

Development and Application of
Large-Scale Protein Folding Stability Analysis in
Drug Target Identification and Disease Biomarker Discovery

by

Renze Ma

Department of Chemistry
Duke University

Date: _____

Approved:

Michael C. Fitzgerald, Advisor

Dewey G. McCafferty

Terrence G. Oas

Kevin D. Welsher

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

2020

ABSTRACT

Development and Application of
Large-Scale Protein Folding Stability Analysis in
Drug Target Identification and Disease Biomarker Discovery

by

Renze Ma

Department of Chemistry
Duke University

Date: _____

Approved:

Michael C. Fitzgerald, Advisor

Dewey G. McCafferty

Terrence G. Oas

Kevin D. Welsher

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

2020

Copyright by
Renze Ma
2020

Abstract

In the past decade, several mass spectrometry-based proteomic techniques have been developed for the large-scale analysis of protein folding stabilities. The main focus of this dissertation is to develop and apply these large-scale protein folding stability approaches in drug target identification and disease biomarker discovery. One goal of this work is to develop a novel chemo-selection strategy to improve the bottom-up proteomics readout in proteome-wide limited proteolysis experiments. Another goal of this work is to apply these methods to the target identification of two drugs with known mode of action, and to the biomarker discovery of Parkinson's disease.

Described in the first part of the dissertation is the development of a chemo-selective enrichment strategy to isolate the semi-tryptic peptides generated in mass spectrometry-based applications of limited proteolysis methods. The method is termed Semi-Tryptic Peptide Enrichment Strategy for Proteolysis Procedures (STEPP). The STEPP-PP workflow was evaluated in two proof-of-principle drug target identification experiments involving two well-studied drugs, cyclosporin A and geldanamycin. The STEPP-LiP workflow was evaluated in one proof-of-principle experiment on identification of protein conformational changes between a breast cancer cell line, MCF-7, and a normal cell line, MCF-10A. The STEPP protocol increased the number of semi-tryptic peptides detected in the LiP and PP experiments by 5- to 10-fold. The STEPP protocol not only increases the proteomic coverage, but also increases the amount of structural information that can be gleaned from limited proteolysis experiments. Moreover, the protocol also enables the quantitative determination of ligand binding

affinities. When coupled to a one-pot data acquisition strategy, the one-pot STEPP-PP technique was found to have a very low false positive rate (i.e., 0.09%) in a proof-of-principle drug target identification experiments involving cyclosporin A and a yeast lysate.

The second part of this dissertation describes the application of protein folding stability approaches to the identification of protein targets of subglutinin A (a natural immunosuppressant) and manassantin A (a natural product with anti-cancer activity).

In the subglutinin A study, a combination of SPROX, TPP, CPP and STEPP-PP strategies was used to identify two consistent protein hits, deoxycytidine kinase (DCK) and exportin-2 (XPO2), from more than 2000 assayed proteins in a 2B4T cell lysate. The binding of DCK with subglutinin A was validated using a targeted gel-based pulse proteolysis experiment. A set of chemical biology experiments were performed to uncover the relation of this interaction with subglutinin A's mode of action. It was shown that neither of the kinase activity, expression level or phosphorylation modification level of DCK was altered by subglutinin A. However, the nuclear transportation of DCK was blocked by subglutinin A. This reduction of DCK level in the cell nucleus possibly leads to the observed reduction of nuclear dCMP pool and the halted proliferation of subglutinin A treated T cells.

In the manassantin A study, a combination of STEPP-LiP, STEPP-PP, one-pot STEPP-PP, one-pot SPROX and one-pot TPP strategies were performed to identify the protein target of the drug in a hypoxia-treated HEK293T cell lysate. These experiments

assayed over 4000 proteins and found 4 protein hits for further validation of their interaction with manassantin A.

The third part of describes the utilization of the SPROX method to characterize the progression of PD in a mouse model of the disease in which the human α -synuclein protein with an A53T mutation was overexpressed. The thermodynamic stabilities of proteins in brain tissue cell lysates from Hu α -Syn(A53T) transgenic mice were profiled at three time points including at 1 Month (n=9), at 6 Months (n=7), and at the time (between 9 and 16 Months) a mouse became Symptomatic (n=8). The thermodynamic stability profiles generated here on over 300 proteins were compared to the thermodynamic stability profiles generated on the same proteins from similarly aged wild-type mice using a two-way ANOVA analysis. A group of 22 proteins were identified with age-related protein stability changes, and a group of 11 proteins were found to be differential stabilized in the Hu α -Syn(A53T) transgenic mouse model. The proteins differentially stabilized in the disease mouse model could potentially be used as Parkinson's disease biomarkers upon further validation.

Dedication

I would like to dedicate this dissertation to my selfless mother, Ping Liu. Thank you for raising me, supporting me and trusting me through all my life's highs and lows. I would never be the person I am today without your standing by my side.

Contents

Abstract	iv
Dedication	vii
Contents.....	viii
List of Tables	xiii
List of Figures	xiv
List of abbreviations	xvi
Acknowledgments	xviii
1. Introduction	1
1.1 Protein Thermodynamic Folding and stability.....	1
1.1.1 Overview	1
1.1.2 Protein Folding Stability Measurements	2
1.1.2.1 Overview of Protein Folding Stability Measurements	2
1.1.2.2 Mass Spectrometry Techniques Used in Protein Folding Stability Measurements	3
1.1.2.3 Stability of Proteins from Rate of Oxidation (SPROX).....	5
1.1.2.4 Thermal Proteome Profiling (TPP)	7
1.1.2.5 Chemical Denaturation and Protein Precipitation (CPP).....	8
1.1.2.6 Limited Proteolysis Methods.....	9
1.1.2.7 One-Pot Strategy for Large-scale Protein Stability Measurements	12
1.2 Drug Target Identification	13
1.2.1 Overview of Drug Discovery	13
1.2.2 Target Identification Approach.....	14
1.2.2.1 Overview of Drug Target Identification Approach	14

1.2.2.2 Affinity Chromatography	16
1.2.2.3 Expression-cloning-based Approaches.....	17
1.2.2.4 Protein Stability Approaches.....	17
1.3 Parkinson's Disease	18
1.3.1 Overview of Parkinson's Disease	18
1.3.2 Parkinson's Disease Biomarker Discovery	20
1.4 Dissertation Focus.....	21
2. Development of a Chemo-Selection Strategy for Limited Proteolysis Experiments on the Proteomic Scale	23
2.1 Introduction	23
2.2 Experimental Section.....	24
2.2.1 Materials.....	24
2.2.2 Cell culture and lysis	25
2.2.3 STEPP-PP analysis	26
2.2.4 STEPP-LiP analysis	28
2.2.5 One-Pot STEPP-PP Analysis.....	29
2.2.6 Quantitative LC-MS/MS Analysis	30
2.2.7 STEPP-PP Quantitative Proteomic Data Analysis.....	32
2.2.8 STEPP-LiP Quantitative Proteomic Data Analysis.	33
2.2.9 One-Pot STEPP-PP Quantitative Proteomic Data Analysis.....	35
2.2.10 K_d Value Determination in STEPP-PP experiment.....	35
2.3 Results and Discussion.....	39
2.3.1 STEPP Workflow.....	39
2.3.2 STEPP-PP Analysis of CsA and Geldanamycin Binding to Proteins in a Yeast Cell Lysate.....	41

2.3.3 STEPP-PP Generates Domain Specific Ligand-Binding Information.....	46
2.3.4 Dissociation Constant Determination in STEPP-PP.....	47
2.3.5 STEPP-LiP Study of Breast Cancer-Related Protein Conformational Changes	48
2.3.6 One-Pot STEPP-PP Analysis of CsA Binding to Proteins in a Yeast Cell Lysate	53
3. Investigation of Subglutinin A's Mode of Action.....	56
3.1 Introduction.....	56
3.2 Experimental Section.....	58
3.2.1 Materials.....	58
3.2.2 Cell Culture and Lysis.....	59
3.2.3 SPROX analysis.....	60
3.2.4 TPP analysis.....	61
3.2.6 STEPP-PP analysis.....	62
3.2.7 Protein Expression Level and Phospho-proteome Analysis.....	63
3.2.8 Quantitative LC-MS/MS Analysis for Proteomics Samples.....	64
3.2.10 Proteomic Data Analysis.....	65
3.2.11 Gel-Based Pulse Proteolysis on DCK – Sub A Binding.....	69
3.2.12 DCK Protein Activity Assay.....	70
3.2.13 Gel-Based DCK Localization Analysis.....	72
3.2.14 LC-MS/MS dCMP Concentration Assay.....	72
3.3 Results and Discussion.....	74
3.3.1 Identification of Sub A Protein Targets.....	74
3.3.2 Biochemical Validation of Sub A Binding to DCK.....	81
3.3.3 Evaluation of DCK Activity with Sub A Treatment.....	83

3.3.4	Proteome-wide Evaluation of Protein Expression Level and Phosphorylation Level Changes in Sub A-Treated Cells.....	85
3.3.5	Identification of DCK Nuclear transportation Blocked by Sub A Treatment ...	88
3.3.6	Identification of Reduced dCMP Level in Sub A Treated Cell Nuclei	90
3.4	Conclusion	93
4.	Search for Protein Targets of Manassantin A.....	95
4.1	Introduction	95
4.2	Experimental Section.....	97
4.2.1	Materials.....	97
4.2.2	HEK293T Cell Culture and Lysis.....	97
4.2.3	STEPP-LiP Analysis	98
4.2.4	STEPP-PP and one-pot STEPP-PP Analyses	98
4.2.5	one-pot TPP and one-pot SPROX Analyses	99
4.2.6	Quantitative LC-MS/MS Analysis for Proteomics Samples.....	101
4.2.7	Proteomic Data Analysis.....	102
4.3	Results and Discussion.....	103
4.3.1	Identification of Manassantin A Protein Target	103
5.	Global Analysis of Protein Folding Stability Changes During Parkinson’s Disease in a Hua-Syn(A53T) transgenic Mouse Model	109
5.1	Introduction	109
5.2	Experimental Section.....	110
5.2.1	Mouse Euthanasia and Tissue Lysis.....	110
5.2.2	iTRAQ-SPROX Protocol.....	111
5.2.3	Proteomic Sample Preparation.....	112
5.2.4	LC-MS/MS Analyses.....	113

5.2.5 Data Analysis.....	116
5.2.6 Hit Identification.....	117
5.3 Results.....	118
5.3.1 Experimental Design	118
5.3.2 Proteomic Coverage and Hit Selection	119
5.3.3 Biological Variability	126
5.4 Discussion	127
5.4.1 Age-Related Proteins Hits	127
5.4.2 PD- and PD-Age Interaction-Related Protein Hits.....	129
6. Conclusions.....	132
Appendix Determination of Pathogenic Mutations on the Folding Stability of Human Transthyretin	136
Introduction	136
Experimental Section.....	137
HNSB Analysis.....	137
Time Course Oxidation Analysis.....	138
Results and Discussion.....	139
Folding Stability Analysis of Amyloidogenic TTRs.....	139
Analysis of Strand A and D Local Environment	141
Conclusion	142
References	144
Biography.....	159
Publications.....	160

List of Tables

Table 1 Summary of proteomic coverages and numbers of hits observed in the STEPP-PP CsA and geldanamycin ligand binding experiments	43
Table 2 Summary of proteomic coverages and numbers of hits observed in the STEPP-LiP MCF-7 vs MCF-10A comparative analysis	50
Table 3 Summary of Proteomic Coverages and Hits Observed in CsA Ligand Binding Experiments Using the one-pot STEPP-PP approach.....	54
Table 4 Gradient used in LC-MS/MS quantification of dCMP and dTMP	74
Table 5 Summary of proteomic coverages and hits observed in the Sub A binding SPROX, TPP, CPP and STEPP-PP experiments on proteins from 2B4 T cell lysate	77
Table 6 The four overlapped protein hits and their behavior in the techniques applied here in this work	78
Table 7 Summary of proteomic coverages and hits observed in the Man A binding experiments on proteins from HEK293T cell lysate.....	105
Table 8 Protein hits observed in the proteomics experiments performed in this work..	107
Table 9 summaries the LC-MS/MS instrument used for analysis of each mouse brain samples in this work.....	115
Table 10 Summary of the proteomic data obtained in the iTRAQ-SPROX experiments performed on the brain tissue cell lysates from each group in this study	120
Table 11 Summary of the hit proteins and their hit methionine-containing peptide probes	122

List of Figures

Figure 1 Schematic representation of the SPROX, TPP and CPP approaches.....	6
Figure 2 Schematic representation of the PP, LiP and DARTS approaches.	10
Figure 3 Schematic representation of the STEPP protocol workflow developed in this work.	41
Figure 4 Schematic representation of the STEPP-PP workflow used in this work.....	42
Figure 5 Representative STEPP-PP data obtained in the CsA binding study..	44
Figure 6 STEPP-PP data obtained in the geldanamycin binding experiment.	45
Figure 7 Schematic representation of the STEPP-LiP workflow used in this work.....	49
Figure 8 Comparison of the STEPP-LiP and LiP data generated on HSP90AA1 (heat shock protein HSP 90-alpha).	51
Figure 9 Schematic representation of the one-pot STEPP-PP workflow used in this study.	53
Figure 10 Volcano plots of the average log ₂ (fold change) values and p-values generated using a Student's t test to analyze the one-pot STEPP-PP data in this work.....	55
Figure 11 Structure of Sub A.	57
Figure 12 Schematic representation of the workflow of proteomics methods used in this work for identification of Sub A protein target.	76
Figure 13 Denaturation curves of DCK obtained from CPP and TPP experiments.	80
Figure 14 Denaturation curves of XPO2 obtained from SPROX and TPP experiments. ...	81
Figure 15 Pulse proteolysis with immunoblot readout workflow for target validation...	82
Figure 16: Immunoblot readout of pulse proteolysis on Sub A and DCK binding.	82
Figure 17 Purified DCK kinase activity assay result.	84
Figure 18 Radiochemical DCK activity assay workflow.	84
Figure 19 Radiochemical DCK activity assay results.....	85

Figure 20 Workflows of expression proteomics and phosphoproteomics experiments performed in this study.....	86
Figure 21 Volcano plots showing the results of expression proteomics and phosphoproteomics experiments in this work.	87
Figure 22 Workflow of DCK localization study with western blot readout.....	89
Figure 23 Western plot readout of cytosolic and nuclear DCK level in cells treated with and with Sub A for 24 or 48 hours.....	90
Figure 24 Workflow of nuclear dCMP quantification experiments used in this study. ...	91
Figure 25 Results of the LC-MS/MS-based dCMP and dTMP quantification in this study.	92
Figure 26 Structure of manassantin A.....	95
Figure 27 Schematic representation of the workflow of proteomics methods used in this work for identification of manassantin A protein target.....	104
Figure 28 Volcano plots of the average \log_2 (fold change) values and t test p-values in one-pot SPROX, one-pot TPP, one-pot STEPP, and STEPP-LiP data generated in this work.	106
Figure 29 Schematic representation of the experimental workflow utilized in this work. The data of wild-type mice sacrificed at 6 month and 18 month are from a previous study. ⁵³	119
Figure 30 Demonstration of $C_{1/2}$ value determination in this work: the dashed arrow indicates the average $C_{1/2}$ value of the fitted curve.	121
Figure 31 Example plots of $C_{1/2}$ changes in selected proteins with non-hit and age-related hit methionine-containing peptide probes.	124
Figure 32 Example plots of $C_{1/2}$ changes in selected proteins with PD-related and PD-age interaction hit methionine-containing peptide probes.	125
Figure 33 Frequency distributions of standard deviations associated with the $C_{1/2}$ values determined in this work.....	126

List of abbreviations

MS	mass spectrometry
LC-MS/MS	liquid chromatography tandem mass spectrometry system
QQQ	triple quadrupole mass spectrometer
MRM	multiple reaction monitoring
HCD	higher energy collision-induced dissociation
TMT	tandem mass tags
iTRAQ	isobaric tags for relative and absolute quantitation
SILAC	stable isotope labeling with amino acids in cell culture
SPROX	stability of proteins from rates of oxidation
LiP	limited proteolysis
PP	pulse proteolysis
CPP	chemical denaturation and protein precipitation
TPP	thermal protein profiling
STEPP	semi-tryptic peptide enrichment strategy for proteolysis procedures
FASP	filter-aided sample preparation
TEAB	triethylammonium bicarbonate
SDS	sodium dodecyl sulfate
TCEP-HCl	tris(2-carboxyethyl)phosphine hydrochloride
PBS	phosphate-buffered saline
MMTS	methylmethane thiosulfonate
TFA	trifluoroacetic acid

CsA	cyclosporin A
GdmCl	guanidine hydrochloride
TCA	trichloroacetic acid
DCK	deoxycytidine kinase
XPO2	exportin-2
Sub A	subglutinin A
Man A	manassantin A
PD	Parkinson's disease
CSF	Cerebrospinal fluid

Acknowledgments

First, I would like to give my most sincere thanks to my advisor, Professor Michael C. Fitzgerald, for his guidance and motivation throughout my PhD research. He has probably heard this for more than thirty times in his career, but this dissertation would never have been possible without his support.

I also would like to thank my committee members, Professor Dewey G. McCafferty, Professor Terrence G. Oas, and Professor Kevin D. Welsher, for their valuable advice and guidance of my research towards the right direction.

I would like to acknowledge both current and past Fitzgerald lab members. In particular, I am grateful to Dr. Julia H. R. Johnson for her mentoring when I joined the lab, and Dr. Fang Liu, Dr. Lorrain Jin and Dr. Yingrong Xu for their great help in discussing the ideas and experimental approaches with me. I shall ever be grateful to Dr. He Meng for being my mentor in all respects, for helping me through my PhD path and always providing insightful suggestion in my research. I am also thankful to the present lab members Aurora Cabrera, Nancy Corsi, Baiyi Quan, Yun Tang, Morgan Bailey and former lab member Dr. Rylene N. Ogburn for all the enlightening discussions and support. I would like to thank Ryan Fader for his help in experiments and being the best mentee ever.

I would like to express my gratitude to the director of Instrument Operations of Duke University Chemistry Department, Dr. Peter Silinski for always giving the most reliable advice and guidance in instrument operations.

I would like to give special thanks to the Duke University Proteomics and Metabolomics Shared Resource, especially Sarah Hiles and Dr. Will Thompson, for acquiring most of the LC-MS/MS data in this dissertation.

Finally, I would like to express my deepest gratitude to my parents for their love and support throughout my life. Thank you for always believe in me and stand by my side. I would like to thank my grandparents for their love and for their lessons on virtues of life. Also, to my girlfriend, Liqi, thank you for being the peaceful harbor in my life, and thank you for all the hope you ignite and support you offer.

1. Introduction

Portions of this chapter were taken from the introductory materials in references 1-3.¹⁻

3

1.1 Protein Thermodynamic Folding and stability

1.1.1 Overview

Proteins are the major actors in living organism, performing an amazingly diverse set of functions including but not limited to biochemical reaction catalysis, signaling, ligand binding, molecule transportation and structure formation. This kaleidoscope of protein functions is the result of the diversity of protein structures. From the plainness of just one or several unstructured long chains of amino acids, a protein is able to fold into its characteristic three-dimensional structure that take on its predesigned function(s). In physiological condition, this folding of a protein is in equilibrium with the reverse process - the unfolding of itself. This equilibrium of folding and unfolding determines the thermodynamic stability of the folded protein, which can be described by the folding free energy (ΔG_f). Factors that alternates protein functions, such as mutations, modifications, and binding also induce changes of the protein folding stability. Therefore, an observable change in the folding free energy can be an indication of important protein functional changes. Indeed, more and more research have incorporated measurements of protein stabilities to understand important biomolecular questions, such as identification of protein drug target and discovery of disease protein biomarkers. This dissertation is

focusing on the development and application of protein folding stability measurements to these topics.

1.1.2 Protein Folding Stability Measurements

1.1.2.1 Overview of Protein Folding Stability Measurements

Traditionally, measurements of protein folding stability are accomplished using either spectroscopic or calorimetric methods.⁴⁻¹¹ Spectroscopic methods, including fluorescence spectroscopy and circular dichroism (CD) spectroscopy, use thermal or chemical denaturation or changes of solvent pH to drive the protein into its unfolded state, where the intrinsic fluorescence and polarization properties of the protein are lost. By quantification of the change in fluorescence or light polarization of the protein in a series of denaturing conditions (i.e., increasing temperature, chemical denaturant or pH), a protein denaturation curve can be constructed. From this protein denaturation curve, the folding free energy (ΔG_f) can be determined. Calorimetric methods include the differential scanning calorimetry (DSC) and the isothermal calorimetry (ITC). The DSC experiment measures the heat capacity of a protein solution as a function of the temperature. As the protein of interest transit from the folded state into the unfolded state, the increase of heat capacity is monitored and the protein folding enthalpy (ΔH_f) is determined. The ITC experiment is generally not for determination of the protein folding free energy (ΔG_f), but for ligand binding free energy ($\Delta\Delta G_f$). By monitoring the heat generated or absorbed during the addition of ligand into the protein solution, thermodynamic parameters describing the ligand binding (i.e., $\Delta\Delta G_f$, K_d and $\Delta H_{\text{binding}}$) can be determined by ITC.

Unfortunately, all of these traditional techniques require a large amount of purified protein. For example, a typical DSC experiment requires at least 0.1 mg of the purified protein of interest.¹² The need of cumbersome purification procedures for this large amount of protein hinders the application of these traditional methods in studies where high throughput is needed, such as protein drug target identification and disease protein biomarkers discovery.

In the past decade, several mass spectrometry-based protein thermodynamic stability measurements have been developed for probing protein folding stabilities on the proteomics scale. Beginning with similar chemical or thermal denaturation of proteins, these methods incorporate a variety of modification reactions to distinguish the folded and unfolded proteins for latter mass spectrometry (MS) readout. By coupling with MS, the protein folding thermodynamics of thousands of proteins can be determined from a cell lysate containing only tens of micrograms of proteins. The ability of these methods to make high throughput and proteomic-scale measurements make them ideal methods for drug target identification and disease biomarkers discovery. The details of these methods, including SPROX, TPP, CPP and the limited proteolysis methods, are to be introduced in latter sections. Before going into these methods, a brief overview of the mass spectrometry techniques associated with them is provided to illustrate the technical background that enables their proteomic-scale measurements.

1.1.2.2 Mass Spectrometry Techniques Used in Protein Folding Stability Measurements

To achieve high throughput identification and quantification of thousands of proteins in one analysis, a mass spectrometry-based approach called bottom-up

proteomics (also known as shotgun proteomics) is widely used. It involves a trypsin digestion of the proteins in a complex mixture of interest, e.g., a cell lysate, into smaller tryptic peptide segments that can be efficiently separated with liquid chromatography (LC). The LC system feeds the eluted peptides to a tandem mass spectrometry system over time. A general tandem mass spectrometry system (i.e., a LC-MS/MS system) consists of two mass analyzers and a collision cell in between. The eluted peptides from the LC are ionized, selected by the first mass analyzer (MS1) and guided to the collision cell where they are fragmented. These fragments are detected by the second mass analyzer (MS2) and are used to identify (i.e., map the amino acid sequence) and quantify their parent peptide. In bottom-up proteomics, the identified tryptic peptides are further used to identify and quantify their parent proteins that appear in the protein mixture of interest.

To quantify the amounts of peptides in protein folding stabilities measurements, a technique called isobaric labeling is commonly used.¹³ There are several commercially available isobaric labeling reagents including tandem mass tags (TMT) and isobaric tags for relative and absolute quantification (iTRAQ).^{14,15} Both tags can undergo a N-hydroxysuccinimide-facilitated click reaction to covalently label the primary amine groups of peptides.^{16,17} When a labeled peptide goes through the collision cell of the LC-MS/MS system, a reporter ion will be cleaved from the labeled peptide. The signal intensity of this report ion provides the quantification of the labeled protein.

The isobaric mass tags also provide multiplexing of different samples. For example, the TMT isobaric labeling reagent is available in 10-plex where there are 10 different tags with the same structure, same nominal mass (i.e., isobaric) but different

reporter ion masses.¹⁴ After being labeled with 10 different TMT tags separately, ten different samples can be combined together and quantify in a single LC-MS/MS analysis. The identical peptides originally from different samples will still have the same mass after labeling and are able to be selected together by MS1. After collision, the ten different reporter ions are released for quantification of the amount of the labeled peptide in ten different samples.

The bottom-up proteomics and isobaric labeling techniques provide the foundation for the following protein folding stability measurements to generate superior performance over conventional methods.

1.1.2.3 Stability of Proteins from Rate of Oxidation (SPROX)

SPROX is a protein folding stability measurement that involves a H₂O₂ oxidation reaction to mark the unfolded proteins (Figure 1). Similar to conventional methods, the initial step of the SPROX approach involves the distribution of protein samples (e.g., cell lysates with or without the ligand of interest, or organ lysates) into increasing concentrations of chemical denaturants, usually guanidium chloride (GdmCl) or urea. The proteins samples are incubated with the chemical denaturant for sufficient amount of time (usually 1-2 hours) to allow the rebalance of folding/unfolding equilibria. Hydrogen peroxide is then introduced into the system to oxidize the thioester groups of the proteins' methionine residues. The oxidation reaction time and H₂O₂ concentration are controlled to allow the pseudo-first order oxidation reaction to proceed for about 3 half-lives of a solvent-exposed methionine residue under room temperature. Because the hydrophobic methionine residues are generally buried inside the proteins' hydrophobic cores in folded

proteins, only methionine residues of the unfolded proteins are solvent-accessible and oxidized. The oxidation reaction is quenched with excess amount of either free methionine or a reducing reagent, triscarboxyethylphosphine (TCEP). Bottom-up sample preparation procedures and LC-MS/MS analysis are then performed on the samples. Protein folding thermodynamics parameters can be determined from quantification of methionine-containing peptides as a function of chemical denaturant concentrations. The peptide quantification in SPROX are generally facilitated by either isobaric mass tag labeling (i.e., iTRAQ or TMT) during bottom-up sample preparation or stable isotope labeling by amino acid (SILAC) during cell cultures.

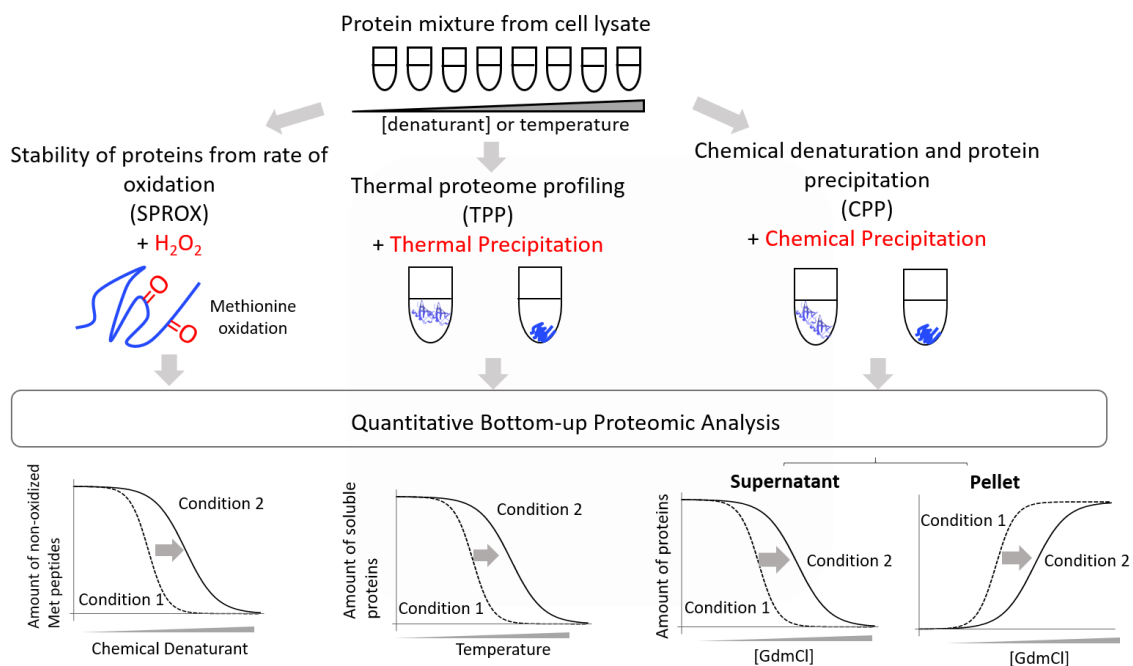


Figure 1 Schematic representation of the SPROX, TPP and CPP approaches.

One major advantage of SPROX is its ability to determine the folding thermodynamics parameters such as ligand binding-induced $\Delta\Delta G_f$ values. This feature is particularly useful in drug target identification where the K_d value of binding can be

determined from $\Delta\Delta G_f$ and compare to IC_{50} . In ligand-binding studies, SPROX is also preferred over thermal denaturation methods (i.e., TPP) in situations where heating of samples should be avoided, such as gas-binding studies. Also, by quantification of the denaturant response on the peptide level, SPROX is able to determine folding thermodynamics for specific protein domains where the peptides are located. However, unlike protein-level quantification where signals from peptides are averaged, this quantification by only one peptide in SPROX can lead to a higher false discovery rate. In order for the folding thermodynamics properties of a protein or protein domain to be determined, the SPROX method requires identification of corresponding methionine-containing peptide(s). Although most of proteins have at least one globally protected methionine residue, only about 20% of peptides in a trypsin-digested protein lysate contain methionine residues, limiting the proteomic coverage of a SPROX experiment. This limitation can be partially remedied with a methionine-containing peptide enrichment step which boost the percentage of methionine-containing peptide to >70%.¹⁸ If SPROX is used in combination with a similar tryptophan-modification strategy, the proteomic coverage can be further increased by 50%.¹⁹ When coupled with a so-called one-pot data acquisition strategy, the proteomic coverage can be two-fold higher, but the ability to generate thermodynamic properties (e.g., K_d value) is sacrificed in one-pot SPROX.³ The details of this one-pot strategy is introduced in details in section 1.1.2.5.

1.1.2.4 Thermal Proteome Profiling (TPP)

The TPP method, resembling DSC, uses a thermal denaturation to profile protein stabilities. The initial step of the TPP protocol involves incubation of the protein samples

at a series of increasing temperatures for 3 minutes. The samples are ultra-centrifugated to remove aggregated proteins from the sample solutions (Figure 1). The samples are then subject to bottom-up samples preparations procedures and isobaric mass tag (i.e., TMT-10plex) labeling before LC-MS/MS quantification. For each protein, a thermal denaturation curve of relative amount of soluble protein as a function of temperature is generated.

The advantages of TPP method are related to its incorporation of thermal denaturation rather than chemical denaturation in other methods. As heat easily penetrate cells membranes while chemical denaturant does not, TPP is the only method able to identify protein folding stability characteristics in intact cells. Another advantage of using thermal denaturation is that it does not perturb protein-protein binding as chemical denaturant does in high concentrations. This feature makes TPP ideal to study protein interaction network. It also enables TPP to identify drug-induced stabilization or destabilization of a whole protein complex. However, as the thermal denaturation in TPP is irreversible, important thermodynamic parameters such as ΔG_f or K_d values cannot be obtained.

1.1.2.5 Chemical Denaturation and Protein Precipitation (CPP)

The CPP method is a hybridization of SPROX-like chemical denaturation and TPP-like precipitation for folded and unfolded protein separation (Figure 1). The initial step of a CPP experiment involves incubation of protein samples at increasing concentration of GdmCl. After the folding and unfolding of proteins in the denaturation-contracting buffers have been equilibrated, the protein samples are abruptly diluted 25-folds with

water. The abrupt dilution deprives GdmCl molecules from stabilizing unfolded proteins and cause unfolded proteins to aggregate. The aggregated proteins are pelleted by centrifugation and separated from soluble proteins in the supernatant. Both protein pellets and supernatants are collected in CPP and subject to bottom-up samples preparation procedures, isobaric mass tag labeling and LC-MS/MS quantification. Soluble proteins in the supernatant fraction decreases as a function of the GdmCl concentration and the aggregated proteins in the pellet fraction show a reverse trend. The two denaturation curves of the same protein are expected to generate similar thermodynamics parameters and can be used to confirm each other and reduce false positive rates.

The major advantage of CPP is its ability to provide protein folding thermodynamic parameters as SPROX, while generating a 50% higher proteomic coverage.²⁰ If both supernatant and pellet fractions of a CPP experiment is analyzed, the false positive can be effectively lower to 0.²⁰ However, due to the need of efficient protein precipitation, CPP can only use GdmCl but not urea denaturation. As GdmCl disrupts ionic interactions in ligand-protein binding, it limits the use of CPP to target identification studies where the ionic interactions are not involved.

1.1.2.6 Limited Proteolysis Methods

Limited proteolysis methods include Limited Proteolysis (LiP), Drug Affinity Responsive Target Stability (DARTS), and Pulse Proteolysis (PP) techniques. Unlike SPROX, TPP and CPP experiments where denaturation conditions are incorporated to probe the global protein folding stabilities, LiP and DARTS experiments focus on local folding in buffer conditions under which proteins are generally folded into their native

three-dimensional structure (Figure 2). The LiP approach involves the cleavage of solvent exposed regions of protein structure by a nonspecific protease, proteinase K. Analogous to SPROX, TPP and CPP, the PP approach is measurement of global protein folding stabilities (Figure 2). The pulse proteolysis (PP) approach, which involves measuring the denaturant dependence of a limited proteolysis reaction using a nonspecific protease (i.e., thermolysin), has proven useful for the quantitative analysis of protein folding stabilities and protein-ligand binding affinities.²¹⁻²⁴

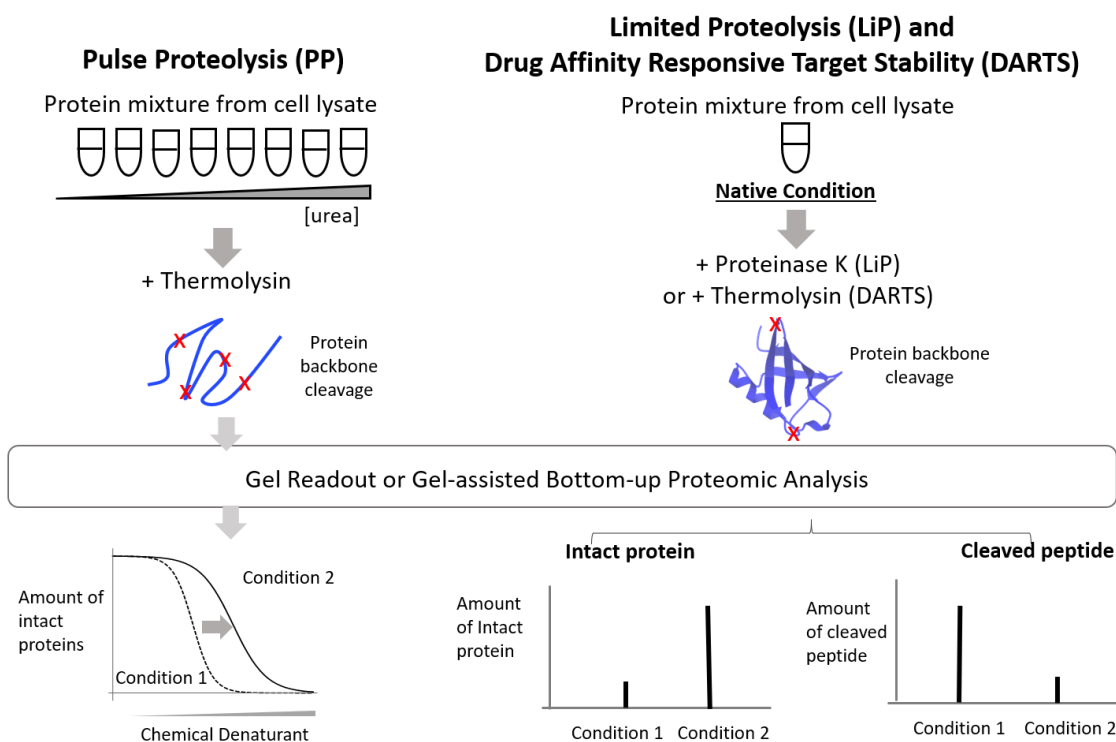


Figure 2 Schematic representation of the PP, LiP and DARTS approaches.

Comparing to other methods mentioned in the previous sections, limited proteolysis methods have unique advantages. The LiP and DARTS experiments can identify protein conformational prosperities too subtle to be measured on the global folding stability level. Subtle local conformational changes can be captured by LiP and

DARTS but may not show a strong denaturant dependence to be captured in SPROX, TPP or CPP experiments. Although LiP and DARTS sacrifice the power to generate folding thermodynamic parameters (e.g., ΔG_f and K_d values), they have been proven useful for identifying protein conformational changes related to environmental changes, disease states, post-translational modifications as well as small molecule binding.²⁵⁻³⁰ While unable to prove local folding properties, the PP approach can generate folding thermodynamic parameters specifically to the protein domain level, which is not available in TPP and CPP.

While limited proteolysis methods for studying the conformational properties of purified proteins and protein-ligand complexes are well established, their recent extension to the proteomic-scale LC-MS/MS-based experiments has been challenging. Before this work, proteome-wide applications of these techniques have typically relied on either gel- or solution-based bottom-up proteomics strategies.²²⁻³⁰ Unfortunately, the relatively low resolving power of gel electrophoresis methods can complicate applications of the LiP, DARTS and PP techniques to complex protein mixtures. Higher resolution LC-MS/MS strategies using bottom-up proteomics methods have been utilized to read out the results of proteome-wide limited proteolysis experiments.^{24,25,27-30} However, the large number of tryptic peptides generated in the bottom-up proteomics readout can make identification of the cleavage sites in the limited proteolysis experiment especially challenging, due to the relatively low abundance of the semi-tryptic peptides (i.e., peptides that have both a terminal resulting from trypsin cleavage and a terminal resulting from non-trypsin cleavage, such as thermolysin or proteinase K cleavage) that

result from such cleavages. The semi-tryptic peptides generated in a limited proteolysis experiment are those that have a non-tryptic cleavage site at one end as a result of the limited proteolysis reaction and a tryptic cleavage site at the other end as a result of the trypsin digestion step used to prepare the sample for bottom-up proteomics analysis.

Recently, attempts have been made to improve the bottom-up proteomics readout in proteome-wide limited proteolysis experiments by separating the intact proteins from the smaller peptides generated in the limited proteolysis reaction. For example, the use of high molecular weight (e.g., 100,000 Da) cut-off filters has been reported to isolate the intact proteins prior to their trypsin digestion for bottom-up proteomics analysis.³¹ A protein precipitation reaction involving trichloroacetic acid (TCA) has also been used to remove the undigested, intact proteins from the limited proteolysis reaction so that the remaining peptide fragments could be digested with trypsin and subject to a quantitative bottom-up proteomics analysis.³² A drawback to the proteomics readouts employed in the filtration and precipitation approaches is that they do not provide any structural information about the site(s) of conformational changes, as they are both protein-centered readouts.

Therefore, a better strategy for coupling limited proteolysis methods and large - scale LC-MS/MS-based proteomics analysis is needed. To fulfill this gap, a new strategy is developed as part of the work of this dissertation and described in Chapter 2.

1.1.2.7 One-Pot Strategy for Large-scale Protein Stability Measurements

Recently, a one-pot data acquisition and analysis strategy has been developed for coupling to the large-scale protein folding stability measurements including TPP and

PP.^{33,34} Instead of quantification of the modification reactions at each denaturation point, the one-pot strategy only reports the average of the modification level across the series of denaturation conditions used in the experiment. Although this simplification sacrifices the power to generate protein folding thermodynamic parameters and ligand binding affinities, it reduces the number of isobaric mass tags needed for a single analysis. With TMT-10plex labeling, five replicate experiments can be analyzed in a single LC-MS/MS quantification. The statistical analysis is also streamlined with the one-pot strategy. As data from five biological replicates are available all at once, more statistically significant results (i.e., lower false positive and false negative rates) can be generated for hit selection.

1.2 Drug Target Identification

1.2.1 Overview of Drug Discovery

Modern drug discovery is an expensive and slow process with high failure rate. It takes an average of 2.6 billion US dollars and more than 10 years to launch a new drug into market.³⁵ Generally, the process of drug discovery takes place with either of two strategies: target-based drug discovery (TDD) or phenotypic drug discovery (PDD).

The starting point of the TDD strategy is the identification of a potential drug target, typically a protein, that appears to play a key role in disease pathogenesis. Then an activity assay is designed to monitor the response of this target protein to a library of small molecules. The so-called “lead” molecules that induce the desired response of the target protein are further studied in cell and animal models to determine their impact of key aspects such as efficacies, pharmacokinetics, tolerability, and safety issues.³⁶ During this

process, the molecular structures of lead molecules are constantly being optimized for these aspects to achieve a higher probability in clinical success.

Unlike the TDD approach, the PDD approach does not prespecifies a key protein of interest. In the PDD approach, a library of small molecules is screened for their alternation of disease phenotypes in a cell or animal model. After a set of lead molecules are generated from the process, their protein target must be identified. The identification of their protein target(s), also known as target deconvolution, is essential for lead optimization and prediction of potential side effects.^{37,38} If the optimization is successful, candidate molecules optimized based on these lead molecules will move forward to clinical trials.

In both PDD and TDD approaches, large-scale characterization of protein-drug interaction is essential. As mentioned above, in PDD approach, the target protein must be identified before fully understand the molecular mode of actions that underlines the efficacy of drug candidates.^{37,38} In TDD approach, identification of off-target proteins is crucial for determination of drug specificity and prediction of potential side effects.³⁹ Indeed, in the last decade, major advances have been made in the innovation of new methods to fulfill the ever-growing needs of large-scale drug target identification methods. Several selected methods are reviewed in the following sections.

1.2.2 Target Identification Approach

1.2.2.1 Overview of Drug Target Identification Approach

The identification of drug binding target(s) in the human proteome with 10's of thousands of proteins can be as challenging as finding a needle in a haystack.⁴⁰

Fortunately, major advances in computation power and mass spectrometry technology has been made in the past several decades for a convenient infrastructure to develop new methods towards achieving the goal. However, even with the development of various methods in this purpose, target identification can still be a difficult task. Part of this difficulty stems from the ubiquitous errors come along with hypothesis testing – it is very difficult to eliminate false positives and false negatives in a target identification approach. The high the error rates, the less likely to successfully identify the protein targets, and the more complex and time-consuming it will be for following biochemical target validations. Another part of this difficulty in drug target identification stems from the limitations of these methods inherited from their experimental designs. A detailed review of the limitations of MS-based methods are described in section 1.1.2. The need for a method to overcome some limitations is a major driving force of new method development in drug target identification. For example, the STEPP strategy to be described in chapter 2 is developed to meet the need for a large-scale quantification method that retains the advantages of the limited proteolysis experiments.

One possible solution for the struggle in target identification is just multiple methods together. A recent proof-of-principle study of several state-of-the-art protein stability-based methods has shown that even for the characterization of the tight binding between cyclosporine A (CsA) and the cyclophilins proteins (K_d of CsA to the cyclophilins are around 40nM, 100nM and 10nM for cyclophilin A, C and D, respectively)^{41,42} in a yeast lysate, up to two-third of the hit proteins in a single method can be false positives.³ However, if more than one methods are applied, the false positive rate can be reduce to 0.

This suggests that for a drug with unknown mode of action, it is probably a better way to apply more methods rather than trying to dig the true positives from the false positives in one method, while ignoring the possibility of false negatives. However, the choice of methods should not be random. It should be considered accordingly to the situation and the optimal choice of method can be different from case to case, although some methods are generally preferred over others. The limitations of methods should be taken into account and the chemical properties of the drug should also be considered. For example, TPP may not be a good choice when the drug is heat sensitive.

Generally, these target identification strategies include affinity chromatography, expression-cloning-based approaches, and protein stability approaches.

1.2.2.2 Affinity Chromatography

Affinity chromatography is the conventional technique to identify drug binding target. In an affinity chromatography experiment, the small molecule drug is first chemically linked to a solid support. Then the immobilized drug is incubated with a protein solution and any non-binding proteins will be subsequently be removed in extensive washing steps. After non-binding proteins are removed, the target proteins that bind to the solid phase reversibly are eluted from the system and subject to immunodetection or bottom-up LC-MS/MS characterization.

Although the idea is straightforward, there are major limitations of affinity chromatography in practice. The link that is covalently attached to the drug molecule can disrupt the interaction with its protein target by steric hinderance.⁴³ The introduction of

hydrophobic environment into the system also increases the chance of nonspecific binding and elevate false positive rate.³⁸

1.2.2.3 Expression-cloning-based Approaches

Expression-cloning-based approaches include three hybrid systems, phage display and mRNA display. All of these methods involve the testing of binding between the small molecule drug and a library of recombinant proteins.³⁷ In three hybrid systems, the recombinant proteins in the library are each fused to a DNA transcription activation domain. When binding to the drug of interest, which is attached to a DNA, the DNA transcription activation domain that is fused to the target protein will start transcription of the DNA. Thus, the target protein is indicated from monitoring the expected reporter gene. In phage display and mRNA display, the recombinant proteins are fused to surface of a phage or a mRNA (respectively) that encode transcription information of the same protein. Proteins that bind to the covalently immobilized small molecule drug are collected and amplified by replication of the phage or mRNA.

The major limitation of these expression-cloning-based method is the requirement of a sophisticated recombinant protein library. These methods will not be productive when the protein targets are not present in the library. Like the affinity chromatography, a covalent labeling of the drug molecule could disrupt the protein-drug interaction. Also, the lack of important PTMs on the recombinant proteins could affect the binding.

1.2.2.4 Protein Stability Approaches

In the past decades, several protein folding stability-based techniques have been developed and applied to drug target identification. These strategies involve large-scale

identification of either global or local folding stability changes of target proteins as a results of drug binding. Principles of these methods, including SPROX, TPP, CPP, LiP, DARTS and PP, are reviewed in detail in section 1.1 of this dissertation.

Comparing to other methods of drug target identification, these stability-based methods have significant advantages. Unlike other methods, protein stability approaches do not require any labeling of either the drug or the proteins of interest. These experiments can be performed with proteins extracts or even in intact cells that maximally preserve the protein-drug binding environment as in vivo. Indeed, these advantages have contributed to their gain of popularity in the past decade. However, limitations of each of these methods, which are reviewed in detail in section 1.1, should not be overlooked when dealing with a drug target identification study.

Due to the diversity of mechanisms used for identification of protein stability changes in these methods, a combination of several of these methods together in a study may remedy the limitation of each single method and increase the chance of successful discovery.³ In light of this, in chapter 3 and 4, several methods of this categories are used in a combined effort to identify protein targets of two drugs with unknown mode of action.

1.3 Parkinson's Disease

1.3.1 Overview of Parkinson's Disease

The presentation of Parkinson's Disease (PD) has been phenotypically described for centuries. The disease is known by its characteristic disorder of movement including bradykinesia, tremors and muscle rigidity.⁴⁴ A collection of other symptoms, such as

cognitive impairment, depression, dysautonomia, sleep disorders, and hyposmia are also found in PD patients.⁴⁵ Parkinson's disease (PD) has become the second most common neurodegenerative disease following Alzheimer's disease (AD). In the year 2016, there were about 6.2 million PD patients worldwide and about 200,000 died from PD.⁴⁶

The cause of PD is still unknown. Age has been identified as the most important risk factor, with most of the cases occur in elder population. Environmental factors such as pesticides may play a role in development of PD as well. Study on familial PD also revealed that point mutations of several protein, such as α -synuclein (SNCA), parkin (PARK2), PTEN-induced kinase 1 (PINK1), Leucine-rich repeat kinase 2 (LRRK2) are linked to PD.⁴⁷

At the cell level, PD presents progressive loss of neuron cells, particularly dopaminergic neurons, in the central nervous system. The degeneration of neurons cells in PD is accompanied by a deposit of Lewy bodies inside these cells. The Lewy body is an abnormal aggregation of proteins, primarily α -synuclein. The relation of α -synuclein and PD is a subject of intensive research during the last two decades.^{48,49} However, the exact role of α -synuclein in PD is still unknown. Nevertheless, when mice are genetically modified to express human α -synuclein mutants that are found in familial PD cases, similar phenotypes are observed. For example, in the Hu α -Syn(A53T) transgenic mice model, intensive formation of Lewy bodies and onset of motor disorders are observed as earlier as 8 months of age. Studies on these mouse models can provide important information on the role of α -synuclein and the biomolecular basis of PD.

Apart from our limited knowledge on PD development, the diagnosis of PD is also challenging. Definitive diagnosis of PD can only be made postmortem. The premortem diagnosis of PD is not only prone to error (i.e., ~20% false discovery rate)⁵⁰, but also usually too late. With current diagnostic tools which is based on the assessment of motor functions, up to 70% of final neuron degradation has already materialized at the time of diagnosis.⁵¹ Unfortunately, the neuron cell damage in PD is irreversible. Therefore, better diagnosis tools are needed for early-stage PD patients. The earlier the confirmation of PD, the earlier medical intervention can be performed to delay the process of neurodegeneration.

1.3.2 Parkinson's Disease Biomarker Discovery

The identification of PD biomarkers could potentially facilitate the early diagnosis of the disease. In the past twenty years, many new analytical tools have been utilized in PD biomarker discovery.⁵² Among them, the application of -omics methods (i.e., transcriptomics, proteomics and metabolomics) are considered very promising. They have the capability of high-throughput screening of thousands of biomolecules in one analysis. However, the current -omics studies have not been as useful as expected at identifying PD biomarkers for early stage diagnosis. One possible reason for this could be the indirect link between the level of a biomolecule and the disease biological functional change; that is, the disease-related changes in biological function may not always be reflected in the changes of the abundance of biomolecules, either proteins, mRNAs or metabolite, especially at the early stage.

Rather than screening for biomolecule abundance changes, focusing on properties directly related to functional changes could be a better approach. Factors that alternate protein functions, such as point mutations, modifications, and differential transcription also induce changes of the protein thermodynamic stability. Therefore, proteome-wide analysis of protein folding stability changes in PD could generate useful information for understanding and biomolecular basis of disease and identify biomarker for disease diagnosis. Indeed, this approach has been used in several disease state model studies and uncovered useful biomarkers of breast cancers and aging.^{27,53-55} The successful applications of large-scale protein folding stability measurements in disease studies suggest PD biomarker discovery could also benefit from the application of these methods.

1.4 Dissertation Focus

This dissertation focuses on the development and application large-scale protein folding stability measurements to drug target identification and disease biomarker discovery.

In the method development part of this work, a novel semi-tryptic peptide enrichment strategy for proteolysis procedures (STEPP) is described in chapter 2. The novel strategy not only increases the proteomic coverage, but also increases the amount of structural information that can be gleaned from limited proteolysis experiments. Moreover, the strategy also enables the quantitative determination of ligand binding affinities. The efficacy of STEPP-PP strategy was evaluated with two proof-of-principle experiment on drug target identification using two drugs, cyclosporine A and geldanamycin, in a yeast lysate. The efficacy of the STEPP-LiP strategy was evaluated

with one proof-of-principle experiment on identification of protein conformational changes between a breast cancer cell line, MCF-7, and a normal cell line, MCF-10A. The advantages of coupling the STEPP-PP technique to a one-pot data acquisition strategy was also evaluated with a proof-of-principle experiment on drug target identification using cyclosporine A in a yeast lysate.

In the application part of this work, large-scale protein folding stability measurements are applied to drug target identification and disease biomarker discovery. In chapter 3 and 4, a combination of several of these measurements are performed to identify the target protein of small molecule drugs subglutinol A and manassantin A. In the subglutinol A target identification study, deoxycytidine kinase was identified as a binding protein in large-scale methods and the protein-drug interaction was confirmed by targeted analysis. Further biochemical experiments were conducted on DCK to uncover the exact mode of action of subglutinol A. Chapter 5 described the application of one large-scale protein folding stability measurement, iTRAQ-SPROX, to identify protein stabilities changes in a Parkinson's disease iTRAQ-SPROX mouse model to discover potential biomarkers of the disease.

2. Development of a Chemo-Selection Strategy for Limited Proteolysis Experiments on the Proteomic Scale

The work described in this chapter comes largely from the research papers titled “Chemo-Selection Strategy for Limited Proteolysis Experiments on the Proteomic Scale” that was published in 2018 in the *Analytical Chemistry* (vol. 90(23), p. 14039-14047), and “Comparative Analysis of Mass-Spectrometry-Based Proteomic Methods for Protein Target Discovery Using a one-pot Approach” that was published in 2019 in *the Journal of the American Society for Mass Spectrometry* (vol 31(2), p. 217-226).^{2,3}

2.1 Introduction

Described here is a general strategy termed, semi-tryptic peptide enrichment for proteolysis procedures (STEPP), to facilitate detection of the limited proteolysis cleavage sites generated in proteome-wide limited proteolysis experiments. The semi-tryptic peptide-centered proteomics readout in the STEPP protocol significantly enhances the information content of limited proteolysis experiments by facilitating the acquisition of more detailed structural information about the site(s) of conformational changes observed in the limited proteolysis experiment. The results of the proof-of-principle studies performed here demonstrate the ability of the STEPP protocol to generate domain- and amino acid-specific structural information about the conformational properties of proteins using the PP and LiP techniques, respectively. In combination with PP the STEPP protocol can also be used to evaluate the dissociation constants, K_d values, of protein-ligand complexes. As part of this work the STEPP technique is interfaced with two different proteolysis methods including the Pulse Proteolysis (PP) and Limited

Proteolysis (LiP) methods. The STEPP-PP workflow is evaluated in two proof-of-principle experiments involving the proteins in a yeast cell lysate and two well-studied drugs, cyclosporin A and geldanamycin. The STEPP-LiP workflow is evaluated in a proof-of-principle experiment involving the proteins in two cell culture models of human breast cancer, MCF-7 and MCF-10A cell lines.

Recently, a one-pot data acquisition and analysis strategy has been developed for coupling to the protein folding stability measurements including PP and TPP.^{33,34} These one-pot workflows also have the potential to improve the STEPP-PP strategy by facilitating the acquisition of data from more biological replicates and ultimately increasing the statistical significance of the results. The one-pot STEPP-PP workflow is evaluated in a proof-of-principle experiment involving the proteins in a yeast cell lysate and cyclosporin A.

2.2 Experimental Section

2.2.1 Materials

The following materials were from Sigma-Aldrich (St. Louis, MO): guanidine hydrochloride (GdmCl), S-methylmethanesulfonate (MMTS), dimethyl sulfoxide (DMSO), urea, centrifugal filter units (Amicon Ultra-0.5, 10k), tris(hydroxymethyl)aminomethane hydrochloride (Tris•HCl), thermolysin from *geobacillus stearothermophilus*, proteinase K from *tritirachium album*, phenylmethylsulfonyl fluoride (PMSF), trifluoroacetic acid (TFA), triethylammonium bicarbonate buffer (TEAB, 1 M, pH 8.5), methacrylic acid N-hydroxysuccinimide(NHS) ester, cytochrome c from equine heart. The following materials were from Thermo-

Scientific (Waltham, MA): acetonitrile (ACN, LC-MS grade), 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF), Bestatin, E-64, Leupeptin, Pepstatin A, EDTA solution (Corning, 0.5M, pH 8.0), TMT10plex isobaric label reagent set, N-hydroxysuccinimide(NHS)-activated agarose dry resin (Pierce), trypsin protease MS grade (Pierce). Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) was from Santa Cruz Biotechnology (Dallas, TX). Phosphate-buffered saline (PBS, pH 7.4, 1X) was from Gibco (Gaithersburg, MD). Geldanamycin (≥ 98 wt%) was from Chem-Impex International, Inc. (Wood Dale, IL). Cyclosporin A (CsA) was from LKT Laboratories, Inc. MacroSpin columns (silica C4 and C18) were from Nest Group (Southborough, MA).

2.2.2 Cell culture and lysis

Yeast strain YDR155C was obtained from Open Biosystems (Lafayette, CO) and cultured in YEPD medium according to standard protocols. Briefly, The yeast GAL1 overexpression strain YDR155C from Y258 host library was purchased from Open Biosystems. A yeast colony was incubated in 50 mL YEPD medium at 30 °C overnight to reach an OD₆₀₀ of ~1.6. A 20 mL portion of the overnight culture was inoculated with the YEPD medium (1 L) to give an OD₆₀₀ of ~0.3. The inoculated medium was incubated at 30 °C until the OD₆₀₀ of the solution was between 1.2-2.0. Yeast pellets were generated by centrifuging 250 mL fractions of the final YEPD medium. The pellets were stored at -20 °C.

The MCF-7 and MCF-10A cell lines were purchased from American Type Culture Collection (ATCC) (Manassas, VA) and cultured according to ATCC guidelines. The yeast cells in the STEPP-PP experiments were lysed in PBS in the presence of the following

protease inhibitors: 1 mM AEBSF, 20 μ M leupeptin, 10 μ M pepstatin A, 500 μ M bestatin and 15 μ M E-64. The human cells in the STEPP-LiP experiments were lysed in PBS. Cell lysis was accomplished using 0.5 mm glass beads (in the case of yeast) or 1mm zirconia/silica beads (in the case of the human cell lines) and mechanical disruption at 4 °C with 25 s of disruption and 1 min intervals on ice for 15 cycles. The lysates were centrifuged at 14,000 x g for 10 min at 4 °C. The supernatants were separated and used for subsequent analyses. The total concentration of protein in each lysate was determined using a Bradford assay and adjusted to 5.0 mg/ml. For ligand-binding experiments using PP, the yeast lysate was divided into two equal aliquots. One aliquot was spiked with a solution of the test ligand (CsA or geldanamycin) prepared in DMSO and the other one was spiked with DMSO as vehicle. The concentration of drug in each cell lysate aliquot was 500 μ M and 50 μ M in the CsA and geldanamycin binding experiments, respectively. The final concentration of DMSO in each cell lysate aliquot was 5% v/v in all experiments. The solutions were equilibrated for either 1 (CsA) or 24 hours (geldanamycin).

2.2.3 STEPP-PP analysis

In both CsA and geldanamycin-binding experiments, 10 μ L aliquots of the (+) and (-) ligand-containing lysate samples were distributed into a series of 30 μ L urea-containing buffers (PBS buffer, pH 7.4) where the final urea concentration ranged from 0 to 6.8 M. The final concentration of ligand was 125 μ M in the CsA-binding experiment, 12.5 μ M in the geldanamycin binding experiment. The samples in the urea-containing buffers were equilibrated for ~2 h at room temperature. The proteolysis reaction was initiated by adding thermolysin to the protein and protein–ligand samples in each denaturant-

containing buffer. The thermolysin to total protein ratio was approximately 1:10 (w/w). The thermolysin proteolysis reactions were allowed to proceed for 1 min at room temperature before each reaction was quenched with the addition of 60 μ L urea/EDTA solution (~0.2 M EDTA, 9 M urea, pH 8.0). These thermolysin proteolysis reaction conditions were similar to those previously described.^{23,24}

The proteolyzed samples were subject to the STEPP protocol developed here. As part of the STEPP protocol, the proteolyzed samples were reacted with 1.5 mM TCEP for 1 h at 30 °C and then with 2.5 mM MMTS for 10 min at room temperature. Subsequently, the thermolysin digested samples were labeled with TMT reagents by adding 41 μ L of ACN containing 0.5 unit of each TMT reagent and incubating for 1.5 hours at room temperature. In the CsA-binding (+) ligand samples, the urea concentrations, from low to high, were labeled with TMT tags 130N, 130C, 126, 127N, 127C, 128N, 128C, 129N, 129C, 131. The (-) ligand samples were labeled in the reverse manner (i.e., the urea concentrations, from high to low, were labeled with TMT tags 130N, 130C, 126, 127N, 127C, 128N, 128C, 129N, 129C, 131). In the geldanamycin-binding experiment (+) ligand samples, the urea-concentrations, from low to high, were labeled with TMT tags 126, 127N, 127C, 128N, 128C, 129N, 129C, 130N, 130C, 131. The (-) ligand samples were labeled in the reverse manner. The labeling reaction was quenched by the addition of 4 μ L of a 5% v/v hydroxylamine for 15 min at room temperature. In both experiments, the (+) and (-) ligand samples of the 1st, 3rd, 5th, 7th and 9th (i.e., the odd) urea concentration (from low to high) were combined into one sample, and the samples from the other 5 concentrations (i.e., the even urea concentrations) were combined into another sample.

The resulting samples were dried, re-dissolved in 2% v/v TFA solution, and desalted using C18 columns according to the manufacturer's protocol. The desalted samples were dried, re-dissolved in 100 μ L of 0.1 M TEAB solution (pH 8.5), and digested with trypsin overnight at 37 °C. NHS-activated agarose resin and 50 μ L of 0.5M NaCl were added to the trypsin-digested samples such that the NHS-activated agarose resin to total peptide ratio was approximately 150:1 (w/w). The tryptic peptides mixture was allowed to react with the NHS-activated agarose resin for 1.5 hours at room temperature before the samples were acidified with 2% v/v TFA and the soluble peptides desalted using C18 columns. The desalted samples were subject to the LC-MS/MS analysis described below.

2.2.4 STEPP-LiP analysis

Cell lysates generated from five MCF-7 and five MCF-10A cell cultures were each divided into two 10 μ L aliquots and each aliquot was combined with 30 μ L of PBS. Five of the 10 resulting samples generated from each cell line were treated with proteinase K at an enzyme/substrate ratio of 1/100 (w/w) for 5 min at room temperature. The proteinase K proteolysis reaction was quenched with 5 mM PMSF and the samples were subjected to the same STEPP protocol used in the STEPP-PP Analysis (with the exception that C4 columns were used to desalt the TMT labelled samples instead of C18 columns), which included a trypsin digestion step, to generate 5 "double digestion" samples for each cell line. The other 5 aliquots generated for each cell line were reduced with 1.5 mM TCEP for 1 h at 55 °C, modified with 2.5 mM MMTS for 10 min at room temperature and then treated with trypsin at 37 °C overnight to generate 5 "single digestion" samples for each cell line. In total, five biological replicates of the double and single digestion samples were

generated from the MCF-7 and MCF-10A cell lines. The 10 double digested samples were labeled with TMT tags in the STEPP procedure such that one TMT 10-plex contained 5 biological replicates of the MCF-7 double digested samples and 5 biological replicates of the MCF-10A double digest samples. The 10 singly digested samples were also labelled with TMT tags and combined in the same manner as that described above for the double digested samples.

2.2.5 One-Pot STEPP-PP Analysis

The (+) and (-) ligand samples were subjected to a similar STEPP-PP analysis with the exception that a one-pot analysis strategy was employed in the mass spectrometry-based proteomics readout.^{33,34} Briefly, aliquots of the (+) and (-) ligand samples were distributed into a series of 12 urea-containing buffers (PBS, pH 7.4) where the final concentrations of urea were equally spaced at 0.4 M intervals between 1.0 and 5.4 M. The final concentration of CsA in the samples was 120 μ M, and the total amount of protein in each sample was 100 μ g. The samples in the urea-containing buffers were incubated for 2 h at room temperature before 5 μ g of thermolysin was added to each of the (+) and (-) ligand samples in the denaturant-containing buffers. The thermolysin proteolysis reactions were allowed to proceed for 1 min at room temperature before they were quenched upon addition of 60 μ L of a urea/EDTA solution (~0.2 M EDTA, 8 M Urea, pH 8.0). Equal aliquots of the (+) ligand sample solutions were combined, as were equal aliquots of the (-) ligand samples. The above procedure was performed on 5 separate yeast lysates. This ultimately generated 5 (+) and (-) ligand sample pairs for a total of 10 samples. Aliquots containing ~80 μ g of total protein from each of the 5 (+) and (-) ligand

pairs were subjected to the STEPP protocol we recently reported.²⁹ Before the STEPP protocol, an additional 20 μ L of a urea/EDTA solution (\sim 0.2 M EDTA, 8 M Urea, pH 8.0) was added to each sample to ensure proper unfolding for labeling with isobaric mass tags. As part of the STEPP protocol, samples were reacted with 1.5 mM TCEP for 1 h at 30 °C and then with 2.5 mM MMTS for 10 min at room temperature. The protein material in the (+) and (-) ligand samples from each of the 5 biological replicates was labeled with a TMT 10-Plex reagent kit according to the protocol outlined in ref 29. The TMT-labelled samples were combined to generate a single protein sample that was lyophilized, redissolved in 2% v/v TFA, and desalted using C18 columns according to the manufacturer's protocol. The desalted sample was lyophilized, redissolved in 0.1 M TEAB solution (pH 8.5), and digested with trypsin overnight at 37 °C. The ratio of trypsin to total peptide was 1:100 (w/w). NHS-activated agarose resin and 50 μ L of 0.5 M NaCl were added to the digested sample, such that the NHS-activated agarose resin to total peptide ratio was approximately 150:1 (w/w). Samples were reacted for 1.5 h at room temperature, acidified with 2% v/v TFA, and transferred to C18 Macrospin columns for desalting prior to LC-MS/MS analysis.

2.2.6 Quantitative LC-MS/MS Analysis

LC-MS/MS analyses of STEPP-PP and STEPP-LiP samples were performed on a Q-Exactive HF high-resolution mass spectrometer (Thermo) with a nano-Acquity UPLC system (Waters) and a nano-electrospray ionization source fitted with a SilicaTip emitter (New Objective). Samples were trapped on a 2D Symmetry C18 trapping column with dimensions 180 μ m x 20 mm and particle diameter of 5 μ m, pore size 100 Å. The trapping

time was 5 minutes at 5 $\mu\text{L}/\text{minute}$ (99.9:0.1 v/v water/ACN 0.1% formic acid). The samples were separated on a 75 μm \times 250 mm high strength silica (HSS) T3 column with 1.8 μm particle diameter (Waters) heated to 55 $^{\circ}\text{C}$. Peptides were eluted using a gradient of 3-30% acetonitrile with 0.1% formic acid over 90 minutes at a flow rate of 0.3 $\mu\text{L}/\text{min}$. LC-MS/MS data were collected using a top 20 data-dependent acquisition (DDA) method including MS1 at 120k and MS2 at 30k resolution. The AGC target for MS1 was 3×10^6 ions with a max IT of 50 msec. The AGC target for MS2 was 1×10^5 ions with a max IT of 45 msec. The normalized collision energy (NCE) was set to 30 V and the scan range was 375-1600 m/z. The isolation window was 0.7 m/z and the dynamic exclusion time was set to 20.0 seconds.

The LC-MS/MS analyses of one-pot STEPP-PP samples were performed using a nanoAcquity UPLC system (Waters) coupled to Thermo Orbitrap Fusion Lumos mass spectrometer systems. The dried peptide material generated from each analysis was reconstituted in 20 μL (CPP), 15 μL (TPP), 12 μL (PP), and 10 μL (SPROX) of 1% TFA, 2% acetonitrile in H_2O , and a 1 μL aliquot was injected into the UPLC system. The peptides were first trapped on a Symmetry C18 20 mm \times 180 μm trapping column (5 $\mu\text{L}/\text{min}$ at 99.9/0.1 water/acetonitrile, v/v). The analytical separation was performed using an Acquity 75 μm \times 250 mm high-strength silica (HSS) T3 C18 column with a 1.8 μm particle size (Waters); the column temperature was set to 55 $^{\circ}\text{C}$. Peptide elution was performed using a 90 min linear gradient of 3– 30% ACN with 0.1% formic acid at a flow rate of 400 nL/min. The MS data were collected using a top 20 data-dependent acquisition (DDA) method which included MS1 at 120k and MS2 at 50k resolution. The MS1 AGC target was

4.0 × 10⁵ ions with a max injection time of 50 ms. For MS₂, the AGC target was 1.0 × 10⁵ ions with a max injection time of 105 ms. The collision energy was set to 38%, and the scan range was 375–1500 m/z. The isolation window was 0.7, and the dynamic exclusion duration was 60 s. The peptide sample generated in each of the four techniques was subjected to 3 LC-MS/MS analyses. The raw MS data generated in this work have been uploaded to the PRIDE database (accession number PXD014309).

2.2.7 STEPP-PP Quantitative Proteomic Data Analysis

The raw LC-MS/MS files were searched against the yeast UniProt Knowledgebase (2017-06-07 release) using Proteome Discoverer 2.2 (Thermo). The searches were performed with fixed MMTS modification on cysteine, fixed TMT-10plex labeling of lysine side chains, fixed TMT10plex labeling of peptide N-termini, variable deamidation of asparagine and glutamine, variable oxidation of methionine, and variable acetylation of the protein N-terminus. The precursor ion mass tolerance was set at 10 ppm. The fragment ion mass tolerance was set at 0.02 Da. Trypsin(semi) was set as the enzyme, and up to three missed cleavages were allowed. For peptide quantification, reporter ion abundance was set as "intensity" and the normalization mode and scaling mode were each set as "none." The peptide FDR confidence cutoff was set as "medium" (i.e., FDR <5%).

The data from different technical replicates of the same sample was exported and combined into one file. PSMs corresponding to tryptic peptides mapping to protein N-termini were filtered out. The TMT reporter ion intensities in all the PSMs recorded for a semi-tryptic peptide sequence were summed. For each unique semi-tryptic peptide, the summed (-) ligand TMT reporter ion intensity was divided by the summed (+) ligand TMT

reporter ion intensity to generate a fold-change value for each urea concentration at which TMT reporter ion intensities were recorded. Hit peptides were identified with two criteria: i) the peptide must have a significantly altered $\log_2(\text{fold-change})$ ($>3\sigma$ deviations from the mean $\log_2(\text{fold-change})$), for at least two consecutive urea concentrations; ii) these significantly altered $\log_2(\text{fold-change})$ values must have a fold-change in the same direction for at least two consecutive urea concentrations (e.g., a peptide with fold-changes of $>3\sigma$, $<-3\sigma$ and $>3\sigma$ at three consecutive urea concentrations was not a hit).

2.2.8 STEPP-LiP Quantitative Proteomic Data Analysis.

The raw LC-MS/MS files generated in the STEPP-LiP experiment were searched against the human UniProt Knowledgebase (2017-06-07 release) with the same searching parameters as in STEPP-PP data analysis except that the raw file from the single digestion experiment was searched with enzyme set to trypsin(full). For the double digestion sample, similar to STEPP-PP data analysis, PSMs corresponding to tryptic peptides mapping to protein N-termini were filtered out. The TMT reporter ion intensities in all the PSMs recorded for a semi-tryptic peptide sequence were summed. The sum of each TMT reporter ion recorded for a given semi-tryptic peptide was divided by the average of all 10 of the summed TMT reporter ion intensities for that semi-tryptic peptide to generate a single set of so-called relative TMT reporter ion intensities for each semi-tryptic peptide. For the single digestion sample, TMT reporter ion intensities for each identified protein were generated from Proteome Discoverer and these TMT reporter ion intensities for each protein were also divided by the average of all 10 of the summed TMT reporter

ion intensities for that protein to generate a single set of so-called relative TMT reporter ion intensities for each protein.

The data normalization and hit selection procedures used in the STEPP-LiP experiments were directly analogous to those in a previously described LiP experiment using SILAC quantitation.²⁷ Briefly, for each semi-tryptic peptide, the relative TMT reporter ion intensities from the double digestion experiment were divided by the parent protein's relative TMT reporter ion intensities from the single digestion experiment in order to generate normalized TMT reporter ion intensities, which accounted for protein abundance differences in the two cell lines. These normalized values were used in subsequent analyses and selection of semi-tryptic peptide hits. For each semi-tryptic peptide, a median normalized TMT reporter ion intensity value was determined from the five MCF-7 biological replicates. This value was divided by the median normalized TMT reporter ion intensity value determined from the five MCF-10A biological replicates group to generate a MCF-7/MCF-10A fold-change. The fold-change value was used to determine the proteolysis susceptibility change behavior for each semi-tryptic peptide (e.g., a MCF-7/MCF-10A fold change < 1 meant the semi-tryptic peptide site was less susceptible to proteolysis in MCF-7). Semi-tryptic peptide hits were only selected from those that were identified in at least two biological replicates of both cell lines. For peptide hit selection, a Student's two-tailed t test was used to identify hit peptides with significantly different normalized TMT reporter ion intensities between the MCF-7 and MCF-10A groups. Hit peptides were selected as those with a p-value less than 0.05.

2.2.9 One-Pot STEPP-PP Quantitative Proteomic Data Analysis

Raw LC-MS/MS files of the one-pot STEPP-PP samples were searched with Proteome Discoverer 2.2 (Thermo) with the same searching parameters as in section 2.2.3 for STEPP-PP samples with the exception that only peptides with peptide FDR confidence labelled as “high” (i.e., FDR < 1%) were allowed.

For each biological replicate, a normalization factor was calculated using the ratio of the summed signal intensities of all semi-tryptic peptides recorded in the (+) and (-) ligand samples from each biological replicate. For each identified semi-tryptic peptide, a ratio of the observed reporter ion intensities in the (+) ligand sample to the (-) ligand sample was generated for each biological replicate. The resulting ratio was divided by the normalization factor for each of the 5 replicates. These normalized ratios (fold change) were then log₂-base transformed, averaged, and tested by two-tailed Student’s t test comparing with a mean of 0.

2.2.10 K_d Value Determination in STEPP-PP experiment

The median fold change of the four semi-tryptic peptides from HSP82 binding-domain in the geldanamycin binding STEPP-PP experiment were fit to equation (1) using a Mathematica 11.0 script:

$$\log_2 \text{Fold change} = \log_2 \frac{\frac{1}{1 + \text{Exp}\left(-\frac{\Delta G_f + m \times [\text{Urea}]}{RT}\right)} + A}{\frac{1}{1 + \text{Exp}\left(-\frac{\Delta G_f + \Delta\Delta G_f + m \times [\text{Urea}]}{RT}\right)} + A} \quad (1)$$

In equation (1), the *Fold change* is the (-)/(+) semi-tryptic peptide TMT reporter ion intensity ratio at each denaturant concentration, ΔG_f is the folding free energy, $\Delta\Delta G_f$ is the observed ligand-induced protein folding free energy change (i.e., the binding free energy), m is $\frac{d\Delta G_f}{d[Urea]}$, R is the ideal gas constant; T is temperature in Kelvin, A is a customized constant representing the experimental noise (see below for explanation and derivation of equation (1)). Ultimately, the $\Delta\Delta G_f$ value obtained from fitting the $\log_2(\text{fold-change})$ versus $[Urea]$ data to equation (1) was used in equation (2) to calculate the dissociation constant, K_d :

$$K_d = \frac{[L]}{e^{-\Delta\Delta G_f/RT} - 1} \quad (2)$$

In equation (2), $[L]$ is the free ligand concentration, $\Delta\Delta G_f$ is as described above, R is the ideal gas constant, and T is the temperature in K.

Equation (1) and equation (2) were derived from the following steps:

As described by Myers and co-workers,⁵⁶ the folding free energy of a protein, $\Delta G_{f,(-)}$, in chemical denaturant $[Urea]$ can be expressed as:

$$\Delta G_{f,(-)} = \Delta G_f + m \times [Urea] = -RT \ln K_{f,(-)} \quad (3)$$

, where $\Delta G_{f,(-)}$ is the folding free energy of the protein in the absence of ligand, and $K_{f,(-)}$ is the equilibrium constant between the folded and unfolded state of the protein in the absence of ligand, which is defined as

$$K_{f,(-)} = \frac{[Folded]_{(-)}}{[Unfolded]_{(-)}} \quad (4)$$

where $[Folded]$ is the concentration of the folded protein, and $[Unfolded]$ is the concentration of the unfolded protein.

Equations (3) and (4) can be combined and we get

$$\frac{[Folded]_{(-)}}{[Unfolded]_{(-)}} = \text{Exp}\left(-\frac{\Delta G_f + m \times [Urea]}{RT}\right) \quad (5)$$

The total concentration of the protein can be expressed as, $[Protein] = [Folded]_{(-)} + [Unfolded]_{(-)}$. Therefore, equation (5) can then be expressed as,

$$\frac{[Unfolded]_{(-)}}{[Protein]} = \frac{1}{1 + \text{Exp}\left(-\frac{\Delta G_f + m \times [Urea]}{RT}\right)} \quad (6)$$

Similarly, the folding free energy of a protein in the presence of ligand, $\Delta G_{f,(+)}$, in chemical denaturant can be expressed as,

$$\Delta G_{f,(+)} = \Delta G_f + \Delta\Delta G_f + m \times [Urea] = -RT \ln K_{f,(+)} \quad (7)$$

Similarly, we can have

$$\frac{[Unfolded]_{(+)}}{[Protein]} = \frac{1}{1 + \text{Exp}\left(-\frac{\Delta G_f + \Delta\Delta G_f + m \times [Urea]}{RT}\right)} \quad (8)$$

Dividing equation (6) by (8) and taking the logarithm of both sides yields,

$$\log_2 \frac{[Unfolded]_{(-)}}{[Unfolded]_{(+)}} = \log_2 \frac{\frac{1}{1 + \text{Exp}\left(-\frac{\Delta G_f + m \times [Urea]}{RT}\right)}}{\frac{1}{1 + \text{Exp}\left(-\frac{\Delta G_f + \Delta\Delta G_f + m \times [Urea]}{RT}\right)}} \quad (9)$$

As the all unfolded protein are digested by thermolysin and result in semi-tryptic peptides, the ratio $\frac{[Unfolded]_{(-)}}{[Unfolded]_{(+)}}$ is the same as the (-)/(+) semi-tryptic peptide TMT reporter ion intensity ratios. As the STEPP-PP is a TMT-based experiment, there is a constant noise in target peptide quantification from co-isolation and co-fragmentation of other interfering TMT-labeled peptides. To account for this noise in TMT quantification, we add the constant, A , to both numerator and denominator. This constant term is set to be the median isolation interference of the data to be fitted (e.g. in fitting of the HSP82 hit peptides in geldanamycin-binding experiment, this constant A is set to be the median isolation interference of the four HSP82 hit peptides, which is 0.17).

By substituting *fold change* for $\frac{[Unfolded]_{(-)}}{[Unfolded]_{(+)}}$ and adding the constant A , we get equation (1) for K_d determination of STEPP-PP experiment:

$$\log_2 \text{Fold change} = \log_2 \frac{\frac{1}{1 + \text{Exp}\left(-\frac{\Delta G_f + m \times [Urea]}{RT}\right)} + A}{\frac{1}{1 + \text{Exp}\left(-\frac{\Delta G_f + \Delta\Delta G_f + m \times [Urea]}{RT}\right)} + A} \quad (1)$$

2.3 Results and Discussion

2.3.1 STEPP Workflow

Shown in Figure 3 is the STEPP protocol developed here to enrich the semi-tryptic peptides generated in limited proteolysis experiments performed on the proteomic scale. In the STEPP protocol, any intact proteins and the non-tryptic peptide fragments generated in the limited proteolysis reaction are directly labeled with a set of isobaric mass tags (e.g., TMT 10-plex), which react with the ϵ -amino groups in lysine side-chains and with the proteins' and non-tryptic peptides' N-termini. The isobaric mass tag labeling reactions are quenched, and the samples labeled with different TMT tags are combined. The combined is digested with trypsin. This exposes new N-terminal amino groups at the trypsin cleavage sites, which only occur at arginine residues, as lysine cleavage sites are blocked due to the presence of the isobaric mass tag. The trypsin digestion shortens the non-tryptic peptide fragments and generates semi-tryptic peptides, that are suitable for bottom-up shotgun proteomics analysis. However, the trypsin digestion also creates a large number of tryptic peptides. These tryptic peptides are removed from solution through reaction with a N-hydroxysuccinimide (NHS) activated agarose resin. This chemo-selection reaction, while immobilizing tryptic peptides, leaves semi-tryptic peptides in solution and ready for subsequent LC-MS/MS analysis. In theory, each cleavage site resulting from the limited proteolysis reaction in the first digestion step of the STEPP protocol generates two non-tryptic "daughter" peptides. Upon trypsin digestion, these two non-tryptic "daughter" peptides each produce one semi-tryptic peptide that terminates at the non-tryptic digestion site, where one of the semi-tryptic

peptides has a tryptic N-terminus and non-tryptic C-terminus and the other semi-tryptic peptide has a non-tryptic N-terminus and tryptic C-terminus. We note that the chemo-selection strategy outlined here only enriches for one of these semi-tryptic peptides (i.e., the semi-tryptic peptide that has a semi-tryptic N-terminus and tryptic C-terminus).

We also note that although the mechanism for this enrichment is similar to a previously published technique, TAILS, for enriching proteolyzed peptides in degradomics studies, this STEPP methodology has unique novelty and advantages.⁵⁷ The STEPP method uses a NHS ester click chemistry for amine group protection and immobilization, enabling the reaction to be much faster (i.e., the amine group reaction time in STEPP is less than 2 h, comparing to the 16 h reaction time in TAILS) than TAILS where aldehyde reaction are used.⁵⁷ Unlike in TAILS where synthesis of the immobilization polymer is needed, the immobilization agent in STEPP, i.e., NHS-activated agarose, is commercially available. It is also in this work that the enrichment of semi-tryptic peptides is first established for the proteolysis methods including LiP and PP.

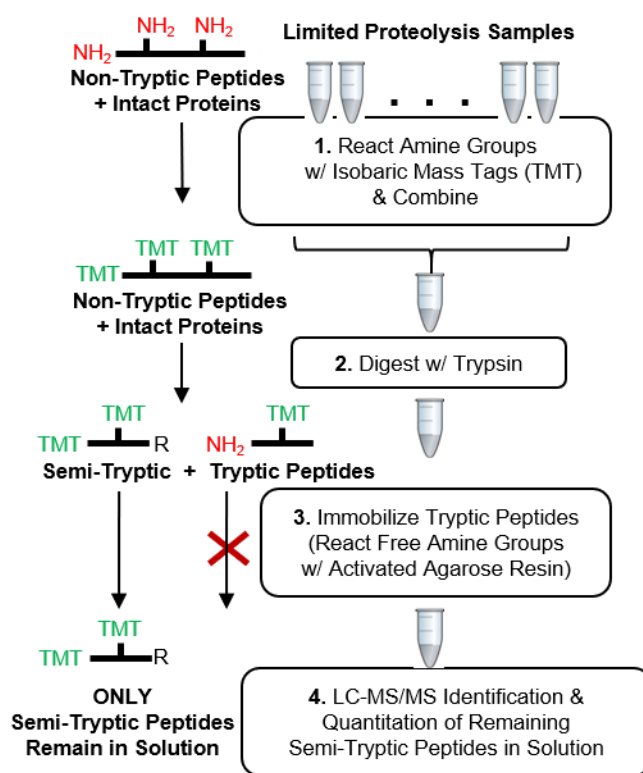


Figure 3 Schematic representation of the STEPP protocol workflow developed in this work.

2.3.2 STEPP-PP Analysis of CsA and Geldanamycin Binding to Proteins in a Yeast Cell Lysate

The STEPP protocol was interfaced with PP as illustrated in Figure 4, and the STEPP-PP workflow was used in a proof-of-principle study to identify the protein targets of two well-studied small molecule drugs, CsA and geldanamycin, using the proteins in a yeast cell lysate.^{58,59} These two model drugs were chosen for these proof-of-principle studies because the binding interactions with their protein targets have been well-studied by conventional methods and by other energetic-based studies, such as chemical denaturation and protein precipitation (CPP), SPROX and SILAC-PP methodology.^{24,60-64}

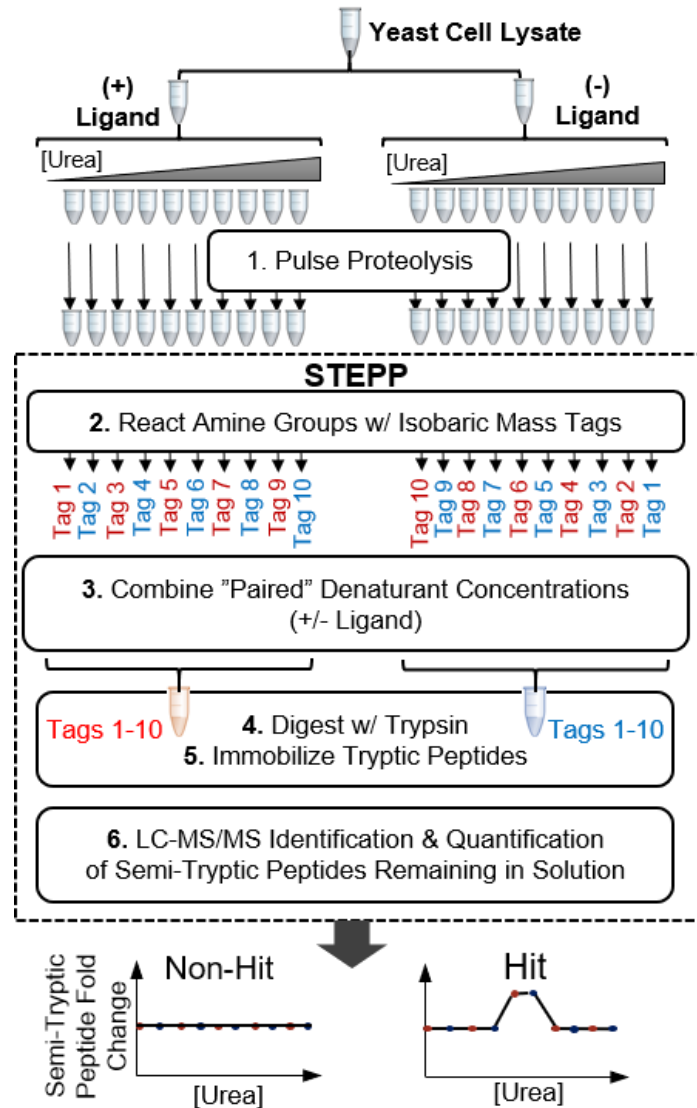


Figure 4 Schematic representation of the STEPP-PP workflow used in this work.

In both the CsA and geldanamycin binding experiments close to 900 yeast proteins were identified by more than 4000 unique semi-tryptic peptides (see Table 1). The total number of semi-tryptic peptides was about 10 times more than that observed in a typical proteome-wide LiP experiment with the same number of LC-MS/MS runs.²⁷ The proteomic coverages observed in these STEPP-PP experiments were also significantly

greater (~40% larger) than the proteomic coverage observed in similar drug-mode-of-action study using SPROX to identify protein targets in the yeast proteome.⁶⁴

Table 1 Summary of proteomic coverages and numbers of hits observed in the STEPP-PP CsA and geldanamycin ligand binding experiments using the proteins in a yeast cell lysate

Drug Experiment	Binding Assay Criteria	Protein Assay Criteria	Assayed Peptides (Proteins)	Hit Peptides (Proteins)	Peptides from Known Protein Target(s)	Protein False Positive Rate
CsA	One peptide	One	4265 (812)	24 (13)	11 (CPR1), 2 (CPR3)	1.4%
	At least two peptides	At least two	3914 (461)	13 (2)		0%
Geldanamycin	One peptide	One	4955 (901)	39 (33)	4 (HSP82)	3.6%
	At least two peptides	At least two	4533 (492)	10 (4)		0.6%

In the CsA binding study, 24 semi-tryptic peptides from 13 proteins were identified as hits, including 13 semi-tryptic peptides from 2 cyclophilin proteins (CPR1 and CPR3, see Figure 5 A and B), both of which are known to bind or be inhibited by CsA.^{65,66} STEPP-PP data from a typical non-hit protein is shown in Figure 5C.

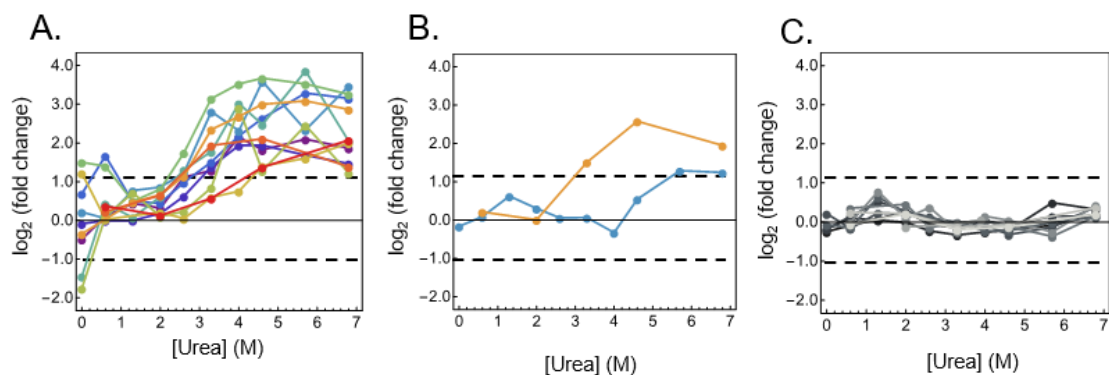


Figure 5 Representative STEPP-PP data obtained in the CsA binding study. (A) Data from the 11 semi-tryptic peptides assayed from the known CsA binding protein, CPR1 (peptidyl-prolyl cis-trans isomerase). (B) Data from the 2 semi-tryptic peptides assayed from the known CsA binding protein, CPR3 (peptidyl-prolyl cis-trans isomerase C, mitochondrial). (C) Data from the 10 semi-tryptic peptides assayed from a non-hit protein, RPL38 (60S ribosomal protein L38). The dashed lines indicate the $\log_2(\text{fold-change})$ values that are $\pm 3\sigma$ deviations from the average $\log_2(\text{fold-change})$ values of all the semi-tryptic peptides assayed in the CsA binding study.

In the geldanamycin binding experiment, 39 semi-tryptic peptides from 33 proteins were identified as hits, including 4 semi-tryptic peptides from the known geldanamycin binding protein, HSP82 (yeast HSP90 isoform) (Figure 6 A).⁶⁷ We note that a total of 28 semi-tryptic peptides from HSP82 were assayed. All 23 of the semi-tryptic peptides mapping to the middle and C-terminal domains were not identified as hits, and 4 out of the 5 semi-tryptic peptides mapping to the N-terminal domain of HSP82 were identified as hits. In both the geldanamycin and CsA binding experiments the known binding targets were successfully identified as hits, giving a 0% protein false negative rate.

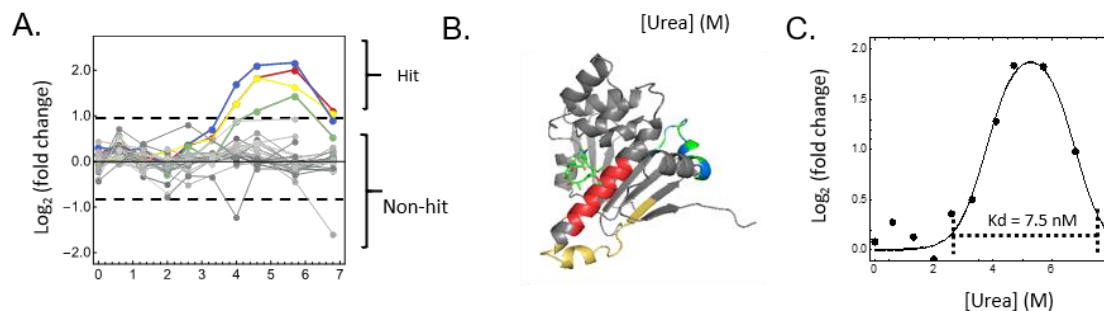


Figure 6 STEPP-PP data obtained in the geldanamycin binding experiment. (A) Data obtained on the 28 semi-tryptic peptides assayed HSP82 peptides in the geldanamycin binding experiment. The 4 hit peptides are shown in blue, red, yellow and green, other non-hit peptides are shown in grey. The dashed lines indicate the $\log_2(\text{fold-change})$ values that are $\pm 3\sigma$ deviations from the average $\log_2(\text{fold-change})$ values of all the semi-tryptic peptides assayed in the geldanamycin binding study. (B) Schematic representation of the HSP82 geldanamycin binding domain with the geldanamycin bound (PDB ID: 1A4H).⁶⁷ The four hit peptides from HSP82 (highlighted with the same colors as in A, respectively) all mapped to this known geldanamycin binding domain. (C) Data used to calculate the K_d value associated with the HSP82-geldanamycin complex. The data points represent the median data from the four hit peptides in (A). The solid line is the best fit of the data to equation (1).

Unfortunately, it is difficult to determine if the peptide hits from proteins other than the cyclophilins in the CsA binding experiment and HSP82 in the geldanamycin binding experiment are from novel CsA or HSP82 binding proteins or are false positives. However, if all such proteins are considered to be false positives, an upper limit for the false positive rates of 0.3-0.7% and 1.4 – 3.6% for peptide and protein hit discovery (respectively) can be established for the STEPP-PP protocol. The false positive rates of protein hit discovery observed here for STEPP-PP are about the same as that previously reported for SILAC-PP with gel fractionation (2.1 - 3.6%);²⁴ and the false positive rates of peptide hit discovery observed here for STEPP-PP are slightly lower than that previously reported for SPROX (~1-2%).⁶⁸

The false positive rate of protein hit discovery in STEPP-PP can be further reduced if hit proteins are also required to have at least 2 semi-tryptic peptide hits. While this reduced the proteome coverage in our experiments by about 45% (see Table 1), it significantly improved the false positive rate. In the CsA binding experiment, only the two cyclophilins CPR1 and CPR3, passed this more stringent hit selection criteria, giving a 0% false positive rate of protein hit discovery. In the geldanamycin binding experiment, only HSP82 and 3 other proteins passed the multiple-peptide hit selection criteria, reducing the false positive rate of protein hit discovery by 6-fold from 3.6% to 0.6%. Thus, requiring just two-peptide hits for protein hit selection can greatly reduce the false positive rate of protein hit discovery using the STEPP-PP approach. However, proteomic coverage is sacrificed. This is because, even with the STEPP-PP protocol to enrich for semi-tryptic peptides, only a fraction (~55%) of the proteins identified in the bottom-up proteomics analysis are identified with two or more semi-tryptic peptides.

2.3.3 STEPP-PP Generates Domain Specific Ligand-Binding Information

The semi-tryptic peptides generated in the STEPP-PP experiment report on the biophysical properties of the protein folding domains to which they map. In the case of single domain proteins all the semi-tryptic peptides generated in the STEPP-PP experiment should display similar behavior. This was, for example, observed with the semi-tryptic peptides from the cyclophilin proteins in the CsA binding experiment (Figure 5 A and B). In the case of large multi-domain proteins, the semi-tryptic peptides from different domains can display different behaviors if they have different interactions with

the ligand. This was, for example, observed with the semi-tryptic peptides from HSP82 in the geldanamycin experiment (Figure 6 A).

HSP82 has three domains, including the N-terminal ATP-binding domain (residues 1-210 that also encompass the geldanamycin binding site), the middle domain (residues 267-547), the C-terminal dimerization domain (residues 548-705), and a flexible linker (residues 211-266).^{59,67} In the geldanamycin binding experiment, the 4 semi-tryptic peptide hits were all from the N-terminal ATP-binding domain, which contains the known geldanamycin binding site.⁶⁷ We note that the 1 semi-tryptic peptide from the N-terminal domain which was not identified as a hit, appears to be a false-negative. Indeed, the fold-change values recorded for this peptide (1.84 and 1.90 at 4 M and 5.7 M urea concentrations, respectively) were very close to the cut-off value (1.93) used for hit selection. As expected, all 23 semi-tryptic peptides from the other domains were not identified as hits (Figure 6 A). In the case of large multi-domain proteins like HSP82, the STEPP-PP protocol can provide domain-specific information about ligand-induced conformational changes. Such domain specific information has not been reported using previous PP protocols.^{22-24,31,32} Presumably, the protein-centered readouts employed in these previous protocols complicates the acquisition of such domain-specific binding information.

2.3.4 Dissociation Constant Determination in STEPP-PP

An inherent advantage of PP over other energetics-based approaches such as the thermal proteome profiling (TPP) approach, is that it can be used to evaluate the dissociation constants (K_d) of protein-ligand complexes.⁶⁹ This is because the chemical

denaturation behavior of proteins determined in the presence and in the absence of ligand can be used to evaluate a binding free energy (i.e., a $\Delta\Delta G_f$ value). In particular, the denaturant dependence of the fold-change values generated in the STEPP-PP experiment can be fit to equation (1) to generate a $\Delta\Delta G_f$ value, which can ultimately be used to calculate a K_d value using equation (2). Indeed, the STEPP-PP data can be used to obtain a K_d of the geldanamycin-HSP82 binding of ~ 7.5 nM (Figure 6 C), which is in good agreement with (i.e., within 20% of) the previously reported value of 9 nM.⁷⁰ However, the determination of a K_d value from STEPP-PP data does require data points at denaturant concentrations that are above the transition mid-point of the most stable form (i.e., that the fold-change values return to the baseline value of ~ 0) (Figure 6 C). For example, the lack of data at higher denaturant concentrations, where the fold-change values are expected to return to 0, precluded the determination of a meaningful K_d value for the CsA - cyclophilin binding interactions (see Figure 5 A and B).

2.3.5 STEPP-LiP Study of Breast Cancer-Related Protein Conformational Changes

The STEPP protocol was also interfaced with the LiP method according to the scheme outlined in Figure 7. The STEPP-LiP protocol was used in a proof-of-principle study of breast cancer-related protein conformational changes using a human breast cancer cell line (MCF-7) and a human normal mammary epithelial cell line (MCF-10A). To evaluate the utility of the STEPP-LiP protocol, the STEPP-LiP results reported here were directly compared to the results generated in a previous analysis of the MCF-7 and MCF-10A cell lines using the normal LiP procedure (i.e., without STEPP).²⁷

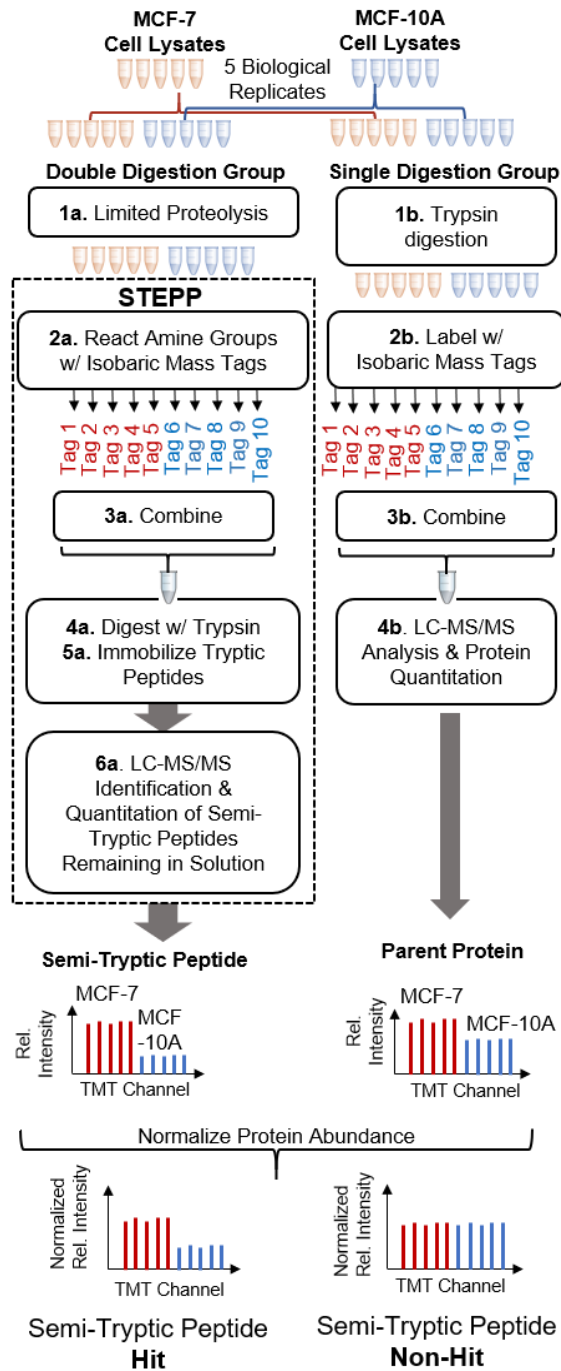


Figure 7 Schematic representation of the STEPP-LiP workflow used in this work

In the STEPP-LiP experiment about 3000 semi-tryptic peptides from 700 proteins (see Table 2) were assayed. This is 5 times more semi-tryptic peptides and 1.25 times more

proteins than that assayed in the normal LiP study of the MCF-7 and MCF-10A cell lines. The protein hits identified in the STEPP-LiP experiment were also in good agreement with those in the normal LiP experiment. A total of 241 proteins were assayed in both the LiP and STEPP-LiP experiments. These 241 proteins included 115 hit proteins in LiP and 181 hit proteins in STEPP-LiP, with a large fraction (80%) of the LiP hits overlapping with the STEPP-LiP hits.

Table 2 Summary of proteomic coverages and numbers of hits observed in the STEPP-LiP MCF-7 vs MCF-10A comparative analysis

	Double Digestion ^a	Single Digestion ^b
Total peptide (protein) identified	3650 (1050)	16656 (1604)
Assayed peptide (protein)	2977 (688) ^c	
Peptide (protein) hits	1240 (440) ^d	

^a The identified peptides in the double digestion group are all semi-tryptic.

^b The identified peptides in the single digestion group are all tryptic.

^c The number of peptides (proteins) identified in least two biological replicates of both the MCF-7 and MCF-10A cell line analyses and in both the double and single digestion groups.

^d The number of semi-tryptic peptides (proteins) with significantly different normalized TMT reporter ion intensities between the MCF-7 and MCF-10A groups ($p < 0.05$).

The increased peptide coverage in the STEPP-LiP experiment not only increases the confidence in protein hits (i.e., protein hits identified with multiple semi-tryptic peptide hits are more likely to be true positives) but it also increases the amount of information that can be gleaned about the conformational properties of hit proteins in the LiP experiment (e.g., see data on HSP90AA1 in Figure 8A and B). The STEPP-LiP data on HSP90AA1 (Figure 8 A) indicate a clear picture of the conformational differences in this

protein in the MCF-7 and MCF-10A cell lines. The STEPP-LiP data reveal that the N-terminal ATP-binding domain of the protein is more protected against proteolysis in the MCF-7 cell line than in the MCF-10A cell line. This is consistent with HSP90 adopting its closed and active form (Figure 8 C) in the MCF-7 cells.^{71,72} This result is also in agreement with the results of other studies that have shown HSP90 is in its closed and active form in cancer cells and in its open and inactive form in normal cells.⁷³

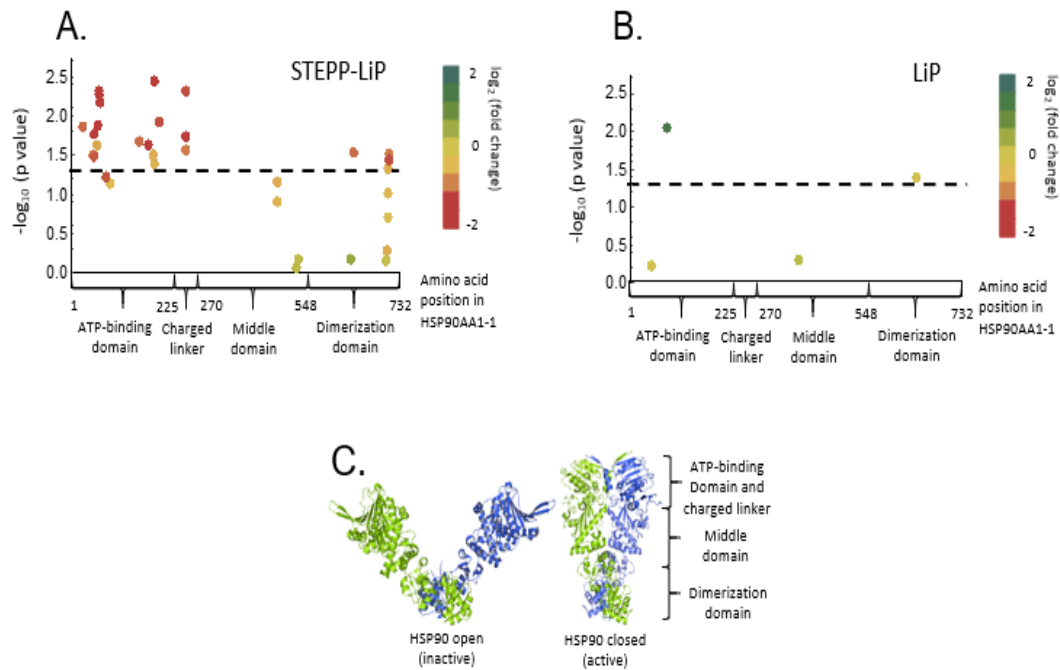


Figure 8 Comparison of the STEPP-LiP and LiP data generated on HSP90AA1 (heat shock protein HSP 90-alpha) in this work and in reference 5 (respectively). (A) The amino acid positions and p-values of the 37 semi-tryptic peptides from HSP90AA1 isoform 1 that were assayed in the STEPP-LiP study of MCF-7 vs MCF-10A protein conformation are plotted. (B) The amino acid positions and p-values of the 4 semi-tryptic peptides from HSP90AA1 isoform 1 that were assayed in the LiP study of MCF-7 vs MCF-10A protein conformation in reference 5 are plotted. In both (A) and (B), the dashed lines correspond to a p-value = 0.05. The color of each data point indicates the MCF-7/MCF-10A TMT reporter ion intensity fold-change of the given semi-tryptic peptide, with more proteolytic susceptibility implicated by a greener color. (C) Schematic representations of the three-dimensional structures associated with the open and closed conformations of the HSP90 homodimer (generated from PDB ID: 2IOQ and 2CG9).^{71,72}

The HSP90AA1 protein was identified as a hit in our previously published normal LiP analysis of the MCF-7 and MCF-10A cell lines. However, the normal LiP data indicated that HSP90AA1 was more protected against proteolysis in the MCF-10A cell line than in the MCF-7 cell line (Figure 6B).²⁷ One explanation for the apparent discrepancy in the normal LiP and STEPP-LiP results on HSP90AA1 is that the semi-tryptic peptide detected in the normal LiP experiment was actually a “tryptic” semi-tryptic peptide generated from non-specific cleavage during the trypsin digestion step. Such “tryptic” semi-tryptic peptides have been reported to constitute 4-40% of the peptides generated in a trypsin digestion, depending on the trypsin manufacturer.⁷⁴ While the trypsin used in the normal LiP is expected to only generate a small fraction of semi-tryptic peptides (<5% of total peptides), the presence of such peptides cannot be ignored in a normal LiP experiment where the number of semi-tryptic peptides is relatively low (i.e., around 13% of the total peptides assayed in the LiP experiment).²⁷

In the normal LiP experiment the presence of “tryptic” semi-tryptic peptides can lead to errant interpretations of the data. Semi-tryptic peptides can be correctly identified as hits in the normal LiP experiment whether they came from the proteinase K or trypsin digestion. However, it becomes ambiguous as to whether or not the hit arises from increased or decreased protection (e.g., a >1-fold change means more susceptible to proteolysis if the semi-tryptic peptide is from proteinase K cleavage but less susceptible if it is from the trypsin digestion step). One advantage of the STEPP-LiP protocol described here is that it is not subject to such ambiguity, as only the N-terminally-TMT-labeled

peptides from the proteinase K digestion step are left in solution after the chemo-selection. Therefore, the above ambiguity is effectively eliminated in the STEPP-LiP experiment.

2.3.6 One-Pot STEPP-PP Analysis of CsA Binding to Proteins in a Yeast Cell Lysate

The STEPP-PP protocol was interfaced with a one-pot data acquisition and data analysis strategy as illustrated in Figure 9, and the one-pot STEPP-PP workflow was used in a proof-of-principle study to identify the protein targets CsA using the proteins in a yeast cell lysate.

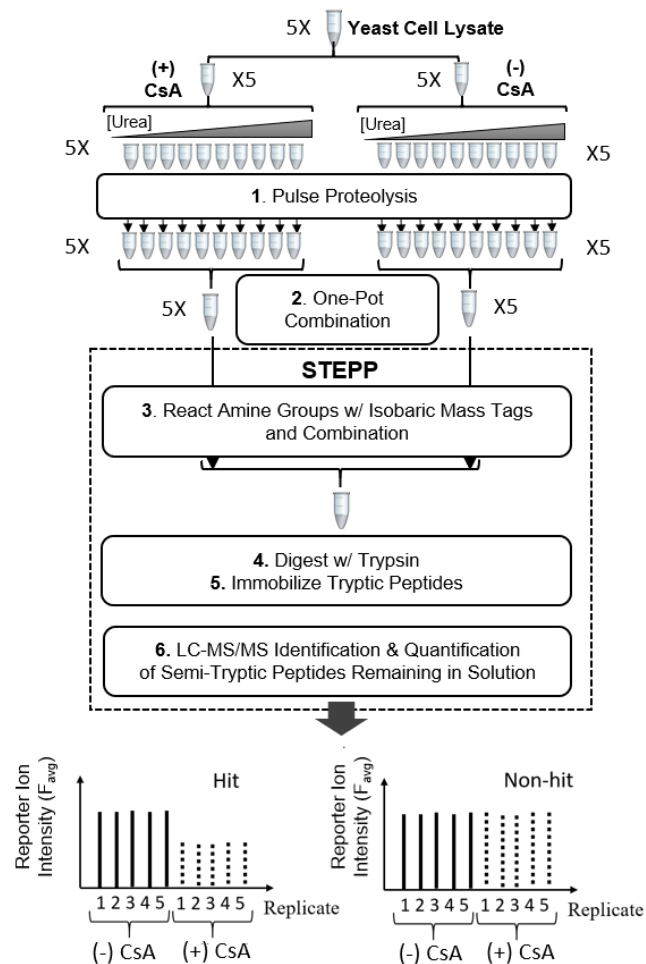


Figure 9 Schematic representation of the one-pot STEPP-PP workflow used in this study.

In the one-pot STEPP-PP workflow, equal amounts of the (+) ligand or (-) ligand samples from a given biological replicate were combined into a single sample after quenching of the proteolysis reaction. This ultimately generated 5 pairs of (+) and (-) ligand samples. These 5 pairs of samples were then subject to a STEPP protocol and analyzed with LC-MS/MS. The TMT reporter ion intensities measured in the LC-MS/MS analyses represent the $F_{\text{avg}(+)}$ and $F_{\text{avg}(-)}$ values for the (+) and (-) ligand denaturation curves in each biological replicate. The selection of semi-tryptic peptide “hits” using one-pot STEPP-PP was based on the two criteria including the following: (i) the p-value from a two-tailed Student’s t test was less than 0.001, and (ii) the average $F_{\text{avg}(+)}/F_{\text{avg}(-)}$ ratio was either greater or smaller than 3 standard deviations from the mean of all the average $F_{\text{avg}(+)}/F_{\text{avg}(-)}$ ratios determined for each technique.

In total, the $F_{\text{avg}(+)}/F_{\text{avg}(-)}$ ratios for 6435 semi-tryptic peptides from 866 proteins were quantified and considered for hit selection in experiment. The details of proteomic coverage and hit are summarized in Table 3. Shown in Figure 10 are volcano plots of the p-values and $F_{\text{avg}(+)}/F_{\text{avg}(-)}$ ratios generated for one-pot STEPP-PP for hit selection.

Table 3 Summary of Proteomic Coverages and Hits Observed in CsA Ligand Binding Experiments Using the one-pot STEPP-PP approach

Technique	Peptide (protein) coverage	Peptide (protein) hits	Known protein (peptide) targets detected as hits	Protein false positive rate
One-pot STEPP-PP	6435 (866)	24 (7)	CypA (17), CypC (1)	0.6%

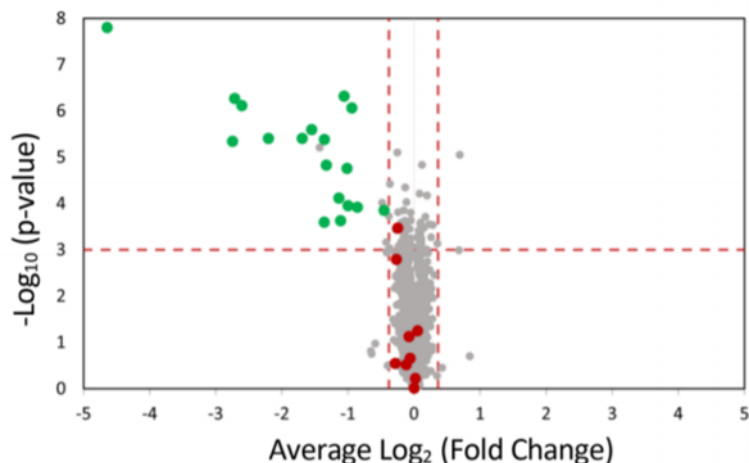


Figure 10 Volcano plots of the average $\log_2 (F_{\text{avg}(+)}/F_{\text{avg}(-)})$ values and p-values generated using a two-tailed Student's t test to analyze the one-pot STEPP-PP data generated in this work. The vertical and horizontal dotted lines mark the hit selection criteria cut-off values for the average $\log_2 (F_{\text{avg}(+)}/F_{\text{avg}(-)})$ values and $-\log_{10}(\text{p-values})$, respectively.

The green data points indicate true positives (i.e., known CsA binding proteins selected as hits), and the red data points indicate false negatives (i.e., known CsA binding proteins or peptides that were not selected as hits).

The one-pot STEPP-PP experiment generate similar proteomic coverage as the normal STEPP-PP experiment (see Table 1), but provide an improved false positive rate. The capability of analyze five biological replicates in one experiment enables a more statistically significant t test hit selection and reduces the false positive rate from 1.4% to 0.6% in one-pot STEPP-PP. However, it should be noted that the K_d value cannot be generated from the one-pot experiment because of the limited information (i.e., only one-fold change value per biological replicate) generated.

3. Investigation of Subglutinol A's Mode of Action

This chapter describes the application of TPP, CPP, SPROX and PP to identify the protein targets of subglutinol A. Subsequent biochemical studies designed to further validate the identified protein targets are also described. This study on subglutinol A is a collaboration with professors Jiyong Hong and Qijing Li at Duke University. Part of the work discussed in this chapter (i.e., the CPP experiment) was performed by H. Meng and was previously reported in his dissertation *Proteomic Methods and Applications for Protein Folding Stability Measurements* (page 92-95).⁷⁵

3.1 Introduction

Autoimmune diseases are a group of more than one hundred chronic inflammatory diseases that affect approximately 3-5% of world population and 8% of the US population.⁷⁶⁻⁷⁸ Current treatments of autoimmune diseases have relied on immunosuppressant or cytotoxic drugs, which are limited in their effectiveness due to the lack of selectivity.⁷⁹ Even with selective immunosuppressive drugs such as cyclosporine A (CsA) and FK506, the required levels of medication can cause serious side effects.⁷⁸ Thus, the continued search for potent and less cytotoxic immunomodulants is critical for reducing the side effects of current autoimmune therapies.

Subglutinol A (Sub A, Figure 11) is a immunosuppressive α -pyrone diterpenoid natural product isolated from an endophytic fungus from the vine *Tripterygium wilfordii*.⁸⁰ For more than two thousand years, *Tripterygium wilfordii* has been used in Chinese traditional medicine as anti-inflammation and immunosuppressive herbal remedy.⁸¹ The

of action. This work on the understanding of Sub A's mode of action could ultimately facilitate future clinical applications of the drug, increase our understanding of the mechanism of the immune process, as well as foster development of better immunosuppressant drugs.

3.2 Experimental Section

3.2.1 Materials

The following materials were from Sigma-Aldrich (St. Louis, MO): guanidine hydrochloride (GdmCl), S-methylmethanesulfonate (MMTS), dimethyl sulfoxide (DMSO), urea, centrifugal filter units 10k (Amicon Ultra-0.5), tris(hydroxymethyl)-aminomethane hydrochloride (Tris·HCl), thermolysin from *Geobacillus stearothermophilus*, and light- and heavy-labeled dCMP and dTMP. The following materials were from Thermo Scientific (Waltham, MA): 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF), bestatin, E-64, leupeptin, pepstatin A, and trypsin protease MS grade. Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) was from Santa Cruz Biotechnology (Dallas, TX). MacroSpin columns (Silica C18) were from Nest Group (Southborough, MA). Phosphopeptide enrichment kit was from GL Sciences (Tokyo, Japan). Phosphate-buffered saline (PBS, pH 7.4, 1x) was from Gibco (Gaithersburg, MD). Ethylenediaminetetraacetic acid (EDTA, 0.5 M, pH = 8.0) was from Mediatech, Inc (Manassas, VA). Sub A was synthesized and provided by Hyeri Park from Professor Jiyong Hong's Laboratory at Duke University. DCK, ACTB, XPO2 and HDAC antibodies were from Santa Cruz Biotechnology, Inc (Dallas, TX). Laemmli sample buffer (4x) was from Bio-Rad (Hercules, CA). Purified recombinant DCK protein and DCK activity assay

were purchased from Novocib (Lyon, France). [5-³H] deoxycytidine was from Moravek (Brea, CA). MN 616 LSB-50 anion exchange filters were from Macherey-Nagel (Düren, Germany). The Oasis WAX 1cc Cartridges were from Waters (Milford, MA).

3.2.2 Cell Culture and Lysis

Mouse 2B4 T hybridoma cells were cultured according to standard protocols, except in the case of the 2B4 T hybridoma cells used for the protein expression level, phosphoproteome, and DCK localization analyses. In these analyses, the cell cultures were supplemented with 100 µM of Sub A (treatment group) in DMSO or only DMSO (control group) for 24 hours or 48 hours during the cell culture before lysis. Cell pellets were lysed in PBS with protease inhibitor cocktail (1 mM AEBSF, 10 µM pepstatin A, 20 µM leupeptin, 15 µM E-64, and 500 µM bestatin) except for cell pellets used for the in-lysate DCK activity assay. In this case the cell pellets were lysed in Tris-HCl pH 8.0 buffer with protease inhibitor cocktail as mentioned above, 2mM DTT, 5mM MgCl₂, 10mM NaF and 1mM NaVA (sodium orthovanadate). Cell lysis was accomplished using zirconia/silica beads (1 mm) at 4 °C with 20 s of disruption 20 times with 1 min intervals on ice in between except for TPP analysis, where cells were lysed by 4 times of liquid nitrogen freeze thaw cycles. The cell lysate was centrifuged at 14 000 × g and 4 °C for 10 min, and the supernatant was saved for subsequent analysis.

The total protein concentration in all lysates was determined using a Bradford assay and adjusted to 10 mg/mL with PBS pH 7.4 or Tris-HCl pH 8.0 for the in-lysate DCK activity assay. For each analysis (except for the protein expression level analysis, phosphoproteome analysis, in-lysate DCK activity assay and DCK localization analysis)

the test lysate was divided into two equal portions. One portion was spiked with a solution of the Sub A prepared in DMSO and one was spiked with DMSO as vehicle. The solutions were equilibrated for 1 hour before analysis.

3.2.3 SPROX analysis

The SPROX experiments were conducted similar to that previously described using a filter aided sample preparation and iTRAQ labeling (iFASP) protocol for bottom-up sample preparation.⁸³

The (+) and (-) ligand cell lysate samples were distributed into a series of urea-containing buffers (PBS pH 7.4) with the final urea concentrations ranging from 0 to 7 M. The final volume in each buffer was 95 μ L. The final concentration of Sub A was 150 μ M. The solutions were equilibrated at room temperature for 1 hour before 5 μ L of 30% (v/v) hydrogen peroxide solution was added into each solution to initiate methionine oxidation. After 6 min, the oxidation reactions were quenched by addition of 500 μ L of 500 mM TCEP solution. An aliquot of 500 μ L of each of the resulting sample solution was transferred into a 10k centrifugal filter unit and was subjected to iFASP protocol that involved reduction with TCEP, reaction with MMTS, digestion with trypsin, and labeling with a iTRAQ-8plex reagent.⁸⁴ The iTRAQ-8plex labeling scheme involved labeling the protein samples derived from each of the denaturant concentrations in the (-) ligand samples with the reagents from one iTRAQ-8plex and labeling each of the denaturant concentrations in the (+) ligand samples with the reagents from another iTRAQ-8plex. The (+) ligand samples were combined and so were the (-) ligand samples. The two combined samples were desalted using a C18 column (The Nest Group) according to the

manufacturer's protocol. These were the 'non-enriched' samples. An aliquot containing about 100-200 ug of iTRAQ-labeled peptides were removed from the desalt samples and enriched for methionine residue containing peptides using a Pi3™ Methionine Reagent kit according to the manufacturer instructions (The Nest Group). These were the 'enriched' samples.

3.2.4 TPP analysis

The TPP experiments were conducted similar to that previously described⁸⁵ but with filter aided sample preparation (FASP) protocol for bottom-up sample preparation.⁸⁵ The total protein concentration of (+) and (-) ligand cell lysate samples were diluted to 1 mg/ml with PBS pH 7.4 solution. The final concentration of Sub A was 100 μM. The solutions were equilibrated at room temperature for 1 hour before ten aliquots of 100 μL of each sample were taken out and incubate at ten different temperatures from 40 to 67 °C. After 3 min of incubation, the samples were cooled down at room temperature for 3 min before further cooling down on ice. The cooled samples were centrifuged at 100 000 × g and 4 °C for 20 min. The same volume of supernatant in each (+) and (-) ligand sample containing no more than 300 μg proteins from each centrifuged sample was transferred into a 10k centrifugal filter unit. All the samples were subjected to a FASP protocol that involved reduction with TCEP, reaction with MMTS, digestion with trypsin and labeling with a TMT-10plex reagent kit according to the manufacturer's protocol. The TMT-10plex labeling scheme involved labeling the protein samples derived from each of the denaturant concentrations in the (-) ligand samples with the reagents from one TMT-10plex and labeling each of the denaturant concentrations in the (+) ligand samples with

the reagents from another TMT-10plex. The TMT labelling reactions were performed according to the manufacturer's instruction, with the exception that 0.5 unit of each TMT reagent was used in each reaction instead of 1 unit.

3.2.6 STEPP-PP analysis

The STEPP-PP analyses applied on this work is identical to that outlined in Chapter 2 of this thesis and previously reported in reference.² Briefly, 10 μ L aliquots of the (+) and (-) ligand-containing lysate samples were distributed into a series of 30 μ L urea-containing buffers (PBS buffer, pH 7.4) where the final urea concentration ranged from 0 to 6.8 M. The final concentration of ligand was 100 μ M. The samples in the urea-containing buffers were equilibrated for ~2 h at room temperature. The proteolysis reaction was initiated by adding thermolysin to the protein and protein–ligand samples in each denaturant-containing buffer. The thermolysin to total protein ratio was approximately 1:10 (w/w). The thermolysin proteolysis reactions were allowed to proceed for 1 min at room temperature before each reaction was quenched with the addition of 60 μ L urea/EDTA solution (~0.2 M EDTA, 9 M urea, pH 8.0).

The proteolyzed samples were subject to the STEPP protocol. As part of the STEPP protocol, the proteolyzed samples were reacted with 1.5 mM TCEP for 1 h at 30 °C and then with 2.5 mM MMTS for 10 min at room temperature. Subsequently, the thermolysin digested samples were labeled with TMT reagents by adding 41 μ L of ACN containing 0.5 unit of each TMT reagent and incubating for 1.5 hours at room temperature. For (+) ligand samples, the urea concentrations, from low to high, were labeled with TMT tags from 126 to 131. The (-) ligand samples were labeled in the reverse manner (i.e., the urea

concentrations, from high to low, were labeled with TMT tags from 131 to 126). The labeling reaction was quenched by the addition of 4 μL of a 5% v/v hydroxylamine for 15 min at room temperature. The (+) and (-) ligand samples of the 1st, 3rd, 5th, 7th and 9th (i.e., the odd) urea concentration (from low to high) were combined into one sample, and the samples from the other 5 concentrations (i.e., the even urea concentrations) were combined into another sample. The resulting samples were dried, re-dissolved in 2% v/v TFA solution, and desalted using C18 columns according to the manufacturer's protocol. The desalted samples were dried, re-dissolved in 100 μL of 0.1 M TEAB solution (pH 8.5), and digested with trypsin overnight at 37 °C. NHS-activated agarose resin and 50 μL of 0.5M NaCl were added to the trypsin-digested samples such that the NHS-activated agarose resin to total peptide ratio was approximately 150:1 (w/w). The tryptic peptides mixture was allowed to react with the NHS-activated agarose resin for 1.5 hours at room temperature before the samples were acidified with 2% v/v TFA and the soluble peptides desalted using C18 columns. The desalted samples were subject to the LC-MS/MS analysis.

3.2.7 Protein Expression Level and Phospho-proteome Analysis

A total of 4 biological replicates of 2B4 T cells treated with 100 nM Sub A in DMSO (treatment group) and DMSO only (control group) were harvested and lysed as described above in cell culture and lysis. An aliquot containing 100 μg of total protein from each cell lysate was transferred into a 10k centrifugal filter unit and was subjected to a iFASP protocol and labeling with a iTRAQ 8plex similar to that described above in the SPROX analysis section. The resulting iTRAQ 8plex-labeled samples were combined and desalted

using a C18 column (The Nest Group) according to the manufacturer's protocol. These were the protein expression level samples. An aliquot containing about 200 ug of iTRAQ-labeled peptides were removed from the desalt samples and enriched for phosphopeptides using a Titansphere Phos-Tio kit (GL Sciences) according to the manufacturer instructions. These were the phosphoproteomic samples.

3.2.8 Quantitative LC-MS/MS Analysis for Proteomics Samples

LC-MS/MS analyses were performed on a Q Exactive HF high-resolution mass spectrometer (Thermo) with a nano-Acquity UPLC system (Waters) and a nanoelectrospray ionization source fitted with a SilicaTip emitter (New Objective). Samples were trapped on a 2D Symmetry C18 trapping column with dimensions of 180 $\mu\text{m} \times 20 \text{ mm}$, particle diameter of 5 μm , and pore size of 100 \AA . The trapping time was 5 min at 5 $\mu\text{L}/\text{min}$ (99.9:0.1 v/v water/acetonitrile 0.1% formic acid). The samples were separated on a 75 $\mu\text{m} \times 250 \text{ mm}$ high strength silica (HSS) T3 column with 1.8 μm particle diameter (Waters) heated to 55 $^{\circ}\text{C}$. Peptides were separated using a gradient of 3–30% acetonitrile with 0.1% formic acid over 90 min at a flow rate of 0.3 $\mu\text{L}/\text{min}$. Data collection was performed in a data-dependent acquisition (DDA) mode with a resolution of 120 000 (at m/z 200) for full MS scan from m/z 375–1600. The target AGC value was 3×10^6 ions; the maximum ion trap fill time was 50 ms, and the normalized collision energy was 27 V. This MS scan was followed by 20 product ion scans at a resolution of 30 000 (at m/z 200), using a minimum AGC target value of 2.25×10^6 ions, an isolation window of 1.2 m/z , and a dynamic exclusion time of 20.0 s.

The raw LC-MS/MS data from TPP analysis was searched by Proteome Discoverer 2.2 against mouse (*mus musculus*) in the 2017-06-07 release of the UniProt Knowledgebase. The searches were performed with fixed MMTS modification on cysteine, TMT-10plex labeling of lysine side chains and peptide N-termini, variable oxidation of methionine, variable deamidation of asparagine and glutamine, and acetylation of the protein N-terminus. The precursor mass tolerance was set at 10 ppm. The fragment mass tolerance was set at 0.02 Da. Trypsin was set as the enzyme, and up to two missed cleavages were allowed. The raw LC-MS/MS data from SPROX and protein expression level analysis were searched with the same searching parameters except for iTRAQ 8plex labeling instead of TMT-10plex labeling. The raw LC-MS/MS data from the phosphoproteome analysis were searched using the same search parameters used in the SPROX and protein expression level analyses except for allowing flexible phosphorylation on serine, threonine and tyrosine. The raw LC-MS/MS data from STEPP-PP analysis were searched with the same parameters described above in section 2.2.7 of this thesis.

The two biological replicates of the (+) ligand and (-) ligand samples generated in the SPROX experiments were each analyzed by LC-MS/MS in triplicate. All other proteomic samples were analyzed with single LC-MS/MS run.

3.2.10 Proteomic Data Analysis

The iTRAQ-SPROX analysis data was analyzed based on reporter ion intensities generated for each methionine-containing peptide as previously described.¹⁸ Briefly, the iTRAQ reporter ion intensities recorded in a given product ion mass spectrum were normalized to the average intensity of the 8 reporter ions in each product ion mass

spectrum to generate N1 normalized values. The N1 values for all the non-methionine containing peptides were averaged for each reporter ion in order to generate an N2 normalization factor for each isobaric mass tag. Ultimately, the N1-normalized values generated for the methionine-containing peptides were divided by the corresponding N2 normalization factor to determine the N2-normalized reporter ion intensity. The resulting chemical denaturation data sets (i.e., the N2-normalized iTRAQ reporter ion intensities generated for each unoxidized methionine-containing peptide) were normalized to a maximum of 1 by dividing each intensity by the maximum intensity and fitted to a four parameter sigmoidal equation, equation 1, using a JAVA-based program (developed in house) that utilized the Nelder and Mead Simplex method for regression analysis.⁸⁶

$$y = A + \frac{(B - A)}{1 + e^{-\frac{C_{1/2} - x}{b}}} \quad (10)$$

In equation (10), A is the pre-transition baseline, B is the post-transition baseline, $C_{1/2}$ is the transition midpoint and b is a measure of the steepness of the transition. The program fit each set of data nine times, once with all eight points and then eight additional times systematically leaving one point out. The fit with the highest R^2 value was chosen as the final output. Subsequent analyses of the data only utilized the chemical denaturation data sets that were determined to be high quality ($R^2 \geq 0.8$). If a peptide was identified multiple times within the same iTRAQ-SPROX experiment, the iTRAQ reporter ion intensities from the high-quality data were summed together to generate one set of iTRAQ reporter ion intensities at the 8 denaturant concentrations per peptide. The

summed data was N2-normalized and fit to equation 1, as described above, to extract a single $C_{1/2}$ value for each identified peptide.

For each biological replicate of the SPROX analysis, peptide hit determination was identical to previously described by Ogburn et al.⁸³ Briefly, peptide hits were selected based on the magnitude of the $\Delta C_{1/2}$ value determined between the (-) and (+) ligand samples and the magnitude of the N2 normalization value differences at or between the transition regions of the chemical denaturation curves generated for the (-) and (+) ligand samples. Statistically significant $\Delta C_{1/2}$ values were taken to be those greater than or equal to 2 times the pooled standard deviation of all the $C_{1/2}$ values generated from all the peptides. Statistically significant N2 reporter ion differences were taken as those with $-\log_{10}(\text{Diff}_{\text{prob}}) > 1$, where $\text{Diff}_{\text{prob}}$ corresponds to the probability that the measured N2 reporter ion difference would randomly occur.

The TPP data was analyzed based on reporter ion intensities generated for each protein as previously described.⁶⁹ Briefly, the exported protein intensities at the same denaturation temperature were summed for the (-) ligand and (+) ligand samples in each experiment. These summed intensities from the (-) ligand and (+) ligand samples in each experiment were normalized by the highest intensities to rescale the intensities from 0 to 1, which are hereafter referred to as super intensities. These averaged super intensities were fit to equation (11),

$$y = \frac{1 - A}{1 + e^{-\frac{a}{x-b}}} \quad (11)$$

and the ratio of the predicted intensities from the best fit to the actual super intensities in the (-) ligand and (+) ligand group was calculated. These ratios obtained from a global analysis of the data were used to normalize the intensity data generated for each individual protein. This normalization accounted for systematic errors between the (-) and (+) ligand groups. The normalized protein intensities from all three biological replicates were input to a R-based fitting program developed previously by Franken et al. and generated a p-value for each protein assayed in at least one of the biological replicate.⁸⁷ Proteins with a p-value less than 0.001 were selected as hit proteins of the TPP analysis.

The STEPP-PP data analysis and hit selection was accomplished similar to previous described in section 2.2.6. An assayed protein was required to be identified with at least one semi-tryptic peptide and a hit protein was required to have at least one semi-tryptic peptide hit. Each assayed protein was required to be identified with at least one semi-tryptic peptide and a hit protein was required to have at least one semi-tryptic peptide hit. Hit peptides were identified with two criteria: 1) the peptide must have a significantly altered $\log_2(\text{fold-change})$ ($>3\sigma$ deviations from the mean $\log_2(\text{fold-change})$), for at least two consecutive urea concentrations; 2) these significantly altered $\log_2(\text{fold-change})$ values must have a fold-change in the same direction for at least two consecutive urea concentrations.

The expression proteomics data was analyzed based on reporter ion intensities generated for each protein. The total exported protein intensities of each iTRAQ channel were summed. These summed intensities from the four replicates of (-) ligand and (+)

ligand samples were normalized by the highest intensities to rescale the intensities from 0 to 1, which are the super intensities. These ratios obtained from a global analysis of the data were used to normalize the intensity data generated for each individual protein. This normalization accounted for difference in total protein expression levels between the four replicates of (-) and (+) ligand samples. A (+)/(-) fold change value was generated for each individual protein, by dividing the mean normalized intensities of the (+) ligand group to the (-) ligand group. A protein was determined as a hit protein if it passes both two criteria: 1) the \log_2 (fold change) of the protein is large than 3σ from the global mean value; and 2) the t test p-value < 0.01 .

The phosphoproteomics data was analyzed based on reporter ion intensities generated for phosphopeptide. Peptides without phosphorylation(s) were filtered out from the datasets. The iTRAQ intensities of each phosphopeptide were normalized by the raw iTRAQ intensities of its parent protein obtained in the expression proteomics experiment. This normalization accounted for difference in protein amounts across different samples. The hit selection criteria was identical to expression proteomics, requiring 1) the \log_2 (fold change) larger 3σ from the global mean value; and 2) the t test p-value < 0.01 .

3.2.11 Gel-Based Pulse Proteolysis on DCK – Sub A Binding

Pulse proteolysis was done collaboratively with Dr. He Meng and Kun Xiang from Duke University. The (+) and (-) Sub A cell lysate samples were distributed into a series of GdmCl-containing buffers (PBS pH 7.4) with the final urea concentrations ranging from 1.25 to 3 M. The final volume in each buffer was 30 μ L. The final concentration of Sub A

in each buffer was 100 μM , 1 μM , or 0 μM depending on the experiment. The solutions were equilibrated at room temperature for 1 hour before 0.5 μg thermolysin in 1 μL PBS was added into each solution. The final ratio of thermolysin to total protein is 1:10 (w/w). The reaction was quenched by adding 1 μL 0.5 M EDTA after 1 min. Each protein solution was mixed with Laemmli sample buffer, heated to 95 $^{\circ}\text{C}$ for 5 min, separated by 10% SDS-PAGE and then transferred to a Hybond membrane (Amersham). The membranes were incubated with DCK antibody in 5% milk/TBST buffer (25 mM Tris pH 7.4, 150 mM NaCl, 2.5 mM KCl, 0.1% Triton-X100) overnight, and then probed for 1 h with secondary horseradish peroxidase-conjugated anti-mouse IgG (Santa Cruz). After extensive washing with PBST, the target proteins were detected on membrane by enhanced chemiluminescence (Pierce).

3.2.12 DCK Protein Activity Assay

The purified DCK protein activity assay was conducted following the manufacturer's protocol with a PRECICE[®] DCK Phosphorylation assay kit (Novocib). Briefly, this activity assay uses deoxyinosine (dI) as a substrate of DCK to form deoxyinosine monophosphate (dIMP), which is then oxidized to deoxyxanthosine monophosphate (dXMP) by a highly active inosine monophosphate dehydrogenase in the presence of NAD. By monitoring the speed of NADH_2 formation from NAD, which was limited by the activity of DCK in the first step of the coupled reaction, activities of purified DCK enzyme under different ligand concentration were able to be compared. In this work, 2 μg of purified DCK protein was incubated with pure DMSO (control) or Sub A in DMSO at a series of final concentration from 0.1 nM to 100 μM in the provided reaction buffer.

After 1 hour of incubation at room temperature, the assay was initiated by addition of ATP. The formation of NADH₂ in the reaction mixtures was monitored with a plate reader at 340 nm absorption.

The endogenous DCK activity assay was performed in-lysate according to a previously established radiochemical assay protocol.⁸⁸ The 2B4 T cell lysate prepared as described in section 3.2.2 Cell Culture and Lysis of this thesis, was distributed into four Sub A containing buffers (50mM Tris-HCl pH 7.6 with 2mM DTT, 5mM MgCl₂ and 10mM NaF and 2% v/v DMSO) with final Sub A concentrations of 0 M, 1 nM, 100 nM and 10 μM. The solutions were equilibrated at room temperature for 1 hour before the radioactive DCK substrate [5-³H] deoxycytidine was added into each solution at final concentration of 1 μM (about 0.02 mCi/mL) and incubated at 37 °C for 5 min. The phosphorylation of the DCK substrate was initiated by addition of ATP at final concentration of 5 mM. The time of addition of ATP was marked as time 0. The solutions were incubated in a thermomixer at 37 °C and constant shaking as a 50 μL aliquot of each solution was taken out at time 5 min, 10 min, 20 min, 30 min and 60 min to monitor the formation of dCMP. The aliquots taken out were first combined with 10 μL 10mM unlabeled dC to quench the phosphorylation reaction of [5-³H] deoxycytidine, then applied to an anion exchange filter that trapped the product dCMP but not dC. The solution-loaded filters were dried first and then washed with water and 1.5 mM ammonium formate before dried with air again. The filters were each soaked in 8 mL of scintillation fluid and analyzed for radioactivity using a LS6500 liquid scintillation counter (Beckman).

3.2.13 Gel-Based DCK Localization Analysis

2B4 T cells cultured with and without Sub A were fractionated to separate nuclear and cytoplasmic proteins with NE-PER™ Nuclear and Cytoplasmic Extraction Reagents (Thermo Scientific) according to the manufacturer's protocol. The total protein concentration both nuclear and cytosolic fractions were tested with Bradford assay. The cytoplasmic protein concentration was tested to be about 10mg/mL and reduced ten-fold with PBS to the same level as the nuclear protein concentration, which is about 1mg/mL. SDS-PAGE gel electrophoresis of the cytosolic and nuclei fraction samples were performed same as described in chapter 3.2.11 Gel-Based Pulse Proteolysis on DCK – Sub A Binding. The resulting SDS-PAGE membranes were also subject to similar western blot readout as described in the above chapter, targeting DCK, HDAC, ACTB and XPO2 with corresponding antibodies.

3.2.14 LC-MS/MS dCMP Concentration Assay

The number of cells in the (+) and (-) Sub A cell cultures were counted with a hemocytometer. The cells were harvested and 10 million cells from each culture were taken out and pelleted. Cell nuclei were separated from these cell pellets and lysed according to an established protocol described by Suzuki et al.⁸⁹ Briefly, the harvested cell pellets were first resuspended in PBS buffer with 0.1% NP-40 to dissolve the cell membrane but not the nucleus membrane. The resulting suspension was then centrifuged with a tabletop mini centrifuge to pellet the intact cell nuclei. The supernatant containing cytosolic fraction was removed and the nuclei pellet was washed with PBS buffer with

0.1% NP-40 again before collection. The collected nuclei were lysed with four liquid nitrogen freeze-thaw cycles.

The cell lysates were spiked with 10 μL of a 50 $\mu\text{g}/\text{mL}$ solutions of heavy labeled dCMP (2'-Deoxy-5'-cytidylic- $^{13}\text{C}_9,^{15}\text{N}_3$) and heavy labeled dTMP (5'-Thymidylic- $^{13}\text{C}_{10},^{15}\text{N}_2$), combined with 200 μL of water, and heated at 99 $^\circ\text{C}$ for 10 min to denature proteins from the cell nuclei. The heated samples were then centrifugated at 14,000 g for 15 min at room temperature. The supernatants were removed from the samples, combined with 1 mL of 100 mM ammonium acetate pH 4.5 solution and subject to a weak anion-exchange solid phase extraction with WAX cartridge (Waters) following the manufacture's protocol. The resulting samples were dried and subjected to LC-MS/MS analysis.

The LC-MS/MS quantification of dCMP and dTMP was performed on an 6460 Triple Quad LC/MS system (Agilent) equipped with an 1260 HPLC (Agilent). The mass spectrometer was operated with an Agilent Jet Stream (AJS) ESI source (Agilent). Mobile phase A1 was comprised of a 10 mM solution of ammonium acetate (pH 7.7) and 10% (v/v) ACN. Mobile phase B was comprised of 97% ACN in water. The stationary phase was a SeQuant ZIC-HILIC column (Millipore Sigma) that was 100 mm long and 2.1 mm in diameter and packed with 3.5 μm particles with 100 angstrom pores. The gradient used in this work is summarized in Table 4.

Table 4 Gradient used in LC-MS/MS quantification of dCMP and dTMP

Time (min)	0	5	15	19	21.5	22	27
Mobile Phase	10%	25%	25%	60%	60%	10%	10%
A							
Mobile Phase	90%	75%	75%	40%	40%	90%	90%
B							

The LC-MS/MS instrument was operated in multiple-reaction-monitoring (MRM) mode with transitions set to be 320 → 119 (positively charged heavy-labeled dCMP), 312 → 112 (positively charged dCMP light), 335 → 86 (positively charged heavy-labeled dTMP) and 323 → 81 (positively charged dTMP light). The collision energy was set to 15 eV for all transitions. The in-source parameters were set as follows: gas temperature 350 °C, gas flow 8 l/min, nebulizer 45 psi, shear gas temperature 400 °C, shear gas flow 11 l/min, capillary voltage 3500 V, nozzle voltage 0 V. Each HILIC-MS/MS run was divided into three sections, in which the eluent from the column was diverted to waste from 0 to 0.5 min and from 21.5 to 27 min in order to minimize the contamination of the ion source.

The LC-MS/MS settings for dTMP quantification was further optimized by Ryan Fader from Duke University. The optimized method altered the in-source parameters as follows: gas temperature 250 °C, gas flow 5 l/min, nebulizer 30 psi, shear gas temperature 350 °C, shear gas flow 12 l/min, capillary voltage 3500 V, nozzle voltage 0 V.

3.3 Results and Discussion

3.3.1 Identification of Sub A Protein Targets

Four mass spectrometry-based proteomics methods, including iTRAQ-SPROX, TPP, CPP and STEPP-PP were employed to identify the protein target of Sub A in a T cell

lysate. The CPP experiment was conducted by Dr. He Meng from the Fitzgerald Laboratory at Duke University and is described in detail in his doctoral thesis.⁷⁵ All of the SPROX, TPP, CPP, and STEPP-PP experiments involved incubating the 2B4 T cell lysate with and without Sub A before aliquots of each sample are distributed into a series of buffers containing increasing chemical denaturant (iTRAQ-SPROX, CPP and STEPP-PP experiments) or incubated at a series of increasing temperature for thermal denaturation (TPP experiment). The iTRAQ-SPROX, TPP, STEPP-PP and CPP experiments were conducted according to published protocols.^{83,90,91} The general workflow used for each technique is summarized in Figure 12.

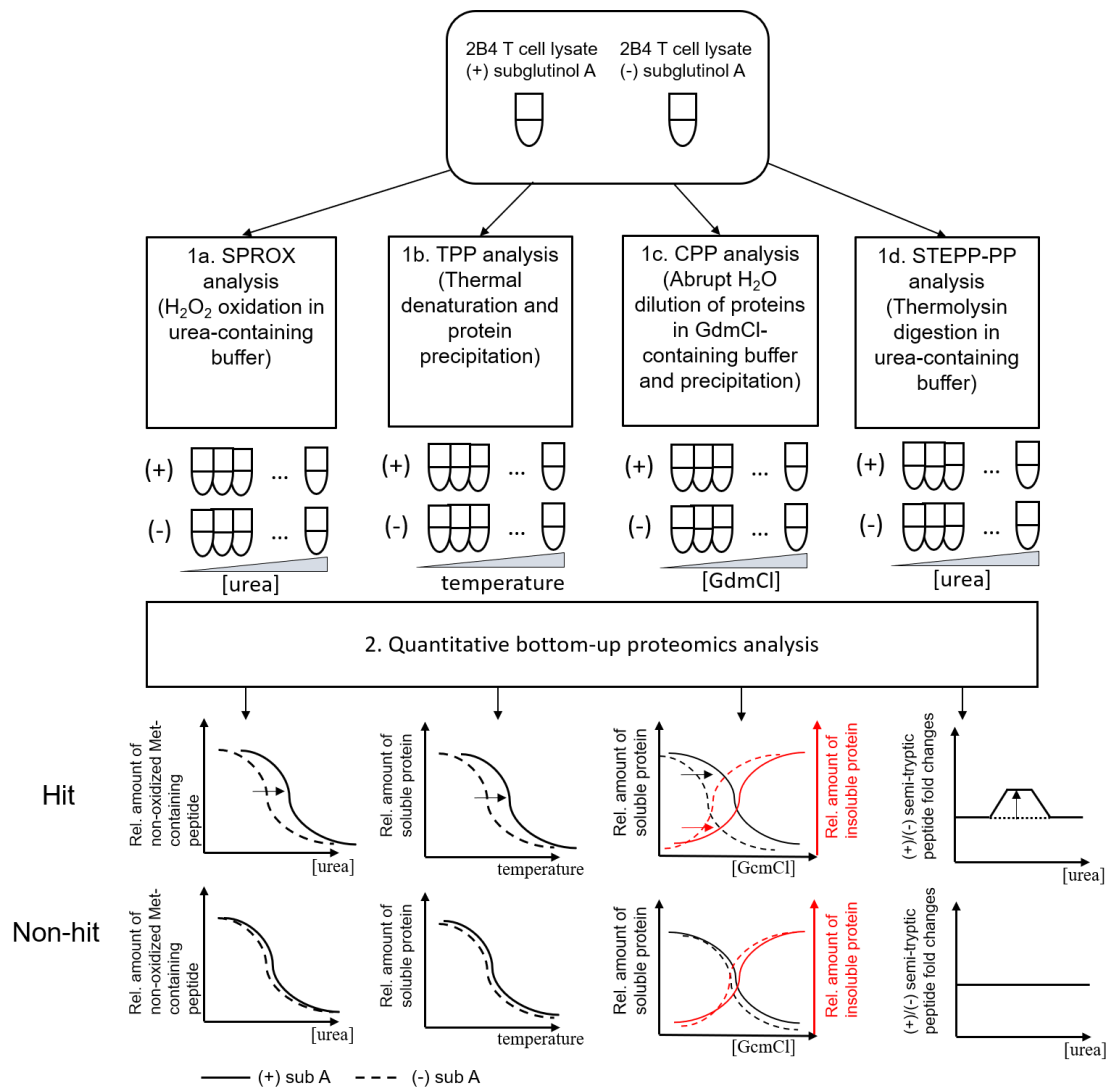


Figure 12 Schematic representation of the workflow of proteomics methods used in this work for identification of Sub A protein target.

The proteomic coverages observed in each experiment are summarized in Table 5. More than 2000 total unique proteins were assayed in the four sets of experiments. The proteomic coverages of the four techniques ranged between 800 and 1200 proteins per biological replicate. The TPP experiment enable the most coverage per biological replicate, with an average of about 1200 proteins assayed per biological replicate, about 20% higher than SPROX and STEPP-PP (about 1000 proteins per replicate). This is likely because the

protein-based TPP experiment does not need to involve complex peptide modification and enrichment strategy such as oxidation and methionine-containing peptide enrichment in SPROX or no-specific proteinase digestion and semi-tryptic peptide enrichment in STEPP-PP. The lower coverage, about 800 proteins per biological replicate, of the CPP experiment is due to the experimental design of CPP where only proteins identified in both supernatant and pellet in a biological replicate are considered as assayed proteins (Figure 12).

Table 5 Summary of proteomic coverages and hits observed in the Sub A binding SPROX, TPP, CPP and STEPP-PP experiments on proteins from 2B4 T cell lysate

Technique	Biological replicate	Number of assayed peptides (proteins)	Number of hit peptides (proteins)
SPROX	1	2049 (1080)	134 (126)
	2	2814 (1091)	123 (114)
	Total	3747 (1434)	201 (186)
TPP	1	(1369)	NA*
	2	(1424)	NA*
	3	(900)	NA*
	Total	(1621)	(35)
CPP	1	(811)	(1)
	2	(708)	(1)
	Total	(882)	(1)
STEPP-PP	1	2604 (918)	8 (8)
Total	/	(2324)	(227)

* For TPP experiment, according to the protocol described by Franken et al., hit selection is not conducted individually for each biological replicate on data of all biological replicates at once.⁸⁷ So, no hit peptide (protein) number is available for individual biological replicates.

In total 227 unique protein hits were identified using the four techniques. Four of these proteins, including deoxycytidine kinase (DCK), exportin 2 (XPO2), moesin (MSN)

and T-complex protein 1 subunit eta (CCT7), were identified as hits in two techniques (Table 6). No protein has been identified as hit in more than two techniques. Of the four overlapping protein hits, the DCK and XPO2 are particularly interesting because they have consistent behavior across techniques, i.e., DCK is stabilized with Sub A in both TPP and CPP, and XPO2 is destabilized in the same techniques (Figure 13 and Figure 14). It is also noteworthy that DCK is the only protein hit in the CPP experiment. Previous CPP experiments indicate that requiring a hit to appear both in the supernatant and the pellet analysis, which is the case for DCK here, brings the false positive rate to 0.⁹¹ This suggests the DCK protein is the most confident protein hit in the hit list. The other two proteins, MSN and CCT7, are hits in SPROX and STEPP-PP but did not show significant stability changes with Sub A in TPP and CPP experiments.

Table 6 The four overlapped protein hits and their behavior in the techniques applied here in this work

Overlapped hit protein	SPROX	TPP	CPP	STEPP-PP
Deoxycytidine kinase (DCK)	Not assayed	Hit (stabilization)	Hit (stabilization)	Not assayed
Exportin 2 (XPO2)	Hit (destabilization)	Hit (destabilization)	Not assayed	Not assayed
Moesin (MSN)	Hit (destabilization)	Not hit	Not hit	Hit (stabilization)
T-complex protein 1 subunit eta (CCT7)	Hit (stabilization)	Not hit	Not hit	Hit (stabilization)

The DCK protein controls the first and rate limiting step in deoxycytidine (dC) salvage. It phosphorylates dC in the presence of dUTP or dATP to form deoxycytidine monophosphate (dCMP), which is further phosphorylated by other kinases to form the DNA building block dCTP.^{92,93} Although dCTP can also be synthesized from the *de novo* pathway, the salvaging pathway limited by DCK activity is important for maintaining the cell dCTP pool especially for T cell development and activation.^{93,94} Both CPP and TPP experiment indicate that the DCK protein is stabilized with Sub A, which is expected for a direct ligand binding. This finding provides one hypothesis of Sub A mode-of-action -- Sub A directly binds to DCK protein and reduced its activity, leading to a reduced dCTP level that blocks DNA synthesis and the fast proliferation of activated T cells. To validate this hypothesis, a gel-based pulse proteolysis experiment and two DCK activity assays are conducted. The results of these validation experiments are described in the following sections.

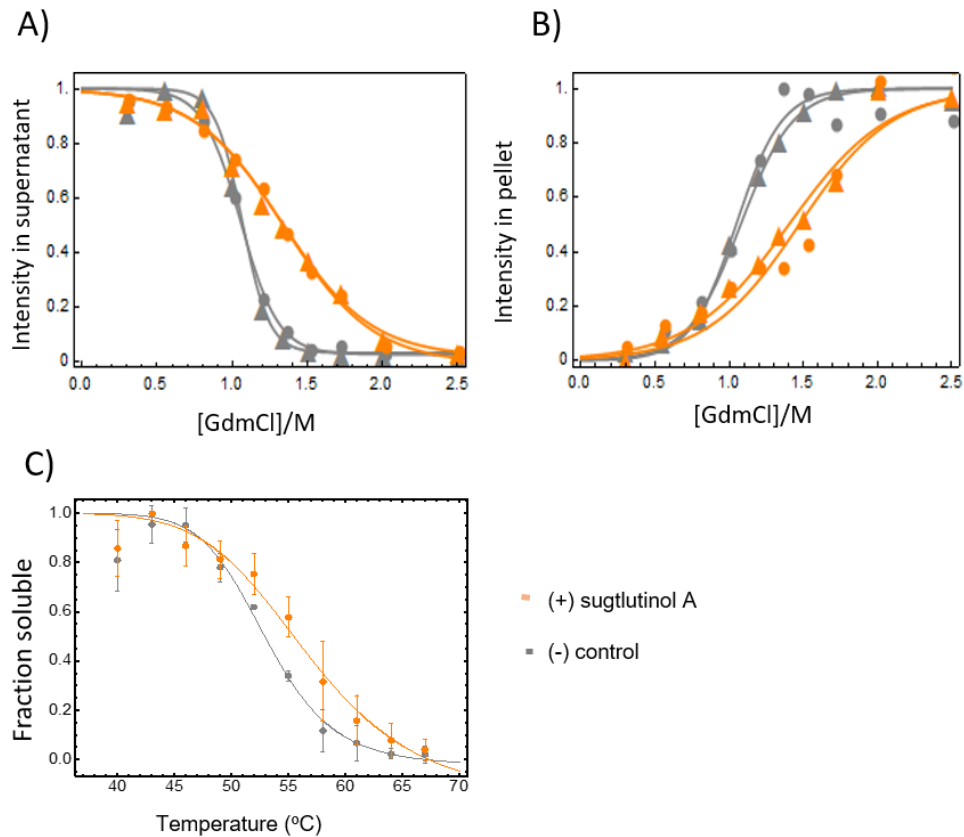


Figure 13 Denaturation curves of DCK obtained from CPP and TPP experiments. A) Denaturation curves acquired for DCK with Sub A (orange) and without Sub A (gray) using quantification on supernatant of the two CPP biological replicates (represented with triangles and disks, respectively). B) Denaturation curves acquired for DCK with Sub A (orange) and without Sub A (gray) using quantification pellet of the two CPP biological replicates (represented with triangles and disks, respectively). The solid lines represent the best fit curve for data from each CPP experiment. C) Denaturation curves acquired for DCK with Sub A (orange points) and without Sub A (gray points). The solid lines represent the best fit curve for the averaged data from 3 biological replicates of the ligand and control TPP experiment.

The other most confident protein hit, XPO2, is involved in transportation of a specific protein, importin- α , from cell nucleus to cytoplasm. It is unclear how Sub A's mode of action could be involved with the function of XPO2. As both SRPOX and TPP experiments indicate the XPO2 protein is destabilized with Sub A, it is not that likely the XPO2 is a direct subglutininol A target comparing to DCK.

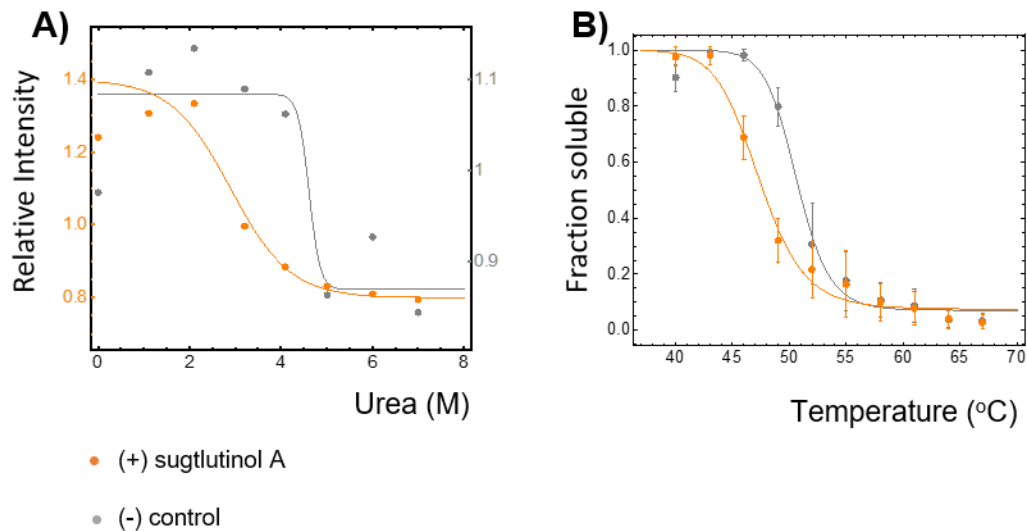


Figure 14 Denaturation curves of XPO2 obtained from SPROX and TPP experiments. **A)** Denaturation curves acquired for XPO2 with Sub A (orange points) and without Sub A (gray points). The solid lines represent the best fit curve for the ligand and control SPROX experiment biological replicate 2. **B)** Denaturation curves acquired for XPO2 with Sub A (orange points) and without Sub A (gray points). The solid lines represent the best fit curve for the averaged data from 3 biological replicates of the ligand and control TPP experiment.

3.3.2 Biochemical Validation of Sub A Binding to DCK

Further validation of the binding of Sub A to DCK was performed with a gel-based targeted pulse proteolysis experiment of a 2B4 T cell lysate. The amount of intact DCK protein after the proteolysis digestion was readout by immunoblot. The workflow of this experiment is shown in Figure 15.

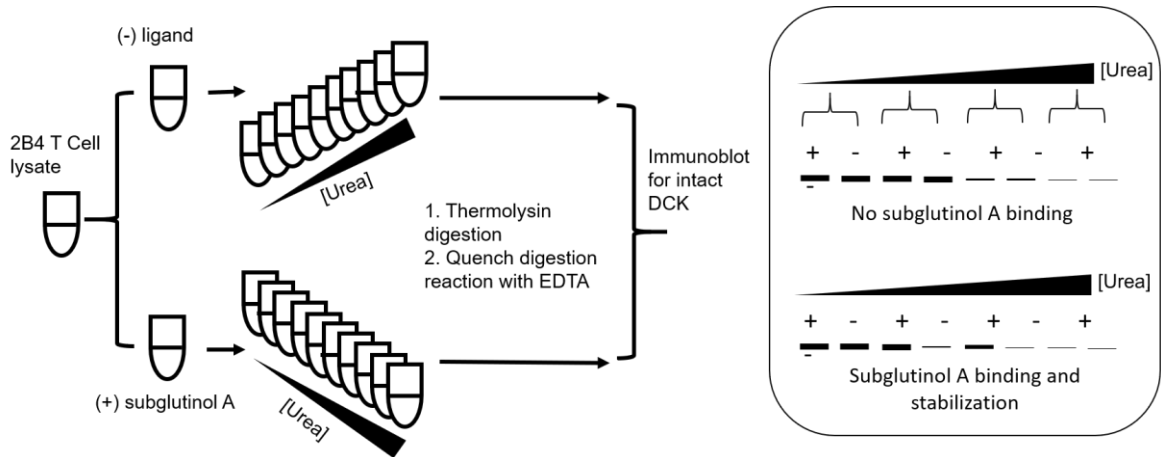


Figure 15 Pulse proteolysis with immunoblot readout workflow for target validation.

The final pulse proteolysis results were shown in Figure 16. With 100 μM of Sub A, DCK was clearly stabilized between 1.5 to 2.75 M urea, indicating there was indeed a binding interaction between DCK and Sub A. The Sub A-DCK binding K_d can also be estimated to be 0.21 μM based on the interval of urea concentration and the Δ value calculated from DCK amino acid sequence.⁵⁶ This value is in parallel (i.e., within 3-fold difference) with the cell-based IC_{50} (i.e., 0.1 μM) and the K_d calculated from CPP experiment (i.e., 0.27 μM).⁸²

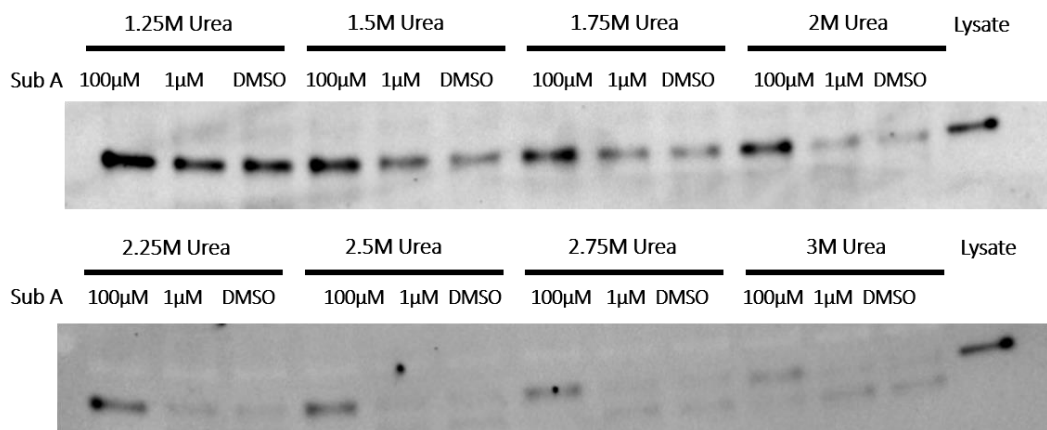


Figure 16: Immunoblot readout of pulse proteolysis on Sub A and DCK binding.

To confirm that the stabilization was a result of a direct DCK-Sub A binding rather than an indirect binding between DCK and a Sub A-protein complex, a low concentration of Sub A was used to study the Sub A concentration dependence of the stabilization. In a direct binding event, assume the K_d is similar to the IC_{50} (i.e., 0.1 μM) of Sub A, the stabilization effect of 1 μM Sub A will be hardly observable and the proteolysis digestion pattern of DCK will be similar to the control experiment. In an indirect binding event, 1 μM Sub A, which should still be in large excess to the other protein (i.e., (direct binding target), will still induce an observable stabilization similar to 100 μM because the limiting factor is the concentration of the other protein. The result indicated in Figure 16 clearly shows the pulse proteolysis pattern with 1 μM Sub A resembles the pattern with DMSO vehicle, indicating there is a direct binding between Sub A and DCK.

3.3.3 Evaluation of DCK Activity with Sub A Treatment

To check if DCK's activity is altered by the presence of Sub A, two DCK activity assays have been conducted in this work, including an assay of purified DCK protein with a commercially available assay kit and a radiochemical assay of endogenous DCK activity in lysate.

The purified DCK activity assay couples the DCK phosphorylation reaction with another rapid enzymatic reaction to form NADH_2 , which can be monitored by a UV-vis spectrometer. If the activity of DCK is inhibited by the ligand added into the reaction mixture, the speed of NADH_2 monitored at 340 nm will be different to the control experiment without ligand. The final purified DCK kinase assay results presented in Figure 17 shows no difference in DCK activity when incubated with 1 nM, 100 nM or 1

μM Sub A. The activity results are also similar to the results of a control experiment with DMSO vehicle. Together these data suggest that Sub A does not alter DCK's activity.

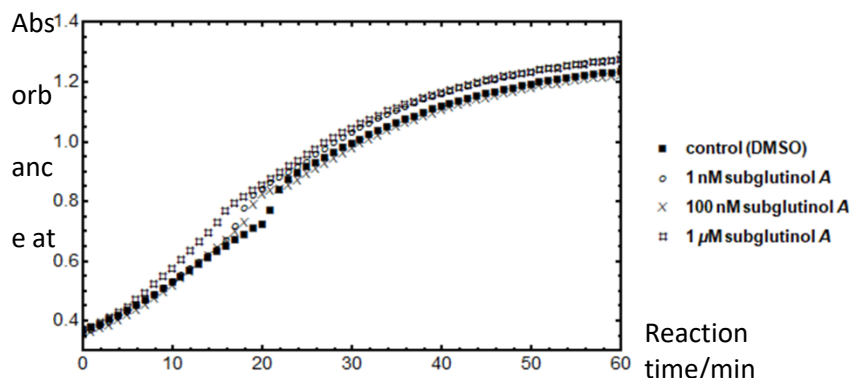


Figure 17 Purified DCK kinase activity assay result.

To rule out the possibility that the purified recombinant DCK does not behave differently than the endogenous DCK protein in a cell lysate, another radiochemical DCK activity assay was performed. The radiochemical assay was performed according to a previously established protocol and the workflow is summarized in Figure 18.⁸⁸

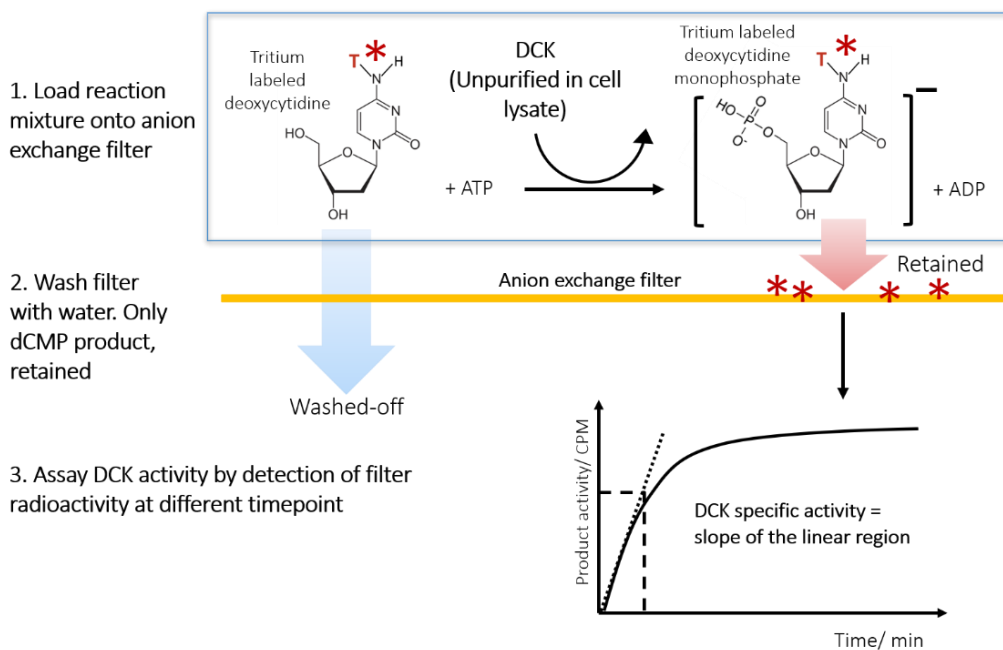


Figure 18 Radiochemical DCK activity assay workflow.

In the case of a ligand binding-induced inhibition of the DCK kinase activity, the formation rate of the product dCMP monitored by scintillation will be different from that in the control experiment without the ligand. Again, the radiochemical DCK kinase assay results presented in Figure 18 showed no difference in DCK activity with up to 10 μ M Sub A. The results obtained were indeed consistent with those obtained using the DMSO vehicle. The results obtained using both DCK kinase assays described above suggest Sub A does not directly inhibit DCK's kinase activity.

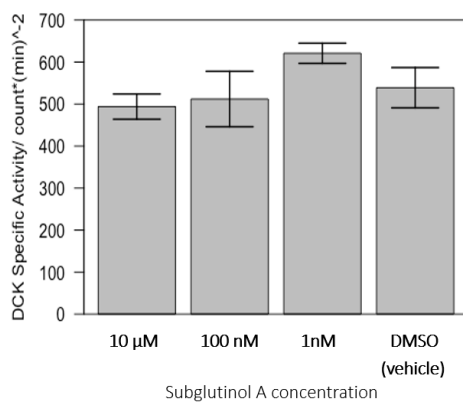


Figure 19 Radiochemical DCK activity assay results.

3.3.4 Proteome-wide Evaluation of Protein Expression Level and Phosphorylation Level Changes in Sub A-Treated Cells

Research on DCK's post-translational modification has shown that a specific phosphorylation at serine 74 enables a ten-fold increase in the DCK enzymatic rate on dC.⁹⁵ The binding of Sub A to DCK could block or allosterically changes the conformation of this important phosphorylation site, leading to a decreased enzymatic activity of DCK. To test if the phosphorylation of serine 74 is altered with Sub A, a phosphoproteomic

study is performed on 2B4 T cells cultured with and without Sub A (Figure 20). The phosphoproteomic study conducted here not only provided information on DCK phosphorylation but also information about phosphorylation level changes on other proteins as well. To normalize the protein expression level changes that can contribute to the observed phosphorylation level alternation, a mass spectrometry-based protein expression level study is also performed on the same cell cultures (Figure 20). Like the phosphoproteomics study, this proteome-wide expression level study generated information on other protein expression level changes in subglutunol A treated cells.

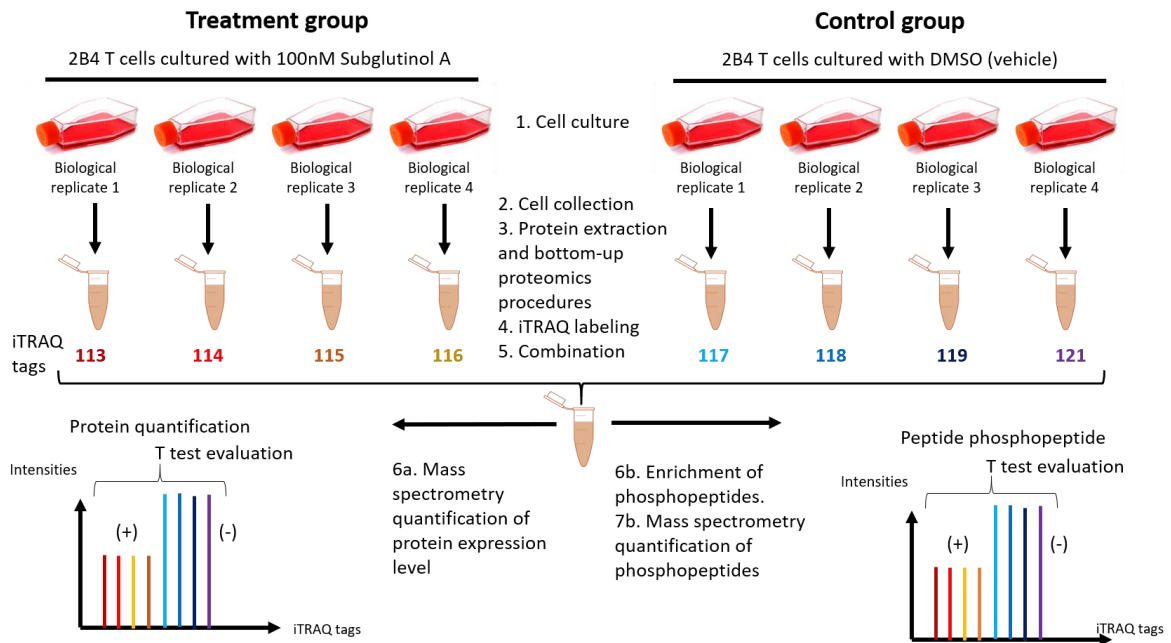


Figure 20 Workflows of expression proteomics and phosphoproteomics experiments performed in this study.

The results in Figure 21 suggest that neither the DCK expression level or the phosphorylation of S74 are significantly altered with addition of Sub A. None of the changes are of both $p < 0.01$ and fold change $> 3\sigma$ deviation from global mean value. Also, for either expression level or phosphorylation level, the fold changes are not consistent in

different time points, i.e., the expression and phosphorylation of DCK are up in 24h but down in 48h with Sub A.

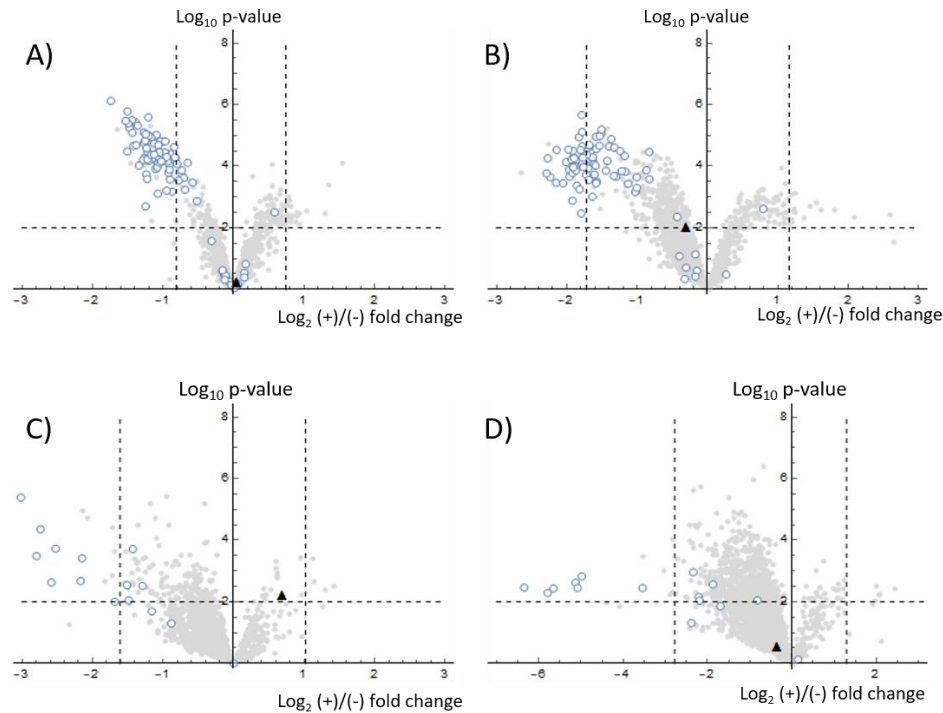


Figure 21 Volcano plots showing the results of expression proteomics and phosphoproteomics experiments in this work. A) and B) are results of protein expression level changes with 100nM Sub A for 24 and 48 hours (respectively). C) and D) are results of changes of phosphopeptide amounts with 100nM Sub A for 24 and 48 hours (respectively). The phosphopeptide amounts are normalized for parent protein expression level changes. The vertical dashed lines represent 3 sd from the global average $\log_2 (+)/(-)$ fold changes. The horizontal dashes lines represent $p = 0.01$. The dark triangles in A) and B) represent the DCK protein, and in C) and D) represent DCK S74 phosphorylation indicated by amounts of phosphopeptide WCNVQSTQEEFEELTTSQK. The open circles in A) and B) represent ribosomal proteins, and in C) and D) represent phosphopeptides of 60S acidic ribosomal protein R0, R1 and R2.

While expression and phosphorylation of DCK is not significantly altered, the results suggest ribosomal proteins could play an important role in Sub A mode of action. The ribosomal proteins make up the majority of proteins that are consistently down-expressed in both 24h and 48h Sub A treated cells (Figure 21 A and B). Interestingly, the

60S acidic ribosomal P proteins, which is three highly phosphorylated proteins including RPL0, RPL1 and RPL2, are also consistently down-phosphorylated in both time points (Figure 21 A and B). The phosphorylation of many of these phosphopeptides are lower by more than 10 folds in 48h of Sub A treatment, after normalization of protein expression level changes. Previous research has shown the phosphorylation is important for the 60S acidic ribosomal P proteins to associate and form the “stalk” of ribosomes.^{96,97} The reduction of phosphorylation might be related with the down-regulation of ribosomal protein expression. However, it is still unclear how Sub A induces these changes.

3.3.5 Identification of DCK Nuclear transportation Blocked by Sub A Treatment

The DCK protein is normally localized in the cytoplasm, but when overexpressed, it is found predominantly in the nucleus.⁹⁸ The flexibility in DCK localization might be related to the functional compartmentalization of the dNTP pools and the regulation of the dNTP metabolism including DNA synthesis.⁹⁸ To detect if the localization of DCK is altered with the ligand, a gel-based DCK localization experiment was performed on cells treated with Sub A for 24h and 48h. The workflow of the experiment is outlined in Figure 22.

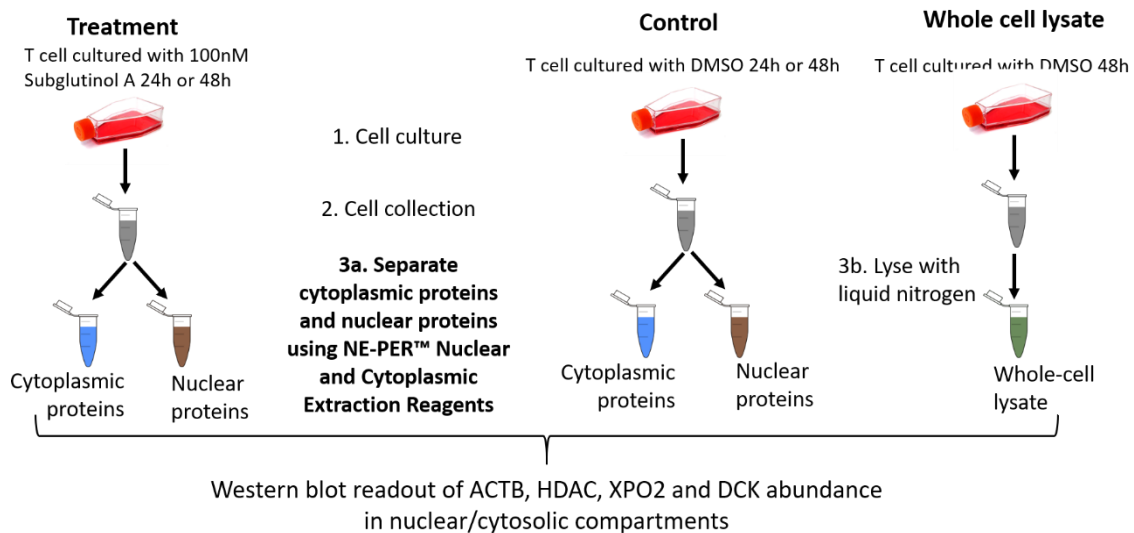


Figure 22 Workflow of DCK localization study with western blot readout.

Shown in Figure 23 is the final western blot readout obtained using the experiment workflow shown in Figure 22. Consistent with findings from previous research, DCK is primarily localized in the cytoplasm in both control and treated cells.⁹⁸ Interestingly, while the cytosolic amount of DCK is not significantly influenced by the ligand (i.e., within 1.3-fold change), the DCK protein in the nuclear fraction is reduced by about 4-fold in both 24h and 48h Sub A treatments, after normalizing to a loading control (i.e., HDAC protein). The results suggest the in-cell transportation of DCK to the nuclei is blocked by Sub A, although the exact mechanism controlling this transportation is still unknown.

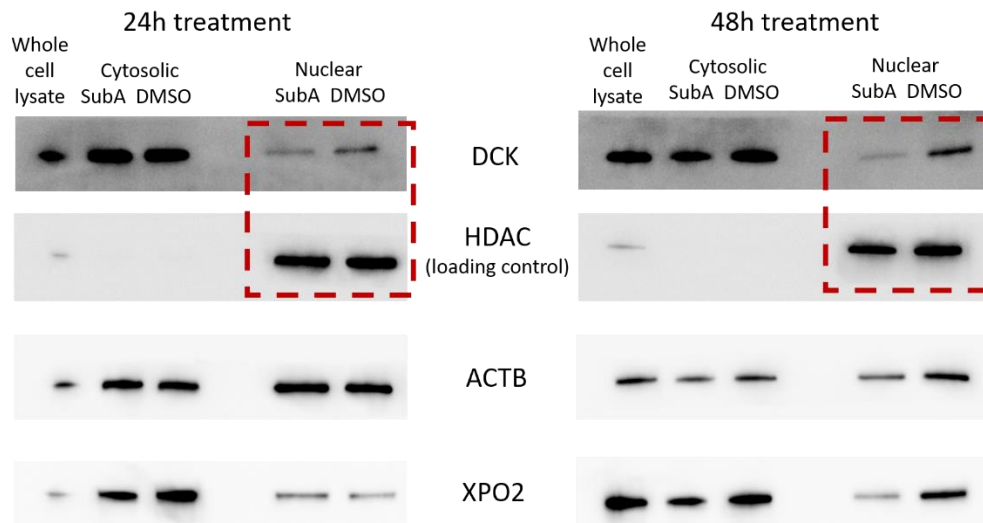


Figure 23 Western plot readout of cytosolic and nuclear DCK level in cells treated with and with Sub A for 24 or 48 hours. Histone deacetylase (HDAC) is a nuclear protein that serves as a loading control. Results of β -actin (ACTB) and exportin-2 (XPO2) are also shown here.

3.3.6 Identification of Reduced dCMP Level in Sub A Treated Cell Nuclei

A LC-MS/MS-based quantification of nuclear dCMP levels was performed on Sub A-treated and control cells to determine if the reduction of nuclear DCK level result in reduced nuclear dCMP concentration. The workflow is outlined in Figure 24. The LC-MS/MS instrumental quantification of dCMP based on a multiple reaction monitoring (MRM) was developed from a previous protocol on dNTP analysis.⁹⁹

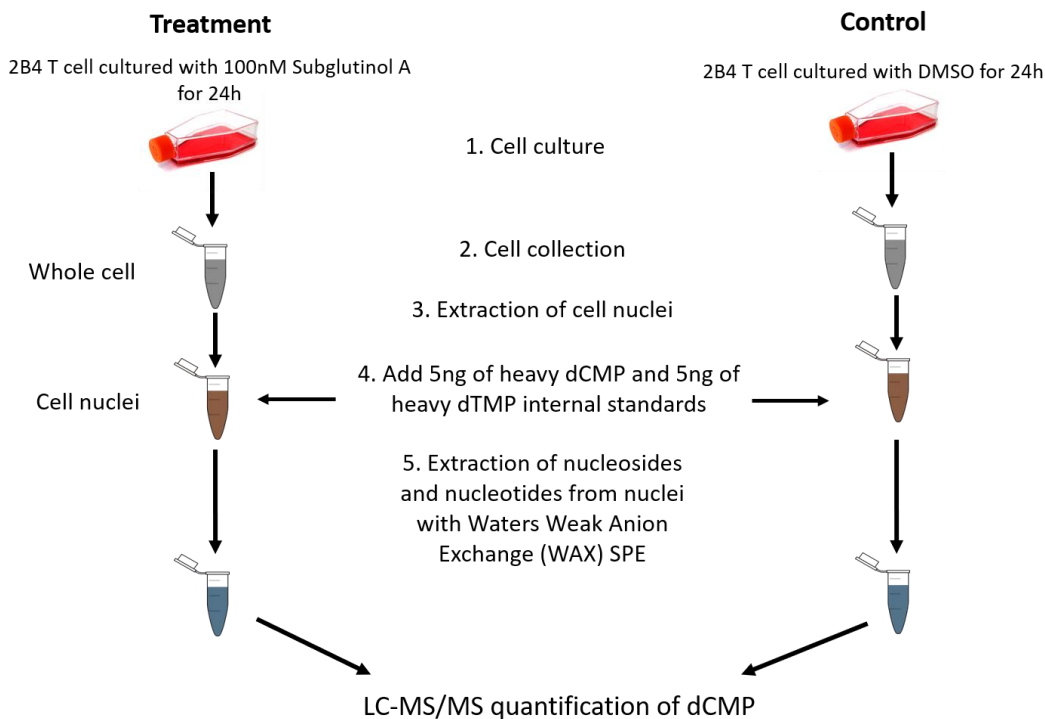


Figure 24 Workflow of nuclear dCMP quantification experiments used in this study.

The quantification results are shown in Figure 25. The LC-MS/MS-based method enables a good quantification ($R^2 = 0.999$) of dCMP extracted from cell nuclei. The nuclear dCMP amount is observed to be reduced by more than 5-fold with 24h of 100 nM Sub A treatment. It is likely that the reduction of nuclear DCK level limits the nuclear salvaging of dC to produce dCMP. However, further study is needed to confirm the lowering of dCMP level is not a general outcome of the cell death brought about by drug treatment. For example, it is possible that Sub A induces cell death by some mechanism unrelated to DCK binding, and this process simply lower every dNMP pool in the cell nucleus.

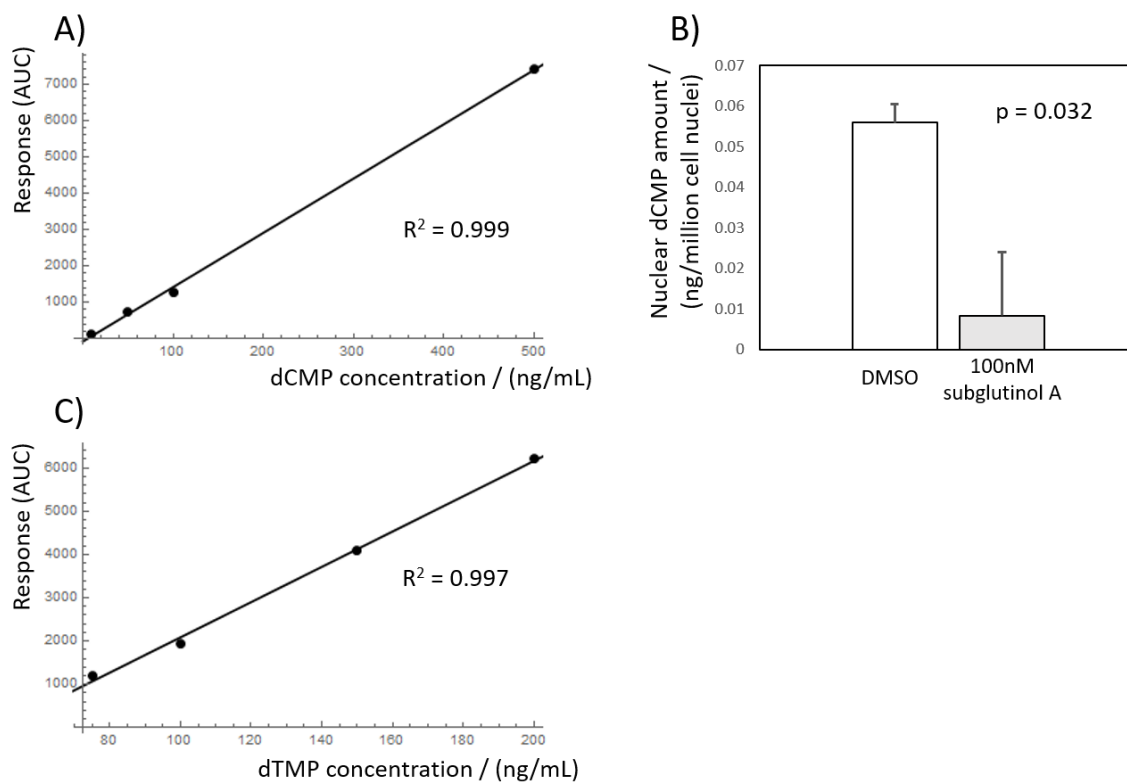


Figure 25 Results of the LC-MS/MS-based dCMP and dTMP quantification in this study. A) shows the calibration curve of dCMP standard solutions; B) shows the calculated nuclear dCMP amount from two replicates of control and treated cells. C) shows the the calibration curve of dTMP standard solutions.

To rule out this possibility, quantification of nuclear dTMP can be performed. Because dTMP is not a product of DCK, its concentration will not be influenced by DCK but general cell death. If the nuclear dTMP level does not drop with Sub A incubation, it indicates the dCMP level reduction is indeed an outcome of blocked DCK nuclear translocation. A LC-MS/MS-based dTMP quantification method was developed by Ryan Fader from the Fitzgerald Laboratory at Duke University. The method was adopted from the dCMP LC-MS/MS quantification but modified specifically for better detection and quantification of dTMP at low concentrations (i.e., lower than 200 ng/mL). The calibration curves of this method using dTMP standard solution is shown in Figure 25. Further

quantification of nuclear dTMP level using this technique in Sub A-treated cells is planned. It is expected to contribute to the understanding of the effect of Sub A-DCK binding.

3.4 Conclusion

In this study, four different mass spectrometry-based proteomics methods have been used to discover the protein binding target of Sub A. More than 2000 proteins have been assayed from these methods. Four overlapping protein hits have been identified but only DCK show consistent stabilization behavior with Sub A in these methods. Further gel-based targeted pulse proteolysis method supports that DCK is a direct binding target of Sub A.

While how DCK-Sub A binding leads to the observed inhibition of T cell proliferation, survival, and cytokine production still largely unknown, this study has demonstrated that Sub A does not alter DCK's activity or phosphorylation level but blocks its nuclear transportation. This reduction of DCK level in the cell nucleus possibly leads to a reduced nuclear dCMP pool. It is necessary to determine if the observed reduction in nuclear dCMP pool is indeed related to DCK or just a phenomenon of cell dying.

A follow-up study on quantification of nuclear dTMP level is in progress. Unlike other dNMPs, dTMP is not a product of DCK phosphorylation.¹⁰⁰ If the dTMP level is stable in Sub A treated cells, it can be deduced that the reduced amount of nuclear DCK level limits the amount of nuclear dCMP. If this is the case, some further studies on DCK nuclear transportation and the nuclear dNMP/dNTP pool are needed before fully understand the mode of action of Sub A. It is still not clear how the DCK localization is

determined by the cell or how the drainage of the dNMP pool could arrest the T cells from proliferation, survival and cytokine production. Perhaps the DCK protein and the underlining dN salvage pathway is uniquely important to the T cells and thus the Sub A's targeting of DCK contributes to its low cytotoxicity.

The expression proteomics and phosphoproteomics experiments show ribosome protein level and phosphorylation of acid ribosomal P proteins are significantly reduced. It is still unclear where this piece of observation fits into the picture of subglutinin A's mode of action.

4. Search for Protein Targets of Manassantin A

The work described in this Chapter was performed in collaboration with Professor Jiyong Hong and his group at Duke University.

4.1 Introduction

Manassantin A (Man A, Figure 26) is a natural dineolignan isolated from an aquatic plant *Saururus cernuus*, which has been traditionally used in treatment of edema, gonorrhoea and jaundice in Asia countries.^{101,102} First isolated in the 1980s, Man A has been identified as an attractive anti-cancer drug candidate.^{102,103} It exhibits cytotoxicity against a wide range of cancer cell lines, but exhibits weak cytotoxicity against noncancer cell lines.^{102,104,105}

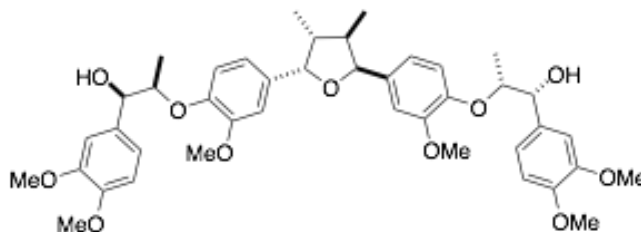


Figure 26 Structure of manassantin A.

While biological activities of Man A have been documented, the mode of action of Man A is still largely unclear. Previous research has indicated Man A as a potent hypoxia inducible factor 1 α (HIF-1 α) inhibitor with an unclear mechanism.¹⁰¹ HIF-1 α is a subunit of transcription factor HIF-1, which is generated by cells to cope with the stress of hypoxia.¹⁰⁶ In normoxia (well-oxygenated environment), the constitutively synthesized HIF-1 α is maintained at a very low level by continuous degradation.¹⁰⁷ In hypoxia (low oxygen level), HIF-1 α binds to HIF-1 β in the nucleus to form HIF-1. The activation of

transcription factor HIF-1 drives the expression of hundreds of genes involved in important biological processes. HIF-1 is particularly important in tumor cell growth. With mediation of HIF-1, the tumor cells adapt to hypoxia condition and become aggressive and resistant to chemotherapy and radiation.¹⁰⁸

Understanding the how Man A inhibits HIF-1 α could be very important in delineation of the hypoxia signaling regulation in tumor progression. Previously, Wallace et al. utilized the iTRAQ-SPROX and SILAC-PP techniques to uncover the protein targets of Man A.¹⁰⁹ This previous work assayed about 1100 proteins and discovery 23 protein hits. However, most of the hit proteins are likely false positives. It is still challenging to identify the true binding protein from the list.

Recently, several additional energetics-based proteomics strategies have been developed for drug protein target discovery. As part of the work in this chapter, four novel approaches, including the STEPP-PP, STEPP-LiP, one-pot STEPP-PP and one-pot TPP approaches, are used to discover the protein targets of Man A in a hypoxia-treated HEK293T cell lysate. The goal of this work was to assay more proteins in the proteome than was possible using the iTRAQ-SPROX and SILAC-PP techniques and to identify more confident protein hits that show Man A-induced stability changes in multiple methods (i.e., "true positives"). Such "true positives" are expected to guide future research about hypoxia signaling regulation.

4.2 Experimental Section

4.2.1 Materials

The following materials were from Sigma-Aldrich (St. Louis, MO): S-methylmethanesulfonate (MMTS), dimethyl sulfoxide (DMSO), urea, centrifugal filter units 10k (Amicon Ultra-0.5), tris(hydroxymethyl)-aminomethane hydrochloride (Tris·HCl), thermolysin from *Geobacillus stearothermophilus*, proteinase K from *Tritirachium album* and phenylmethylsulfonyl fluoride (PMSF). The following materials were from Thermo Scientific (Waltham, MA): N-hydroxysuccinimide(NHS)-activated agarose dry resin (Pierce), 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF), bestatin, E-64, leupeptin, pepstatin A, and trypsin protease MS grade. Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) was from Santa Cruz Biotechnology (Dallas, TX). MacroSpin columns (Silica C18) were from Nest Group (Southborough, MA). Phosphate-buffered saline (PBS, pH 7.4, 1x) was from Gibco (Gaithersburg, MD). Ethylenediaminetetraacetic acid (EDTA, 0.5 M, pH = 8.0) was from Mediatech, inc (Manassas, VA). Manassantin A was synthesized and provided by professor Jiyong Hong at Duke University.

4.2.2 HEK293T Cell Culture and Lysis

The HEK293T cells from ATCC were cultured in a humidified 37 °C incubator with 95% air and 5% CO₂ according to ATCC guideline before moved to a hypoxia chamber with 1% O₂, 5% CO₂, and balanced with N₂. After 24h under the hypoxia condition, cells were harvested on ice and were immediately frozen with dry ice. The harvested cells were lysed in PBS with protease inhibitor cocktail (1 mM AEBSF, 10 μM pepstatin A, 20 μM

leupeptin, 15 μ M E-64, and 500 μ M bestatin). The cell lysis was facilitated by four rounds of liquid nitrogen freeze-thaw cycles. The cell lysates were centrifuged at 14 000 \times g and 4 $^{\circ}$ C for 15 min, and the supernatants were collected. The total protein concentration in the lysates were determined using a Bradford assay and adjusted to 5mg/mL with PBS containing the same concentration of protease inhibitor cocktail as mentioned above.

4.2.3 STEPP-LiP Analysis

Five biological replicates of the hypoxia-treated HEK293T lysates were each divided into two equal portions. One portion was spiked with a solution of the Man A prepared in DMSO and one was spiked with DMSO as vehicle. The solutions were equilibrated for 1 hour before analysis. An aliquot of 10 μ L from each of the 5 (+) ligand and 5 (-) ligand cell lysates was each combined with 30 μ L of PBS. The final Man A concentration in each lysate was 100 μ M. The 10 resulting samples were treated with proteinase K at an enzyme/substrate ratio of 1/100 (w/w) for 5 min at room temperature. The proteinase K proteolysis reaction was quenched with 5 mM PMSF and the samples were subjected to the same STEPP protocol described in chapter 2 and previously reported.² Briefly, the proteinase K-digested samples were TCEP reduced, modified with MMTS, labeled with TMT10-plex, combined, digested with trypsin and chemo-selected for semi-tryptic peptides with an NHS-activated agarose resin. The resulting samples were desalted using C18 columns and subject to the LC-MS/MS analysis.

4.2.4 STEPP-PP and one-pot STEPP-PP Analyses

The STEPP-PP analysis applied in this work was identical to that described in chapter 2 and reported previously.² Briefly, one tube of the hypoxia-treated HEK293T

lysate sample was divided into two equal portion, one was spiked with a solution of the Man A prepared in DMSO and one was spiked with DMSO as vehicle. The (+) and (-) ligand solutions were each distributed into a series of 30 μ L urea-containing buffers (PBS buffer, pH 7.4) where the final urea concentration ranged from 0 to 6.8 M. The final concentration of Man A was 100 μ M. The samples in the urea-containing buffers were equilibrated for 2 h at room temperature before subject to thermolysin pulse proteolysis reaction according to those previously described.^{23,24} The proteolyzed samples were subjected to the identical STEPP protocol described in chapter 2 and the resulting samples were subject to LC-MS/MS analysis.

The one-pot STEPP-PP analysis applied in this work was identical to that described in previous work by Cabrera and co-workers, which involved the combination of a normal STEPP-PP workflow with a one-pot strategy.³ Five biological replicates of the hypoxia-treated HEK293T lysates were each divided into two equal portions, one spiked with Man A solution and one with DMSO as vehicle. Five (+) and (-) ligand pairs of samples were subjected to a pulse proteolysis procedure identical to that mentioned in the above paragraph. The proteolyzed samples originally from the same (+) or (-) ligand sample were combined and subjected to the STEPP protocol and LC-MS/MS analysis.

4.2.5 one-pot TPP and one-pot SPROX Analyses

The one-pot TPP and one-pot SPROX protocols were performed identically to that described Cabrera and co-workers.³ Like the one-pot STEPP-PP protocol, the one-pot TPP and one-pot SPROX protocols were a combination of a one-pot strategy of sample combination and the original TPP and SPROX workflows, respectively.

In the one-pot TPP analysis, five pairs of (+) and (-) ligand samples were subjected to TPP procedures similar to that described in chapter 3. The total protein concentration of (+) and (-) ligand cell lysate samples were diluted to 1 mg/ml with PBS pH 7.4 solution. The final concentration of Man A was 100 μ M. The solutions were equilibrated at room temperature for 1 hour before twelve aliquots of 100 μ L of each (+) and (-) ligand sample were taken out and incubated at twelve different temperatures from 37 to 70 $^{\circ}$ C with a thermocycler. After 3 min of incubation, the samples were cooled down at room temperature for 3 min before further cooling down on ice. The cooled samples originally from the same (+) or (-) ligand sample were combined and ultracentrifuged at 100 000 \times g and 4 $^{\circ}$ C for 20 min. An aliquot of 300 μ L of supernatant from each of the ten centrifuged samples was transferred into a 10k centrifugal filter unit and subjected to a FASP protocol, which involved reduction with TCEP, modification with MMTS, digestion with trypsin and labeling with a TMT-10plex reagent kit according to the manufacturer's protocol. The TMT-labeled samples were combined, C18-cleaned and subject to LC-MS/MS analysis.

In the one-pot SPROX analysis, five pairs of (+) and (-) ligand samples were subjected to SPROX procedures similar to that described in chapter 3. The (+) and (-) ligand cell lysate samples each were distributed into a series of GdmCl-containing buffers (PBS pH 7.4) with the final urea concentrations ranging from 0 to 3 M. The final volume in each buffer was 50 μ L. The final concentration of Man A was 100 μ M. The solutions were equilibrated at room temperature for 1 hour before 5 μ L of 30% (v/v) hydrogen peroxide solution was added into each solution to initiate methionine oxidation. After 3 min, the oxidation reactions were quenched by addition of 500 μ L of 500 mM TCEP

solution. An aliquot of 80 μL from solutions originally from the same (+) or (-) ligand sample were combined and transferred into a 10k centrifugal filter unit before subjected to FASP protocol that involved reduction with TCEP, reaction with MMTS, digestion with trypsin, and labeling with a TMT-10plex reagent. The TMT-labeled samples were combined, C18-cleaned enriched for methionine residue containing peptides using a Pi3TM Methionine Reagent kit according to the manufacturer instructions (The Nest Group). The resulting sample was C18-cleaned and subject to LC-MS/MS analysis.

4.2.6 Quantitative LC-MS/MS Analysis for Proteomics Samples

LC-MS/MS analyses were performed on a Q-Exactive HF high-resolution mass spectrometer (Thermo) with a nano-Acquity UPLC system (Waters) and a nano-electrospray ionization source fitted with a SilicaTip emitter (New Objective). Samples were trapped on a 2D Symmetry C18 trapping column with dimensions 180 μm x 20 mm and particle diameter of 5 μm , pore size 100 Å. The trapping time was 5 minutes at 5 $\mu\text{L}/\text{minute}$ (99.9:0.1 v/v water/ACN 0.1% formic acid). The samples were separated on a 75 μm x 250 mm high strength silica (HSS) T3 column with 1.8 μm particle diameter (Waters) heated to 55 °C. Peptides were eluted using a gradient of 3-30% acetonitrile with 0.1% formic acid over 90 minutes at a flow rate of 0.3 $\mu\text{L}/\text{min}$. LC-MS/MS data were collected using a top 20 data-dependent acquisition (DDA) method including MS1 at 120k and MS2 at 30k resolution. The AGC target for MS1 was 3×10^6 ions with a max IT of 50 msec. The AGC target for MS2 was 1×10^5 ions with a max IT of 45 msec. The normalized collision energy (NCE) was set to 30 V and the scan range was 375-1600 m/z. The isolation window was 0.7 m/z and the dynamic exclusion time was set to 20.0 seconds.

The raw LC-MS/MS data was searched against the human proteome using Proteome Discoverer 2.2. The STEPP-PP, STEPP-LiP, one-pot STEPP-PP data searches were performed with fixed MMTS modification on cysteine, fixed TMT-10plex labeling on peptide N-termini, variable TMT-10plex labeling of lysine side chains, variable oxidation of methionine, variable deamidation of asparagine and glutamine, and variable acetylation of the protein N-termini. The precursor mass tolerance was set at 10 ppm. The fragment mass tolerance was set at 0.02 Da. Trypsin(semi) was set as the enzyme, and up to three missed cleavages were allowed. The one-pot TPP data search was performed with the same searching parameters except for 1) fixed instead of variable TMT-10plex labeling of lysine side chains; 2) trypsin instead of trypsin(semi) was set as the enzyme; and 3) up to two missed cleavages were allowed instead of three.

4.2.7 Proteomic Data Analysis

The STEPP-PP data analysis and hit selection was accomplished similar to previous described in section 2.2.6 of this thesis. Each assayed protein was required to be identified with at least one semi-tryptic peptide and a hit protein was required to have at least one semi-tryptic peptide hit. Hit peptides were identified with two criteria: 1) the peptide must have a significantly altered $\log_2(\text{fold-change})$ ($>3\sigma$ deviations from the mean $\log_2(\text{fold-change})$), for at least two consecutive urea concentrations; 2) these significantly altered $\log_2(\text{fold-change})$ values must have a fold-change in the same direction for at least two consecutive urea concentrations.

The STEPP-LiP, one-pot STEPP-PP and one-pot TPP data analysis and hit selection was accomplished similar to previously described by Cabrera and co-workers.³ A super

intensity was generated for each TMT channel by summing all semi-tryptic peptide signal intensities (for STEPP-LiP and one-pot STEPP-PP experiments) or protein signal intensities (for one-pot TPP experiment) of that channel. Each super intensity was divided by the mean of the ten super intensities to generate a normalization factor for the corresponding TMT channel. For each semi-tryptic peptide in the STEPP-LiP and one-pot STEPP-PP experiments or each protein in the one-pot TPP experiment, the raw TMT intensities were normalized by the corresponding normalization factors to adjust for the difference in total protein amounts in the original samples. For each biological replicate, the normalized (+) ligand intensity was divided by the (-) ligand intensity and \log_2 transformed to generate a \log_2 (fold change) value. The group of five resulting \log_2 (fold change) values were compared to a mean of 0 using a two-tailed t test. Hit peptides in STEPP-LiP and one-pot STEPP-PP experiments and hit proteins in the one-pot TPP experiment were identified with two criteria: 1) t test p-value was less than 0.001; 2) the average \log_2 (fold change) of the peptide or protein was at least 3σ away from the global average \log_2 (fold change).

4.3 Results and Discussion

4.3.1 Identification of Manassantin A Protein Target

Five mass spectrometry-based proteomics experiments, including STEPP-PP, STEPP-LiP, one-pot TPP, one-pot SPROX and one-pot STEPP-PP were employed to identify the protein targets of Man A in a HEK293T cell lysate. The HEK293T cells were incubated under hypoxia (1% O₂) condition for 24 h before they were harvested in order to identify interacting proteins expressed under hypoxia. The workflows of the

experiments here are adopted from previous reports and outlined in Figure 27.^{3,110} The one-pot strategy used in one-pot STEPP-PP, one-pot TPP and one-pot SPROX experiments here enables a streamlined statistical analysis for these experiments.

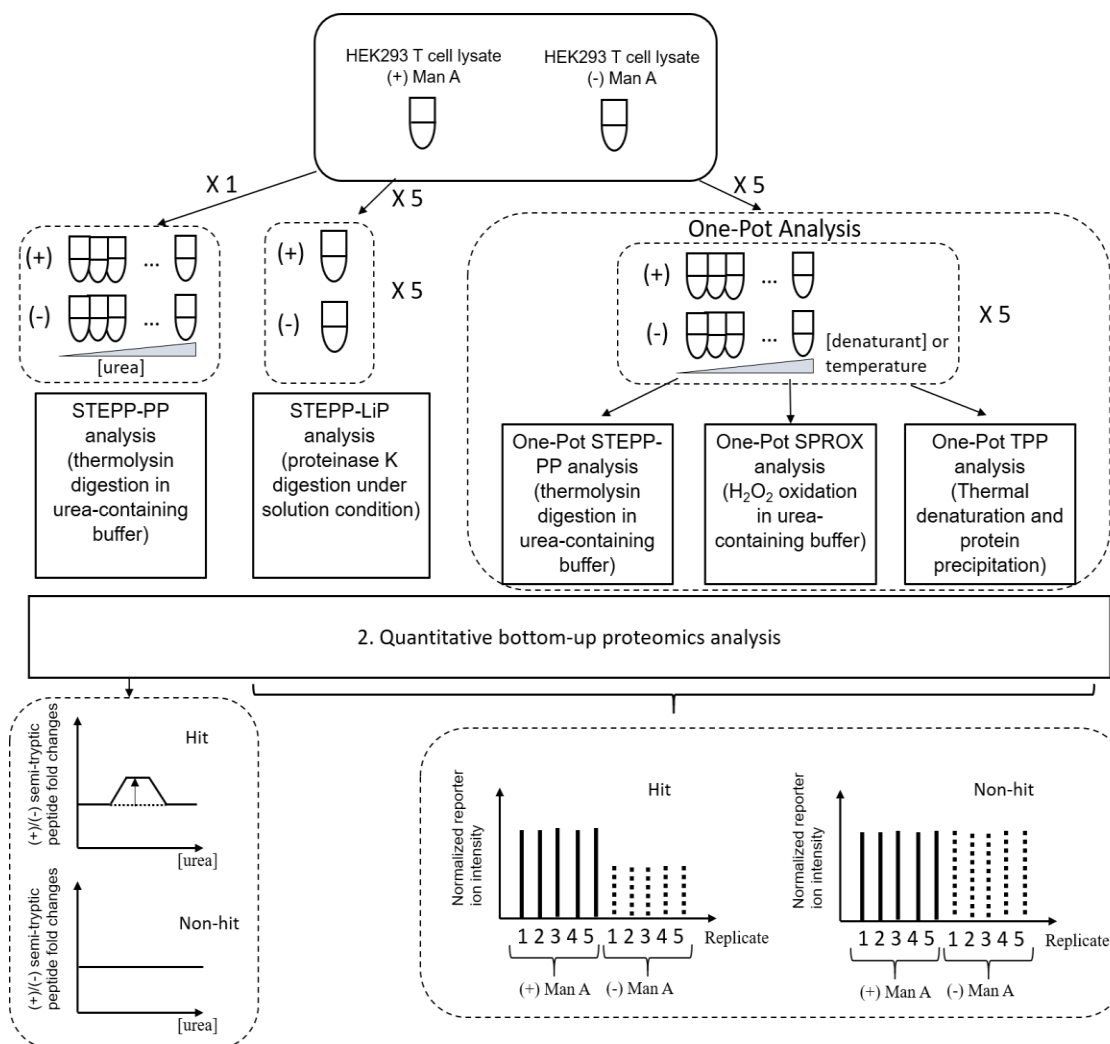


Figure 27 Schematic representation of the workflow of proteomics methods used in this work for identification of manassantin A protein target. The STEPP-LiP analysis and the one-pot analyses use identical hit selection criteria based on t tests

The proteomic coverages obtained in each experiment are summarized in Table 7.

Over 4500 unique proteins were assayed for thermodynamic stability changes induced by

Man A from the five experiments applied here. The hit selection criterion in the STEPP-PP experiment is identical to that described in Chapter 2, which requires a semi-tryptic peptide to have a significantly altered $\log_2(\text{fold-change})$ ($>3\sigma$ deviations from the mean $\log_2(\text{fold-change})$) in the same direction for at least two consecutive urea concentrations. Because the STEPP-LiP experiment and one-pot experiments (i.e., one-pot SPROX, one-pot STEPP-PP and one-pot TPP experiments) all involve a t-test analysis for five biological replicates at once, identical hit selection criteria could be applied to these experiments (see Figure 28).

Table 7 Summary of proteomic coverages and hits observed in the Man A binding experiments on proteins from HEK293T cell lysate

Experiment	Protein (peptide) coverage	Protein (peptide) hits
STEPP-PP	2538 (8477)	26 (45)
STEPP-LiP	1272 (3137)	5 (5)
one-pot SPROX	3187 (9680)	1 (1)
one-pot TPP	2015	6
one-pot STEPP-PP	966 (5032)	12 (17)
Total	4636	49

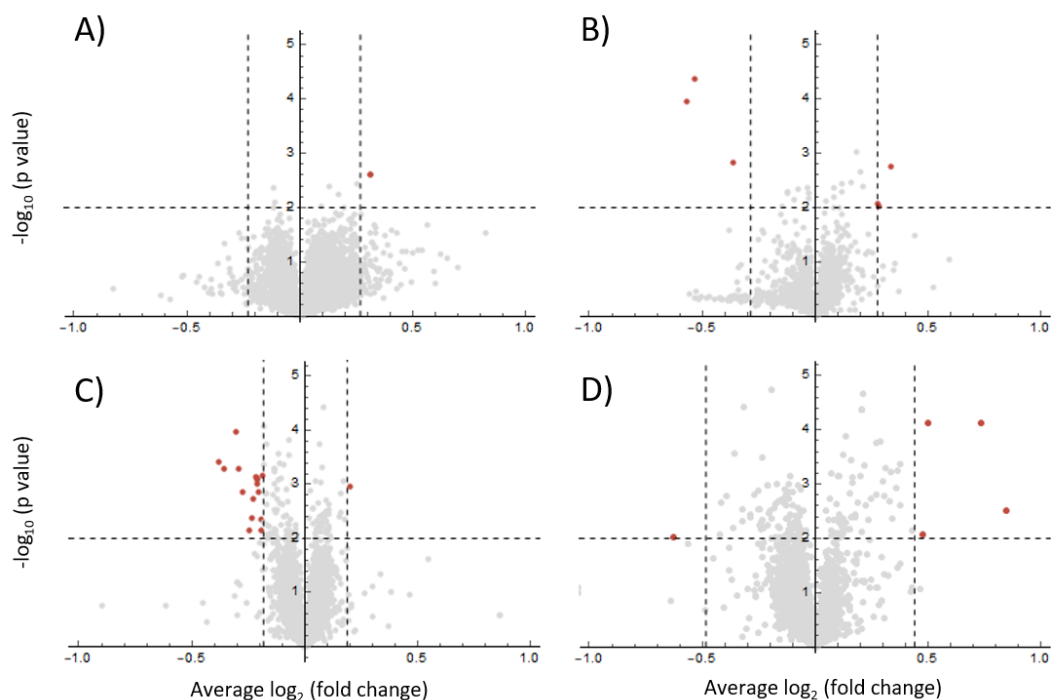


Figure 28 Volcano plots of the average \log_2 ((+)/(-) fold change) values and t test p-values in A) one-pot SPROX, B) one-pot TPP, C) one-pot STEPP, and D) STEPP-LiP data generated in this work. The red dots are the hit protein or peptides which passes two hit selection criteria 1) average \log_2 (fold change) larger than 3σ (represented by the vertical dotted lines) away from the global average \log_2 (fold change); 2) the t test p value is smaller than 0.01 (represented by the horizontal dotted lines).

In total 49 proteins were identified as hits in these experiments. A detailed list of these proteins is shown in Table 8. Only one protein, Peroxiredoxin-1 (PRDX1) has been identified as a hit protein in at least two methods, i.e., STEPP-LiP and STEPP-PP. However, it is assayed but not a hit protein in the other three experiments, which means that PRDX-1 is likely another false positive.

There are four hit proteins, including cAMP-dependent protein kinase inhibitor beta (PKIB), Liprin-alpha-3 (PPFIA3), E3 SUMO-protein ligase (KIAA1586) and Two pore calcium channel protein 2 (TPCN2), that are only assayed in the one experiment in which

they are hits. As these proteins have not been "ruled-out" as false positives by the other proteomics experiments applied here, further experiments like target pulse proteolysis or ITC are needed to substantiate these hit proteins. Although it should be noted that none of these proteins have been previously found to bind HIF-1 α , this lack of relation to HIF-1 α may just be because of our limited knowledge of the cell hypoxia regulation.

Table 8 Protein hits observed in the proteomics experiments performed in this work.

"NA" in the column of a specific experiment means it is not assayed in that experiment; "S" or "D" mean the protein is a hit and stabilized or destabilized, respectively, in the experiment; "NH" means the protein is not a hit. "NA" in the "consistency" column means the protein is only assayed in one experiment; "No" means the protein is not a hit in at least one experiment, i.e., the protein behavior is not consistent across experiments

Protein name	Gene name	STEPP-PP	STEPP-LIP	one-pot STEPP-PP	one-pot TPP	one-pot SPROX	Consistency
cAMP-dependent protein kinase inhibitor beta	PKIB	NA	NA	NA	S	NA	NA
Liprin-alpha-3	PPFIA3	S	NA	NA	NA	NA	NA
E3 SUMO-protein ligase KIAA1586	KIAA1586	S	NA	NA	NA	NA	NA
Two pore calcium channel protein 2	TPCN2	D	NA	NA	NA	NA	NA
Thioredoxin-dependent peroxide reductase, mitochondrial	PRDX3	NH	D	NH	NH	NH	No
Heterogeneous nuclear ribonucleoproteins A2/B1	HNRNPA2B1	NH	D	NH	NH	NH	No
Peroxiredoxin-1	PRDX1	S	D	NH	NH	NH	No
Peroxiredoxin-6	PRDX6	NH	D	NH	NH	NH	No
Heterogeneous nuclear ribonucleoprotein D0	HNRNPD	NH	S	NH	NH	NH	No
26S proteasome regulatory subunit 4	PSMC1	NH	NA	NH	NH	S	No
Tubulin alpha-1B chain	TUBA1B	NA	NH	S	NH	NH	No
Ubiquitin-like modifier-activating enzyme 1	UBA1	NH	NH	S	NH	NH	No
Elongation factor 1-gamma	EEF1G	NH	NH	S	NH	NH	No
L-lactate dehydrogenase B chain	LDHB	NH	NA	S	NH	NH	No
Lactoylglutathione lyase	GLO1	NH	NH	S	NH	NH	No
L-lactate dehydrogenase A chain	LDHA	NH	NA	S	NH	NH	No
Elongation factor 2	EEF2	NH	NH	S	NH	NH	No
Actin, cytoplasmic 1	ACTB	NH	NA	S	NH	NH	No

Heat shock protein HSP 90-beta	HSP90AB1	NH	NH	D	NH	NH	No
Tubulin beta chain	TUBB	NH	NH	S	NH	NH	No
Elongation factor 1-alpha 1	EEF1A1	NH	NH	S	NH	NH	No
Creatine kinase B-type	CKB	NH	NH	S	NH	NH	No
DAZ-associated protein 1	DAZAP1	NA	NA	NA	D	NH	No
Nucleolysin TIAR	TIAL1	NA	NA	NA	D	NH	No
Nucleolysin TIA-1 isoform p40	TIA1	NA	NA	NA	D	NH	No
Alpha-synuclein	SNCA	NA	NH	NA	S	NA	No
GTP-binding protein SAR1a	SAR1A	NA	NA	NA	S	NH	No
Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	D	NH	NH	NH	NH	No
Glutathione S-transferase omega-1	GSTO1	D	NA	NA	NH	NH	No
Glycogen phosphorylase, brain form	PYGB	D	NA	NA	NH	NH	No
Glutamate--cysteine ligase regulatory subunit	GCLM	D	NA	NA	NH	NH	No
Pre-mRNA-splicing regulator WTAP	WTAP	D	NA	NA	NA	NH	No
Proteasomal ubiquitin receptor ADRM1	ADRM1	S	NH	NH	NH	NH	No
Unconventional myosin-Vc	MYO5C	D	NH	NA	NA	NA	No
Exosome complex component RRP40	EXOSC3	S	NA	NH	NA	NH	No
AT-rich interactive domain-containing protein 1A	ARID1A	S	NH	NA	NH	NH	No
Structural maintenance of chromosomes protein 3	SMC3	S	NH	NH	NH	NH	No
Nuclear receptor coactivator 5	NCOA5	S	NA	NA	NA	NH	No
Spectrin beta chain, non-erythrocytic 1	SPTBN1	D	NH	NA	NH	NH	No
Jupiter microtubule associated homolog 2	JPT2	S	NH	NH	NH	NH	No
Bifunctional purine biosynthesis protein PURH	ATIC	S	NA	NH	NH	NH	No
60S ribosomal protein L34	RPL34	D	NH	NH	NH	NA	No
Nuclear pore complex protein Nup153	NUP153	S	NH	NA	NH	NH	No
Y-box-binding protein 1	YBX1	D	NH	NA	NH	NH	No
Protein 4.1	EPB41	S	NA	NA	NH	NH	No
Protein scribble homolog	SCRIB	S	NH	NA	NH	NH	No
Prefoldin subunit 2	PFDN2	S	NH	NH	NH	NH	No
D-3-phosphoglycerate dehydrogenase	PHGDH	D	NH	NH	NH	NH	No
Amidophosphoribosyltransferase	PPAT	D	NA	NA	NH	NH	No

5. Global Analysis of Protein Folding Stability Changes During Parkinson's Disease in a Hu α -Syn(A53T) transgenic Mouse Model

5.1 Introduction

The work described here utilizes thermodynamic measurements of protein stability to identify specific proteins that play a role in the development and progression of PD in a mouse model of the disease. Because of the close link between protein folding thermodynamics and protein function, proteins with stabilities that change during PD progression are likely to have PD related functions or dysfunctions. The mouse model used in this work is one in which the human α -synuclein protein with an A53T mutation is overexpressed at roughly 6-fold over that of the endogenous mouse α -synuclein protein (α Syn). This PD mouse model has been shown to present with the symptoms of PD when mice reach 9-16 months of age with a mean age of onset of 13 months.¹¹¹ The main goal of this work was to generate thermodynamic stability profiles for brain derived proteins from this mouse model at 1 and 6 months when the mice were pre-symptomatic and at a third time point when the mice first presented symptoms. The identification of proteins with altered thermodynamic stabilities during the progression of the disease is expected to elucidate specific pathways and downstream effects of the formation of abnormal α -synuclein fibrils and aggregation.

As part of this work, we also extended our previous protein folding and stability studies of brain-derived proteins in wild-type mice aged 6- and 18-months⁵³ to include a 1-month time point. This enabled a direct comparison of the thermodynamic stability

profiles of brain proteins from the wild-type and PD mouse models at three different time points. The protein hits identified in the comparative analyses described here not only help elucidate the molecular basis of PD, but also have the potential to be useful PD biomarkers that could aid in early diagnosis of the disease.

The thermodynamic stabilities of the mouse brain-derived proteins in this work were profiled using the Stability of Proteins from Rates of Oxidation (SPROX) technique in combination with an isobaric mass tagging strategy employing iTRAQ reagents.^{18,112} The experimental workflow employed here was identical to that which we previously used to profile the thermodynamic stability of brain proteins derived from wild-type mice aged 6- and 18-months.⁵³ Ultimately, the data generated in this work and in earlier earlier work on the wild-type mice were used here in a series of comparative analyses to identify proteins with thermodynamic stability changes during the process of normal aging and the course of PD progression, as well as to identify proteins with thermodynamic stability differences in age-matched transgenic and wild-type mice.

5.2 Experimental Section

5.2.1 Mouse Euthanasia and Tissue Lysis

The mice used in this study were purchased from The Jackson Laboratory and euthanized by cervical dislocation at Duke University. Mice were handled in accordance with the handling procedures approved by Duke University Institutional Animal Care and Use Committee approval for Protocol A196-13-08. All Hu α -Syn(A53T) transgenic mice were females with the α -synuclein A53T mutation (strain: B6.Cg-Tg(Prnp-SNCA*A53T)23Mkle/J). All the 1-month WT mice were females of strain: C57BL/6J. PD

mice were euthanized either at 1 month of age (n=9), 6 months of age (n=7), or upon becoming symptomatic (n=8). The WT mice were euthanized at 1 month of age (n=8). The 8 mice in the PD symptomatic group became symptomatic at 10.5 months, 15 months, 16 months, 9 months, 11.5 months, 13 months, 14.5 months and 16 months. Mice were determined to be symptomatic upon indicating signs previously determined,⁴ such as: tremors, lack of righting response and difficulty walking due to tremors. Immediately following euthanasia, the brains were removed and stored at -20°C until the cell lysis was performed as previously described.⁵³ Immediately prior to iTRAQ-SPROX analysis a Bradford Assay was used to quantify the total protein concentration of the lysates. Lysates were ultimately diluted to a final protein concentration of 5 mg/mL for iTRAQ-SPROX analysis.

5.2.2 iTRAQ-SPROX Protocol

The iTRAQ-SPROX experiments were conducted similar to that previously described⁵³ with the exception of a few modifications that were incorporated into the protocol to improve the proteomic coverage. These modifications included using TCEP instead of methionine for quenching the hydrogen peroxide oxidation reaction and using an iFASP protocol instead of a TCA precipitation protocol for bottom-up proteomics sample preparation. Briefly, ~800 µg of total protein from each mouse brain tissue cell lysate was equally distributed into a series of 8 GdmCl-containing SPROX buffers comprised of 20 mM phosphate buffer, pH 7.4, and increasing concentrations of GdmCl. Hydrogen peroxide was added into each protein-containing buffer to selectively oxidize solvent exposed methionine residues in each denaturant-containing SPROX buffer. The

final concentrations of GdmCl in the reaction buffers for all the SPROX experiments were 0.8 M, 1.1 M, 1.2 M, 1.4 M, 1.6 M, 1.8 M, 2.0 M, and 2.5 M. The final hydrogen peroxide concentration (3% v/v) and reaction time (3 min) was also the same in each SPROX buffer and in each SPROX analysis. The oxidation reaction in each SPROX buffer was either quenched with methionine (PD 1 month mouse number 6-9, PD 6 month mouse numbers 5-7, PD symptomatic mouse numbers 6-8), or quenched with TCEP (WT 1 month mouse numbers 1-8, PD 1 month mice 1-5, PD 6 month mouse number 1-4, PD symptomatic mice numbers 1-5).

5.2.3 Proteomic Sample Preparation

The quenched samples were either subject to a trichloroacetic acid protein precipitation (for methionine-quenched samples) or a filter-aided sample preparation (for TCEP quenched samples) as previously described,⁵³ prior to subjecting them to a quantitative bottom-up shotgun proteomics analysis using isobaric mass tags. Briefly, the trichloroacetic acid-precipitated protein pellets were dissolved in 30 μ L of 0.5 M triethylammonium bicarbonate (TEAB) buffer (pH 7.5) containing 0.1% sodium dodecylsulfate (SDS). Disulfide bonds were reduced using 5 mM tris(2-carboxyethyl)phosphine (TCEP) for 1 hour at 60°C, and the free cysteine residues were reacted 10 mM methylmethanethiosulfonate (MMTS) for 10 minutes at RT. The protein samples from each SPROX buffer were digested with 1 μ g of trypsin at 37°C overnight. The samples subjected to filter aided sample preparation (FASP) protocol also reduced with TCEP, reacted with MMTS, and digested with trypsin using the above reaction conditions.

The resulting tryptic peptides were labeled with iTRAQ reagents from an iTRAQ 8-plex (AB SCIEX, Framingham, MA), according to the manufacturer's protocol with the exception that 0.5 units of each reagent were used in each labelling reaction instead of 1.0 unit. The eight iTRAQ-labelled samples corresponding to the eight denaturant concentrations in each unfolding curve for each mouse were combined. The combined samples were desalted using a C18 column (The Nest Group) according to the manufacturer's protocol. These were the 'non-enriched' samples. An aliquot containing about 100-200 µg of iTRAQ-labeled peptides was removed from the desalt samples and enriched for methionine residue containing peptides using a Pi3™ Methionine Reagent Kit according to the manufacturer instructions (The Nest Group). These were the 'enriched' samples.

5.2.4 LC-MS/MS Analyses

The iTRAQ-labeled peptide generated in the SPROX analyses of the mouse brain derived samples were analyzed on one of two LC-MS/MS mass spectrometer instruments, Instruments 1 and 2. Instrument 1 was comprised of an Orbitrap Fusion mass spectrometer with an EASY-nLC 1000 system (ThermoFischer Scientific Inc.). The trapping column used in these experiments was a 100 µm x 2 cm Integrafrit column (New Objective) that was packed with 200 Å Magic C18 AQ 5 µm material from Michrom. The column in these runs was a 75 µm x 25 cm PicoFrit column (New Objective) packed with 100 Å Magic C18 AQ 5 µm material from Michrom. A flow rate of 300 nL/min was used for all runs. Solvent A is 0.1% formic acid in water and solvent B is 0.1% formic acid in acetonitrile (ACN). The chromatography gradient was as follows: 3% to 6% solvent B in

A over 2 minutes, 6% to 33% solvent B in A over 92 minutes, 33% to 50% solvent B in A over 7 minutes, 50% solvent B in A for 5 minutes, 50% to 90% solvent B in A over 3 minutes and 95% solvent B in A for the final 8 minutes. The LC-MS/MS data were collected in a data dependent MS/MS mode (top 24) using a m/z range of 350-1500. Resolution of the instrument was set to 120,000. Cycles were 3 seconds long, the most abundant ions from the precursor scan were selected for MS/MS analysis and selected ions were dynamically excluded for 30 seconds. Collision induced dissociation (CID) was conducted with HCD and a normalized collision energy of 40% analyzed via Orbitrap with a resolution of 15,000.

The second LC-MS/MS instrument, Instrument 2, was a Q Exactive HF high-resolution mass spectrometer (Thermo) equipped with a nano-Acquity UPLC system (Waters) and a nano-electrospray ionization source fitted with a SilicaTip emitter (New Objective). Samples were trapped on a 2D Symmetry C18 trapping column with dimensions of 180 μm \times 20 mm, particle diameter of 5 μm , and pore size of 100 Å. The trapping time was 5 min at 5 $\mu\text{L}/\text{min}$ (99.9:0.1 v/v water/acetonitrile 0.1% formic acid). Sample separation was performed on a 75 μm \times 250 mm high strength silica (HSS) T3 column with 1.8 μm particle diameter (Waters) at 55 °C, using a gradient of 3–30% acetonitrile with 0.1% formic acid over 90 min at a flow rate of 0.3 $\mu\text{L}/\text{min}$. Data collection was performed in a data-dependent acquisition (DDA) mode (TOP 20) with a resolution of 120 000 (at m/z 200) for full MS scan from m/z 375–1600. The resolution was 30 000 (at m/z 200), using a minimum AGC target value of 3×10^6 ions. Summarized in Table 9, are

the mouse-brain samples and the LC-MS/MS mass spectrometer instruments used in each SPROX analysis.

Table 9 summaries the LC-MS/MS instrument used for analysis of each mouse brain samples in this work

Mouse group	Mouse brain sample number	LC-MS/MS instrument
WT 1mo	1-8	Q Exactive HF
PD 1mo	1-5	Q Exactive HF
	6-9	Orbitrap Fusion
PD 6mo	1-4	Q Exactive HF
	5-7	Orbitrap Fusion
PD Symptomatic	1-5	Q Exactive HF
	6-8	Orbitrap Fusion

Peak lists were extracted from the LC-MS/MS data and were searched against the proteins in the SwissProt Mus musculus database version 2019-03-20, downloaded on 04/11/19 using Proteome Discoverer 2.2. Cysteine residue modification by MMTS and N-termini and lysine residues modified by iTRAQ 8-plex were fixed modifications in the search. Oxidized methionine residues and deamidation on asparagine and glutamine and arginine were variable modifications in the search. Up to two missed tryptic cleavages after R and K were allowed. The parameters included a 10 ppm mass tolerance window for precursor masses and 0.6 Da for fragment mass tolerance. Only peptide spectra with FDR <1%, isolation interference <30% and iTRAQ reporter ion intensities that summed to 1000 were used in subsequent analyses of the data. This iTRAQ intensity cutoff has been previously shown to effectively filter out poor quality SPROX data.¹⁸

The LC-MS/MS analyses of the iTRAQ-SPROX samples analyzed here included 3 technical replicates of each methionine-containing peptide enriched sample and 1 analysis

of each non-enriched sample generated in the SPROX analysis of each brain tissue cell lysate. The proteomics data collected on these samples, including raw LC-MS/MS data and search outputs, can be found at the ProteomeXchange Consortium via the Pride repository with the identifier PXD016985.^{113,114}

5.2.5 Data Analysis

The iTRAQ reporter ion intensities generated for each methionine-containing peptide were used to construct chemical denaturation curves as previously described.¹⁸ Briefly, the iTRAQ reporter ion intensities recorded in a given product ion mass spectrum were normalized to the average intensity of the 8 reporter ions in each product ion mass spectrum to generate N1 normalized values. The N1 values for all the non-methionine containing peptides were averaged for each reporter ion in order to generate an N2 normalization factor for each isobaric mass tag. Ultimately, the N1-normalized values generated for the methionine-containing peptides were divided by the corresponding N2 normalization factor to determine the N2-normalized reporter ion intensity. The resulting chemical denaturation data sets (i.e., the N2-normalized iTRAQ reporter ion intensities generated for each unoxidized methionine-containing peptide) were normalized to a maximum of 1 by dividing each intensity by the maximum intensity and fitted to a four parameter sigmoidal equation, equation 1, using a JAVA-based program (developed in house) that utilized the Nelder and Mead Simplex method for regression analysis.⁸⁶

$$y = A + \frac{(B - A)}{1 + e^{-\frac{C_{1/2} - x}{b}}} \quad (12)$$

In equation (12), A is the pre-transition baseline, B is the post-transition baseline, $C_{1/2}$ is the transition midpoint and b is a measure of the steepness of the transition. The program fit each set of data nine times, once with all eight points and then eight additional times systematically leaving one point out. The fit with the highest R² value was chosen as the final output. Subsequent analyses of the data only utilized the chemical denaturation data sets that were determined to be high quality (R² ≥ 0.8). If a peptide was identified multiple times within the same iTRAQ-SPROX experiment on a particular mouse brain sample, the iTRAQ reporter ion intensities from the high-quality data were summed together to generate one set of iTRAQ reporter ion intensities at the 8 denaturant concentrations per peptide per mouse. The summed data was N²-normalized and fit to equation (12), as described above, to extract a single $C_{1/2}$ value for each identified peptide identified from each mouse.

5.2.6 Hit Identification

Only peptides that were assayed in at least two mice in all of the six groups (i.e., WT 1mo, WT 6mo, WT 18mo, PD 1mo, PD 6mo and PD Symptomatic) were included in the ANOVA analysis used for hit identification. For each methionine-containing peptide, a two-way unbalanced ANOVA test was employed to identify significant differences between the assigned $C_{1/2}$ values across the 6 groups. Each two-way ANOVA analysis generated three p-values including: a p-value for disease factor, a p-value for aging factor, and a p-value for disease-age interaction factor. Peptides with any of the above p-values lower than 0.000085 were selected as hits. The 0.000085 p-value is the Bonferroni-adjusted p value for p = 0.05. Hit peptides were categorized as PD-related hits, if the p-value for

the disease factor or the p-value for the disease-age interaction factor passed the cutoff ($p < 0.000085$). Hit peptides were categorized as age-related hits if only the age factor passed the cutoff ($p < 0.000085$).

5.3 Results

5.3.1 Experimental Design

The experimental workflow used in this work is summarized in Figure 29. The experiments performed here were designed to identify protein folding stability differences that resulted from PD development and progression induced by the α -synuclein mutation A53T in mouse brain samples. The transgenic mice were sacrificed at three time points including at 1 month ($n=9$) to provide a baseline reading of protein thermodynamic stability in transgenic mice, at 6 months ($n=7$) that is roughly a half-way point to disease presentation, and at the age when each mouse became symptomatic ($n=8$). The eight mice at the later time point became symptomatic at 9, 10.5, 11.5, 13, 14.5, 15, 16 and 16 months. The eight wild-type mice used in this experiment were all sacrificed at 1 month. These 1 month old mice together with wild-type mice sacrificed at 6 month ($n=7$) and 18 month ($n=9$) from a previous study⁵³ served as the control mice in this study. Each mouse brain lysate was subjected to one iTRAQ-SPROX analysis that included 1 LC-MS/MS analysis of the non-enriched and 3 LC-MS/MS analyses (technical replicates) of the methionine-containing peptide enriched samples.

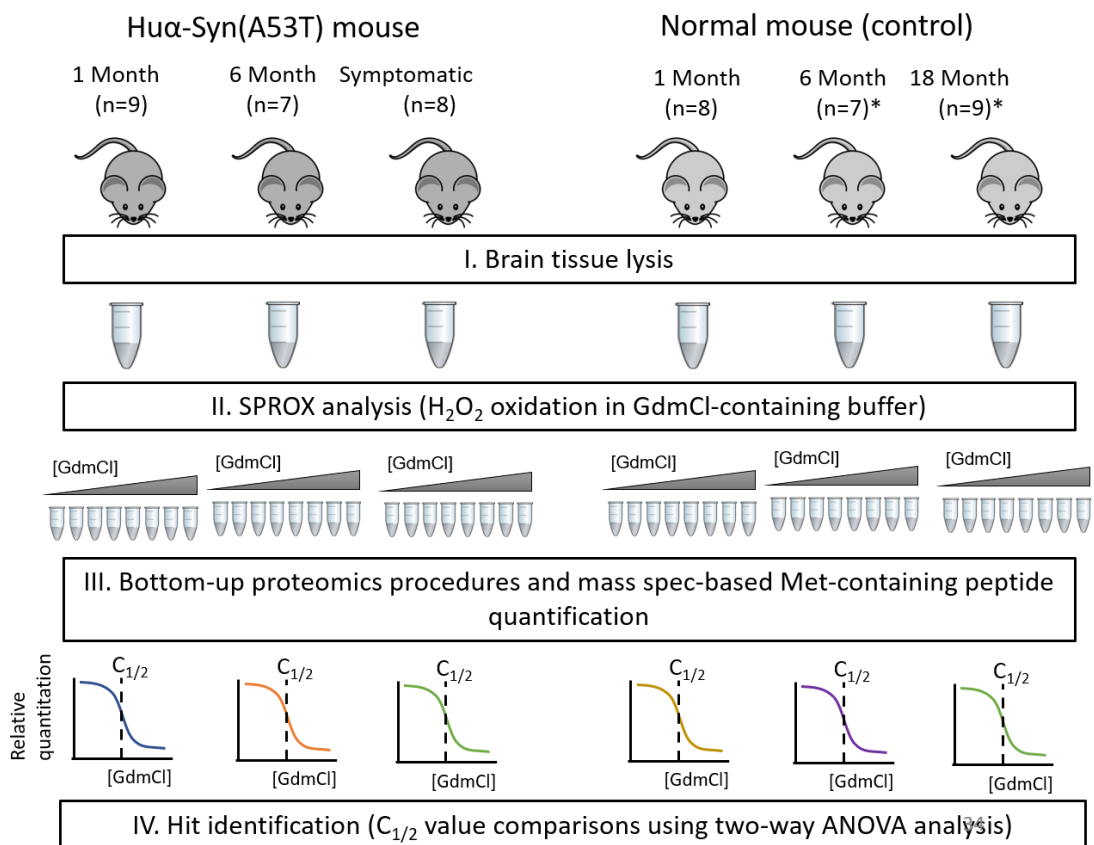


Figure 29 Schematic representation of the experimental workflow utilized in this work. The data of wild-type mice sacrificed at 6 month and 18 month are from a previous study.⁵³

5.3.2 Proteomic Coverage and Hit Selection

The iTRAQ-SPROX analyses performed on the brain tissue cell lysates from experiments on the 32 mice in this study generated high quality chemical denaturation data for 350-2000 proteins using 600-5400 methionine-containing peptide probes in each lysate, depending on the mouse. Our comparative analysis only included methionine-containing peptide probes that were assayed in 2 or more mice in all of the six mouse groups (i.e., WT 1mo, WT 6mo, WT 18mo, PD 1mo, PD 6mo and PD Symptomatic). The number of methionine-containing peptide probes and proteins that were assayed in each

cohort is summarized in Table 10. This requirement reduced the number of methionine-containing peptide probes in this study to 590 and the number of proteins to 332.

Table 10 Summary of the proteomic data obtained in the iTRAQ-SPROX experiments performed on the brain tissue cell lysates from each group in this study

Age and disease state	Number of mice	Assayed (proteins)¹	peptides
WT 1mo	8	2436 (1000)	
WT 6mo	7	2049 (962)	
WT 18mo	9	2206 (1033)	
PD 1mo	9	2730 (1103)	
PD 6mo	7	4681 (1733)	
PD Symptomatic	8	6356 (2054)	
Assayed for hit identification		590 (332) ²	

¹ The assayed methionine-containing peptide probe must be identified in at least two mice in every cohort (e.g., WT 1mo). The proteins are identified by these assayed peptides.

² The assayed peptide probes for hit selection were the ones that have been identified in at least two mice in every age and disease state studied. The proteins are identified by these assayed peptides.

For each methionine-containing peptide probe in this study, a $C_{1/2}$ value was generated for each mouse brain lysate where the probe was identified. A typical unfolding curve generated for a methionine-containing peptide probe in this study is highlighted in Figure 30.

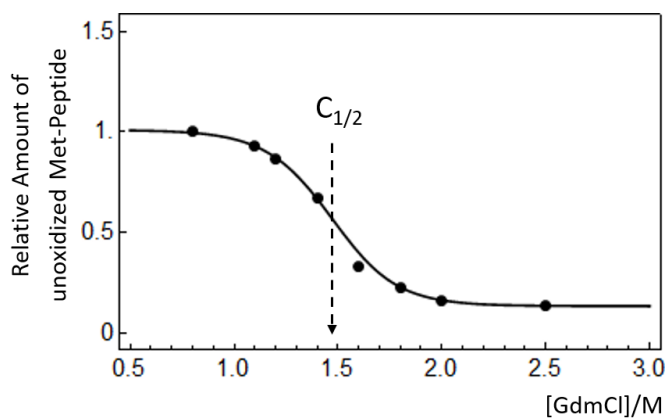


Figure 30 Demonstration of $C_{1/2}$ value determination in this work: the dashed arrow indicates the average $C_{1/2}$ value of the fitted curve. The denaturation data is from the methionine-containing peptide probe YIKDYMK) from Tpt1 protein in PD symptomatic mouse number 3.

Ultimately, a two-way unbalanced ANOVA analysis was performed to identify methionine-containing peptides with significant $C_{1/2}$ -value shifts related to the disease state and/or aging. A Bonferroni-adjusted p-value cutoff of $p = 0.000085$ (i.e., $0.05/590$) was used to reduce false positives stemming from multiple testing. The identified peptide and protein hits are summarized in Table 11. Three categories of hits including: those related to the disease, aging, or a disease and aging interaction were identified. Highlighted in Figure 31 and Figure 32 are the $C_{1/2}$ changes recorded for selected protein hits and nonhits.

Table 11 Summary of the hit proteins and their hit methionine-containing peptide probes

Hit category	Gene ID	Protein name	Hit methionine-containing peptide probe	Connection to aging or PD
PD-related protein hits	Pgm1 ³	Phosphoglucomutase-1	SMPTSGALDR	115–117
	Ca2 ^{2,3}	Carbonic anhydrase 2	TLNFNEEGDAEEAMVDNWRPAQPLK	
	Psmc5	26S proteasome regulatory subunit 8	VIMATNR	
PD-age-interaction-related protein hits	Ap2m1	AP-2 complex subunit mu	SISFIPPDGEFELMR	118 118 119 120 121 122
	Gdi2 ^{2,4}	Rab GDP dissociation inhibitor beta	FKLPGQPASMR	
	Gdi2 ²	Rab GDP dissociation inhibitor beta	MTGSEDFEEMKR	
	Cs ²	Citrate synthase, mitochondrial	TVVGQITVDMMYGGMR	
	Tpt1 ^{2,4}	Translationally-controlled tumor protein	YIKDYMK	
	Nrcam	Neuronal cell adhesion molecule	VMAENSIGR	
	Nsfl1c	NSFL1 cofactor p47	LAHGGQVNLDMEDHRDEDFVKPK	
	Rbbp7 ²	Histone-binding protein RBBP7	YMPQNPPIATK	
Age-related protein hits	Clpp	ATP-dependent Clp protease proteolytic subunit, mitochondrial	GQATDIAIQAEIIMK	122
	Acot7	Cytosolic acyl coenzyme A thioester hydrolase	MIEEAGAIISTR	123
	Glud1 ²	Glutamate dehydrogenase 1, mitochondrial	FTMELAK	124
	Ywhaq ²	14-3-3 protein theta	EMQPTHPIR	125
	Syn2	Synapsin-2	MNQLLSR	126
	Hspa4 ^{1,2}	Heat shock 70 kDa protein 4	QIQQYMK	127
	Lap3	Cytosol aminopeptidase	HMESPANEMTPTR	125
	Ywhag	14-3-3 protein gamma	YDDMAAAMK	
	Gad2	Glutamate decarboxylase 2	GMAAVPR	128
	Ube2i ^{1,2}	SUMO-conjugating enzyme UBC9	MLFKDDYPSSPPK	129
	Psm1	Proteasome subunit alpha type-1	AMSIGAR	
	Eno1 ²	Alpha-enolase	LAMQEFMILPVGASSFR	130
	Add3	Gamma-adducin	VGEIEFGLMR	131
	Tnr ¹	Tenascin-R	TTLTDLVPGTEYGVGISAVMNSK	
	Psmb1	Proteasome subunit beta type-1	AGGSASAMLQPLLDNQVGFK	132
	Atp6v0d1	V-type proton ATPase subunit d 1	MVVEFR	
Cntn1 ¹	Contactin-1	YVHKDETMTSTAFQVK	133	
Pfkip	ATP-dependent 6-phosphofructokinase, platelet type	LRPIMK		
P4hb ²	Protein disulfide-isomerase	LITLEEEMTK		
Pgm1 ³	Phosphoglucomutase-1	SMPTSGALDR		

Ca2 ^{2,3}	Carbonic anhydrase 2	TLNFNEEGDAEEAMVDNWRPAQLK	
Gdi2 ^{2,4}	Rab GDP dissociation inhibitor beta	FKLPGQPPASMGR	118
Tpt1 ^{2,4}	Translationally-controlled tumor protein	YIKDYMK	119

¹ Proteins that previously known to have age-related stability changes⁵³

² Proteins that have known carbonylation or oxidation sites¹³⁴⁻¹³⁶

³ Proteins that were categorized by two-way ANOVA analysis as both PD-related hits and age-related hits.

⁴ Proteins that were categorized by two-way ANOVA analysis as both PD-age-interaction-related hits and age-related hits.

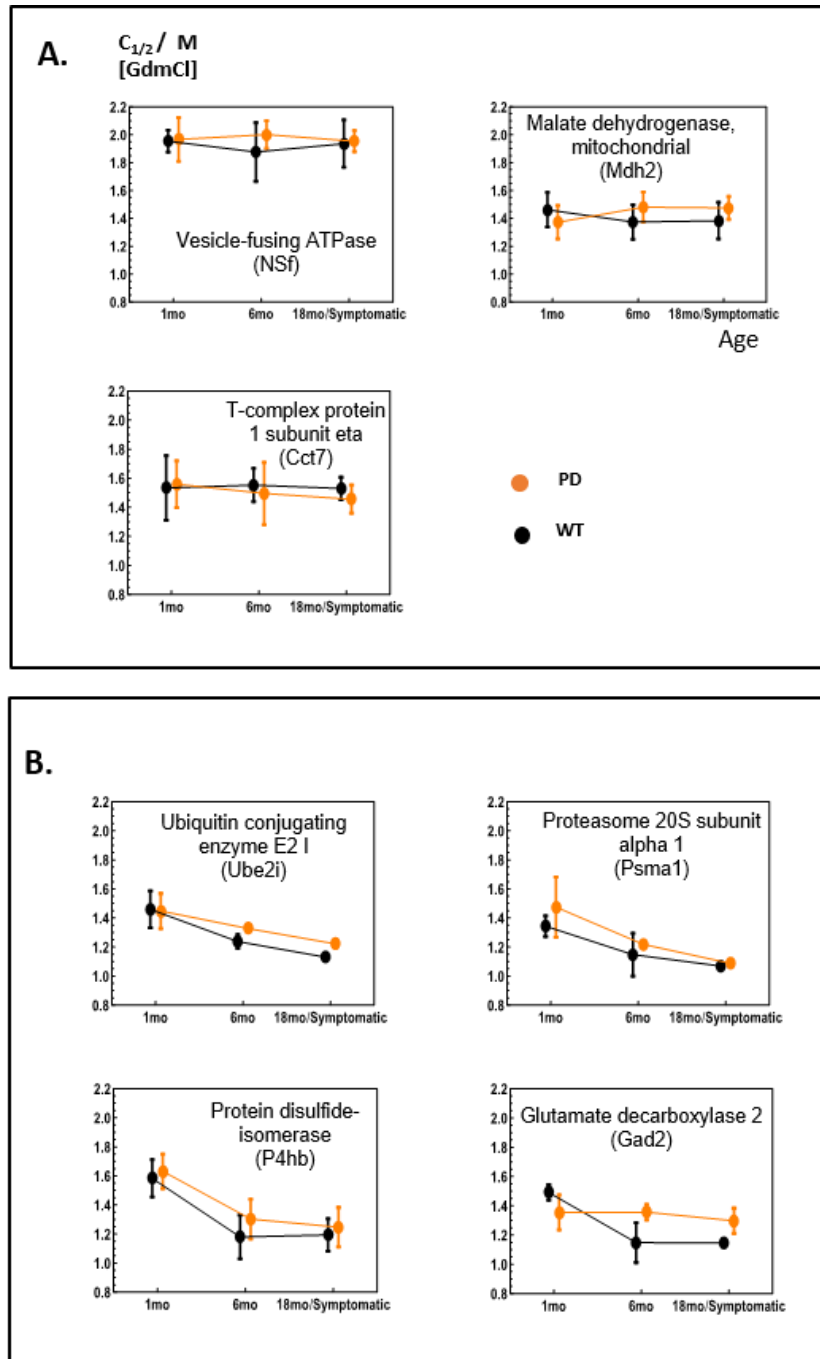


Figure 31 Example plots of $C_{1/2}$ changes in selected proteins with non-hit and age-related hit methionine-containing peptide probes. Shown in (A) are typical data obtained on proteins with non-hit methionine-containing peptide probes that have no significant age-related or PD-related stability changes. Shown in (B) are examples of proteins with age-related methionine-containing peptide probe hits that have age-induced stability changes in WT and PD mice.

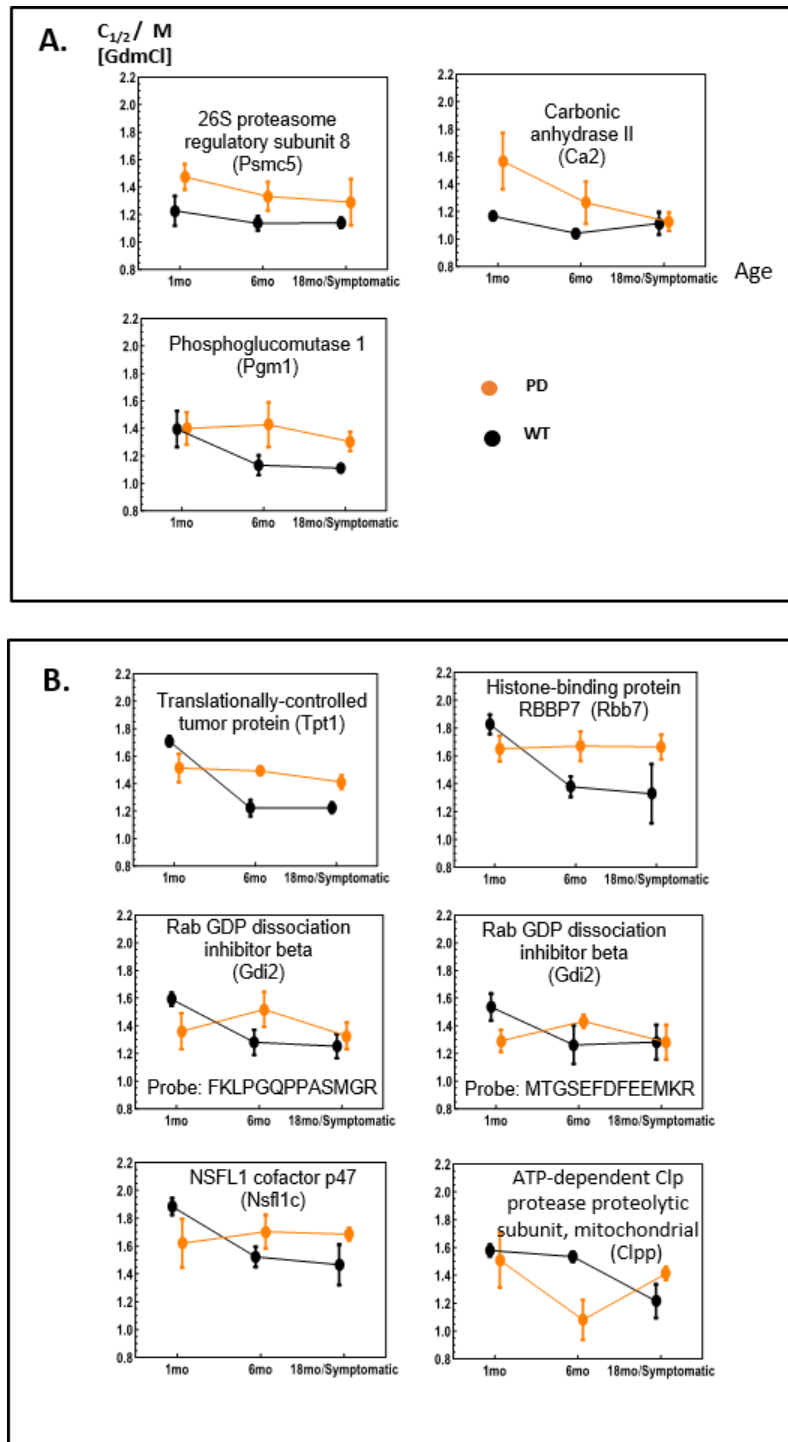


Figure 32 Example plots of $C_{1/2}$ changes in selected proteins with PD-related and PD-age interaction hit methionine-containing peptide probes. Shown in (A) are PD-related protein hits with methionine-containing peptide probes that have consistent stabilization or destabilization in PD and WT mice. Shown in (B) are examples of proteins with methionine-containing peptide probes displaying a PD-age interaction with different age-induced stability trends in PD and WT mice.

5.3.3 Biological Variability

The standard deviations of the $C_{1/2}$ value measurements recorded for a given methionine-containing peptide probe across the mice in each cohort were used to assess the biological variability of the chemical denaturation curves generated in this work. The frequency distributions of the standard deviations of the $C_{1/2}$ values observed for peptides identified in the wild-type and transgenic mice are shown in Figure 33 A and B, respectively. The frequency distributions in Figure 33 are all similar. In each case the median value was 0.15 M GdmCl; and the 95th percentile was 0.35 M GdmCl.

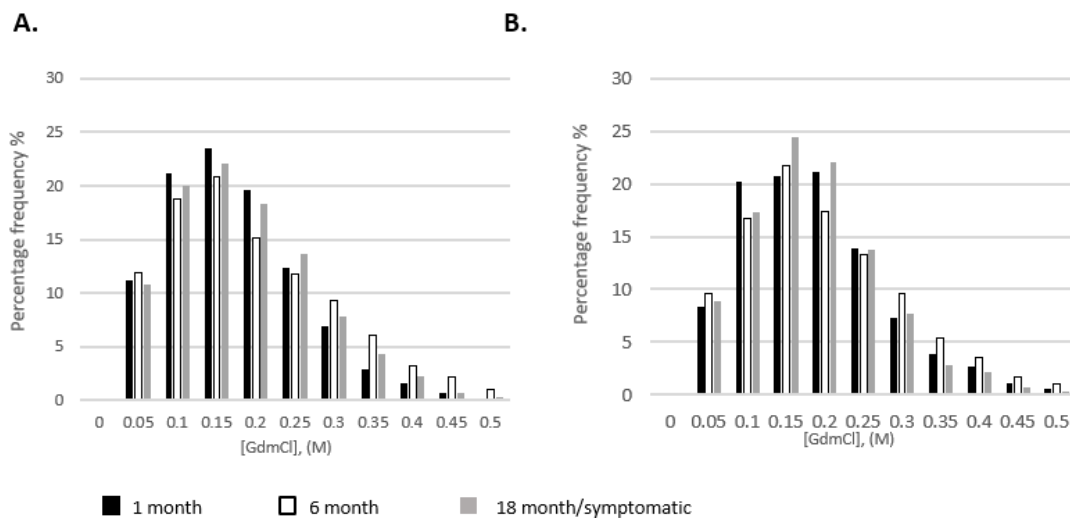


Figure 33 Frequency distributions of standard deviations associated with the $C_{1/2}$ values determined in this work. Shown in (A) are the results for the assayed peptides in the different time points studied in this work. The black, white, and grey bars in (A) are from 1 Month old, 6 Month old, and symptomatic mice, respectively. Shown in (B) are the results on normally aged mice. The black, white, and grey bars in (B) represent the data from the normally aged 1 Month old, 6 Month old and 18 Month old mice, respectively. In both (A) and (B) each bar represents the percent of the total peptides in that state that have a standard deviation within the bin, and the x-axis values indicates the high limit of each bin. The median standard deviation of all the six groups in A and B are around 0.15M GdmCl.

5.4 Discussion

5.4.1 Age-Related Proteins Hits

The ANOVA analysis identified a total of 22 unique methionine-containing peptide probes from 22 proteins with age-related $C_{1/2}$ differences (see Table 11). Consistent with the results of an earlier study on the 6 and 18 Month wild-type mice, the age-related protein hits identified in this work were predominantly destabilized (i.e., over 75% of the hit methionine-containing peptide probes were destabilized with age). It is noteworthy that 4 of these 22 hit proteins (see Table 11), were also found to have age-related stability changes in our earlier study on the 6 and 18 Month wild-type mice.⁵³ The other 18 hits proteins here were not hits in the previous study largely because the previous study only probed late-stage aging (i.e., from 6 to 18 month) but not the early aging (i.e., 1 month to 6 month) effects that were also monitored in this study.

A total of 12 of the 22 age-related protein hits (see Table 11) have previously been tied to the aging process. For example, a gene knockout study has shown that cytosolic acyl coenzyme A thioester hydrolase (Acot7) regulates fatty acid metabolism in brain and protects against neurotoxicity.¹²³ A transgenic mouse study also showed an overexpression of glutamate dehydrogenase 1 (Glud1) in neural system induces age-associated losses of dendritic spines, nerve terminals, and neurons.¹²⁴ Another study of synapsin-2 (syn2) knockout mice showed age-related progression of behavioral defect due to neuronal loss and gliosis in the cerebral cortex and hippocampus.¹²⁶ These previous studies and our protein hits together could provide an interesting insight into the biomolecular basis of age-related loss of functionality in the identified protein hits. Our

work suggests that the specific age-related functions of these proteins may be compromised with age due to a loss of thermodynamic stabilities.

One explanation for the age-related destabilizations observed here is that they result from age-related PTMs such as carbonylations and/or oxidations induced by reactive oxygen species (ROS). Indeed, 9 of the 18 hit proteins have been previously found to have carbonylation or cysteine oxidation sites (see Table 11).^{134,135} For example, protein disulfide isomerase (P4h), which is a ROS target in endoplasmic reticulum (ER), is known to have increased level of carbonylation and reduction in activity during aging.¹³⁷ Another explanation for the observed stability changes in the hit proteins identified here is that during the aging process, neuronal cells produce alternate forms of a protein, that are differentially stabilized. Such an age-related alteration has been observed for one of our hit proteins, glutamic acid decarboxylase 2 (Gad2). Previous studies on human and rat brain samples observed the presence of both the 65 kDa full length Gad2 and a 59 kDa truncated Gad2 resulting from factor Xa cleavage, Interestingly truncated Gad2 was found to be 2-3 times more active than the full length Gad2.^{138,139} Thus, the detected Gad2 protein stability changes in our mouse model during aging process could be the result of the relative abundance changes of the full-length and truncated protein, which are likely to have different thermodynamic stabilities.

The results obtained here suggest that the changes in thermodynamic stabilities could be closely related to the structural and functional changes of proteins in the aging process. These hit proteins detected here can be further exploited for aging biomarkers or therapies against aging. However, it should be noted that the molecular basis for the

detected stability changes in hit proteins are not directly illustrated from the data generated here. More studies on the proteins hits here are needed to understand the exact biophysical changes that induce the observed stability changes. For example, the stabilities of the different forms of the Gad2 proteins can be monitored in a targeted study to uncover if the observed aging-related stability changes of Gad2 are a result of the differential truncation.

5.4.2 PD- and PD-Age Interaction-Related Protein Hits

Identified in this work were 11 proteins with 12 methionine-containing peptide probes displaying PD-induced thermodynamic stability changes (Table 11). These hit proteins include those with peptide probes that passed the ANOVA test p-value cutoff for either the disease factor (3 probes from Pgm1, Ca2 and Psmc5) or the disease-age interaction factor (9 probes from Apem1, Gdi2, Cs, Tpt1, Nrcam, Nsfl1c, Rbbp7 and Clpp). Two of these methionine-containing peptide probes were from the same protein, Rab GDP dissociation inhibitor beta (Gdi2), and showed very similar trends in protein stability changes (Figure 32 B). The two-way ANOVA test identifies peptides with consistent transgenic and WT differences by the disease factor and identifies age-dependent transgenic and WT differences by the age-disease interaction factor. Thus, hits in both categories are likely to aid in the understanding of the biomolecular basis of PD, while the proteins hits screened out by the disease factor may be potentially better PD biomarker due to their consistent PD-related stability alteration.

Examination of the thermodynamic stability changes of the above 11 proteins revealed a majority of the proteins (7 out of 11) were more and more destabilized with age in the WT mice (i.e., average C1/2 values decreased from 1mo to 6mo and 6 mo to 18 mo).

In total, 5 of the 11 hit proteins related to PD have been previously connected to PD (see Table 11). For example, Rab GDP dissociation inhibitor beta (Gdi2) inhibits Rab GTPases, which regulates the cellular membrane dynamics and is already known to be a key player in PD pathology.¹¹⁸ Translationally-controlled tumor protein (Tpt1) is also known to be downregulated in the striatum of PD mouse model.¹¹⁹ Also interesting is that the ATP-dependent Clp protease (Clpp) is a PD-related hit. Clpp is a major contributor to the mitochondrial protein quality control system, as it is responsible for removing damaged or misfolded proteins.¹⁴⁰ Hu et al found that in both α Syn A53T mutant-carrying patient samples and in the Hu α -Syn(A53T) transgenic mouse model used here, α Syn interacts with Clpp causing it to aggregate and lose its ability to relieve mitochondrial oxidative damage.¹²² The destabilization of Clpp observed in the PD mice at 6 mo may be due to this α Syn induced Clpp structural change (Figure 32 B).

The group of PD-related protein hits identified here may be useful early stage PD biomarkers. For example, carbonic anhydrase 2 (Ca2) has increased in thermodynamic stability in our transgenic mice at both 1mo and 6mo age stages comparing to WT mice. Of particular interest, is that Ca2 has been identified in proteomic studies of human CSF.¹¹³ It is possible that the PD-related stability change observed here for Ca2 could also be detected in CSF-derived Ca2. One explanation for the detected stability changes of Ca2 could be related to the carbonylation of this protein that has been previously found to

occur in A30P α Syn transgenic mice,¹³⁶ a very similar α -Syn transgenic mouse model to the mouse model used in this study. Another protein, 26S proteasome regulatory subunit 8 (Psmc5) may also be useful as an early biomarker of PD because of its observed stabilization in the PD mice at both 1mo and 6mo age stages. Psmc5 is has also been detected in proteomic studies of human CSF. In total, 9 of the 11 disease-related protein hits have been previously identified in cerebrospinal fluid (CSF).¹¹³ The altered stabilities of these 9 protein, if detectable in patient-derived CSF samples, could be potential biomarkers for PD.

6. Conclusions

This dissertation focuses on the development and application of mass spectrometry-based method for the large-scale analysis of protein folding stability. One goal of this work involved further advancing the capabilities of several existing methods by developing new protocols. The second goal of this work involved using these newly developed protocols in combination with other more established protocols to identify the protein targets of two drugs with unknown modes of action and to discovery disease biomarkers of Parkinson's disease.

Chapter 2 of the dissertation describes the development of a chemo-selective enrichment strategy to isolate the semi-tryptic peptides generated in mass spectrometry-based applications of limited proteolysis methods. The motivation of this development is to improve the bottom-up proteomics readout in proteome-wide limited proteolysis experiments. The coupling of the STEPP protocol to LiP, DARTS and PP enables these methods to generate large-scale protein folding stability information without the interference of tryptic peptides which drastically reduces the information density and quality of these experiments. The STEPP-LiP method could be particularly useful since it provides a tool of proteome-wide identification of subtle protein conformational changes that are ignored in other methods such as TPP, CPP and SPROX. The STEPP-LiP should be widely adopted to study protein structural changes related to environmental perturbation and protein aggregation where the protein structural changes are usually subtle and not denaturant dependent.

In Chapter 3 and 4, the STEPP-PP and STEPP-LiP methods are used in combination with SPROX, TPP and CPP strategies to identify the protein target of two drugs of unknown mode of action, including Sub A and Man A. The goal is to identify consistent protein hits that are identified in more than two methods to reduce the chance of false discovery.

In the Sub A target identification study described in Chapter 3, only DCK is consistently stabilized with Sub A in a total of more than 2000 assayed proteins. Further gel-based targeted pulse proteolysis method supports that DCK is a direct binding target of subglutinin A with an estimated K_d of 0.21 μM . To understand the downstream effect of Sub A-DCK binding and its role in Sub A's mode of action, a series of chemical biology experiments were employed. These experiments demonstrated that Sub A does not alter DCK's activity, expression level or phosphorylation level but blocks its nuclear transportation. This reduction of DCK level in the cell nucleus possibly leads to the observed reduction of nuclear dCMP pool. Further study is planned to determine if the observed reduction in nuclear dCMP pool is indeed related to DCK but not just a general phenomenon of cell death.

In the Man A target identification study described in Chapter 4, more than 4000 proteins were assayed but unfortunately, no consistent protein hit is found. There were 4 protein hits with unknown consistency because they were only assayed in a single method. Future work on Man A target identification could either focus on validation of these 4 proteins or application of more large-scale experiments to search for potential targets not covered by experiments of this work.

In biomarker discovery part of this dissertation, the SPROX method is utilized in Chapter 5 to characterize the progression of PD in brain tissue cell lysates from Hu α -Syn(A53T) transgenic mice. Thermodynamic stabilities of 332 proteins in brain tissue cell lysates from these transgenic mice were compared to similarly aged wild-type mice using a two-way ANOVA analysis. A group of 11 proteins were found to be differential stabilized in the Hu α -Syn(A53T) transgenic mouse model. The potential of using these proteins as PD biomarkers could be further tested in patient-derived samples, for example cerebrospinal fluids. This study also encourages the application of other protein folding stability approaches, such as TPP, in disease biomarker discovery. Future studies could be employed to determine which method is the best in protein biomarker discovery, or a combination of these methods are needed.

In conclusion, this dissertation demonstrated the development of a novel chemoselection strategy, STEPP, for large-scale limited proteolysis analyses, as well as application of protein folding stability approaches to drug target identification and disease biomarker discovery. The drug target identification works have demonstrated that even with these state-of-the-art methods in hand, it is not easy to pin down the target proteins from a complex proteome. Using a combination of approaches helps in reducing the list of false positives, which is indeed very beneficial in freeing researchers from the time-consuming and difficult drug binding validation experiments. However, future work is still needed to understand one key questions in drug target identification. First, what combination of these methods should be the general choice of studying a novel drug or lead molecule of known mode of action. Should one method be more preferred than

the other and being always included? And should the one-pot strategy be used rather than use the original non-one-pot version? These questions should be answered by more application of these protein folding stability approaches in drug target identification studies. Being able to find a general strategy towards drug target identification is indeed very important in improving our ability to discovery new drugs and advancing the pharmaceutical industry.

Appendix Determination of Pathogenic Mutations on the Folding Stability of Human Transthyretin

The work described in this appendix comes largely from the research paper titled “Pathogenic Mutations Induce Partial Structural Changes in Native β -Sheet Structure of Transthyretin and Accelerate Aggregation” that was published in 2017 on *Biochemistry* (vol. 56(36), p. 4808–4818).¹⁴¹ This work was performed in collaboration with Professor Kwang Hun Lim and his group at Eastern Carolina University.

Introduction

Transthyretin (TTR, Figure S 1) is a tetrameric protein that undergo aberrant conformational transitions from natively folded states into insoluble β -structured aggregates *in vivo*.^{142–144} Extracellular deposition of TTR aggregates in various organs including heart, lung, and peripheral nerves is associated with senile systemic amyloidosis, familial amyloidotic polyneuropathy, familial amyloid cardiomyopathy, and rarely central nervous system selective amyloidosis.^{145–147} The key to understand the amyloidogenesis of TTR is to illustrate the structure of TTR in its amyloidogenic state.



Figure S 1 Protein structure of WT TTR monomer in its native form.^{141,148}

To uncover the structure, our collaborators' preliminary NMR studies suggested the WT TTR aggregates contained an amyloid core consisting of a native-like folded structure. The mutant TTR amyloids adopted a similar structure, but the strand D might be unfolded in amyloid state, which may fully expose strand A for intermolecular interactions.

The goal of the work in this appendix was to collect biophysical data to confirm that the WT and mutant TTRs in their amyloidogenic states have stable folded structures. To confirm this, the thermodynamic folding stability of these proteins were measured. The measurements were accomplished with a previously described MALDI-TOF-based protocol that utilizes the reaction of dimethyl(2-hydroxy-5-nitrobenzyl)sulfonium bromide (HNSB) with globally protected tryptophan (Trp) residues.¹⁹ In order to probe the local environment of the strand A and D of amyloidogenic V30M TTR, oxidation rates of a methionine sidechain on strand A were analyzed using mass spectrometry.

Experimental Section

HNSB Analysis

The chemical denaturation curves of WT and mutant TTR proteins generated here were obtained using a previously established tryptophan modification protocol involving dimethyl(2-hydroxy-5-nitrobenzyl)sulfonium bromide (HNSB).¹⁹ Briefly, the HNSB experiments were performed by reacting the protein (1 μ M) with HNSB (3.68mM) in a series of different buffers (pH 4.4) that contained increasing concentration of guanidinium chloride (GdmCl). The HNSB modification reactions in each buffer were allowed to proceed for 10 min at 37 °C. The reactions were quenched during a desalting step using

Millipore C18 Zip-Tips. The desalted protein samples were each analyzed by MALDI-TOF mass spectrometry using sinapinic acid as the matrix. The MALDI mass spectra were acquired on an UltraFlex II TOF/TOF (Bruker Daltonics) mass spectrometer in the reflective and positive ion mode using a smartbeam Nd:YAG laser (355 nm). Spectra were collected using the following instrument parameters: 25kV ion source 1 voltage, 21.9kV ion source 2 voltage, 9.9kV lens voltage, 26.3kV reflector voltage, 13.7kV reflector 2 voltage, 100ns pulsed ion extraction, and matrix gating to 4000 Da. The protein ion signal in the resulting mass spectra were used to extract $\Delta\text{Mass}_{\text{wt,av}}$ values as we have previously described.¹⁹

Time Course Oxidation Analysis

The time course studies of Met-30 oxidation in V30M TTR were performed by reacting the protein (1 μM) with H_2O_2 (100 mM) at pH 4.4 (200 mM acetate buffer) or pH 7.4 (200 mM phosphate buffer) for a specific of time at 37 °C. The reaction was quenched by adding a 6-fold molar excess of L-methionine over H_2O_2 . A buffer exchange was performed using Amicon Ultra 10K centrifugal filters to obtain protein solutions in 100 mM triethylammonium bicarbonate buffer (pH 8.5). The protein solutions were digested with Lys-C protease, and the resulting peptide solutions were analyzed by MALDI-TOF mass spectrometry using α -cyano-4-hydroxycinnamic acid (HCCA) as the matrix for MALDI-MS analysis. For determination of the oxidation status of Met-30 in the V30M mutant, the MALDI analysis was focused on the ion signals from the Met-30 containing peptide VLDAVRGSPAINVAMHVFRK in both its wild-type and oxidized forms. The MALDI ion signals of the wild-type and oxidized forms of the

VLDAVRGSPAINVAMHVFRK peptide were identified by their masses. MALDI mass spectra for the time-course experiment were acquired on a Voyager-DE PRO (Applied Biosystems) mass spectrometer in the linear and positive ion mode using a nitrogen laser (337 nm). Spectra were collected using the following instrument parameters: manual control mode, 20kV accelerating voltage 93% grid, 0.05% guide wire, 100 nsec delay time, 60 shots/spectrum, and low mass gate 500 Da.

The time-course data were analyzed as previously described.¹⁹ Briefly, the increases in mass as a function of reaction times were obtained by calculating $\Delta\text{Mass}_{\text{wt,av}}$ values, using a weighted average of the intensities of the ion signals from the wild-type and oxidized peptide, and by fitting the data to a single exponential equation.

Results and Discussion

Folding Stability Analysis of Amyloidogenic TTRs

The amyloidogenic TTRs were provided by our collaborators at Eastern Carolina University. Thermodynamic stabilities of the amyloidogenic states of TTR were examined using a chemical modification and mass spectrometry-based approach, which has been previously described.¹⁹ In this mass spectrometry-based analysis, we monitored the unfolding transition of the precursor states by evaluating the chemical denaturant dependence of the reaction of dimethyl(2-hydroxy-5-nitrobenzyl)sulfonium bromide (HNSB) with tryptophan (Trp) residues. At increasing concentrations of chemical denaturant, the TTR protein is unfolded and the Trp residues are more exposed to solvent and more readily react with HNSB, resulting in an increased mass of the protein that can be detected by using mass spectrometry. In the natively folded state of TTR, both Trp

residues (W41 and W79) appear to be largely protected from the reaction with HNSB, as evidenced by the relatively low $\Delta\text{Mass}_{\text{wt,av}}$ values of the pre-transition baselines of the chemical denaturation curves in Figure S 2. In the presence of increasing concentrations of chemical denaturant (GdmCl in this work), the protected Trp residues in the native state become available for the chemical reaction with HNSB, leading to an increase in the protein mass, as demonstrated in Figure S 2. Thus, the chemical denaturant dependence of the HNSB modification reaction can report on the thermodynamic stability of the WT and mutant forms of TTR. The cooperative unfolding transitions observed in Figure S 2 suggest that the amyloidogenic precursor states adopt folded conformations. A qualitative analysis of the transition midpoints of the chemical denaturation curves generated in Figure S 2 also indicates that the precursor states of the TTR variants are less stable than that of WT TTR. The less stable amyloidogenic precursor states of the TTR variants might be linked to the stronger aggregation propensity of the TTR variants.

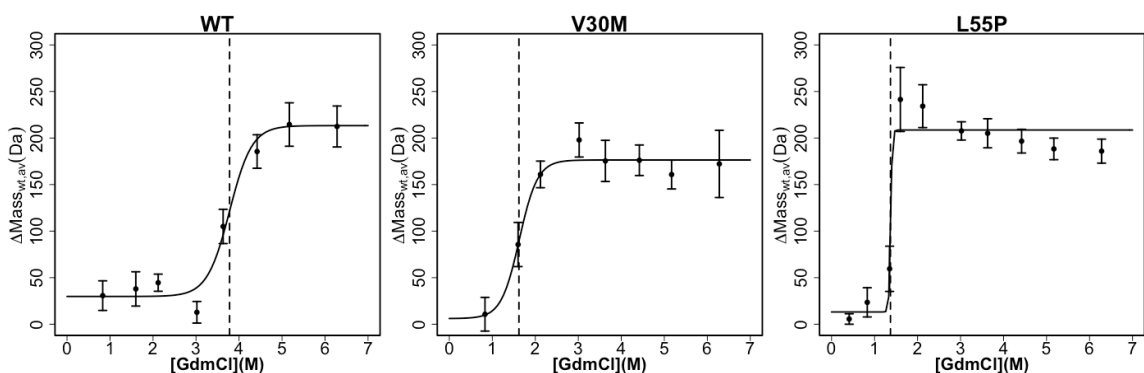


Figure S 2 Chemical denaturation curves generated for WT, V30M, and L55P TTR using the tryptophan modification protocol. The vertical dotted lines indicate the concentrations of guanidinium chloride (GdmCl) at the transition midpoints, which were 3.8, 1.6 and 1.4 M for WT, V30M, and L55P, respectively. The data points and error bars represent the average and standard deviation of the $\Delta\text{Mass}_{\text{wt,av}}$ values determined from the 8 mass spectra collected on each sample.

Analysis of Strand A and D Local Environment

The preliminary solid-state NMR experiments of V30M amyloid done by our collaborators suggest that strand D might be unfolded in amyloid state, which may fully expose strand A for intermolecular interactions. In order to probe the local environment of the methionine sidechain, oxidation rates of the Met-30 sidechain were analyzed using mass spectrometry.⁶³ Previous crystal structure of native tetrameric V30M TTR shows that the Met-30 sidechain, which is located on strand A, is protected by strand D in its non-amyloidogenic state.¹⁴⁹

In this mass spectrometry-based approach we measured the time-course of the H₂O₂ oxidation of the Met-30 sidechain. In the native tetrameric V30M TTR at pH 7.4, the sidechain of residue Met-30 is protected from the oxidation (Figure S 3). On the contrary, the Met-30 sidechain is readily oxidized at pH 4.4, suggesting that the M30 sidechain protected in the native state is exposed to solvent in the amyloidogenic precursor state at pH 4.4. These results confirm that strand D is unfolded in amyloid state, which expose strand A for intermolecular interactions.

We note that the H₂O₂ oxidation of an unprotected Met residue is not significantly different at pHs 4.4 and 7.4. Thus, the different reaction rates observed in Figure S 3 appear to be due to structural differences in the protein. We also note that the second-order reaction rate constant calculated from the time-course data on V30M TTR at pH 4.4 and 37 °C (63 M⁻¹h⁻¹) was only slightly smaller than that measured for an unprotected methionine in the model peptide under the identical condition (82 M⁻¹h⁻¹).¹⁵⁰

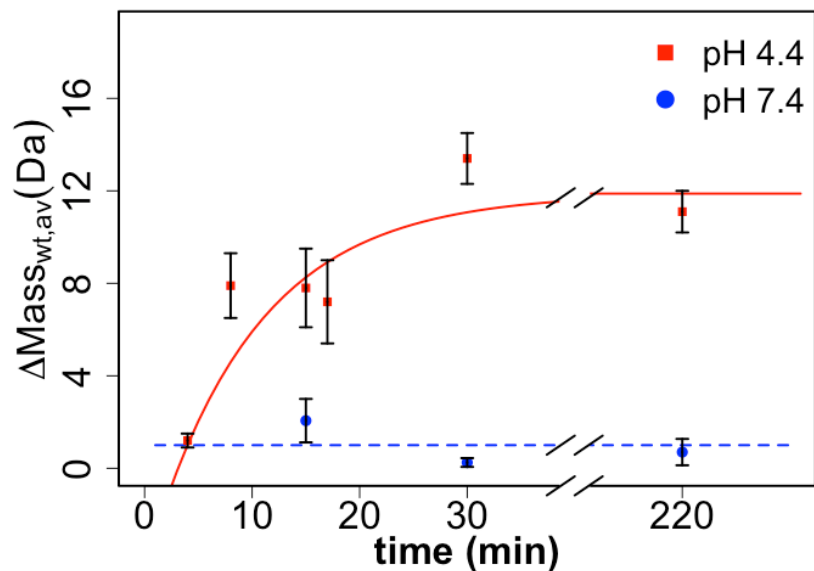


Figure S 3 Time-course analysis of Met-30 side chain oxidation with H₂O₂ in V30M TTR. Data obtained on the Lys-C-generated TTR peptide, VLDAVRGSPAINVAMHVFRK, which contains residue Met-30, are shown. The data points and error bars represent the average and standard deviation of the $\Delta\text{Mass}_{\text{wt,av}}$ values determined from the five mass spectra collected on each sample. The dotted blue line represents the best fit of the pH 7.4 data to a line. The solid red line represents the best fit of the data to a single-exponential equation.

Conclusion

In this study, mass spectrometry-based methods were employed to investigate structural features of amyloidogenic precursor states of TTR using a very low protein concentration (1 μM) to minimize aggregation. HNSB analyses of WT and mutant TTRs at pH 4.4 showed the amyloidogenic precursor states of TTRs adopt largely folded conformations. The analyses also confirmed the mutant TTRs are less stable than the WT. The time-course oxidation of Met-30 sidechain clearly showed that the Met-30 residue at

its native state (i.e., at pH 7.4) is protected from oxidation while in its amyloidogenic precursor state (i.e., at pH 4.4) is not protected. This confirmed our collaborators hypothesis that the local environment of strand DA, where Met-30 is located, has changed in the transition from native state to amyloidogenic precursor state. The findings of the protein conformation of TTR in its amyloidogenic precursor state are important towards the understanding of the amyloidogenesis of TTR.

References

- (1) Kaur, U.; Meng, H.; Liu, F.; Ma, R.; Ogburn, R. N.; Johnson, J. H. R.; Fitzgerald, M. C.; Jones, L. M. Proteome-Wide Structural Biology: An Emerging Field for the Structural Analysis of Proteins on the Proteomic Scale. *J. Proteome Res.* **2018**, *17*, acs.jpoteome.8b00341. <https://doi.org/10.1021/acs.jpoteome.8b00341>.
- (2) Ma, R.; Meng, H.; Wiebelhaus, N.; Fitzgerald, M. C. Chemo-Selection Strategy for Limited Proteolysis Experiments on the Proteomic Scale. *Anal. Chem.* **2018**, acs.analchem.8b04122. <https://doi.org/10.1021/acs.analchem.8b04122>.
- (3) Cabrera, A.; Wiebelhaus, N.; Quan, B.; Ma, R.; Meng, H.; Fitzgerald, M. C. Comparative Analysis of Mass-Spectrometry-Based Proteomic Methods for Protein Target Discovery Using a One-Pot Approach. *J. Am. Soc. Mass Spectrom.* **2020**. <https://doi.org/10.1021/jasms.9b00041>.
- (4) Brandts, J. F.; Lin, L. N. Study of Strong to Ultratight Protein Interactions Using Differential Scanning Calorimetry. *Biochemistry* **1990**. <https://doi.org/10.1021/bi00481a024>.
- (5) Robertson, A. D.; Murphy, K. P. Protein Structure and the Energetics of Protein Stability. *Chem. Rev.* **1997**. <https://doi.org/10.1021/cr960383c>.
- (6) Straume, M.; Freire, E. Two-Dimensional Differential Scanning Calorimetry: Simultaneous Resolution of Intrinsic Protein Structural Energetics and Ligand Binding Interactions by Global Linkage Analysis. *Anal. Biochem.* **1992**. [https://doi.org/10.1016/0003-2697\(92\)90311-T](https://doi.org/10.1016/0003-2697(92)90311-T).
- (7) Saito, Y.; Wada, A. Comparative Study of GuHCl Denaturation of Globular Proteins. II. A Phenomenological Classification of Denaturation Profiles of 17 Proteins. *Biopolymers* **1983**. <https://doi.org/10.1002/bip.360220910>.
- (8) Hill, J. J.; Royer, C. A. Fluorescence Approaches to Study of Protein-Nucleic Acid Complexation. *Methods Enzymol.* **1997**. [https://doi.org/10.1016/S0076-6879\(97\)78021-2](https://doi.org/10.1016/S0076-6879(97)78021-2).
- (9) Julien, O.; Graether, S. P.; Sykes, B. D. Monitoring Prion Protein Stability by NMR. *J. Toxicol. Environ. Heal. - Part A Curr. Issues* **2009**. <https://doi.org/10.1080/15287390903084504>.
- (10) Van Mierlo, C. P. M.; Steensma, E. Protein Folding and Stability Investigated by Fluorescence, Circular Dichroism (CD), and Nuclear Magnetic Resonance (NMR) Spectroscopy: The Flavodoxin Story. *Journal of Biotechnology*; **2000**. [https://doi.org/10.1016/S0168-1656\(00\)00244-3](https://doi.org/10.1016/S0168-1656(00)00244-3).

- (11) Johnson, C. M. Differential Scanning Calorimetry as a Tool for Protein Folding and Stability. *Archives of Biochemistry and Biophysics*. **2013**. <https://doi.org/10.1016/j.abb.2012.09.008>.
- (12) Durowoju, I. B.; Bhandal, K. S.; Hu, J.; Carpick, B.; Kirkitadze, M. Differential Scanning Calorimetry — A Method for Assessing the Thermal Stability and Conformation of Protein Antigen. *J. Vis. Exp.* **2017**. <https://doi.org/10.3791/55262>.
- (13) Rauniyar, N.; Yates, J. R. Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics. *Journal of Proteome Research*. **2014**. <https://doi.org/10.1021/pr500880b>.
- (14) Thermo Fisher Scientific. Tandem Mass Tag (TMT) Systems <https://www.thermofisher.com/us/en/home/life-science/protein-biology/protein-mass-spectrometry-analysis/protein-quantitation-mass-spectrometry/tandem-mass-tag-systems.html> (accessed Jul 3, 2020).
- (15) Sciex. iTRAQ® Reagents <https://sciex.com/products/consumables/itraq-reagent> (accessed Jul 3, 2020).
- (16) Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* **2003**. <https://doi.org/10.1021/ac0262560>.
- (17) Ross, P. L.; Huang, Y. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; et al. Multiplexed Protein Quantitation in *Saccharomyces Cerevisiae* Using Amine-Reactive Isobaric Tagging Reagents. *Mol. Cell. Proteomics* **2004**. <https://doi.org/10.1074/mcp.M400129-MCP200>.
- (18) Strickland, E. C.; Geer, M. A.; Tran, D. T.; Adhikari, J.; West, G. M.; Dearmond, P. D.; Xu, Y.; Fitzgerald, M. C. Thermodynamic Analysis of Protein-Ligand Binding Interactions in Complex Biological Mixtures Using the Stability of Proteins from Rates of Oxidation. *Nat. Protoc.* **2013**. <https://doi.org/10.1038/nprot.2012.146>.
- (19) Xu, Y.; Strickland, E. C.; Fitzgerald, M. C. Thermodynamic Analysis of Protein Folding and Stability Using a Tryptophan Modification Protocol. *Anal. Chem.* **2014**. <https://doi.org/10.1021/ac501278j>.
- (20) Meng, H.; Ma, R.; Fitzgerald, M. C. Chemical Denaturation and Protein Precipitation Approach for Discovery and Quantitation of Protein-Drug Interactions. *Anal. Chem.* **2018**. <https://doi.org/10.1021/acs.analchem.8b01772>.
- (21) Park, C.; Marqusee, S. Pulse Proteolysis: A Simple Method for Quantitative Determination of Protein Stability and Ligand Binding. *Nat. Methods* **2005**, 2 (3),

207–212. <https://doi.org/10.1038/nmeth740>.

- (22) Liu, P. F.; Kihara, D.; Park, C. Energetics-Based Discovery of Protein-Ligand Interactions on a Proteomic Scale. *J. Mol. Biol.* **2011**. <https://doi.org/10.1016/j.jmb.2011.02.026>.
- (23) Chang, Y.; Schleich, J. P.; VerHeul, R. A.; Park, C. Simplified Proteomics Approach to Discover Protein-Ligand Interactions. *Protein Sci.* **2012**. <https://doi.org/10.1002/pro.2112>.
- (24) Adhikari, J.; Fitzgerald, M. C. SILAC-Pulse Proteolysis: A Mass Spectrometry-Based Method for Discovery and Cross-Validation in Proteome-Wide Studies of Ligand Binding. *J. Am. Soc. Mass Spectrom.* **2014**. <https://doi.org/10.1007/s13361-014-0992-y>.
- (25) Feng, Y.; De Franceschi, G.; Kahraman, A.; Soste, M.; Melnik, A.; Boersema, P. J.; De Laureto, P. P.; Nikolaev, Y.; Oliveira, A. P.; Picotti, P. Global Analysis of Protein Structural Changes in Complex Proteomes. *Nat. Biotechnol.* **2014**. <https://doi.org/10.1038/nbt.2999>.
- (26) Lomenick, B.; Hao, R.; Jonai, N.; Chin, R. M.; Aghajan, M.; Warburton, S.; Wang, J.; Wu, R. P.; Gomez, F.; Loo, J. A.; et al. Target Identification Using Drug Affinity Responsive Target Stability (DARTS). *Proc. Natl. Acad. Sci. U. S. A.* **2009**. <https://doi.org/10.1073/pnas.0910040106>.
- (27) Liu, F.; Meng, H.; Fitzgerald, M. C. Large-Scale Analysis of Breast Cancer-Related Conformational Changes in Proteins Using SILAC-SPROX. *J. Proteome Res.* **2017**, *16* (9), 3277–3286. <https://doi.org/10.1021/acs.jproteome.7b00283>.
- (28) Meng, H.; Fitzgerald, M. C. Proteome-Wide Characterization of Phosphorylation-Induced Conformational Changes in Breast Cancer. *J. Proteome Res.* **2018**. <https://doi.org/10.1021/acs.jproteome.7b00795>.
- (29) Geiger, R.; Rieckmann, J. C.; Wolf, T.; Basso, C.; Feng, Y.; Fuhrer, T.; Kogadeeva, M.; Picotti, P.; Meissner, F.; Mann, M.; et al. L-Arginine Modulates T Cell Metabolism and Enhances Survival and Anti-Tumor Activity. *Cell* **2016**. <https://doi.org/10.1016/j.cell.2016.09.031>.
- (30) Kim, D.; Hwang, H. Y.; Kim, J. Y.; Lee, J. Y.; Yoo, J. S.; Marko-Varga, G.; Kwon, H. J. FK506, an Immunosuppressive Drug, Induces Autophagy by Binding to the V-ATPase Catalytic Subunit A in Neuronal Cells. *J. Proteome Res.* **2017**. <https://doi.org/10.1021/acs.jproteome.6b00638>.
- (31) Zeng, L.; Shin, W. H.; Zhu, X.; Park, S. H.; Park, C.; Tao, W. A.; Kihara, D. Discovery of Nicotinamide Adenine Dinucleotide Binding Proteins in the Escherichia Coli

- Proteome Using a Combined Energetic- and Structural-Bioinformatics-Based Approach. *J. Proteome Res.* **2017**, *16* (2), 470–480. <https://doi.org/10.1021/acs.jproteome.6b00624>.
- (32) Trindade, R. V.; Pinto, A. F. M.; Santos, D. S.; Bizarro, C. V. Pulse Proteolysis and Precipitation for Target Identification. *J. Proteome Res.* **2016**, *15* (7), 2236–2245. <https://doi.org/10.1021/acs.jproteome.6b00214>.
- (33) Liu, Y. K.; Chen, H. Y.; Chueh, P. J.; Liu, P. F. A One-Pot Analysis Approach to Simplify Measurements of Protein Stability and Folding Kinetics. *Biochim. Biophys. Acta - Proteins Proteomics* **2019**, *1867* (3), 184–193. <https://doi.org/10.1016/j.bbapap.2018.12.006>.
- (34) Gaetani, M.; Sabatier, P.; Saei, A. A.; Beusch, C.; Yang, Z.; Lundstrom, S.; Zubarev, R. Proteome Integral Stability Alteration Assay Dramatically Increases Throughput and Sensitivity in Profiling Factor-Induced Proteome Changes. *bioRxiv* **2018**. <https://doi.org/10.1101/496398>.
- (35) DiMasi, J. A.; Grabowski, H. G.; Hansen, R. W. Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs. *J. Health Econ.* **2016**, *47*, 20–33. <https://doi.org/10.1016/j.jhealeco.2016.01.012>.
- (36) Hughes, J. P.; Rees, S. S.; Kalindjian, S. B.; Philpott, K. L. Principles of Early Drug Discovery. *Br. J. Pharmacol.* **2011**, *162* (6), 1239–1249. <https://doi.org/10.1111/j.1476-5381.2010.01127.x>.
- (37) Terstappen, G. C.; Schlüpen, C.; Raggiaschi, R.; Gaviraghi, G. Target Deconvolution Strategies in Drug Discovery. *Nature Reviews Drug Discovery.* 2007. <https://doi.org/10.1038/nrd2410>.
- (38) Kubota, K.; Funabashi, M.; Ogura, Y. Target Deconvolution from Phenotype-Based Drug Discovery by Using Chemical Proteomics Approaches. *Biochim. Biophys. Acta - Proteins Proteomics* **2019**, *1867* (1), 22–27. <https://doi.org/10.1016/j.bbapap.2018.08.002>.
- (39) Schenone, M.; Dančik, V.; Wagner, B. K.; Clemons, P. A. Target Identification and Mechanism of Action in Chemical Biology and Drug Discovery. *Nat. Chem. Biol.* **2013**, *9* (4), 232–240. <https://doi.org/10.1038/nchembio.1199>.
- (40) Bateman, A. UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019**. <https://doi.org/10.1093/nar/gky1049>.
- (41) Husi, H.; Zurini, M. G. M. Comparative Binding Studies of Cyclophilins to Cyclosporine A and Derivatives by Fluorescence Measurements. *Anal. Biochem.* **1994**. <https://doi.org/10.1006/abio.1994.1481>.

- (42) Schlatter, D.; Thoma, R.; Küng, E.; Stihle, M.; Müller, F.; Borroni, E.; Cesura, A.; Hennig, M. Crystal Engineering Yields Crystals of Cyclophilin D Diffracting to 1.7 Å Resolution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2005**. <https://doi.org/10.1107/S0907444905003070>.
- (43) Hahn, R.; Berger, E.; Pfliegerl, K.; Jungbauer, A. Directed Immobilization of Peptide Ligands to Accessible Pore Sites by Conjugation with a Placeholder Molecule. *Anal. Chem.* **2003**. <https://doi.org/10.1021/ac025846v>.
- (44) Parkinson, J. *An Essay on Shaking Palsy*; 1817.
- (45) Poewe, W.; Seppi, K.; Tanner, C. M.; Halliday, G. M.; Brundin, P.; Volkmann, J.; Schrag, A. E.; Lang, A. E. Parkinson Disease. *Nat. Rev. Dis. Prim.* **2017**. <https://doi.org/10.1038/nrdp.2017.13>.
- (46) Ray Dorsey, E.; Elbaz, A.; Nichols, E.; Abd-Allah, F.; Abdelalim, A.; Adsuar, J. C.; Ansha, M. G.; Brayne, C.; Choi, J. Y. J.; Collado-Mateo, D.; et al. Global, Regional, and National Burden of Parkinson's Disease, 1990–2016: A Systematic Analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **2018**. [https://doi.org/10.1016/S1474-4422\(18\)30295-3](https://doi.org/10.1016/S1474-4422(18)30295-3).
- (47) Alcalay, R. N.; Caccappolo, E.; Mejia-Santana, H.; Tang, M. X.; Rosado, L.; Ross, B. M.; Verbitsky, M.; Kisselev, S.; Louis, E. D.; Comella, C.; et al. Frequency of Known Mutations in Early-Onset Parkinson Disease: Implication for Genetic Counseling: The Consortium on Risk for Early Onset Parkinson Disease Study. *Arch. Neurol.* **2010**. <https://doi.org/10.1001/archneurol.2010.194>.
- (48) Meade, R. M.; Fairlie, D. P.; Mason, J. M. Alpha-Synuclein Structure and Parkinson's Disease - Lessons and Emerging Principles. *Molecular Neurodegeneration.* 2019. <https://doi.org/10.1186/s13024-019-0329-1>.
- (49) Stefanis, L. α -Synuclein in Parkinson's Disease. *Cold Spring Harb. Perspect. Med.* **2012**. <https://doi.org/10.1101/cshperspect.a009399>.
- (50) Rizzo, G.; Copetti, M.; Arcuti, S.; Martino, D.; Fontana, A.; Logroscino, G. Accuracy of Clinical Diagnosis of Parkinson Disease. *Neurology* **2016**. <https://doi.org/10.1212/WNL.0000000000002350>.
- (51) George, J.; Mok, S.; Moses, D.; Wilkins, S.; Bush, A.; Cherny, R.; Finkelstein, D. Targeting the Progression of Parkinsons Disease. *Curr. Neuropharmacol.* **2009**. <https://doi.org/10.2174/157015909787602814>.
- (52) He, R.; Yan, X.; Guo, J.; Xu, Q.; Tang, B.; Sun, Q. Recent Advances in Biomarkers for Parkinson's Disease. *Frontiers in Aging Neuroscience.* **2018**. <https://doi.org/10.3389/fnagi.2018.00305>.

- (53) Roberts, J. H.; Liu, F.; Karnuta, J. M.; Fitzgerald, M. C. Discovery of Age-Related Protein Folding Stability Differences in the Mouse Brain Proteome. *J. Proteome Res.* **2016**, *15* (12), 4731–4741. <https://doi.org/10.1021/acs.jproteome.6b00927>.
- (54) Adhikari, J.; West, G. M.; Fitzgerald, M. C. Global Analysis of Protein Folding Thermodynamics for Disease State Characterization. *J. Proteome Res.* **2015**. <https://doi.org/10.1021/acs.jproteome.5b00057>.
- (55) Liu, F.; Meng, H.; Fitzgerald, M. C. Large-Scale Analysis of Breast Cancer-Related Conformational Changes in Proteins Using SILAC-SPROX. *J. Proteome Res.* **2017**. <https://doi.org/10.1021/acs.jproteome.7b00283>.
- (56) Myers, J. K.; Nick Pace, C.; Martin Scholtz, J. Denaturant m Values and Heat Capacity Changes: Relation to Changes in Accessible Surface Areas of Protein Unfolding. *Protein Sci.* **1995**. <https://doi.org/10.1002/pro.5560041020>.
- (57) Kleifeld, O.; Doucet, A.; Auf Dem Keller, U.; Prudova, A.; Schilling, O.; Kainthan, R. K.; Starr, A. E.; Foster, L. J.; Kizhakkedathu, J. N.; Overall, C. M. Isotopic Labeling of Terminal Amines in Complex Samples Identifies Protein N-Termini and Protease Cleavage Products. *Nat. Biotechnol.* **2010**. <https://doi.org/10.1038/nbt.1611>.
- (58) Handschumacher, R. E.; Harding, M. W.; Rice, J.; Drugge, R. J.; Speicher, D. W. Cyclophilin: A Specific Cytosolic Binding Protein for Cyclosporin A. *Science (80-.)*. **1984**. <https://doi.org/10.1126/science.6238408>.
- (59) Stebbins, C. E.; Russo, A. A.; Schneider, C.; Rosen, N.; Hartl, F. U.; Pavletich, N. P. Crystal Structure of an Hsp90-Geldanamycin Complex: Targeting of a Protein Chaperone by an Antitumor Agent. *Cell* **1997**. [https://doi.org/10.1016/S0092-8674\(00\)80203-2](https://doi.org/10.1016/S0092-8674(00)80203-2).
- (60) Meng, H.; Ma, R.; Fitzgerald, M. C. Chemical Denaturation and Protein Precipitation Approach for Discovery and Quantitation of Protein-Drug Interactions. *Anal. Chem.* **2018**, *90* (15), 9249–9255. <https://doi.org/10.1021/acs.analchem.8b01772>.
- (61) West, G. M.; Tucker, C. L.; Xu, T.; Park, S. K.; Han, X.; Yates, J. R.; Fitzgerald, M. C. Quantitative Proteomics Approach for Identifying Protein-Drug Interactions in Complex Mixtures Using Protein Stability Measurements. *Proc. Natl. Acad. Sci.* **2010**, *107* (20), 9078–9082. <https://doi.org/10.1073/pnas.1000148107>.
- (62) DeArmond, P. D.; West, G. M.; Huang, H. T.; Fitzgerald, M. C. Stable Isotope Labeling Strategy for Protein-Ligand Binding Analysis in Multi-Component Protein Mixtures. *J. Am. Soc. Mass Spectrom.* **2011**. <https://doi.org/10.1007/s13361-010-0060-1>.

- (63) Tran, D. T.; Adhikari, J.; Fitzgerald, M. C. 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* **2014**. <https://doi.org/10.1074/mcp.M113.034702>.
- (64) Xu, Y.; Wallace, M. A. G.; Fitzgerald, M. C. Thermodynamic Analysis of the Geldanamycin-Hsp90 Interaction in a Whole Cell Lysate Using a Mass Spectrometry-Based Proteomics Approach. *J. Am. Soc. Mass Spectrom.* **2016**. <https://doi.org/10.1007/s13361-016-1457-2>.
- (65) Haendler, B.; Keller, R.; Hiestand, P. C.; Kocher, H. P.; Wegmann, G.; Movva, N. R. Yeast Cyclophilin: Isolation and Characterization of the Protein, cDNA and Gene. *Gene* **1989**. [https://doi.org/10.1016/0378-1119\(89\)90401-0](https://doi.org/10.1016/0378-1119(89)90401-0).
- (66) Friedman, J.; Weissman, I. Two Cytoplasmic Candidates for Immunophilin Action Are Revealed by Affinity for a New Cyclophilin: One in the Presence and One in the Absence of CsA. *Cell* **1991**. [https://doi.org/10.1016/0092-8674\(91\)90123-G](https://doi.org/10.1016/0092-8674(91)90123-G).
- (67) Prodromou, C.; Roe, S. M.; O'Brien, R.; Ladbury, J. E.; Piper, P. W.; Pearl, L. H. Identification and Structural Characterization of the ATP/ADP-Binding Site in the Hsp90 Molecular Chaperone. *Cell* **1997**. [https://doi.org/10.1016/S0092-8674\(00\)80314-1](https://doi.org/10.1016/S0092-8674(00)80314-1).
- (68) Strickland, E. C.; Geer, M. A.; Hong, J.; Fitzgerald, M. C. False-Positive Rate Determination of Protein Target Discovery Using a Covalent Modification- and Mass Spectrometry-Based Proteomics Platform. *J. Am. Soc. Mass Spectrom.* **2014**. <https://doi.org/10.1007/s13361-013-0754-2>.
- (69) Savitski, M. M.; Reinhard, F. B. M.; Franken, H.; Werner, T.; Savitski, M. F.; Eberhard, D.; Molina, D. M.; Jafari, R.; Dovega, R. B.; Klaeger, S.; et al. Tracking Cancer Drugs in Living Cells by Thermal Profiling of the Proteome. *Science (80-.)*. **2014**, *346* (6205). <https://doi.org/10.1126/science.1255784>.
- (70) Gooljarsingh, L. T.; Fernandes, C.; Yan, K.; Zhang, H.; Grooms, M.; Johanson, K.; Sinnamon, R. H.; Kirkpatrick, R. B.; Kerrigan, J.; Lewis, T.; 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.* **2006**. <https://doi.org/10.1073/pnas.0602650103>.
- (71) Shiau, A. K.; Harris, S. F.; Southworth, D. R.; Agard, D. A. Structural Analysis of E. Coli Hsp90 Reveals Dramatic Nucleotide-Dependent Conformational Rearrangements. *Cell* **2006**. <https://doi.org/10.1016/j.cell.2006.09.027>.
- (72) Ali, M. M. U.; Mark Roe, S.; Vaughan, C. K.; Meyer, P.; Panaretou, B.; Piper, P. W.; Prodromou, C.; Pearl, L. H. Crystal Structure of an Hsp90-Nucleotide-P23/Sba1

- Closed Chaperone Complex. *Nature* **2006**. <https://doi.org/10.1038/nature04716>.
- (73) Kamal, A.; Thao, L.; Sensintaffar, J.; Zhang, L.; Boehm, M. F.; Fritz, L. C.; Burrows, F. J. A High-Affinity Conformation of Hsp90 Confers Tumour Selectivity on Hsp90 Inhibitors. *Nature* **2003**. <https://doi.org/10.1038/nature01913>.
- (74) Burkhart, J. M.; Schumbrutzki, C.; Wortelkamp, S.; Sickmann, A.; Zahedi, R. P. Systematic and Quantitative Comparison of Digest Efficiency and Specificity Reveals the Impact of Trypsin Quality on MS-Based Proteomics. *J. Proteomics* **2012**. <https://doi.org/10.1016/j.jprot.2011.11.016>.
- (75) Meng, H. Proteomic Methods and Applications for Protein Folding Stability Measurements, Duke University, 2019.
- (76) Wang, L.; Wang, F. S.; Gershwin, M. E. Human Autoimmune Diseases: A Comprehensive Update. *Journal of Internal Medicine*. 2015. <https://doi.org/10.1111/joim.12395>.
- (77) Balagué, C.; Kunkel, S. L.; Godessart, N. Understanding Autoimmune Disease: New Targets for Drug Discovery. *Drug Discovery Today*. 2009. <https://doi.org/10.1016/j.drudis.2009.07.002>.
- (78) Goodnow, C. C.; Sprent, J.; Barbara, B. F.; Vinuesa, C. G. Cellular and Genetic Mechanisms of Self Tolerance and Autoimmunity. *Nature*. 2005. <https://doi.org/10.1038/nature03724>.
- (79) Sathish, J. G.; Sethu, S.; Bielsky, M. C.; De Haan, L.; French, N. S.; Govindappa, K.; Green, J.; Griffiths, C. E. M.; Holgate, S.; Jones, D.; et al. Challenges and Approaches for the Development of Safer Immunomodulatory Biologics. *Nat. Rev. Drug Discov.* **2013**. <https://doi.org/10.1038/nrd3974>.
- (80) Lee, J. C.; Lobkovsky, E.; Pliam, N. B.; Strobel, G.; Clardy, J. Subglutinols a and b: Immunosuppressive Compounds from the Endophytic Fungus *Fusarium Subglutinans*. *J. Org. Chem.* **1995**, *60* (22), 7076–7077. <https://doi.org/10.1021/jo00127a001>.
- (81) Tao, X.; Lipsky, P. E. The Chinese Anti-Inflammatory and Immunosuppressive Herbal Remedy *Tripterygium Wilfordii* Hook F. *Rheum. Dis. Clin. North Am.* **2000**. [https://doi.org/10.1016/S0889-857X\(05\)70118-6](https://doi.org/10.1016/S0889-857X(05)70118-6).
- (82) Lin, R.; Kim, H.; Hong, J.; Li, Q. J. Biological Evaluation of Subglutininol a as a Novel Immunosuppressive Agent for Inflammation Intervention. *ACS Med. Chem. Lett.* **2014**, *5* (5), 485–490. <https://doi.org/10.1021/ml4004809>.
- (83) Ogburn, R. N.; Jin, L.; Meng, H.; Fitzgerald, M. C. Discovery of Tamoxifen and N-

- Desmethyl Tamoxifen Protein Targets in MCF-7 Cells Using Large-Scale Protein Folding and Stability Measurements. *J. Proteome Res.* **2017**. <https://doi.org/10.1021/acs.jproteome.7b00442>.
- (84) McDowell, G. S.; Gaun, A.; Steen, H. IFASP: Combining Isobaric Mass Tagging with Filter-Aided Sample Preparation. *J. Proteome Res.* **2013**. <https://doi.org/10.1021/pr400032m>.
- (85) Savitski, M. M.; Reinhard, F. B. M.; Franken, H.; Werner, T.; Savitski, M. F.; Eberhard, D.; Molina, D. M.; Jafari, R.; Dovega, R. B.; Klaeger, S.; et al. Tracking Cancer Drugs in Living Cells by Thermal Profiling of the Proteome. *Science (80-.)*. **2014**. <https://doi.org/10.1126/science.1255784>.
- (86) Nelder, J. A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**. <https://doi.org/10.1093/comjnl/7.4.308>.
- (87) Franken, H.; Mathieson, T.; Childs, D.; Sweetman, G. M. A.; Werner, T.; Tögel, I.; Doce, C.; Gade, S.; Bantscheff, M.; Drewes, G.; et al. Thermal Proteome Profiling for Unbiased Identification of Direct and Indirect Drug Targets Using Multiplexed Quantitative Mass Spectrometry. *Nat. Protoc.* **2015**. <https://doi.org/10.1038/nprot.2015.101>.
- (88) Smal, C.; Cardoen, S.; Bertrand, L.; Delacauw, A.; Ferrant, A.; Van Den Berghe, G.; Neste, E. Van Den; Bontemps, F. Activation of Deoxycytidine Kinase by Protein Kinase Inhibitors and Okadaic Acid in Leukemic Cells. *Biochem. Pharmacol.* **2004**, *68* (1), 95–103. <https://doi.org/10.1016/j.bcp.2004.02.031>.
- (89) Suzuki, K.; Bose, P.; Leong-Quong, R. Y.; Fujita, D. J.; Riabowol, K. REAP: A Two Minute Cell Fractionation Method. *BMC Res. Notes* **2010**. <https://doi.org/10.1186/1756-0500-3-294>.
- (90) Savitski, M. M.; Reinhard, F. B. M.; Franken, H.; Werner, T.; Savitski, M. F.; Eberhard, D.; Molina, D. M.; Jafari, R.; Dovega, R. B.; Klaeger, S.; et al. Tracking Cancer Drugs in Living Cells by Thermal Profiling of the Proteome. *Science (80-.)*. **2014**, *346* (6205), 1–33. <https://doi.org/10.1126/science.1255784>.
- (91) Meng, H.; Ma, R.; Fitzgerald, M. C. Chemical Denaturation and Protein Precipitation Approach for Discovery and Quantitation of Protein-Drug Interactions. *Anal. Chem.* **2018**, *90* (15), 9249–9255. <https://doi.org/10.1021/acs.analchem.8b01772>.
- (92) Reichard, P. Interactions Between Deoxyribonucleotide And DNA Synthesis. *Annu. Rev. Biochem.* **1988**. <https://doi.org/10.1146/annurev.biochem.57.1.349>.
- (93) Nathanson, D. A.; Armijo, A. L.; Tom, M.; Li, Z.; Dimitrova, E.; Austin, W. R.;

- Nomme, J.; Campbell, D. O.; Ta, L.; Le, T. M.; et al. Co-Targeting of Convergent Nucleotide Biosynthetic Pathways for Leukemia Eradication. *J. Exp. Med.* **2014**. <https://doi.org/10.1084/jem.20131738>.
- (94) Choi, O.; Heathcote, D. A.; Ho, K.-K.; Muller, P. J.; Ghani, H.; Lam, E. W.-F.; Ashton-Rickardt, P. G.; Rutschmann, S. A Deficiency in Nucleoside Salvage Impairs Murine Lymphocyte Development, Homeostasis, and Survival. *J. Immunol.* **2012**, *188* (8), 3920–3927. <https://doi.org/10.4049/jimmunol.1102587>.
- (95) Hazra, S.; Szewczak, A.; Ort, S.; Konrad, M.; Lavie, A. Post-Translational Phosphorylation of Serine 74 of Human Deoxycytidine Kinase Favors the Enzyme Adopting the Open Conformation Making It Competent for Nucleoside Binding and Release. *Biochemistry* **2011**. <https://doi.org/10.1021/bi2001032>.
- (96) MacConnell, W. P.; Kaplan, N. O. The Role of Ethanol Extractable Proteins from the 80S Rat Liver Ribosome. *Biochem. Biophys. Res. Commun.* **1980**. [https://doi.org/10.1016/0006-291X\(80\)91517-X](https://doi.org/10.1016/0006-291X(80)91517-X).
- (97) MacConnell, W. P.; Kaplan, N. O. The Activity of the Acidic Phosphoproteins from the 80 S Rat Liver Ribosome. *J. Biol. Chem.* **1982**.
- (98) Hatzis, P.; Al-Madhoon, A. S.; Jüllig, M.; Petrakis, T. G.; Eriksson, S.; Talianidis, I. The Intracellular Localization of Deoxycytidine Kinase. *J. Biol. Chem.* **1998**, *273* (46), 30239–30243. <https://doi.org/10.1074/jbc.273.46.30239>.
- (99) Kong, Z.; Jia, S.; Chabes, A. L.; Appelblad, P.; Lundmark, R.; Moritz, T.; Chabes, A. Simultaneous Determination of Ribonucleoside and Deoxyribonucleoside Triphosphates in Biological Samples by Hydrophilic Interaction Liquid Chromatography Coupled with Tandem Mass Spectrometry. *Nucleic Acids Res.* **2018**, *46* (11), e66–e66. <https://doi.org/10.1093/nar/gky203>.
- (100) Chottiner, E. G.; Shewach, D. S.; Datta, N. S.; Ashcraft, E.; Gribbin, D.; Ginsburg, D.; Fox, I. H.; Mitchell, B. S. Cloning and Expression of Human Deoxycytidine Kinase CDNA. *Proc. Natl. Acad. Sci. U. S. A.* **1991**. <https://doi.org/10.1073/pnas.88.4.1531>.
- (101) Hossain, C. F.; Kim, Y. P.; Baerson, S. R.; Zhang, L.; Bruick, R. K.; Mohammed, K. A.; Agarwal, A. K.; Nagle, D. G.; Zhou, Y. D. Saururus Cernuus Lignans - Potent Small Molecule Inhibitors of Hypoxia-Inducible Factor-1. *Biochem. Biophys. Res. Commun.* **2005**. <https://doi.org/10.1016/j.bbrc.2005.05.191>.
- (102) Song, S. Y.; Lee, I.; Park, C.; Lee, H.; Hahm, J. C.; Kang, W. K. Neolignans from Saururus Chinensis Inhibit PC-3 Prostate Cancer Cell Growth via Apoptosis and Senescence-like Mechanisms. *Int. J. Mol. Med.* **2005**. <https://doi.org/10.3892/ijmm.16.4.517>.

- (103) Rao, K. V.; Alvarez, F. M. Manassantins A/B and Saucerneol: Novel Biologically Active Lignoids from *Saururus Cernuus*. *Tetrahedron Lett.* **1983**. [https://doi.org/10.1016/S0040-4039\(01\)99818-1](https://doi.org/10.1016/S0040-4039(01)99818-1).
- (104) Lee, Y. K.; Seo, C. S.; Lee, C. S.; Lee, K. S.; Kang, S. J.; Jahng, Y.; Chang, H. W.; Son, J. K. Inhibition of DNA Topoisomerases I and II and Cytotoxicity by Lignans from *Saururus Chinensis*. *Arch. Pharm. Res.* **2009**. <https://doi.org/10.1007/s12272-009-2010-7>.
- (105) Hahm, J. C.; Lee, I. K.; Kang, W. K.; Kim, S. U.; Ahn, Y. J. Cytotoxicity of Neolignans Identified in *Saururus Chinensis* towards Human Cancer Cell Lines. *Planta Med.* **2005**. <https://doi.org/10.1055/s-2005-864143>.
- (106) Ziello, J. E.; Jovin, I. S.; Huang, Y. Hypoxia-Inducible Factor (HIF)-1 Regulatory Pathway and Its Potential for Therapeutic Intervention in Malignancy and Ischemia. *Yale Journal of Biology and Medicine.* 2007.
- (107) Kaelin, W. G.; Ratcliffe, P. J. Oxygen Sensing by Metazoans: The Central Role of the HIF Hydroxylase Pathway. *Molecular Cell.* 2008. <https://doi.org/10.1016/j.molcel.2008.04.009>.
- (108) Moeller, B. J.; Richardson, R. A.; Dewhirst, M. W. Hypoxia and Radiotherapy: Opportunities for Improved Outcomes in Cancer Treatment. *Cancer and Metastasis Reviews.* 2007. <https://doi.org/10.1007/s10555-007-9056-0>.
- (109) Geer Wallace, M. A.; Kwon, D. Y.; Weitzel, D. H.; Lee, C. T.; Stephenson, T. N.; Chi, J. T.; Mook, R. A.; Dewhirst, M. W.; Hong, J.; Fitzgerald, M. C. Discovery of Manassantin A Protein Targets Using Large-Scale Protein Folding and Stability Measurements. *J. Proteome Res.* **2016**, *15* (8), 2688–2696. <https://doi.org/10.1021/acs.jproteome.6b00237>.
- (110) Ma, R.; Meng, H.; Wiebelhaus, N.; Fitzgerald, M. C. Chemo-Selection Strategy for Limited Proteolysis Experiments on the Proteomic Scale. *Anal. Chem.* **2018**, *90*, 14039–14047. <https://doi.org/10.1021/acs.analchem.8b04122>.
- (111) Lee, M. K.; Stirling, W.; Xu, Y.; Xu, E.; Qui, D.; Mandir, A. S.; Dawson, T. M.; Copeland, N. G.; Jenkins, N. A.; Price, D. L. Human α -Synuclein-Harboring Familial Parkinson's Disease-Linked Ala-53 \rightarrow Thr Mutation Causes Neurodegenerative Disease with α -Synuclein Aggregation in Transgenic Mice. *Proc. Natl. Acad. Sci. U. S. A.* **2002**. <https://doi.org/10.1073/pnas.132197599>.
- (112) Dearmond, P. D.; Xu, Y.; Strickland, E. C.; Daniels, K. G.; Fitzgerald, M. C. Thermodynamic Analysis of Protein-Ligand Interactions in Complex Biological Mixtures Using a Shotgun Proteomics Approach. *J. Proteome Res.* **2011**. <https://doi.org/10.1021/pr200403c>.

- (113) Macron, C.; Lane, L.; Núñez Galindo, A.; Dayon, L. Deep Dive on the Proteome of Human Cerebrospinal Fluid: A Valuable Data Resource for Biomarker Discovery and Missing Protein Identification. *J. Proteome Res.* **2018**. <https://doi.org/10.1021/acs.jproteome.8b00300>.
- (114) Vizcaíno, J. A.; Csordas, A.; Del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; et al. 2016 Update of the PRIDE Database and Its Related Tools. *Nucleic Acids Res.* **2016**. <https://doi.org/10.1093/nar/gkv1145>.
- (115) Bosley, T. M.; Salih, M. A.; Alorainy, I. A.; Islam, M. Z.; Oystreck, D. T.; Suliman, O. S. M.; Malki, S. Al; Suhaibani, A. H.; Khiari, H.; Beckers, S.; et al. The Neurology of Carbonic Anhydrase Type II Deficiency Syndrome. In *Brain*; 2011. <https://doi.org/10.1093/brain/awr302>.
- (116) Pollard, A.; Shephard, F.; Freed, J.; Liddell, S.; Chakrabarti, L. Mitochondrial Proteomic Profiling Reveals Increased Carbonic Anhydrase II in Aging and Neurodegeneration. *Aging (Albany, NY)*. **2016**. <https://doi.org/10.18632/aging.101064>.
- (117) Blanco-Rivero, A.; Shutova, T.; Román, M. J.; Villarejo, A.; Martinez, F. Phosphorylation Controls the Localization and Activation of the Luminal Carbonic Anhydrase in *Chlamydomonas Reinhardtii*. *PLoS One* **2012**. <https://doi.org/10.1371/journal.pone.0049063>.
- (118) Shi, M. M.; Shi, C. H.; Xu, Y. M. Rab GTPases: The Key Players in the Molecular Pathway of Parkinson's Disease. *Frontiers in Cellular Neuroscience*. 2017. <https://doi.org/10.3389/fncel.2017.00081>.
- (119) Zhang, X.; Zhou, J. Y.; Chin, M. H.; Schepmoes, A. A.; Petyuk, V. A.; Weitz, K. K.; Petritis, B. O.; Monroe, M. E.; Camp, D. G.; Wood, S. A.; et al. Region-Specific Protein Abundance Changes in the Brain of MPTP-Induced Parkinson's Disease Mouse Model. *J. Proteome Res.* **2010**, *9* (3), 1496–1509. <https://doi.org/10.1021/pr901024z>.
- (120) Xiao, M. F.; Xu, J. C.; Tereshchenko, Y.; Novak, D.; Schachner, M.; Kleene, R. Neural Cell Adhesion Molecule Modulates Dopaminergic Signaling and Behavior by Regulating Dopamine D2 Receptor Internalization. *J. Neurosci.* **2009**. <https://doi.org/10.1523/JNEUROSCI.4860-09.2009>.
- (121) Wang, K. Z. Q.; Steer, E.; Anthony Otero, P.; Bateman, N. W.; Cheng, M. H.; Scott, A. L.; Wu, C.; Bahar, I.; Shih, Y. T.; Hsueh, Y. P.; et al. PINK1 Interacts with VCP/P97 and Activates PKA to Promote NSFL1C/P47 Phosphorylation and Dendritic Arborization in Neurons. *eNeuro* **2018**. <https://doi.org/10.1523/ENEURO.0466-18.2018>.

- (122) Hu, D.; Sun, X.; Liao, X.; Zhang, X.; Zarabi, S.; Schimmer, A.; Hong, Y.; Ford, C.; Luo, Y.; Qi, X. Alpha-Synuclein Suppresses Mitochondrial Protease ClpP to Trigger Mitochondrial Oxidative Damage and Neurotoxicity. *Acta Neuropathol.* **2019**. <https://doi.org/10.1007/s00401-019-01993-2>.
- (123) Ellis, J. M.; Wong, G. W.; Wolfgang, M. J. Acyl Coenzyme A Thioesterase 7 Regulates Neuronal Fatty Acid Metabolism To Prevent Neurotoxicity. *Mol. Cell. Biol.* **2013**. <https://doi.org/10.1128/mcb.01548-12>.
- (124) Michaelis, E. K.; Wang, X.; Pal, R.; Bao, X.; Hascup, K. N.; Wang, Y.; Wang, W. T.; Hui, D.; Agbas, A.; Choi, I. Y.; et al. Neuronal Glud1 (Glutamate Dehydrogenase 1) over-Expressing Mice: Increased Glutamate Formation and Synaptic Release, Loss of Synaptic Activity, and Adaptive Changes in Genomic Expression. *Neurochem. Int.* **2011**. <https://doi.org/10.1016/j.neuint.2011.03.003>.
- (125) Fan, X.; Cui, L.; Zeng, Y.; Song, W.; Gaur, U.; Yang, M. Age14-3-3 Proteins Are on the Crossroads of Cancer, Aging, and Age-Related Neurodegenerative Disease. *Int. J. Mol. Sci.* **2019**, *20* (14). <https://doi.org/10.3390/ijms20143518>.
- (126) Corradi, A.; Zanardi, A.; Giacomini, C.; Onofri, F.; Valtorta, F.; Zoli, M.; Benfenati, F. Synapsin-I- and Synapsin-II-Null Mice Display an Increased Age-Dependent Cognitive Impairment. *J. Cell Sci.* **2008**. <https://doi.org/10.1242/jcs.035063>.
- (127) Glass, D.; Viñuela, A.; Davies, M. N.; Ramasamy, A.; Parts, L.; Knowles, D.; Brown, A. A.; Hedman, Å. K.; Small, K. S.; Buil, A.; et al. Gene Expression Changes with Age in Skin, Adipose Tissue, Blood and Brain. *Genome Biol.* **2013**. <https://doi.org/10.1186/gb-2013-14-7-r75>.
- (128) Davis, K. N.; Tao, R.; Li, C.; Gao, Y.; Gondré-Lewis, M. C.; Lipska, B. K.; Shin, J. H.; Xie, B.; Ye, T.; Weinberger, D. R.; et al. GAD2 Alternative Transcripts in the Human Prefrontal Cortex, and in Schizophrenia and Affective Disorders. *PLoS One* **2016**. <https://doi.org/10.1371/journal.pone.0148558>.
- (129) Wang, X.; Dong, C.; Sun, L.; Zhu, L.; Sun, C.; Ma, R.; Ning, K.; Lu, B.; Zhang, J.; Xu, J. Quantitative Proteomic Analysis of Age-Related Subventricular Zone Proteins Associated with Neurodegenerative Disease. *Sci. Rep.* **2016**, *6* (October), 1–11. <https://doi.org/10.1038/srep37443>.
- (130) Perluigi, M.; Di Domenico, F.; Giorgi, A.; Schininà, M. E.; Coccia, R.; Cini, C.; Bellia, F.; Cambria, M. T.; Cornelius, C.; Butterfield, D. A.; et al. Redox Proteomics in Aging Rat Brain: Involvement of Mitochondrial Reduced Glutathione Status and Mitochondrial Protein Oxidation in the Aging Process. *J. Neurosci. Res.* **2010**. <https://doi.org/10.1002/jnr.22500>.
- (131) Zhao, Y.; Liu, X.; He, Z.; Niu, X.; Shi, W.; Ding, J. M.; Zhang, L.; Yuan, T.; Li, A.;

- Yang, W.; et al. Essential Role of Proteasomes in Maintaining Self-Renewal in Neural Progenitor Cells. *Sci. Rep.* **2016**. <https://doi.org/10.1038/srep19752>.
- (132) Shimazaki, K.; Hosoya, H.; Takeda, Y.; Kobayashi, S.; Watanabe, K. Age-Related Decline of F3/Contactin in Rat Hippocampus. *Neurosci. Lett.* **1998**. [https://doi.org/10.1016/S0304-3940\(98\)00179-7](https://doi.org/10.1016/S0304-3940(98)00179-7).
- (133) Perri, E. R.; Thomas, C. J.; Parakh, S.; Spencer, D. M.; Atkin, J. D. The Unfolded Protein Response and the Role of Protein Disulfide Isomerase in Neurodegeneration. *Frontiers in Cell and Developmental Biology.* 2016. <https://doi.org/10.3389/fcell.2015.00080>.
- (134) Prasad Rao, R. S.; Zhang, N.; Xu, D.; MaxMøller, I. CarbonylDB: A Curated Data-Resource of Protein Carbonylation Sites. *Bioinformatics* **2018**. <https://doi.org/10.1093/bioinformatics/bty123>.
- (135) Sun, M. A.; Wang, Y.; Cheng, H.; Zhang, Q.; Ge, W.; Guo, D. RedoxDB-a Curated Database for Experimentally Verified Protein Oxidative Modification. *Bioinformatics* **2012**. <https://doi.org/10.1093/bioinformatics/bts468>.
- (136) Poon, H. F.; Frasier, M.; Shreve, N.; Calabrese, V.; Wolozin, B.; Butterfield, D. A. Mitochondrial Associated Metabolic Proteins Are Selectively Oxidized in A30P α -Synuclein Transgenic Mice - A Model of Familial Parkinson's Disease. *Neurobiol. Dis.* **2005**. <https://doi.org/10.1016/j.nbd.2004.12.009>.
- (137) Nuss, J. E.; Choksi, K. B.; DeFord, J. H.; Papaconstantinou, J. Decreased Enzyme Activities of Chaperones PDI and BiP in Aged Mouse Livers. *Biochem. Biophys. Res. Commun.* **2008**. <https://doi.org/10.1016/j.bbrc.2007.10.194>.
- (138) Buss, K.; Drewke, C.; Lohmann, S.; Piwonska, A.; Leistner, E. Properties and Interaction of Heterologously Expressed Glutamate Decarboxylase Isoenzymes GAD65kDa and GAD67kDa from Human Brain with Ginkgotoxin and Its 5'-Phosphate. *J. Med. Chem.* **2001**. <https://doi.org/10.1021/jm010868f>.
- (139) Wei, J.; Jin, Y.; Wu, H.; Sha, D.; Wu, J. Y. Identification and Functional Analysis of Truncated Human Glutamic Acid Decarboxylase 65. *J. Biomed. Sci.* **2003**. <https://doi.org/10.1159/000073527>.
- (140) Franco-Iborra, S.; Vila, M.; Perier, C. Mitochondrial Quality Control in Neurodegenerative Diseases: Focus on Parkinson's Disease and Huntington's Disease. *Frontiers in Neuroscience.* 2018. <https://doi.org/10.3389/fnins.2018.00342>.
- (141) Lim, K. H.; Dasari, A. K. R.; Ma, R.; Hung, I.; Gan, Z.; Kelly, J. W.; Fitzgerald, M. C. Pathogenic Mutations Induce Partial Structural Changes in the Native β -Sheet Structure of Transthyretin and Accelerate Aggregation. *Biochemistry* **2017**, *56* (36),

4808–4818. <https://doi.org/10.1021/acs.biochem.7b00658>.

- (142) Dobson, C. M. Protein Folding and Misfolding. *Nature*. 2003. <https://doi.org/10.1038/nature02261>.
- (143) Kelly, J. W. The Alternative Conformations of Amyloidogenic Proteins and Their Multi-Step Assembly Pathways. *Curr. Opin. Struct. Biol.* **1998**. [https://doi.org/10.1016/S0959-440X\(98\)80016-X](https://doi.org/10.1016/S0959-440X(98)80016-X).
- (144) Jahn, T. R.; Radford, S. E. Folding versus Aggregation: Polypeptide Conformations on Competing Pathways. *Archives of Biochemistry and Biophysics*. 2008. <https://doi.org/10.1016/j.abb.2007.05.015>.
- (145) Jacobson, D. R.; Pastore, R. D.; Yaghoubian, R.; Kane, I.; Gallo, G.; Buck, F. S.; Buxbaum, J. N. Variant-Sequence Transthyretin (Isoleucine 122) in Late-Onset Cardiac Amyloidosis in Black Americans. *N. Engl. J. Med.* **1997**. <https://doi.org/10.1056/NEJM199702133360703>.
- (146) Connors, L. H.; Lim, A.; Prokaeva, T.; Roskens, V. A.; Costello, C. E. Tabulation of Human Transthyretin (TTR) Variants, 2003. *Amyloid*. 2003. <https://doi.org/10.3109/13506120308998998>.
- (147) João, M.; Saraiva, M. Transthyretin Mutations in Health and Disease. *Hum. Mutat.* **1995**. <https://doi.org/10.1002/humu.1380050302>.
- (148) Peterson, S. A.; Klabunde, T.; Lashuel, H. A.; Purkey, H.; Sacchettini, J. C.; Kelly, J. W. Inhibiting Transthyretin Conformational Changes That Lead to Amyloid Fibril Formation. *Proc. Natl. Acad. Sci. U. S. A.* **1998**. <https://doi.org/10.1073/pnas.95.22.12956>.
- (149) Yokoyama, T.; Kosaka, Y.; Mizuguchi, M. Inhibitory Activities of Propolis and Its Promising Component, Caffeic Acid Phenethyl Ester, against Amyloidogenesis of Human Transthyretin. *J. Med. Chem.* **2014**. <https://doi.org/10.1021/jm500997m>.
- (150) Yin, J.; Chu, J. W.; Ricci, M. S.; Brems, D. N.; Wang, D. I. C.; Trout, B. L. Effects of Antioxidants on the Hydrogen Peroxide-Mediated Oxidation of Methionine Residues in Granulocyte Colony-Stimulating Factor and Human Parathyroid Hormone Fragment 13-34. *Pharm. Res.* **2004**. <https://doi.org/10.1007/s11095-004-7692-4>.

Biography

Renze Ma attended Hong Kong University of Science and Technology from 2011-2015 and graduated with a Bachelor of Science in Chemistry with a minor in China Study. He attended graduate school at Duke University (Durham, NC) in August 2015 and worked as a graduate assistant in the lab of Dr. Michael C. Fitzgerald. He is expected to graduate with a Doctor of Philosophy degree in Chemistry in the summer of 2020. While attending Duke University, he served as president of Duke Chinese Students and Scholars Association (DCSSA) graduate student organization for the 2017-2018 academic year.

Publications

1. Ma, R., Roberts, J., Fitzgerald, M. Global Analysis of Protein Folding Stability Changes During Parkinson's Disease in a Mouse Model. (*manuscript in progress*)
2. Cabrera, A., Wiebelhaus, N., Quan, B., Ma, R., Meng, H., Fitzgerald, M. (2020) Comparative Analysis of Mass-Spectrometry-Based Proteomics Methods for Protein Target Discovery Using a one-pot Approach. *J. Am. Soc. Mass Spectrom.* 31(2), pp.217-226
3. Ma, R., Meng, H., Wiebelhaus, N. and Fitzgerald, M. (2018). Chemo-Selection Strategy for Limited Proteolysis Experiments on the Proteomic Scale. *Analytical Chemistry*, 90(23), pp.14039-14047.
4. Kaur, U., Meng, H., Liu, F., Ma, R., Ogburn, R., Johnson, J., Fitzgerald, M. and Jones, L. (2018). Proteome-Wide Structural Biology: An Emerging Field for the Structural Analysis of Proteins on the Proteomic Scale. *Journal of Proteome Research*, 17(11), pp.3614-3627.
5. Meng, H., Ma, R. and Fitzgerald, M. (2018). Chemical Denaturation and Protein Precipitation Approach for Discovery and Quantitation of Protein-Drug Interactions. *Analytical Chemistry*, 90(15), pp.9249-9255.
6. Lim, K., Dasari, A., Ma, R., Hung, I., Gan, Z., Kelly, J. and Fitzgerald, M. (2017). Pathogenic Mutations Induce Partial Structural Changes in the Native β -Sheet Structure of Transthyretin and Accelerate Aggregation. *Biochemistry*, 56(36), pp.4808-4818.
7. Zhu, L., Liu, Y., Ma, R. and Tong, R. (2014). Total Synthesis and Structural Revision of (+)-Uprolide G Acetate. *Angewandte Chemie*, 127(2), pp.637-642.