,1]=mean(Samples$Midpoint)
  Summary[j,2]=sd(Samples$Midpoint)
  Summary[j,3]=mean(Samples$b)
  Summary[j,4]=sd(Samples$b)
  Summary[j,5]=upperbound*(mean(Samples$Intensity0.Fac))
  Summary[j,6]=upperbound*(sd(Samples$Intensity0.Fac))
  Summary[j,7]=upperbound*(mean(Samples$Post_Intensity.Fac))
```

```
   Summary[j,8]=upperbound*(sd(Samples$Post_Intensity.Fac))
  }
  }
SummaryAll=cbind(AllData,Summary)
write.csv(SummaryAll,file="Output file.csv")
```

References

1. A. R. Timerbaev, C. G. Hartinger, S. S. Aleksenko, B. K. Keppler, Interactions of antitumor metallodrugs with serum proteins: advances in characterization using modern analytical methodology. Chem Rev **106**, 2224-2248 (2006).

2. A. Virkamaki, K. Ueki, C. R. Kahn, Protein-protein interaction in insulin signaling and the molecular mechanisms of insulin resistance. J Clin Invest **103**, 931-943 (1999).

3. J. F. Brandts, L. N. Lin, Study of strong to ultratight protein interactions using differential scanning calorimetry. Biochemistry **29**, 6927-6940 (1990).

4. M. Straume, E. Freire, Two-dimensional differential scanning calorimetry: simultaneous resolution of intrinsic protein structural energetics and ligand binding interactions by global linkage analysis. Anal Biochem **203**, 259-268 (1992).

5. B. W. Sigurskjold, Exact analysis of competition ligand binding by displacement isothermal titration calorimetry. Anal Biochem **277**, 260-266 (2000).

6. A. D. Robertson, K. P. Murphy, Protein Structure and the Energetics of Protein Stability. Chem Rev **97**, 1251-1268 (1997).

7. J. J. Hill, C. A. Royer, Fluorescence approaches to study of protein-nucleic acid complexation. Methods Enzymol **278**, 390-416 (1997).

8. C. P. van Mierlo, E. Steensma, Protein folding and stability investigated by fluorescence, circular dichroism (CD), and nuclear magnetic resonance (NMR) spectroscopy: the flavodoxin story. J Biotechnol **79**, 281-298 (2000).

9. D. R. Shortle, Structural analysis of non-native states of proteins by NMR methods. Curr Opin Struct Biol **6**, 24-30 (1996).

10. S. Ghaemmaghami, M. C. Fitzgerald, T. G. Oas, A quantitative, high-throughput screen for protein stability. P Natl Acad Sci USA **97**, 8296-8301 (2000).

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

12. L. Y. Ma, M. C. Fitzgerald, A new H/D exchange- and mass spectrometry-based method for thermodynamic analysis of protein-DNA interactions. Chem Biol **10**, 1205-1213 (2003).

13. L. Tang et al., H/D exchange- and mass spectrometry-based strategy for the thermodynamic analysis of protein-ligand binding. Anal Chem **79**, 5869-5877 (2007).

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

15. E. C. Strickland et al., Thermodynamic analysis of protein-ligand binding interactions in complex biological mixtures using the stability of proteins from rates of oxidation. Nat Protoc **8**, 148-161 (2013).

16. P. F. Liu, D. Kihara, C. Park, Energetics-Based Discovery of Protein-Ligand Interactions on a Proteomic Scale. J Mol Biol **408**, 147-162 (2011).

17. D. M. Molina et al., Monitoring Drug Target Engagement in Cells and Tissues Using the Cellular Thermal Shift Assay. Science **341**, 84-87 (2013).

18. H. Franken et al., Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. Nat Protoc **10**, 1567-1593 (2015).

19. G. M. West et al., Quantitative proteomics approach for identifying protein-drug interactions in complex mixtures using protein stability measurements. Proc Natl Acad Sci U S A **107**, 9078-9082 (2010).

20. P. D. Dearmond, Y. Xu, E. C. Strickland, K. G. Daniels, M. C. Fitzgerald, Thermodynamic analysis of protein-ligand interactions in complex biological mixtures using a shotgun proteomics approach. J Proteome Res **10**, 4948-4958 (2011).

21. M. A. Geer, M. C. Fitzgerald, Characterization of the Saccharomyces cerevisiae ATP-Interactome using the iTRAQ-SPROX Technique. J Am Soc Mass Spectrom, (2015).

22. G. M. West, L. Tang, M. C. Fitzgerald, Thermodynamic analysis of protein stability and ligand binding using a chemical modification- and mass spectrometry-based strategy. Anal Chem **80**, 4175-4185 (2008).

23. F. Baud, S. Karlin, Measures of residue density in protein structures. Proc Natl Acad Sci U S A **96**, 12494-12499 (1999).

24. G. E. Reid, K. D. Roberts, R. J. Simpson, R. A. O'Hair, Selective identification and quantitative analysis of methionine containing peptides by charge derivatization and tandem mass spectrometry. J Am Soc Mass Spectrom **16**, 1131-1150 (2005).

25. M. P. Washburn, D. Wolters, J. R. Yates, 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol **19**, 242-247 (2001).

26. R. J. Suhadolnik, M. B. Lennon, T. Uematsu, J. E. Monahan, R. Baur, Role of adenine ring and adenine ribose of nicotinamide adenine dinucleotide in binding and catalysis with alcohol, lactate, and glyceraldehyde-3-phosphate dehydrogenases. J Biol Chem **252**, 4125-4133 (1977).

27. E. C. Strickland, M. A. Geer, J. Hong, M. C. Fitzgerald, False-positive rate determination of protein target discovery using a covalent modification- and mass spectrometry-based proteomics platform. J Am Soc Mass Spectrom **25**, 132-140 (2014).

28. D. T. Tran, J. Adhikari, M. C. Fitzgerald, Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC)-Based Strategy for Proteome-Wide Thermodynamic Analysis of Protein-Ligand Binding Interactions. Mol Cell Proteomics **13**, 1800-1813 (2014).

29. J. Adhikari, G. M. West, M. C. Fitzgerald, Global analysis of protein folding thermodynamics for disease state characterization. J Proteome Res **14**, 2287-2297 (2015).

30. P. D. DeArmond, G. M. West, H. T. Huang, M. C. Fitzgerald, Stable isotope labeling strategy for protein-ligand binding analysis in multi-component protein mixtures. J Am Soc Mass Spectrom **22**, 418-430 (2011).

31. Y. Xu, I. N. Falk, M. A. Hallen, M. C. Fitzgerald, Mass spectrometry- and lysine amidination-based protocol for thermodynamic analysis of protein folding and ligand binding interactions. Anal Chem **83**, 3555-3562 (2011).

32. G. M. West et al., Mass spectrometry-based thermal shift assay for protein-ligand binding analysis. Anal Chem **82**, 5573-5581 (2010).

33. C. Park, S. Marqusee, Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding. Nat Methods **2**, 207-212 (2005).

34. A. Hnizda et al., Conformational properties of nine purified cystathionine beta-synthase mutants. Biochemistry **51**, 4755-4763 (2012).

35. A. Hnizda et al., Cross-Talk between the Catalytic Core and the Regulatory Domain in Cystathionine beta-Synthase: Study by Differential Covalent Labeling and Computational Modeling. Biochemistry **49**, 10526-10534 (2010).

36. K. Hiraga, V. Derbyshire, J. T. Dansereau, P. Van Roey, M. Belfort, Minimization and stabilization of the Mycobacterium tuberculosis recA intein. J Mol Biol **354**, 916-926 (2005).

37. R. Ingrassia, G. Gerardi, G. Biasiotto, P. Arosio, Mutations of ferritin H chain C-terminus produced by nucleotide insertions have altered stability and functional properties. J Biochem **139**, 881-885 (2006).

38. J. Chiu, D. Tillett, P. E. March, Mutation of Phe102 to Ser in the carboxyl terminal helix of Escherichia coli thioredoxin affects the stability and processivity of T7 DNA polymerase. Proteins **64**, 477-485 (2006).

39. A. L. Mallam, S. E. Jackson, Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins. Nat Chem Biol **8**, 147-153 (2012).

40. J. P. Renn, M. Junker, R. N. Besingi, E. Braselmann, P. L. Clark, ATP-independent control of autotransporter virulence protein transport via the folding properties of the secreted protein. Chem Biol **19**, 287-296 (2012).

41. J. R. Kasper, P. F. Liu, C. Park, Structure of a partially unfolded form of Escherichia coli dihydrofolate reductase provides insight into its folding pathway. Protein Sci **23**, 1728-1737 (2014).

42. V. Castillo, A. Espargaro, V. Gordo, J. Vendrell, S. Ventura, Deciphering the role of the thermodynamic and kinetic stabilities of SH3 domains on their aggregation inside bacteria. Proteomics **10**, 4172-4185 (2010).

43. Y. Chang, C. Park, Mapping transient partial unfolding by protein engineering and native-state proteolysis. J Mol Biol **393**, 543-556 (2009).

44. A. Prudova et al., S-adenosylmethionine stabilizes cystathionine beta-synthase and modulates redox capacity. Proc Natl Acad Sci U S A **103**, 6489-6494 (2006).

45. T. Velkov, Thermodynamics of lipophilic drug binding to intestinal fatty acid binding protein and permeation across membranes. Mol Pharm **6**, 557-570 (2009).

46. M. S. Hanes, K. Ratcliff, S. Marqusee, T. M. Handel, Protein-protein binding affinities by pulse proteolysis: application to TEM-1/BLIP protein complexes. Protein Sci **19**, 1996-2000 (2010).

47. M. S. Kim, J. Song, C. Park, Determining protein stability in cell lysates by pulse proteolysis and Western blotting. Protein Sci **18**, 1051-1059 (2009).

48. N. J. Dawson, K. B. Storey, Regulation of tail muscle arginine kinase by reversible phosphorylation in an anoxia-tolerant crayfish. J Comp Physiol B **181**, 851-859 (2011).

49. N. J. Dawson, K. B. Storey, An enzymatic bridge between carbohydrate and amino acid metabolism: regulation of glutamate dehydrogenase by reversible phosphorylation in a severe hypoxia-tolerant crayfish. Journal of Comparative Physiology B-Biochemical Systemic and Environmental Physiology **182**, 331-340 (2012).

50. K. Abnous, C. A. Dieni, K. B. Storey, Suppression of MAPKAPK2 during mammalian hibernation. Cryobiology **65**, 235-241 (2012).

51. Y. Chang, J. P. Schlebach, R. A. VerHeul, C. Park, Simplified proteomics approach to discover protein-ligand interactions. Protein Sci **21**, 1280-1287 (2012).

52. J. Adhikari, M. C. Fitzgerald, SILAC-Pulse Proteolysis: A Mass Spectrometry-Based Method for Discovery and Cross-Validation in Proteome-Wide Studies of Ligand Binding. J Am Soc Mass Spectr **25**, 2073-2083 (2014).

53. R. Jafari et al., The cellular thermal shift assay for evaluating drug target interactions in cells. Nat Protoc **9**, 2100-2122 (2014).

54. M. M. Savitski et al., Tracking cancer drugs in living cells by thermal profiling of the proteome. Science **346**, 1255784 (2014).

55. Y. R. Xu, E. C. Strickland, M. C. Fitzgerald, Thermodynamic Analysis of Protein Folding and Stability Using a Tryptophan Modification Protocol. Anal Chem **86**, 7041-7048 (2014).

56. K. Gevaert et al., Chromatographic isolation of methionine-containing peptides for gel-free proteome analysis: identification of more than 800 Escherichia coli proteins. Mol Cell Proteomics **1**, 896-903 (2002).

57. D. T. Tran, S. Banerjee, A. I. Alayash, A. L. Crumbliss, M. C. Fitzgerald, Slow Histidine H/D Exchange Protocol for Thermodynamic Analysis of Protein Folding and Stability Using Mass Spectrometry. Anal Chem **84**, 1653-1660 (2012).

58. H. R. Horton, D. E. Koshland, Jr., A Highly Reactive Colored Reagent with Selectivity for the Tryptophan Residue in Proteins. 2-Hydroxy-5-Nitrobenzyl Bromide. Journal of the American Chemical Society **87**, 1126-1132 (1965).

59. H. R. Horton, W. P. Tucker, Dimethyl(2-hydroxy-5-nitrobenzyl)sulfonium salts. Water-soluble environmentally sensitive protein reagents. J Biol Chem **245**, 3397-3401 (1970).

60. M. Irie, M. Harada, F. Sawada, Studies on the state of tryptophan residues in ribonuclease from Aspergillus saitoi. J Biochem **72**, 1351-1359 (1972).

61. S. Lindskog, A. Nilsson, The location of tryptophanyl groups in human and bovine carbonic anhydrases. Ultraviolet difference spectra and chemical modification. Biochimica et biophysica acta **295**, 117-130 (1973).

62. M. Strohalm, J. Santrucek, R. Hynek, M. Kodicek, Analysis of tryptophan surface accessibility in proteins by MALDI-TOF mass spectrometry. Biochemical and biophysical research communications **323**, 1134-1138 (2004).

63. M. Nakazawa, K. Manabe, Differential Exposure of Tryptophan Residues in the Red and Far-Red Light Absorbing Forms of Phytochrome, as Revealed by Chemical Modification. Plant Cell Physiol **34**, 1097-1105 (1993).

64. T. E. Barman, D. E. Koshland, Jr., A colorimetric procedure for the quantitative determination of tryptophan residues in proteins. J Biol Chem **242**, 5771-5776 (1967).

65. M. Burr, D. E. Koshland, Use of Reporter Groups in Structure-Function Studies of Proteins. P Natl Acad Sci USA **52**, 1017-& (1964).

66. Y. M. Peyser, A. Muhlrad, M. M. Werber, Tryptophan-130 is the most reactive tryptophan residue in rabbit skeletal myosin subfragment-1. FEBS letters **259**, 346-348 (1990).

67. R. L. Lundblad, Chemical reagents for protein modification. (CRC Press, Boca Raton, Fla., ed. 3rd, 2005), pp. 339 p.

68. Y. Nozaki, The preparation of guanidine hydrochloride. Methods Enzymol **26**, 43-50 (1972).

69. A. Foettinger, A. Leitner, W. Lindner, Selective enrichment of tryptophan-containing peptides from protein digests employing a reversible derivatization with malondialdehyde and solid-phase capture on hydrazide beads. J Proteome Res **6**, 3827-3834 (2007).

70. K. D. Powell, M. C. Fitzgerald, Accuracy and precision of a new H/D exchange- and mass spectrometry-based technique for measuring the thermodynamic properties of protein-peptide complexes. Biochemistry **42**, 4962-4970 (2003).

71. M. Strohalm, M. Kodicek, M. Pechar, Tryptophan modification by 2-hydroxy-5-nitrobenzyl bromide studied by MALDI-TOF mass spectrometry. Biochemical and biophysical research communications **312**, 811-816 (2003).

72. K. P. Wilson, B. A. Malcolm, B. W. Matthews, Structural and thermodynamic analysis of compensating mutations within the core of chicken egg white lysozyme. J Biol Chem **267**, 10842-10849 (1992).

73. F. Ahmad, C. C. Bigelow, Estimation of the free energy of stabilization of ribonuclease A, lysozyme, alpha-lactalbumin, and myoglobin. J Biol Chem **257**, 12935-12938 (1982).

74. T. Imoto, L. N. Johnson, A. C. T. North, D. C. Phillips, J. A. Rupley, in The Enzymes. (Academic Press, New York, 1972), vol. 7, pp. 666-836.

75. C. T. Supuran, A. Scozzafava, A. Casini, Carbonic anhydrase inhibitors. Medicinal research reviews **23**, 146-189 (2003).

76. Y. Saito, A. Wada, Comparative study of GuHCl denaturation of globular proteins. I. Spectroscopic and chromatographic analysis of the denaturation curves of ribonuclease A, cytochrome c, and pepsinogen. Biopolymers **22**, 2105-2122 (1983).

77. G. A. Mines, T. Pascher, S. C. Lee, J. R. Winkler, H. B. Gray, Cytochrome c folding triggered by electron transfer. Chem Biol **3**, 491-497 (1996).

78. M. M. Santoro, D. W. Bolen, Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. Biochemistry **27**, 8063-8068 (1988).

79. C. Tanford, R. H. Pain, N. S. Otchin, Equilibrium and kinetics of the unfolding of lysozyme (muramidase) by guanidine hydrochloride. J Mol Biol **15**, 489-504 (1966).

80. T. R. Sosnick, L. Mayne, S. W. Englander, Molecular collapse: the rate-limiting step in two-state cytochrome c folding. Proteins **24**, 413-426 (1996).

81. R. W. Henkens, B. B. Kitchell, S. C. Lottich, P. J. Stein, T. J. Williams, Detection and Characterization Using Circular-Dichroism and Fluorescence Spectroscopy of a Stable Intermediate Conformation Formed in the Denaturation of Bovine Carbonic-Anhydrase with Guanidinium Chloride. Biochemistry **21**, 5918-5923 (1982).

82. A. G. Murzin, S. E. Brenner, T. Hubbard, C. Chothia, SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol **247**, 536-540 (1995).

83. L. H. Pearl, C. Prodromou, Structure, function, and mechanism of the Hsp90 molecular chaperone. Adv Protein Chem **59**, 157-186 (2001).

84. R. Zhao et al., Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. Cell **120**, 715-727 (2005).

85. S. H. Millson et al., Chaperone ligand-discrimination by the TPR-domain protein Tah1. Biochem J **413**, 261-268 (2008).

86. S. H. Millson et al., Investigating the protein-protein interactions of the yeast Hsp90 chaperone system by two-hybrid analysis: potential uses and limitations of this approach. Cell Stress Chaperones **9**, 359-368 (2004).

87. K. Eckert et al., The Pih1-Tah1 cochaperone complex inhibits Hsp90 molecular chaperone ATPase activity. J Biol Chem **285**, 31304-31312 (2010).

88. J. L. Johnson, A. Halas, G. Flom, Nucleotide-dependent interaction of Saccharomyces cerevisiae Hsp90 with the cochaperone proteins Sti1, Cpr6, and Sba1. Mol Cell Biol **27**, 768-776 (2007).

89. A. Kamal et al., A high-affinity conformation of Hsp90 confers tumour selectivity on Hsp90 inhibitors. Nature **425**, 407-410 (2003).

90. Y. Fukuyo, C. R. Hunt, N. Horikoshi, Geldanamycin and its anti-cancer activities. Cancer Lett **290**, 24-35 (2010).

91. S. C. Onuoha et al., Mechanistic studies on Hsp90 inhibition by ansamycin derivatives. J Mol Biol **372**, 287-297 (2007).

92. M. E. Cardenas et al., Antifungal activities of antineoplastic agents: Saccharomyces cerevisiae as a model system to study drug action. Clin Microbiol Rev **12**, 583-611 (1999).

93. C. M. Cazal et al., Evaluation of effect of triterpenes and limonoids on cell growth, cell cycle and apoptosis in human tumor cell line. Anticancer Agents Med Chem **10**, 769-776 (2010).

94. L. Whitesell, E. G. Mimnaugh, B. De Costa, C. E. Myers, L. M. Neckers, Inhibition of heat shock protein HSP90-pp60v-src heteroprotein complex formation by benzoquinone ansamycins: essential role for stress proteins in oncogenic transformation. Proc Natl Acad Sci U S A **91**, 8324-8328 (1994).

95. S. J. Uddin et al., Gedunin, a limonoid from Xylocarpus granatum, inhibits the growth of CaCo-2 colon cancer cell line in vitro. Phytother Res **21**, 757-761 (2007).

96. E. R. Geertsma, FX cloning: a simple and robust high-throughput cloning method for protein expression. Methods Mol Biol **1116**, 153-164 (2014).

97. G. E. Brandt, M. D. Schmidt, T. E. Prisinzano, B. S. Blagg, Gedunin, a novel hsp90 inhibitor: semisynthesis of derivatives and preliminary structure-activity relationships. J Med Chem **51**, 6495-6502 (2008).

98. C. A. Patwardhan et al., Gedunin inactivates the co-chaperone p23 protein causing cancer cell death by apoptosis. J Biol Chem **288**, 7313-7325 (2013).

99. J. Franke, S. Eichner, C. Zeilinger, A. Kirschning, Targeting heat-shock-protein 90 (Hsp90) by natural products: geldanamycin, a show case in cancer therapy. Nat Prod Rep **30**, 1299-1323 (2013).

100. S. G. Kamath et al., Gedunin, a novel natural substance, inhibits ovarian cancer cell proliferation. Int J Gynecol Cancer **19**, 1564-1569 (2009).

101. M. M. Ali et al., Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. Nature **440**, 1013-1017 (2006).

102. G. E. Karagoz et al., N-terminal domain of human Hsp90 triggers binding to the cochaperone p23. Proc Natl Acad Sci U S A **108**, 580-585 (2011).

103. C. E. Stebbins et al., Crystal structure of an Hsp90-geldanamycin complex: targeting of a protein chaperone by an antitumor agent. Cell **89**, 239-250 (1997).

104. S. M. Roe et al., Structural basis for inhibition of the Hsp90 molecular chaperone by the antitumor antibiotics radicicol and geldanamycin. J Med Chem **42**, 260-266 (1999).

105. K. Patel et al., Engineered biosynthesis of geldanamycin analogs for Hsp90 inhibition. Chem Biol **11**, 1625-1633 (2004).

106. B. Panaretou et al., ATP binding and hydrolysis are essential to the function of the Hsp90 molecular chaperone in vivo. EMBO J **17**, 4829-4836 (1998).

107. T. W. Schulte, L. M. Neckers, The benzoquinone ansamycin 17-allylamino-17-demethoxygeldanamycin binds to HSP90 and shares important biologic activities with geldanamycin. Cancer Chemother Pharmacol **42**, 273-279 (1998).

108. G. Chiosis et al., 17AAG: low target binding affinity and potent cell activity--finding an explanation. Mol Cancer Ther **2**, 123-129 (2003).

109. L. Neckers, Heat shock protein 90 inhibition by 17-allylamino-17-demethoxygeldanamycin: a novel therapeutic approach for treating hormone-refractory prostate cancer. Clin Cancer Res **8**, 962-966 (2002).

110. G. Chiosis et al., A small molecule designed to bind to the adenine nucleotide pocket of Hsp90 causes Her2 degradation and the growth arrest and differentiation of breast cancer cells. Chem Biol **8**, 289-299 (2001).

111. J. K. Myers, C. N. Pace, J. M. Scholtz, Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. Protein Sci **4**, 2138-2148 (1995).

112. C. C. Lee, T. W. Lin, T. P. Ko, A. H. Wang, The hexameric structures of human heat shock protein 90. PLoS One **6**, e19961 (2011).

113. M. A. Geer, Duke University, (2015).

114. S. F. Altschul et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25**, 3389-3402 (1997).

115. L. T. Gooljarsingh et al., A biochemical rationale for the anticancer effects of Hsp90 inhibitors: slow, tight binding inhibition by geldanamycin and its analogues. Proc Natl Acad Sci U S A **103**, 7625-7630 (2006).

116. WHO, "World Malaria Report 2015," (World Health Organization, Geneva, 2015).

117. M. Prudencio, A. Rodriguez, M. M. Mota, The silent path to thousands of merozoites: the Plasmodium liver stage. Nat Rev Microbiol **4**, 849-856 (2006).

118. L. H. Miller, D. I. Baruch, K. Marsh, O. K. Doumbo, The pathogenic basis of malaria. Nature **415**, 673-679 (2002).

119. M. L. Baniecki, D. F. Wirth, J. Clardy, High-throughput Plasmodium falciparum growth assay for malaria drug discovery. Antimicrob Agents Chemother **51**, 716-723 (2007).

120. F. J. Gamo et al., Thousands of chemical starting points for antimalarial lead identification. Nature **465**, 305-310 (2010).

121. W. A. Guiguemde et al., Chemical genetics of Plasmodium falciparum. Nature **465**, 311-315 (2010).

122. D. Plouffe et al., In silico activity profiling reveals the mechanism of action of antimalarials discovered in a high-throughput screen. Proc Natl Acad Sci U S A **105**, 9059-9064 (2008).

123. E. L. Flannery, A. K. Chatterjee, E. A. Winzeler, Antimalarial drug discovery - approaches and progress towards new medicines. Nat Rev Microbiol **11**, 849-862 (2013).

124. M. Carraz et al., A plant-derived morphinan as a novel lead compound active against malaria liver stages. PLoS Med **3**, e513 (2006).

125. S. Meister et al., Imaging of Plasmodium liver stages to drive next-generation antimalarial drug discovery. Science **334**, 1372-1377 (2011).

126. E. R. Derbyshire, M. Prudencio, M. M. Mota, J. Clardy, Liver-stage malaria parasites vulnerable to diverse chemical scaffolds. Proc Natl Acad Sci U S A **109**, 8511-8516 (2012).

127. F. P. da Cruz et al., Drug screen targeted at Plasmodium liver stages identifies a potent multistage antimalarial drug. J Infect Dis **205**, 1278-1286 (2012).

128. A. Nagle et al., Imidazolopiperazines: lead optimization of the second-generation antimalarial agents. J Med Chem **55**, 4244-4273 (2012).

129. T. Wu et al., Imidazolopiperazines: hit to lead optimization of new antimalarial agents. J Med Chem **54**, 5116-5130 (2011).

130. T. Andersson, L. Weidolf, Stereoselective disposition of proton pump inhibitors. Clin Drug Investig **28**, 263-279 (2008).

131. P. Lindberg, P. Nordberg, T. Alminger, A. Brandstrom, B. Wallmark, The mechanism of action of the gastric acid secretion inhibitor omeprazole. J Med Chem **29**, 1327-1329 (1986).

132. L. Olbe, E. Carlsson, P. Lindberg, A proton-pump inhibitor expedition: the case histories of omeprazole and esomeprazole. Nat Rev Drug Discov **2**, 132-139 (2003).

133. A. Shevchenko, H. Tomas, J. Havlis, J. V. Olsen, M. Mann, In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc **1**, 2856-2860 (2006).

134. H. Mi, A. Muruganujan, J. T. Casagrande, P. D. Thomas, Large-scale gene function analysis with the PANTHER classification system. Nat Protoc **8**, 1551-1566 (2013).

135. H. Mi, A. Muruganujan, P. D. Thomas, PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res **41**, D377-386 (2013).

136. P. Picotti et al., A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. Nature **494**, 266-270 (2013).

137. K. C. Martin, A. Ephrussi, mRNA localization: gene expression in the spatial dimension. Cell **136**, 719-730 (2009).

138. A. R. Buxbaum, G. Haimovich, R. H. Singer, In the right place at the right time: visualizing and understanding mRNA localization. Nat Rev Mol Cell Biol **16**, 95-109 (2015).

139. R. M. Long et al., Mating type switching in yeast controlled by asymmetric localization of ASH1 mRNA. Science **277**, 383-387 (1997).

140. D. L. Beach, E. D. Salmon, K. Bloom, Localization and anchoring of mRNA in budding yeast. Curr Biol **9**, 569-578 (1999).

141. E. Bertrand et al., Localization of ASH1 mRNA particles in living yeast. Mol Cell **2**, 437-445 (1998).

142. Z. Shen, A. St-Denis, P. Chartrand, Cotranscriptional recruitment of She2p by RNA pol II elongation factor Spt4-Spt5/DSIF promotes mRNA localization to the yeast bud. Genes Dev **24**, 1914-1926 (2010).

143. M. Muller et al., A cytoplasmic complex mediates specific mRNA recognition and localization in yeast. PLoS Biol **9**, e1000611 (2011).

144. F. Bohl, C. Kruse, A. Frank, D. Ferring, R. P. Jansen, She2p, a novel RNA-binding protein tethers ASH1 mRNA to the Myo4p myosin motor via She3p. EMBO J **19**, 5514-5524 (2000).

145. R. M. Long, W. Gu, E. Lorimer, R. H. Singer, P. Chartrand, She2p is a novel RNA-binding protein that recruits the Myo4p-She3p complex to ASH1 mRNA. EMBO J **19**, 6592-6601 (2000).

146. P. A. Takizawa, A. Sil, J. R. Swedlow, I. Herskowitz, R. D. Vale, Actin-dependent localization of an RNA encoding a cell-fate determinant in yeast. Nature **389**, 90-93 (1997).

147. A. Niedner, M. Muller, B. T. Moorthy, R. P. Jansen, D. Niessing, Role of Loc1p in assembly and reorganization of nuclear ASH1 messenger ribonucleoprotein particles in yeast. Proc Natl Acad Sci U S A **110**, E5049-5058 (2013).

148. Y. Deng, R. H. Singer, W. Gu, Translation of ASH1 mRNA is repressed by Puf6p-Fun12p/eIF5B interaction and released by CK2 phosphorylation. Genes Dev **22**, 1037-1050 (2008).

149. N. Paquin et al., Local activation of yeast ASH1 mRNA translation through phosphorylation of Khd1p by the casein kinase Yck1p. Mol Cell **26**, 795-809 (2007).

150. W. Gu, Y. Deng, D. Zenklusen, R. H. Singer, A new yeast PUF family protein, Puf6p, represses ASH1 mRNA translation and is required for its localization. Genes Dev **18**, 1452-1465 (2004).

151. K. Irie et al., The Khd1 protein, which has three KH RNA-binding motifs, is required for proper localization of ASH1 mRNA in yeast. EMBO J **21**, 1158-1167 (2002).

152. C. Qiu, K. L. McCann, R. N. Wine, S. J. Baserga, T. M. Hall, A divergent Pumilio repeat protein family for pre-rRNA processing and mRNA localization. Proc Natl Acad Sci U S A **111**, 18554-18559 (2014).

153. X. Wang, J. McLachlan, P. D. Zamore, T. M. Hall, Modular recognition of RNA by a human pumilio-homology domain. Cell **110**, 501-512 (2002).

154. H. T. Jenkins, R. Baker-Wilding, T. A. Edwards, Structure and RNA binding of the mouse Pumilio-2 Puf domain. J Struct Biol **167**, 271-276 (2009).

155. Y. Wang, L. Opperman, M. Wickens, T. M. Hall, Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein. Proc Natl Acad Sci U S A **106**, 20186-20191 (2009).

156. X. Wang, P. D. Zamore, T. M. Hall, Crystal structure of a Pumilio homology domain. Mol Cell **7**, 855-865 (2001).

157. M. T. Miller, J. J. Higgin, T. M. Hall, Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. Nat Struct Mol Biol **15**, 397-402 (2008).

158. M. Wickens, D. S. Bernstein, J. Kimble, R. Parker, A PUF family portrait: 3'UTR regulation as a way of life. Trends Genet **18**, 150-157 (2002).

159. C. G. Cheong, T. M. Hall, Engineering RNA sequence specificity of Pumilio repeats. Proc Natl Acad Sci U S A **103**, 13635-13639 (2006).

160. S. Dong et al., Specific and modular binding code for cytosine recognition in Pumilio/FBF (PUF) RNA-binding domains. J Biol Chem **286**, 26732-26742 (2011).

161. A. Filipovska, M. F. Razif, K. K. Nygard, O. Rackham, A universal code for RNA recognition by PUF proteins. Nat Chem Biol **7**, 425-427 (2011).

162. K. Shahbabian, C. Jeronimo, A. Forget, F. Robert, P. Chartrand, Co-transcriptional recruitment of Puf6 by She2 couples translational repression to mRNA localization. Nucleic Acids Res **42**, 8692-8704 (2014).

163. T. Ito et al., A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci U S A **98**, 4569-4574 (2001).

164. P. Uetz et al., A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature **403**, 623-627 (2000).

165. C. von Mering et al., Comparative assessment of large-scale data sets of protein-protein interactions. Nature **417**, 399-403 (2002).

166. A. C. Gavin et al., Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature **415**, 141-147 (2002).

167. N. J. Krogan et al., Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature **440**, 637-643 (2006).

168. Y. Ho et al., Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature **415**, 180-183 (2002).

169. H. Zhu et al., Global analysis of protein activities using proteome chips. Science **293**, 2101-2105 (2001).

170. A. H. Tong et al., Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science **294**, 2364-2368 (2001).

171. A. H. Tong et al., Global mapping of the yeast genetic interaction network. Science **303**, 808-813 (2004).

172. L. Ma, M. C. Fitzgerald, A new H/D exchange- and mass spectrometry-based method for thermodynamic analysis of protein-DNA interactions. Chem Biol **10**, 1205-1213 (2003).

173. J. M. Cherry et al., Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res **40**, D700-705 (2012).

174. S. Ghaemmaghami et al., Global analysis of protein expression in yeast. Nature **425**, 737-741 (2003).

175. S. R. Collins et al., Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae. Mol Cell Proteomics **6**, 439-450 (2007).

176. I. Lee, Z. Li, E. M. Marcotte, An improved, bias-reduced probabilistic functional gene network of baker's yeast, Saccharomyces cerevisiae. PLoS One **2**, e988 (2007).

177. Z. Li et al., Rational extension of the ribosome biogenesis pathway using network-guided genetics. PLoS Biol **7**, e1000213 (2009).

178. M. Altvater et al., Targeted proteomics reveals compositional dynamics of 60S pre-ribosomes after nuclear export. Mol Syst Biol **8**, 628 (2012).

179. J. Talkish, J. Zhang, J. Jakovljevic, E. W. Horsey, J. L. Woolford, Jr., Hierarchical recruitment into nascent ribosomes of assembly factors required for 27SB pre-rRNA processing in Saccharomyces cerevisiae. Nucleic Acids Res **40**, 8646-8661 (2012).

180. A. Gelman et al., Bayesian data analysis. (Taylor & Francis Group, Boca Raton, 2014).

Biography

Yingrong Xu was born on September 12, 1990 in Nanchang, Jiangxi province, China. She attended Peking University (Beijing, China) from 2007-2011 and graduated in July, 2011 with a B.S. in Materials Chemistry. She attended graduate school in the Department of Chemistry at Duke University (Durham, NC) and worked as a research assistant in Prof. Michael C. Fitzgerald's lab from 2011-present. She is expected to graduate with a Ph.D. in Chemistry in May 2016. While attending Duke University, she was awarded the Kathleen Zielek Fellowship from 2013-2015 and the best talk award at Duke Chemistry Department 2015 Graduate Research Symposium. She has been a member of the American Chemical Society and American Society for Mass Spectrometry from 2012-present.

Publications

1. **Y. Xu**, M. A. G. Wallace, M. C. Fitzgerald, Measurement of Geldanamycin Binding Affinity to Hsp90 in Cell Lysates using the SPROX Technique. Manuscript in preparation (2016).
2. **Y. Xu**, C. Qiu, T. M. T. Hall, M. C. Fitzgerald, Identification of Puf6 Protein-Protein Interactions in a Yeast Cell Lysate using the SPROX Technique. Manuscript in preparation (2016).
3. **Y. Xu**, E. C. Strickland, M. C. Fitzgerald, Thermodynamic Analysis of Protein Folding and Stability using a Tryptophan Modification Protocol. Anal Chem **86**, 7041-7048 (2014).
4. H. Wu, L. Zhang, **Y. Xu**, Z. Ma, Z. Shen, X. Fan, Q. Zhou, Amphiphilic Mesogen-Jacketed Liquid Crystalline Polymers: Design, Synthesis and Self-Assembly Behaviors. J Polym Sci A Polym Chem **50**, 1792-1800 (2012).