THE UNIVERSITY OF CHICAGO

HIGH-THROUGHPUT QUANTITATIVE PROTEOMICS PLATFORMS FOR DISCOVERY OF FUNCTIONAL PROTEIN MODIFICATIONS AND INTERACTIONS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN BIOCHEMISTRY AND MOLECULAR BIOPHYSICS

BY

JUN XIAN HUANG

CHICAGO, ILLINOIS

JUNE 2021

Table of Contents

List of Figures ..... **vi**

List of Tables ..... **ix**

Abstract ..... **x**

Acknowledgements ..... **xii**

CHAPTER 1 Introduction **1**

1.1 Background **1**

1.1.1 Heterogeneity of the proteome 1

1.1.2 Mass spectrometry-based studies of post-translational modifications 8

1.1.3 Deciphering the functions of PTMs 12

1.2 Scope of the dissertation **15**

CHAPTER 2 Hotspot Thermal Profiling (HTP) for the functional discovery post-translational modifications **17**

2.1 Introduction **17**

2.1.1 Cellular Thermal Shift Assay 17

2.1.2 Expansion of CETSA using mass spectrometry for the interrogation of proteome-wide protein thermal stability 20

2.1.3 Post-translational modifications as a low molecular weight covalent ligand 22

2.2 Results **24**

2.2.1 Hotspot Thermal Profiling method development 24

2.2.2 HTP measures the in situ thermal stability changes in response to site-specific phosphorylation 26
2.2.2.1 Defining the qualitative and quantitative filters in the HTP data analysis pipeline 27
2.2.3 HTP detects protein thermal stability shifts in live cells 29
2.2.3.1 Global T_m measurements of unmodified, bulk proteins and phosphomodiform 29
2.2.3.2 Reproducibility of peptide-level and protein level T_m measurements 32
2.2.4 Trends and functional correlations in phosphorylation-induced thermal stability shifts detected by HTP 35
2.2.4.1 HTP detects three types of site-specific thermal stability shifts 35
2.2.4.2 Perturbation of protein stability by phosphorylation is site-specific, not systematic 38
2.2.4.3 Functional significance and degrees of PTM-induced thermal perturbation 40
2.2.5 HTP detects known phosphorylation-mediated protein-protein interaction in 4EBP1-EIF4E 41
2.2.6 HTP detects previously uncharacterized phosphorylation-mediated intramolecular changes in Vinculin 44
2.2.6.1 Significant thermal perturbation by pS721 in Vinculin 44
2.2.6.2 Phenotypic characterization of Vinculin phosphomimetics 45
2.2.7 HTP detects previously uncharacterized phosphorylation-mediated protein-metabolite interaction in GAPDH 48
2.2.7.1 Significant thermal perturbations by pS210 in GAPDH 48

2.2.7.2 Thermal stabilities of GAPDH modiforms under acute glucose-withdrawal....50

2.2.7.3 Phenotypic characterization of GAPDH phosphomimetics 51

2.3 Conclusion and Discussion.....54

2.4 Materials and Methods 59

2.5 Tables.....69

CHAPTER 3 Diverse Hotspot Thermal Profiling methods detect phosphorylation-dependent changes in protein stability 70

3.1 Introduction 70

3.2 Results.....75

3.2.1 "Label-Enrich" Hotspot Thermal Profiling (LE-HTP) workflow 75

3.2.2 "Label-Fractionate-Enrich" Hotspot Thermal Profiling (LFE-HTP) workflow.....77

3.2.3 Global thermal stability profiles comparison between methods 78

3.2.4 Data reproducibility comparison between methods 80

3.2.5 Global trends 82

3.2.6 Site-specific thermal stability profiles comparison between methods 84

3.2.7 LFE-HTP detects uncharacterized phosphorylation sites in ALDOA.....86

3.2.8 Different data analysis pipelines and statistical treatments, but not upfront biochemical workflow, can affect the significance threshold of ΔT_m values.....88

3.3 Conclusion and Discussion.....95

3.4 Materials and Methods 97

CHAPTER 4 Application of thermal profiling methods to interrogate dynamic cell states 107

4.1 Introduction 107

4.1.1 Deregulation of glucose metabolism in diseases 107

4.1.2 Proteome thermal stability under low-glucose cellular environment 109

4.2 Results 112

4.2.1 Deep HTP profiling of proteome under acute glucose withdrawal 112

4.2.2 Deep dynamic PISA profiling of proteome under acute glucose withdrawal 116

4.3 Conclusion and Discussion 120

4.3.1 Future directions 121

4.4 Materials and Methods 126

CHAPTER 5 Summary and Perspective 132

5.1 Summary 132

5.2 Perspectives and Future Directions: integrating interaction profiling methods to elucidate PTM-mediated signaling mechanisms 136

5.2.1 Orthogonal strategies to complement proteome thermal stability profiling of PTMs 136

5.2.2 Proximity profiling strategies to elucidate local protein networks 140

5.2.3 Top-down proteomics to interrogate thermal stability of intact proteoforms 144

5.2.4 Integrating kinetic profiling strategies to derive mechanism of a novel PTM 146

List of References 149

List of Figures

Figure 1.1 Commonly detected protein post-translational modifications and their functions. 3

Figure 1.2 Number of human PTMs reported in databases.....5

Figure 1.3 Schematic of a typical proteomics workflow to study PTMs 11

Figure 2.1 Mechanistic overview MS-CETSA/TPP 21

Figure 2.2 Schematic of Hotspot thermal profiling workflow 25

Figure 2.3 Effect of temperature pulse on global phosphorylation level 27

Figure 2.4 Schematic that showing how phosphopeptides containing the same phosphosites are processed 28

Figure 2.5 Schematic showing the total number of detected peptides and phosphopeptides in the unmodified, bulk proteome and phosphoproteome, and the data processing workflow. 30

Figure 2.6 HTP identifies thousands of unmodified, bulk and phosphomodiform Tm's in HEK293T cells 31

Figure 2.7 Statistical comparison strategy used to analyze ΔT_m values detected using HTP 33

Figure 2.8 HTP dataset reproducibility and error analysis.....34

Figure 2.9 Types of T_m shifts detected by HTP.....37

Figure 2.10 Global relationships between local phosphosite environment and altered phosphomodiform stability.....39

Figure 2.11 Plot of phosphosite ΔT_m and number of annotations in PhosphositePlus database ...40

Figure 2.12 Effects of N-terminal phosphorylation sites on 4EBP1 protein stability and function 42

Figure 2.13 Effects of C-terminal phosphorylation site on 4EBP1 protein stability and function43

Figure 2.14 HTP detects significant phosphosite-induced thermal perturbation in Vinculin 45
Figure 2.15 In situ characterization of Vinculin phosphomimetics 47
Figure 2.16 HTP detects phosphosite-induced thermal perturbations in GAPDH.....49
Figure 2.17 Thermal stability of GAPDH modiforms under acute glucose withdrawal.....50
Figure 2.18 Phenotypic characterization of GAPDH phosphomimetics.....52
Figure 2.19 Proposed model of phosphoserine 210 function in GAPDH 53
Figure 2.20 Simplified schematic of the HTP workflow 55
Figure 3.1 Schematic showing interference contamination from TMT reporter tags using MS2 quantification.....73
Figure 3.2 Comparison of new variant HTP workflows 76
Figure 3.3 Global T_m values measured by three HTP variant workflows 79
Figure 3.4 Representative T_m correlation plots in LFE- and LE-HTP datasets.....81
Figure 3.5 Comparing global relationships between local phosphosite environment and altered phosphomodiform stability of the 3 HTP variants 83
Figure 3.6 Site-specific thermal stability profiles comparison between methods.....85
Figure 3.7 Phosphorylation induced thermal stability shifts and structural correlation in aldolase A 87
Figure 3.8 Percentage of significantly shifted phosphosites reported in different HTP datasets..89
Figure 3.9 Schematic of statistical comparisons used in our HTP methods 90
Figure 3.10 Comparing data analysis pipelines.....91
Figure 3.11 Comparing of anecdotal phosphosites measured by different data analysis pipelines 93

Figure 3.12 Comparing the percentages of significant ΔT_m values based on application of different statistical cutoffs 94
Figure 4.1 Schematic showing the data quantification difference between CETSA and PISA assays 111
Figure 4.2 Schematic of showing thermal proteome profiling workflow with deep bRP fractionation 113
Figure 4.3 Deep global thermal proteome profiles of proteome under acute glucose withdrawal perturbation 114
Figure 4.4 Representative protein melting curves 116
Figure 4.5 Schematic of modified PISA assay using SILAC cells for dynamic profiling of acute glucose withdrawal 117
Figure 4.6 Deep global PISA thermal stability profiles of cells under acute glucose withdrawal perturbation 118
Figure 4.7 Glucose-deficiency induced stability shifts in glycolytic enzymes 119
Figure 4.8 Workflow schematic for the proposed PTM-PISA method 125
Figure 5.1 Schematics of thermal stabilities profiling approaches described in this thesis 135
Figure 5.2 Schematics of limited-proteolysis profiling approaches 139
Figure 5.3 Schematics of “bottom-up” proximity profiling methods 141
Figure 5.4 Schematic depiction of integrating interaction profiling strategies in a kinetic series to discover protein targets and signaling networks 143
Figure 5.5 Differences between top-down and bottom-up proteomics methods 145
Figure 5.6 Global kinetic quantitative profiling of pgK sites in S. elongatus 148

List of Tables

Table 2.1 Primers used for site-directed mutagenesis of GAPDH and vinculin 69

Table 2.2 Sequencing primers used to verify GAPDH and vinculin mutant plasmids 69

Abstract

High-throughput quantitative proteomic platforms for discovery of functional protein modifications and interactions

By Jun X. Huang Under Direction of Professor Raymond. E. Moellering

Protein post-translational modification (PTM) represents an important functional regulatory mechanism in the cell. Technological advances in the field of LC-MS/MS-based quantitative proteomics have enabled massively parallel detection and identification of novel PTMs in the past few decades. However, conventional biochemical approaches to characterize PTM function remain labourious and low throughput, and we lack general methods with which to predict, or even prioritize, which modification sites are likely to be functional in the proteome.

Here we developed a quantitative proteomic method, termed "Hotspot Thermal Profiling" (HTP), to enable the unbiased, global, and high-throughput measurements of protein stability changes in response to site-specific phosphorylation events in live cells. Our proof-of-concept dataset showed that phosphorylation-induced thermal stability shifts can serve as a readout on the biophysical and functional effect of a phosphosite. HTP measured thermal stability profiles of thousands of bulk, unmodified proteoforms and phosphomodiforms, and revealed trends related to protein function and structure. HTP rediscovered well-known phosphorylation-mediated protein-protein interactions in protein 4EBP1 and enabled discovery of previously uncharacterized PTM-mediated interaction GAPDH and Vinculin. *In vitro* functional experiments of GAPDH and Vinculin validated and elucidated the molecular functions of

Next, we developed two HTP method variants with aims to improve precision in sample processing, expand the depth of (phospho)proteome coverage, and increase the accuracy in LC-MS/MS quantitative measurements. We termed these method variants “Label-Enrich”- and “Label-Fractionate-Enrich”-HTP and demonstrated that the thermal stability of different proteoforms be reliably measured in these variant methods.

Lastly, we explored the application of different thermal profiling strategies to profile protein-metabolite interactions in cells. We used a modified HTP workflow with an additional fractionation step and developed a SILAC-based thermal stability method to interrogate the proteome under acute-glucose withdrawal in a kinetic series. The resultant dataset revealed the time-dependent shifts in thermal stabilities of thousands of proteins, including in glycolytic enzymes known to interact with glucose metabolites.

Overall, these new profiling approaches provided novel mechanistic insights into molecular interactions in the cell, and we envision that the extension HTP methods to diverse organisms, cell types, PTMs, and kinetic states can expand our understanding of the role of PTMs in regulating protein structure, function and signal transduction.

Acknowledgements

I would like to thank every single person who supported me in my PhD journey. The work presented in this dissertation would not be possible without their support both inside and outside of the lab.

First and foremost, I would like to thank my doctorate advisor Dr. Raymond E. Moellering for his mentorship, encouragement, and support throughout my graduate career. He has inspired me with his breadth of scientific knowledge and perseverance in the face of obstacles. I am forever grateful for the opportunity to tackle diverse research projects in his lab and for the trust he has placed in me to take on responsibilities, which helped me gain many new skillsets for both in science and life. He has always encouraged critical thinking and provided me with insightful guidance with my research. In addition, I am forever thankful for his incredible support in all of my career endeavours. The past five years would not have been possible without the wonderful mentorship from Ray.

Next, I would like to thank every past and present member of the Moellering lab who helped me through many challenges. Dr. Gihoon Lee was first member of the lab that I worked closely. He taught me many techniques ranging from proteomics sample processing to molecular cloning and helped generate some of the data presented in this thesis. I enjoyed having Dr. Jae Won Chang as my desk neighbour for many years and greatly appreciated his help with the metabolomics work presented in Chapter 2. Many thanks to Dr. Shaghayegh Fathi and Dr. Colin S. Swenson who helped proofread this dissertation. Shay is one of the most organized and responsible lab mates and always first to help. I would also like to thank Kavya Pillai who has been incredibly helpful in maintaining our mass spectrometry instruments and organizing lab

social events. Many thanks to Dr. Gang Li for the insightful help and discussion of microscopy and molecular biology techniques, and Jeffrey E. Montgomery for being my carpool partner during the Covid-19 pandemic lockdown. I am incredibly appreciative of every past and present member of the Moellering lab for all the wonderful scientific discussions we have had and all the help they have extended me throughout my time in the lab.

This dissertation would have not been possible without the help and expertise from talented collaborators, including Dr. Kate Cavanaugh and Dr. Margaret L. Gardel at the University of Chicago who helped with the vinculin microscopy work in Chapter 2. Many thanks to Dr. Benjamin D. Stein and David Wu from the Cantley lab at Weill Cornell Medical who were instrumental in generating LFE-HTP data presented in Chapter 3. I am particularly thankful for Dr. Benjamin D. Stein who flew all the way over form New York to show me his phosphoenrichment techniques. Additionally, I would also like to thank my thesis Committee Members, Dr. Michael J. Rust, Dr. Benoît Roux, and Dr. Tobin R. Sosnick for their insightful feedback and support of my thesis work.

I am forever grateful for the tremendous amount of love and support I have received from family and friends. First, I would like to thank my parents Weiling Wu and Zhiyuan Huang for their love even though I still have not left my rebellious phase at the age of 30. They have supported me unconditionally in whatever I wish to pursue, and their reassuring words have always reminded me how much I am loved. I would like to thank other members of my extended family back in China whom I have not seen in several years for always thinking of and loving me. Lastly, I would like to thank my new families in America for being incredibly welcoming and supportive.

Finally, I would like to thank the pillar of my life, my husband Nicholas A. Stefanski, for his endless love and tireless support. For much of the past 13 years, I have been mostly alone by

myself without any family in this strange country of America, but thanks to him, Chicago is the first city that truly feels like home. His encouragement, patience, and comforting words have helped me through many challenges both inside and outside of lab, and his company is what keeps me going every single day. A wise philosopher once said on Twitter: “Forget the mitochondria, you are the powerhouse of my cell.” This dissertation is dedicated to him.

CHAPTER 1

Introduction

1.1 Background

1.1.1 Heterogeneity of the proteome

Proteins carry out a wide variety of essential processes in the cell including catalytic, regulatory, signaling and structural functions[1, 2]. Typically ranging between 50 to 1000 amino acids per polypeptide chain[3], these small biological machineries are comprised of unique sequences of amino acid residues that determine not only their chemical and biophysical properties, but also three-dimensional architecture[4]. This unique primary sequence directs the folding of the native protein into three-dimensional shapes to enable specific functions, from serving structural purposes as the basic building block of the cytoskeleton in the cell (e.g. actin)[5, 6], to mediating signaling pathways in a variety of signal transduction cascades (e.g. kinase and acetyltransferase)[7-9], and carrying out metabolic processes such as the breakdown of glucose molecules to provide sources of energy and biosynthetic precursors necessary for sustaining life (e.g. glycolytic enzymes)[10, 11].

The size of the human proteome is a core proteomics question. The human genome project estimated the number of protein-coding genes to be around ~20,000 genes[12], but researchers project that there may be several million distinct protein forms in the human proteome[13-15]. The first source of proteome diversity is alternative RNA splicing which enables a single gene to code for multiple proteins (i.e. protein variants or isoforms)[16]. At least 95% of human genes are believed

to be alternative spliced[17, 18] and the Uniprot database has recorded 77,027 distinct protein sequences in the human proteome so far[19]. In addition to splice variants, a process known as post-translational modification (PTM) plays a key role in further expanding the diversity of the proteome. After their assembly on ribosomes and throughout their lifetimes, proteins are regularly modified with small, covalent chemical moieties on their amino acid side chains or N-terminus. PTMs are ubiquitous in all forms of life, and these chemical modifications can be both enzyme- (e.g. protein kinase)[7] or non-enzyme mediated (e.g. reactive metabolites)[20, 21]. PTMs function to modulate protein activity, structural conformation, protein subcellular localization, biomolecular interactions, and many more, and thus play roles in many core aspects of biological processes. The reversible nature of PTMs also enables spatial and temporal control of protein function[22].

PTM is a major source of heterogeneity in the proteome. The very first protein modification was discovered in 1906 by Levene and Alsberg at Rockefeller University who detected a phosphate group on egg yolk protein vitellin while trying to study the chemical composition of amino acids[23]. Since then, more than 500 different protein modifications have been detected across all proteomes and all 20 canonical amino acids[19, 22]. PTMs can range from relatively small (e.g. methylation)[24] to as large as thousands of Daltons in size (e.g. ubiquitination)[25] (Figure 1.1). In addition to the sheer chemical diversities of protein modifications, multisite and different types of PTM can exist on a protein at the same time, which leads to a combinatorial explosion in the number of possible molecular states or proteoforms[26].

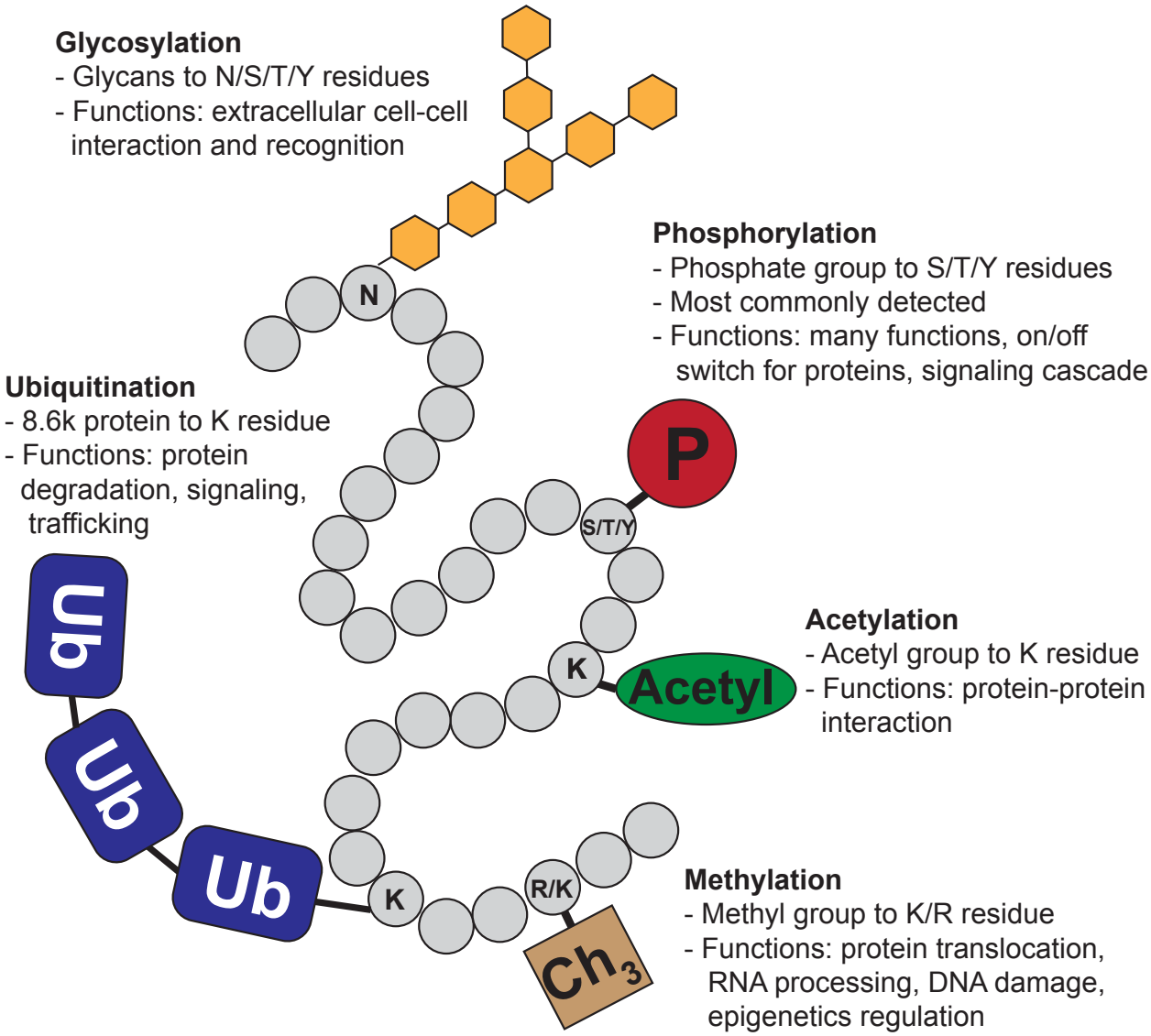

Figure 1.1 Commonly detected protein post-translational modifications and their functions.

The most commonly detected PTM in human is by far protein phosphorylation, clocking in at more than 38,000 entries in the Uniprot database and accounting for more than 70% of all annotated PTM sites in human[2, 19] (Figure 1.2A). Phosphorylation is enzymatically installed onto protein side chains by kinases ("writer") and removed by phosphatases ("eraser")[27]. This reversible addition of a phosphate group PO_4^{3-} can be found on serine, threonine, and tyrosine residues in

mammalian proteomes[4, 28], as well as histidine and aspartate residues in bacterial and fungal proteomes[29]. Protein phosphorylation is a major regulatory mechanism in many cellular processes, acting as "on" or "off" switches for many enzymes and receptors. The pioneering work of Edwin Krebs and Edmond Fischer in 1956 showed that protein phosphorylation can reversibly regulate the activity of phosphorylase in the breakdown of glycogen molecules to glucose subunits in the first known example of allosteric regulation[30, 31]. Many more years of research went into elucidating the regulatory mechanism of protein phosphorylation and dephosphorylation events. As protein phosphorylation drives regulation of cellular pathways, it's no surprise that aberrant activity is implicated in many diseases. For example, in tumour suppressor protein p53, N-terminal residues are phosphorylated in response to cell stress which weakens the protein-protein interaction between p53 and E3 ubiquitin ligase MDM2[32]. This leads to activation of p53 and upregulation of genes involved in cell cycle arrest and DNA damage repair[33, 34]. An imbalance in the phosphorylation/ dephosphorylation mechanism of p53 can lead to aberrant cell proliferation and cancer[35]. Phosphorylation also has important roles in signal transduction cascades. The mitogen-activated protein kinase (MAPK) pathway mediated by Ras kinases transmits signals and stimuli from cell surface receptors to intracellular processes[7]. The phosphate group functions as the "messenger" and communicates the signal as it is transmitted from one protein to the next in a cascading event. Defects in the phosphorylation signaling pathways is implicated in many diseases, making protein kinases a major therapeutic target. Many kinase inhibitors have been developed to great therapeutic success, including the blockbuster drug imatinib (Gleevec) which binds to the ATP binding site of BCR/ABL tyrosine kinase and inhibits autophosphorylation of the protein, effectively shutting down downstream signaling pathways that promote leukomogenesis[36, 37]. Tofocitinib, approved for the treatment of rheumatoid arthritis, binds to a small hydrophobic ATP

pocket of intracellular Janus kinases (JAK) and inhibits intracellular phosphorylation signaling pathways initiated by cytokines that lead to inflammatory responses[38, 39]. These are only a couple examples of the broadening research to develop therapeutics targeting protein kinases and phosphorylation signaling pathways.

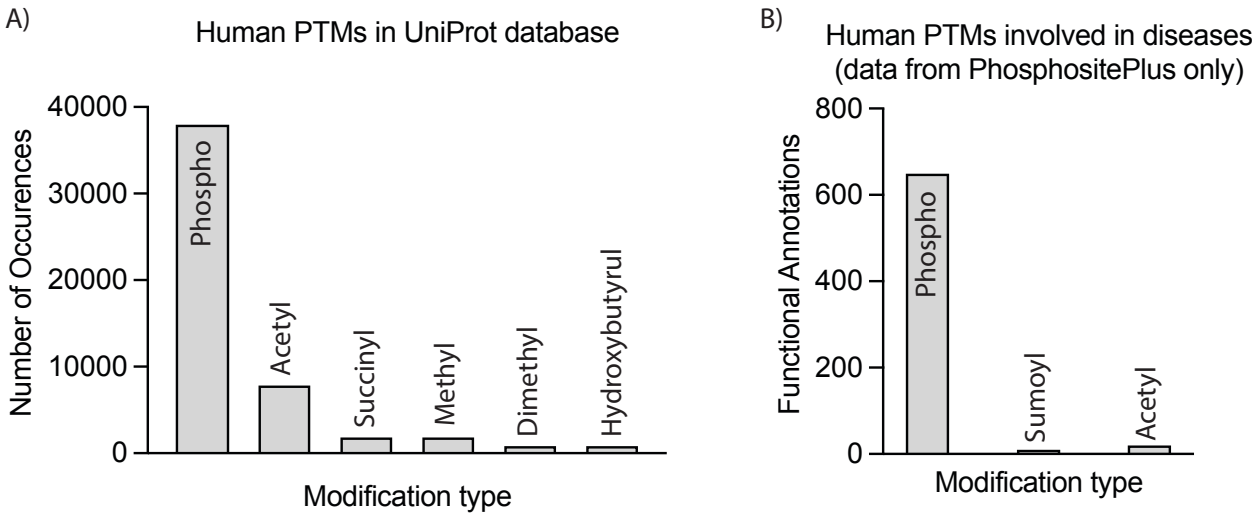

Figure 1.2 Number of human PTMs reported in databases
A) Histogram of human PTMs recorded in UniProt. By far phosphorylation is the most frequently reported PTM at ~38,000 entries. **B**) Histogram of human PTMs with functional annotations involved in diseases. Data from A-B) are adapted from [2],[40].

Other common examples of enzyme-mediated protein modifications are illustrated in Figure 1.1. Currently, the most popular PTM to study is acetylation as the number of publications published annually just exceeded that of phosphorylation in recent years. Acetylation is a highly evolutionarily conserved modification occurring in both prokaryotes and eukaryotes[41]. Interest in this PTM exploded in the late 1990s after a slew of discoveries such as the identification of histone acetyltransferases (HATs) and deacetylases (HDACs)[42, 43]. Early studies of acetylation focused on the link of histone acetylation to epigenetics, though in recent years, acetylation research expanded its scope to non-histone acetylation events after the PTM was detected in hundreds of different

proteins throughout the cell[42, 44]. Acetylation on the C-terminal lysine residues of p53 was found to interfere with DNA binding and fine-tunes p53 function in a context-dependent manner[33]. Acetylation of heat shock protein HSP90 impairs its association with an essential co-chaperone p23, and resulting in the loss of chaperone-assisted protein folding activity of HSP90[45]. Acetylation on K145 of protein TDP43 was found to be implicated in amyotrophic lateral sclerosis (ALS) as it impairs protein-RNA binding and promotes protein aggregation resulting in neuronal damages[46]. This broad functional impact of acetylation resulted in deacetylases becoming desirable therapeutic. For example, vonriostat is a broad HDAC inhibitor that binds to the zinc atom of the catalytic site and is approved for the treatment of advanced primary cutaneous T cell lymphoma[47]. Other examples of well-characterized enzyme-mediated PTM include ubiquitination which is one of the most abundant PTM in the cell[48]. Ubiquitin is unique because it is not a small chemical moiety but a highly conserved regulatory protein containing 76 amino acids and it covalently conjugates to other proteins by ubiquitin ligases[49]. A homologous polyubiquitination chain can serves as a degradation signal for transferring proteins to the 26S proteasome[50] while a mixed and branched linkage induces Epsin1-mediated endocytosis of major histocompatibility complex MHCI[51]. Enzyme-mediated PTMs remain highly sought after for both basic research and therapeutic development because of their tightly choreographed regulatory mechanisms.

A new class of non-enzymatic PTMs called reactive metabolite PTM (rmPTM) is also gaining popularity in the field of protein modification studies in recent years[52, 53]. A representative example is a non-enzymatic lysine acylation discovered by Moellering and Cravatt[20]. The highly reactive primary glycolytic metabolite 1,3-bisphosphoglycerate (1,3-BPG) induces acylation on lysine residues to form 3-phosphoglyceryl-lysine (pgK) in a variety of glycolytic enzymes. The pgK modification has been shown to alter the chemical properties of target proteins and can induce

inhibitory effects on glycolytic proteins in response to glucose level in the cell. Lysine glutarylation is another recent example of non-enzymatic acylation[54-56]. Hyperglutarylation on carbamoyl phosphate synthase 1 (CPS1) results in the loss of enzyme activity and perturbation in mitochondrial nitrogen metabolism[52]. Another example of rmPTM is the reactive glycolytic metabolite methylglyoxal (MGx) which has been found to form a methylimidazole crosslink between proximal cysteine and arginine residues (MICA) in Kelch-like ECH-associated protein 1 (KEAP1)[21]. The KEAP1-Nrf2 pathway is a master regulatory pathway of antioxidant responses[57-60]. The non-enzymatic MICA modification on KEAP1 was shown to induce dimerization of the protein and disrupt the interaction between KEAP1 and the transcription factor Nrf2, resulting in upregulation of antioxidant genes by transcription factor Nrf2. The MICA modification exemplifies the regulatory crosstalk of biological networks mediated by PTMs.

In summary, protein post-translational modification is the primary source of proteome heterogeneity. The chemical diversity and combinatorial presence of PTMs allow them to regulate protein function in a wide variety of mechanisms across all biological processes in the cell. Many PTMs are reversible in nature which enables spatial and temporal control of protein chemistry and activity in response to stimuli. Aberrant modification networks are implicated in many human diseases and are therefore of great interest to the scientific community to elucidate the molecular mechanisms of PTMs.

1.1.2 Mass spectrometry-based studies of post-translational modifications

In recent years, advances in liquid chromatography-tandem mass spectrometry (LC-MS/MS)-based technologies greatly accelerated the study of PTMs[61]. The first explosion of novel PTM discovery began in the 1980s with the development of electrospray ionization (ESI) mass spectrometry to study peptides and proteins[22], which awarded the 2002 Nobel Prize in chemistry to Koichi Tanaka and John[62]. In ESI-MS, a liquid sample solution of biological macromolecules is ionized into charged droplets after passing through an electrically charged capillary before entering the mass spectrometer for mass analysis[63, 64]. In the tandem MS (i.e. MS/MS or MS^2) approach, enzymatically digested peptides are further fragmented in a collision chamber with high energy gas, leading to the cleavages of peptide backbone between different amino acid residues and yielding b- and y-type fragments of varying lengths[65]. The resulting masses from MS/MS fragment ions spectra are matched with a protein sequence database with bioinformatic search engines which allows accurate determination of peptide sequences with amino acid resolution[62]. This peptide-based, "bottom-up" approach – also known as "shotgun proteomics" when performed in a large set of proteins such as cell lysate[66, 67] – has allowed the massively parallel detection and confident identification of post-translationally modified peptides.

The second major advancement in PTM studies was the development of front-end enrichment strategies of modified peptides prior to LC-MS/MS analysis to enable to the detection of low abundant modified peptides. Typically, a modified proteoform only represents a minor member of the protein pool. This inadvertently puts a disadvantage on the detection of PTM-peptides using data-dependent acquisition (DDA) mode which is the most popular method used in MS-based proteomics research today[68, 69]. After the first full MS scan, DDA mode selects the top 10-15 most abundance precursor ions for MS/MS analysis in an attempt to minimize redundant

peptide precursor selection and maximize proteome coverage[70, 71]. As a result, low abundant PTM-peptides often get overlooked as their signals are routinely masked by the more abundant non-modified peptides. Different PTM-specific enrichment strategies have been developed to isolate modified peptides and reduce complexity of the sample. The most traditional method uses antibodies to purify specific PTM-peptides for MS-identification[72]. Immunoprecipitation enrichment was used in the first global acetylation proteomics profiling study[44], and more recently, in the detection of acylation using polyclonal antibodies generated from immunizing rabbits with malonylated BSA[73], as well as other less studied PTMs[20]. On the other hand, affinity chromatography using immobilized titanium oxide (TiO_2) to enrich phosphopeptides is readily and commercially available[74], which has helped the robust and deep detection of the phosphoproteome. Enrichment strategies that rely on the physiochemical properties of the modified peptides, such as strong cation exchange chromatography (SCX) to selectively isolate charge-reduced modified peptides, also enables robust enrichment of a variety of PTMs[75]. Lastly, a variety of chemical derivatization methods has also been developed to tag PTMs[72], including chemical probes with an azide moiety that can be subsequently conjugated to an affinity linker that was used to enrich an O-GlcNac-modified proteome[76].

More recently, novel quantitative proteomics methods have added another dimension to deriving the functional significance of modification sites. In addition to identity information (i.e. whether a PTM is present on a protein or not), quantitative strategies enable researchers to measure and compare the relative abundance of PTMs in different conditions to investigate the regulatory crosstalk of PTM networks in diverse cell states and in response to different stimuli. Several methods have been developed in the past decade to enable these quantitative measurements. In the metabolic isotopic labeling (SILAC) approach which allows multiplexing of up to three conditions,

cells are grown in a medium supplied with amino acids of different weight (light, medium, or heavy isotopes) which get incorporated into the cell proteomes[77-79]. Different metabolically and isotopically labeled peptides have the same chemical properties but a known mass shift allows the MS to differentiate them from one another. The relative signal intensity differences between the different versions of the same peptide can be used derive PTM abundance differences[80]. Using the SILAC approach, Olsen et al obtained a global view of time-dependent dynamic regulation by phosphorylation in HeLa S3 cells during mitosis in response to a growth factor stimulus[79]. In chemical labeling approaches, isobaric reagents can label the proteome at the peptide-level and allows quantitative comparisons of 8 (iTRAQ) to 16 (TMTPro) conditions simultaneously[81-83]. In isobaric tandem mass tagging (TMT), the reagents are comprised of an amine-reactive NHS ester group that reacts with peptide N-termini, a mass normalizing linker, and a mass reporter tag whose signal intensity in MS/MS spectra is used for relative quantification[83]. Using this quantitative approach, Rose et al performed large scale quantification of ubiquitylation across 10 different tissue type to elucidate targets and mechanisms of ubiquitin-driven signaling systems in neurogenerative disease[84].

A typical proteomics experiment using strategies discussed are illustrated in Figure 1.3. Altogether, the increased sensitivity of MS-instrumentation, new enrichment strategies, and refined quantification methods have helped enhance the high-throughput detection of PTMs in a variety of proteomes.

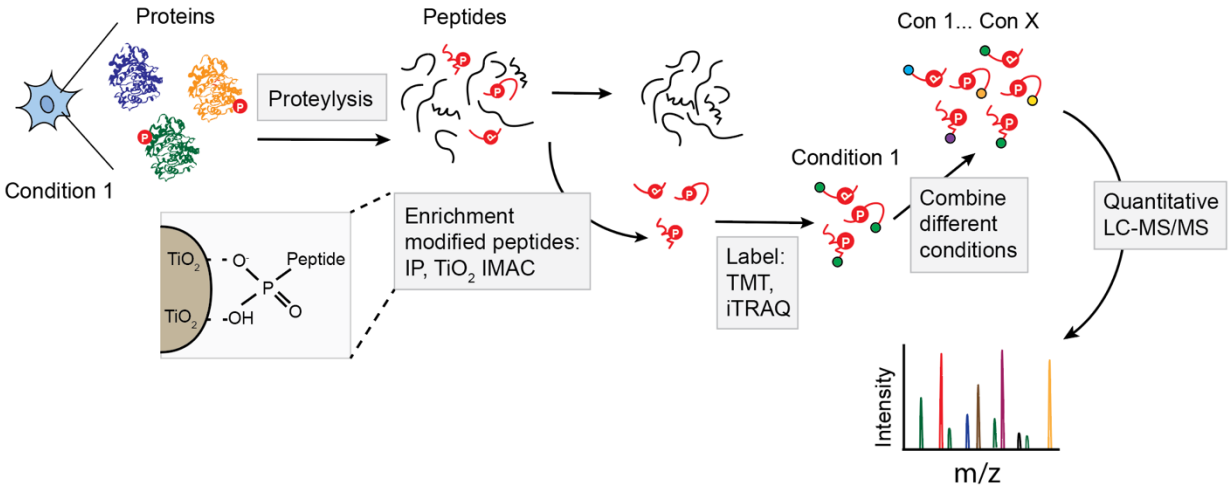

Figure 1.3 Schematic of a typical proteomics workflow to study PTMs
In a typical PTM proteomics experiment (phosphoproteomics is shown here as an example), proteins are isolated from cells and proteolyzed down to peptides. Modified peptides are enriched from the peptide pool using a variety of techniques such as immunoprecipitation, IMAC and TiO_2 (for phosphopeptides). Peptides from different treatment conditions are labeled with different isobaric tags (i.e., TMT, ITRAQ) before multiplexing and analyzed using LC-MS/MS for quantitative proteomics analysis.

1.1.3 Deciphering the functions of PTMs

Despite the ever-growing list of novel PTM sites detected with LC-MS/MS-based proteomics technologies, very few modification sites have been functionally characterized. According to PhosphositePlus database, more than 380,000 unique modification sites from just seven types of PTMs have been detected in the human proteome alone[85, 86]; however, for PTMs that are implicated in human diseases, only ~6500 of them have been functionally validated (Figure 1.2B). This stark contrast presents a major dilemma in the field of PTM studies: how do we prioritize which PTM from this long list to functionally validate?

One common pitfall shared by the current state of quantitative PTM proteomics analyses is that they simply report on the "writing" and "erasing" status of PTMs. That is, whether a PTM is present or not during one cellular state versus another. These global PTM datasets may tell us the identity and stoichiometry of modification sites and help identify pathways that are implicated in diseases, such as global PTM profiles of Tau proteins to identify patient heterogeneity and stages of Alzheimer's Disease[87], they do not offer any mechanistic insight. In other words, LC-MS/MS-based studies can provide "writing" and "erasing" information of PTMs, but they provide no dynamic "reading" information about how the modification site interacts with other biomolecules and induce functional changes in proteins.

Despite the chemical diversity and functional complexity of PTMs, their molecular mechanism in regulating protein activity is rather straightforward. Through electronic and steric effects, PTMs impact intra- and intermolecular interactions of proteins. In phosphorylation, the negatively charged phosphate group can repel or attract other negatively or positively charged amino acid side chains from nearby intramolecular residues or an intermolecular biomolecules, resulting in changes to protein conformation and function[4]. Similarly, histone acetylation

diminishes the electrostatic affinity between histone proteins and DNA and regulates the transcription of genes in response to cellular stimuli[88, 89]. While these molecular mechanisms appear straightforward, conventional biochemical approaches to decipher PTM's mode of action are low-throughput and laborious. For decades, PTM characterization are confined to individual proteins and isolated pathways. Researchers must first identify specific modification sites associated with a phenotype of interest, and then meticulously study the properties of the modified protein one-by-one. Endogenously modified proteins are often too low in abundance for purification to use in functional *in vitro* biochemical applications. On the other hand, it is near impossible to generate constitutively modified protein at a specific location with the desired PTM of interest. To mimic the modified protein, researchers commonly substitute the specific residue of interest with another amino acid that mimic the chemical properties of the modified residue. For phosphorylation, amino acids glutamate and aspartate are often used to mimic phosphoserine/threonine/tyrosine (pS/T/Y) since they provide the same negative charge and are of similar volume to the phosphoresidues (Figure 1)[90]. However, these mutations often fail to truly replicate the chemical properties of the modification, and many residue substitutions for other PTMs are not always possible[91]. To generate mimetic mutants, researchers need to perform lengthy site-directed mutagenesis and protein expression purification for each desired mimetic in order generate a large amount of the recombinant protein to use in follow up biochemical validation experiments. This lengthy functional PTM characterization process is exemplified in the study of phosphorylation of the cytoskeleton regulatory protein cofilin[92]. The researchers in the study generated various phosphomimetic cofilin mutants and characterized the PTM-mediated dynamic conformational and functional shift using cryo-EM. Phosphorylation at the Ser-3 position of cofilin was found to regulate actin binding and severing by "undocking" and repositioning the protein N-

terminus away from the filament axis. In another example, a number of phosphomimetic mutants of translational initiation factor complex 3 (eIF3) were expressed to meticulously decipher the eIF3d-mediated translation of *Jun* during metabolic stress[93]. Although some data science attempts have been made to help identify functional PTMs from primary protein sequences[94-96]. In one report, PTM sites are mapped to proteins with available three-dimensional (3D) structural information in the Protein Data Bank (PDB) to predict sites that may affect protein-protein interactions[96]. However, the reliability of bioinformatics predictions remains to be seen.

In summary, traditional biochemistry approaches used to characterize modifications remain the rate-limiting step in PTM functional characterization. Despite the ever-growing list of novel PTM sites detected by LC-MS/MS-based approaches in massively parallel formats, we lack general methods with which to predict, or even prioritize the studies of, protein modifications that are likely functional.

1.2 Scope of the dissertation

My dissertation will focus novel method developments in the areas of LC-MS/MS-based quantitative proteomics to help identify PTMs with functional significance. The central questions remain at the core of PTM studies are how such small changes to the chemical structure of proteins alter the biophysical properties, and how these changes lead to signaling and phenotypic responses. I was involved in all aspects of the novel proteomics platforms development and applying them to discover previously known and unknown PTM-mediated interactions and pathways.

Chapter 2 discusses our recent development of a high-throughput quantitative proteomics platform termed "Hotspot Thermal Profiling" (HTP) to discover functional PTMs. This method sets out to measure the melting temperature (T_m) of thousands of modified proteoforms in parallel and utilizes the modification-induced thermal perturbation as a readout for functional significance. This section also discusses follow-up functional characterizations of previously uncharacterized modification sites detected by the HTP method.

Chapter 3 discusses the development of two HTP method variants with aims to improve precision in sample processing, expand the depth of (phospho)proteome coverage, and increase the accuracy in LC-MS/MS quantitative measurements. This section compares and contrasts the datasets generated with all three variants of HTP.

Chapter 4 discusses the application of HTP to explore the thermal stability of cells under dynamic metabolic stress. We apply a modified proteome integral stability alteration assay in a kinetic series to track the proteome thermal stability over time.

Chapter 5 summarizes the future aspects of each research project in response to the development thermal profiling strategies in the studies of protein post-translational modifications and envisions the future research directions.

CHAPTER 2

Hotspot Thermal Profiling (HTP) for the functional discovery post-translational modifications*

The work of this chapter is published: Huang et al, *Nature Methods,* 16.9 (2019): 894-901

2.1 Introduction

2.1.1 Cellular Thermal Shift Assay

The phenomenon that low-molecular weight ligands binding to proteins can increase the stability of proteins has been known for over a century. In 1890, O'Sullivan and Tompson first observed that thermal damages to enzyme invertase could be significantly reduced in the presence of cane sugar[97]. The invention of analytical technologies such as differential scanning calorimetry (DSC) and differential scanning fluorimetry (DSF) have allowed more sophisticated and quantitative studies of this biophysical phenomenon across a variety of ligands in complex with a variety of proteins[98-101]. More modern examples of protein-ligand interaction characterizations include thermal denaturation studies of ribonuclease A in complex with small molecule inhibitors and nucleotides[102], and tri-N-aceytylglucosamine (GlcNAc)3 binding to the active site of native lysozyme which led to significant stabilization of the protein to thermal guanidine hydrocholride

* The author's contribution to the studies presented in this chapter: The author conducted all experiments described in this chapter that led to the development of the HTP method, as well as target characterizations. The author designed and performed functional biochemical experiments, cell-based experiments, mass spectrometry experiments and analyzed data. The author wrote the manuscript with Raymond E. Moellering for the publication of this work.

denaturation[103]. In the latter example, the midpoint of the thermal denaturation curve (i.e. melting temperature T_m) of lysozyme increased by 5.3℃ in the presence of (GlcNAc)3.

Protein denaturation, or unfolding, describes the process from which the three-dimensional biologically functional structure of the protein in the native state unfolds into a one-dimensional sequence of amino acids in a random coil[104-107]. A typical protein with a two-state folding model follows a sigmoidal denaturation curve when it undergoes thermally-induced unfolding[107]. A characteristic melting temperature – T_m value – can then be calculated from the midpoint of the thermal denaturation curve. Upon ligand binding, the thermal stability of a protein often changes, resulting in a shift of T_m value[103]. This concept let to the development of the thermal shift assay (TSA) which has been used extensively in the drug discovery industry in the past two decades to perform compound library screens for protein-stabilizing small molecule drugs[108-110], as well as ligands that promote protein stability to help with protein crystallization and structure determination in academic research[111]. Differential scanning fluorimetry (DSF)-based methods measure the temperature at which a protein unfolds accompanied by an increase in the fluorescence of a dye with affinity for hydrophobic parts of the protein, which are exposed as the protein unfolds[101]. An alternative TSA method uses Differential Static Light Scattering (DSLS) which quantifies the amount of unfolded proteins that rapidly aggregate and has been used to interrogate protein-metabolite interactions[111]. After TSA techniques were miniaturized into a microplate format[109], it lead to the explosion of variant methods and platforms aimed at biomarkers and target identifications[112], such family-wide profiling and analysis of inhibitors for protein kinases[113] and poly-ADP-ribose polymerases[114]. In addition to protein-drug interactions studies, TSA has been used extensively to examine a variety of interactions such as protein-protein[115] and protein-nucleic acid[116] interactions.

However, conventional thermal shift assays are often done *in vitro* with purified proteins and cannot be used to measure directly in cells. This requirement limits the utility of the assay in global target discovery studies where the protein target is unknown; it also precludes the studies of proteins that cannot be easily purified. Isolating a protein from its native cellular environment to study protein-target interaction *in vitro* could also remove other endogenous factors such as intracellular localization, post-translational modifications, and interactions with other biomolecules that might be necessary for protein function.

A novel variant of the classic thermal shift assay, termed Cellular Thermal Shift Assay (CETSA), was first introduced by Molina and Nordlund[117] to study protein thermal stability *in situ.* The authors discovered that when cells are heated, many proteins rapidly unfold and precipitate out of solution in a similar fashion as do purified proteins. The amount of proteins that remain soluble correlated to the amount of proteins that remain folded, which could be isolated and quantified using simple techniques such as Western Blot[117] or induced proximity of two target-directed antibodies upon binding to soluble protein in solution[118] to generate classic protein melting curves. Specifically, this assay involves taking aliquots of samples of various origins (i.e. cultured cells or patient-derived materials) and heating them to a range or temperatures before cells are lysed and the soluble proteins are isolated. As opposed to measuring the melting behaviour of purified proteins, CETSA can measure the thermal stabilities of endogenous proteins in their native cellular environment. When ligands or drugs are introduced to live cells, CETSA can measure the thermal stability shifts of not only the direct protein targets, but also the indirect and downstream targets of the ligand-induced interaction, allowing the surveillance of network wide effects[119]. CETSA was first established using Western Blot-mediated detection to characterize protein-drug binding in lysates, cells, and tissues to demonstrate the wide biological applicability of technique.

A dose-dependence measurement performed at constant temperature was later introduced to allow further characterization of target occupancy. This two-prong approach consisting of temperature-varying and isothermal dose-dependent steps was used to discover and characterize the protein target of antifolate methotrexate in drug-resistant cancer cells[117], and thymidylate synthase inhibitors in the first compound library screen based on CETSA[120].

2.1.2 Expansion of CETSA using mass spectrometry for the interrogation of proteome-wide protein thermal stability

Given the simple and transferable nature of the technique, the next evolution of CETSA is the marriage of the elegant assay with quantitative mass spectrometry, a rapidly growing technique and field. LC-MS/MS-based mass spectrometry was the logical next step because it allows high-throughput detection of proteome-wide experiments. The first expansion of mass spectrometry based CETSA (MS-CETSA), termed Thermal Proteome Profiling (TPP), reported the use of isobaric labeling strategy to enable relative quantification of proteins from 10 different temperatures simultaneously[121]. Isobaric tandem mass tagging (TMT) reagents have the same nominal mass and chemical structure composed of a mass reporter, a spacer arm (mass normalization), and an amine-reactive NHS ester group that reacts with peptide N-termini[83] (Figure 2.3A). Two aspects of the TMT labeling strategy are integral to MS-CETSA. First, the isobaric tag allows barcoding of the whole, soluble proteome from different temperature samples. Second, the signal intensity of the mass reporter enables relative quantification of all detected proteins (measured at the peptide level) present at each temperature sample, ultimately allowing protein-specific melting curves to be fitted (Figure 2.3B). The MS-CETSA/TPP assays provide a

way to study target engagement by interrogating global protein stability across the proteome and experimental conditions in a high-throughput manner[122].

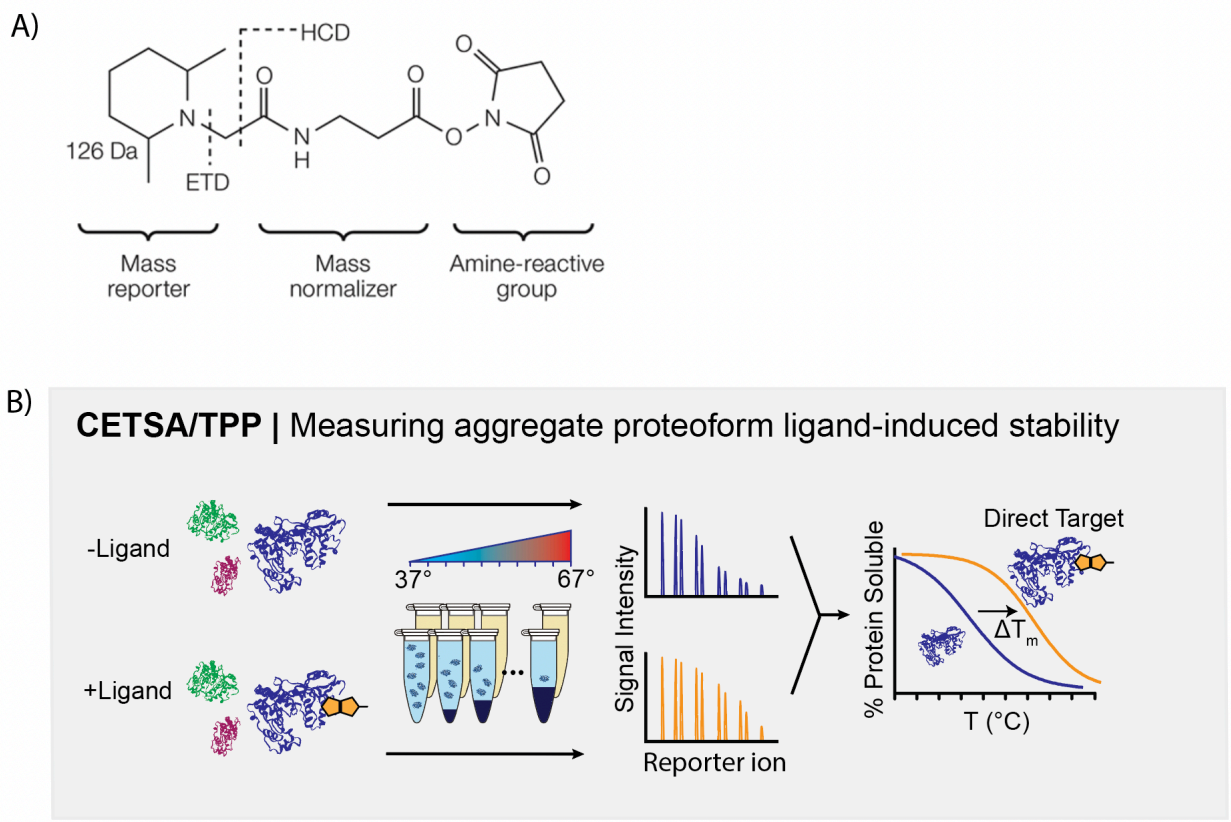

Figure 2.1 Mechanistic overview MS-CETSA/TPP
A) Chemical structure of isobaric labeling reagent tandem mass tag (TMT). **B)** General workflow of MS-based thermal profiling to measure aggregate proteoform stability in live cells (i.e. the bulk population of a protein of interest). Bioactive ligands can cause
altered protein stability, either because of direct ligand engagement or downstream effects on interactions with other biomolecules.

The primary application of MS-CETSA/TPP assays has been the massively parallel interrogation of ligand-induced, protein-specific stabilization or destabilization for thousands of proteins. This was first demonstrated in live jurkat cells using the pan-kinase inhibitor stauorsporine caused significant T_m shifts in dozens of known protein kinases[121]. TPP has subsequently been deployed to identify lipid kinase PIP4K as the target of dual-inhibitory compound "a131"[123], delineate the molecular mechanism of the CDK4/6 inhibitor Palbociclib-

induced proteasome activation[124], and validate the best-in-class MTH1 inhibitor in cancer cells[125]. In addition to mammalian proteomes, TPP has also been used to profile the "meltomes" of prokaryokes such as *E. coli*[126], and *P. falciparum*[127]which led to the identification of nucleoside phosphorylase (PfPNP) as the elusive target of antimalaria drugs quinine and mefloquine. Lastly, TPP has been successfully deployed to detect endogenous biomolecular interactions. TPP metabolite interaction mapping identified cognate transmembrane receptor STING which is involved in immune signaling as the target of metabolite 2'3'-cGAMP in live cells[128]. Meanwhile, proteome-wide solubility and thermal stability profiling of ATP revealed ATP-dependent solubilization of the proteome[129].

2.1.3 Post-translational modifications as a low molecular weight covalent ligand

In many ways, post-translational modifications are very similar to low-molecular weight ligands in how they impact the biophysical properties of the much larger proteins. PTMs are small chemical impingements that covalently modify their protein of interests, leading to changes in protein structure or biomolecular interactions, and eventually signaling and phenotypic responses in the cell. Since small molecule ligand-binding can affect the thermal stability of a protein through impacting its chemical and biophysical properties, it is logical to deduce that PTMs may also have similar effects on thermal stability of proteins. This assumption sparks the idea that MS-CETSA may also be modified and applied to query thermal stability changes impacted by the presence of PTMs on proteins.

There is a major advantage of this potential application of MS-CETSA to the studies of PTMs. In the original TPP workflow, the primary application is to interrogate ligand-induced thermal stability changes in the proteome. Although the method can potentially provide clues to

biomolecular interactions and network effects in response to a stimulus, it is impossible to differentiate whether a protein T_m shift is the result of a direct or indirect interaction with the small molecule. In other words, the ligand cannot be detected alongside its direct protein target and be directly associated with a protein T_m shift. TPP-generated T_m shift datasets, in their current form, simply provide a ranked list of proteins from most to least thermally perturbed in the presence of the added ligand and it is up to the researcher to follow up and decipher the functional significance of said T_m shift. On the hand, PTM *can* be detected alongside its direct protein target using LC-MS/MS-based proteomics technology which provides peptide-level information with amino-acid resolution. The presence of a PTM can be directly correlated to the presence (or the lack) of a protein T_m shift since LC-MS/MS-based proteomics methods are based on a peptide-mapping approach and can differentiate discrete proteoforms.

In this Chapter, we sought to develop a general method to overcome the limitations of current approaches to detect and discover functional PTMs. The quantitative proteomic approach utilizes principles from MS-CETSA/TPP approaches to enable direct, high-throughput interrogation of altered protein stability in response to endogenous, site-specific posttranslational modifications for thousands of proteins in parallel. This method, titled Hotspot Thermal Profiling (HTP), couples PTM-specific enrichments and isotopic labeling with TPP to globally detect and quantitatively measure the thermal stability of endogenous proteins and their site-specific, modified proteoforms, which we refer to herein as "modiforms," from *in vitro* and *in situ* samples. We applied this method to examine phosphorylation, the most commonly detected and well-studied PTM, as a proof-of-concept to demonstrate the utility of our approach.

2.2 Results

2.2.1 Hotspot Thermal Profiling method development

The HTP workflow begins with pulsed exposure of equivalent aliquots of live cells to temperatures ranging from 37 to 67°C for 3 minutes, followed by cooling to room temperature, rapid cell lysis in the presence of protease and phosphatase inhibitors, and centrifugation to remove unfolded proteins (Figure 2.2A). The soluble proteome remaining in the supernatant of each temperature-matched aliquot is then proteolyzed into tryptic peptides, and 5% of each peptide aliquot is labeled with specific tandem mass tag (TMT) isotope channels and reserved for protein-level thermal profiling of the "unmodified, bulk" protein pool. The remaining 95% from each tryptic peptide aliquot is used for phosphopeptide enrichment by TiO_2 resin, followed by TMT isotope labeling (Figure 2.2B). All aliquots of the unenriched and phosphorylated TMT tryptic peptides are subsequently combined into two respective pools, and used for LC-MS/MS proteomic analysis and quantitation of relative peptide levels across the temperature range. Finally, these peptide-level measurements are analyzed en masse with peptide identification, relative quantification and automated curve fitting algorithms to determine peptide-specific melting points (T_m) for tens of thousands of specific peptides in the same sample. A core property of this approach is the fact that these quantitative proteomic measurements are performed on enzymatically digested peptides, which are used to identify and report on their parent proteins. As such, measurements made on a unique modified peptide (e.g., phosphorylated peptide) serve as a 'bar-code' to report on any protein that contains a modification at that site to the thermal challenge in live cells. Therefore, this method does not report on single "proteoforms" as recently defined as the unique combination of modifications found on a protein at any given time[130-132], but instead reports on an inclusive set of the protein pool that are unambiguously phosphorylated at the

detected site, which we define as the "phosphomodiform." This is a key differentiation from measuring the effect of a modification on the bulk protein pool, such that HTP can specifically detect altered thermal stability even if a unique proteoform is only a minor member of all the modified proteoforms present in the cell.

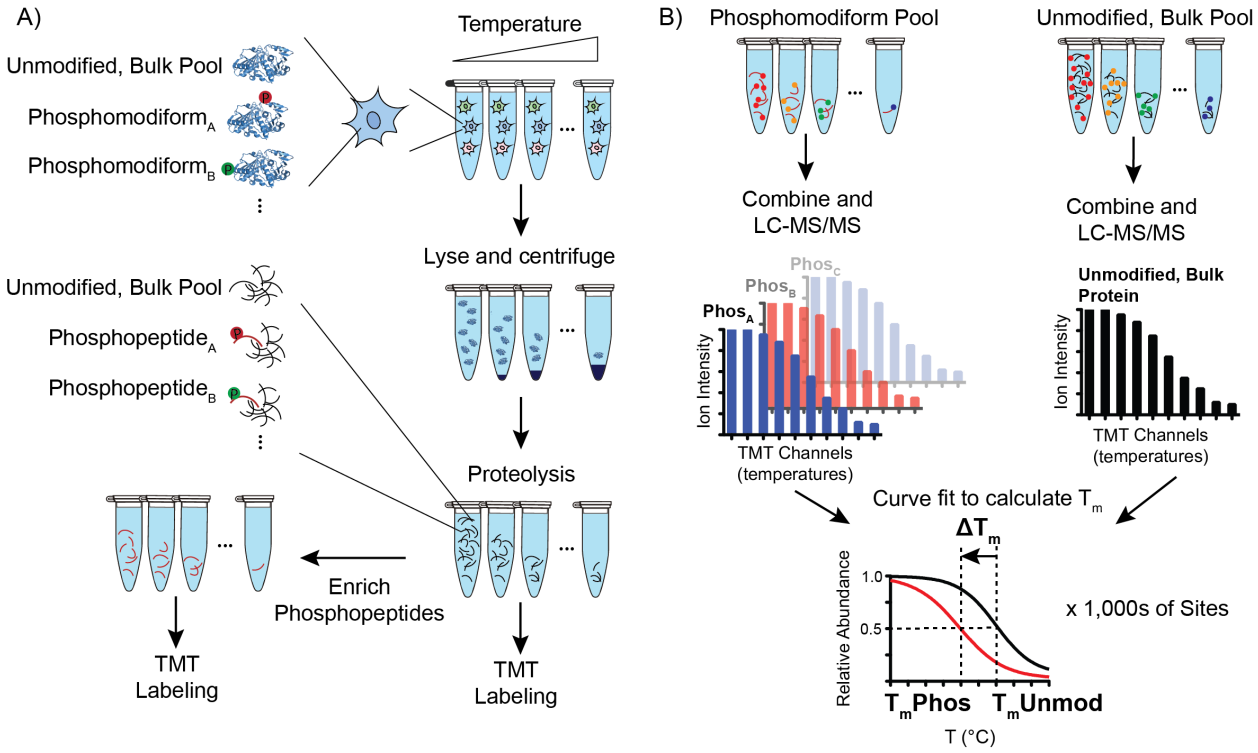

Figure 2.2 Schematic of Hotspot thermal profiling workflow
A) Aliquots of live cells are exposed to short, varied temperature challenge in parallel, lysed by rapid freeze-thawing, and insoluble proteins are removed by ultracentrifugation. Remaining soluble proteome is processed for trypsinization, and split for bulk, unmodified protein T_m measurements (5%), and phosphopeptide enrichment (95%). **B)** Bulk and enriched peptide aliquots are subsequently isotopically labeled with independent tandem-mass tag channels, pooled and analyzed by LC-MS/MS. The output of each mass spectrometry run is searched using an integrated platform for peptide identification and quantification, T_m curve fitting, and integration for bulk, unmodified protein- or modification site-level T_m determination. Unique phosphomodiforms and their daughter phosphopeptides are abbreviated as $Phos_A$, $Phos_B$…$Phos_X$.

2.2.2 HTP measures the in situ thermal stability changes in response to site-specific phosphorylation

We applied the HTP method to detect and compare the T_m values of phosphorylated proteins to their bulk, unmodified protein pool in HEK293T cells. These experiments were performed in live cells in order to capture the effects of phosphorylation in the native cellular environment, which we hypothesized would enable detection of both intrinsic (e.g., structural changes) and extrinsic (e.g., intermolecular interactions) effects of site-specific phosphorylation.

2.2.2.1 Effect of rapid temperature pulses on global phosphorylation level in live cells

To determine whether the 3-minute temperature pulse applied to live cells in the first step of the HTP workflow might have an effect on global phosphorylation level, and thus compound the experimental results derived from this method, we first measured the total phosphorylation level after different temperature pulses in HEK293T cells using Western blot. We exposed aliquots of cells to the same varied temperature series for 3 minutes, then lysed the cells and isolated the whole proteome – both soluble and insoluble fractions – and measured total phosphorylation from each temperature sample using a pan pS/T/Y antibody (Figure 2.3A). We performed densitometry quantification and determined that total phosphorylation level was not systematically affected by the short heat pulse (Figure 2.3B) and therefore any detectable melting curve shifts are the direct results of that phosphorylation site.

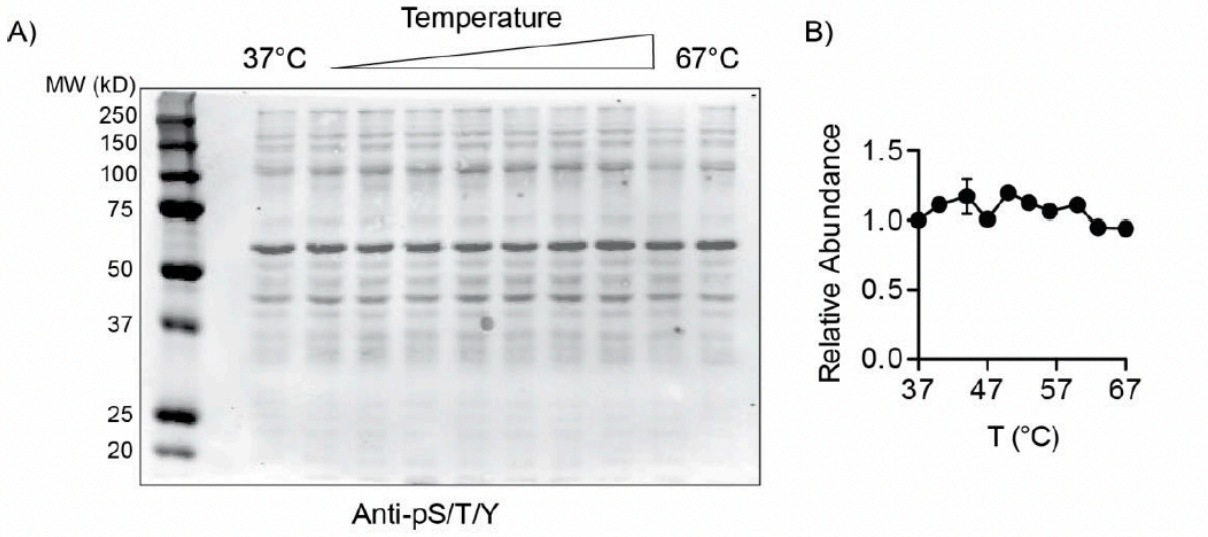

Figure 2.3 Effect of temperature pulse on global phosphorylation level
A-B) Pan phospho-serine, threonine and tyrosine antibody-stained western blot **(A)** on whole cell lysate following indicated temperature pulse for three minutes. Densitometry quantification of immunoblot signal is shown in **(B)**. Data points shown are mean and SEM, n = 2 independent biological replicates).

2.2.2.1 Defining the qualitative and quantitative filters in the HTP data analysis pipeline

Given the complex nature of our quantitative proteomics-based method, it is imperative to define how the HTP data will be qualitatively and quantitatively analyzed. LC-MS/MS-based proteomics technology is based on a peptide-mapping approach; in the case of HTP, we will be comparing the "phosphorylated modiform" (also referred to as phosphomodiform) of a protein to its "unmodified, bulk proteoform". We defined the first as all phosphopeptides that contain the same unique (combination of) phosphorylation site(s), and the latter pool as all unmodified peptides that map to a given protein of interest. The final T_m value assigned to the unmodified bulk protein reflects the bulk "reference" population of the protein present in the cells.

For the phosphomodiform pool, we grouped as all detected phosphopeptides that contain the same modification site(s) as the same modiform (Figure 2.4). This distinction is important because tryptic peptides often contain missed tryptic sites despite containing the exact same modifications. Therefore, the identity of a phosphomodiform is represented by the unequivocal presence of one single unique (or a combination of unique) phosphorylation site(s). All phosphopeptides that contain the same site (or sites) are grouped together and redesignated with a new ID (Gene_pSite) to indicate their unique phosphomodiform status. In addition, all detected phosphopeptides are further analyzed by LuciPHOR algorithm to test the confidence of the modification site localization before they are moved to the next step of the data analysis pipeline. Therefore, the final T_m value assigned to the phosphomodiform represents the melting temperature of the modiform that is unambiguously modified at a specific site regardless of if it is only a minor member of the proteome.

Protein Accesion	Peptide	Spec Count	TMT126	TMT127L	TMT127H	TMT128L	TMT128H	TMT129L	TMT129H	TMT130L	TMT130H	TMT131
P35611	K.SRSPGS(79.9663)PVGEGTGSPPKWQIGEQEFEALMR.M	4	440269	328538	443578	375591	279611	197907	236167	124475	140455	105110
	R.SPGS(79.9663)PVGEGTGSPPKWQIGEQEFEALMR.M	5	1307636	1141377	1439788	1170831	998542	862672	951488	457602	429344	370125
	↓ *Redesignate by Gene_pSite*					↓ *Add up Spec Count and each TMT Channel*						
	ADD1_pS358	9	1747905	1469915	1883366	1546422	1278153	1060579	1187655	582077	569799	475235
P35611	K.YSDVEVPASVTGYSFASDGDS(79.9663)GTCSPLR.H	3	262547	240351	228326	199987	147314	137376	108598	69320	63677	56650
	K.YSDVEVPASVTGYSFASDGDS(79.9663)GTCSPLRHSFQK.Q	1	132462	111486	87296	77293	40795	33827	29782	21088	17554	37154
	↓ *Redesignate by Gene_pSite*					↓ *Add up Spec Count and each TMT Channel*						
	ADD1_pS427	4	395009	351837	315622	277280	188109	171203	138380	90408	81231	93804
	… Repeat for all detected peptides											

Figure 2.4 Schematic that showing how phosphopeptides containing the same phosphosites are processed
Unique phosphopeptides that map to a common phosphosite (i.e., different tryptic status or charge state) are collapsed to a single phosphosite reporter, and subsequently to a phosphomodiform T_m value.

In terms of how the quantitative data from independent peptide-level measurements will be used to generate modiform-level melting curves, TMT signal intensities from all peptides that map to the same proteoform – that is, all unmodified peptides for bulk proteome, or all

phosphopeptides containing the same combination of unique phosphosites – are combined to generate one single protein modiform specific melt curve and T_m measurement. Each melting curve must be quantified in all 10 TMT channels and have a high-quality sigmoidal curve fit ($R^2 > 0.8$). Lastly, in order to obtain a high-quality ΔT_m, an unmodified bulk protein must also be detected for every detected phosphomodiform of interested, and there must be $n \geq 2$ high-quality melting curves on either side (Figure 2.5).

2.2.3 HTP detects protein thermal stability shifts in live cells

We applied the HTP method in live HEK293T cells to explore the global thermal stability shift mediated by phosphorylation.

2.2.3.1 Global T_m measurements of unmodified, bulk proteins and phosphomodiform

Using HTP, we detected >250,000 TMT-labeled peptides from thousands of proteins in HEK293T cells across all LC-MS/MS runs (n = 12 replicates for the unmodified, bulk proteome and n = 10 replicates for the phosphoproteome) (Figure 2.5). Statistical filtering at the peptide identification level identified 4,407 unique unmodified, bulk protein, and 10,800 unique phosphomodiforms after grouping distinct phosphopeptides that map to a common. Of those 10,800 unique phosphomodiforms, 7,769 of them belong to proteins also detected in unmodified bulk proteome (Figure 2.6A). The application of stringent automated T_m curve fitting cutoffs ($R^2 > 0.8$) (and site-localization filter specifically for the phosproteome) resulted in a database of site-specific T_m values for 4,293 unmodified, bulk proteins, and 9,695 distinct phosphomodiforms in the steady state HEK293T proteome (Figure 2.5). Finally, filtering of only high-confidence T_m values detected twice or more in both phospho- and unmodified, bulk proteomes, as well as

stringent modification site localization validation, yielded 2,883 high-quality ΔT_m values (Figure 2.5, Figure 2.6A).

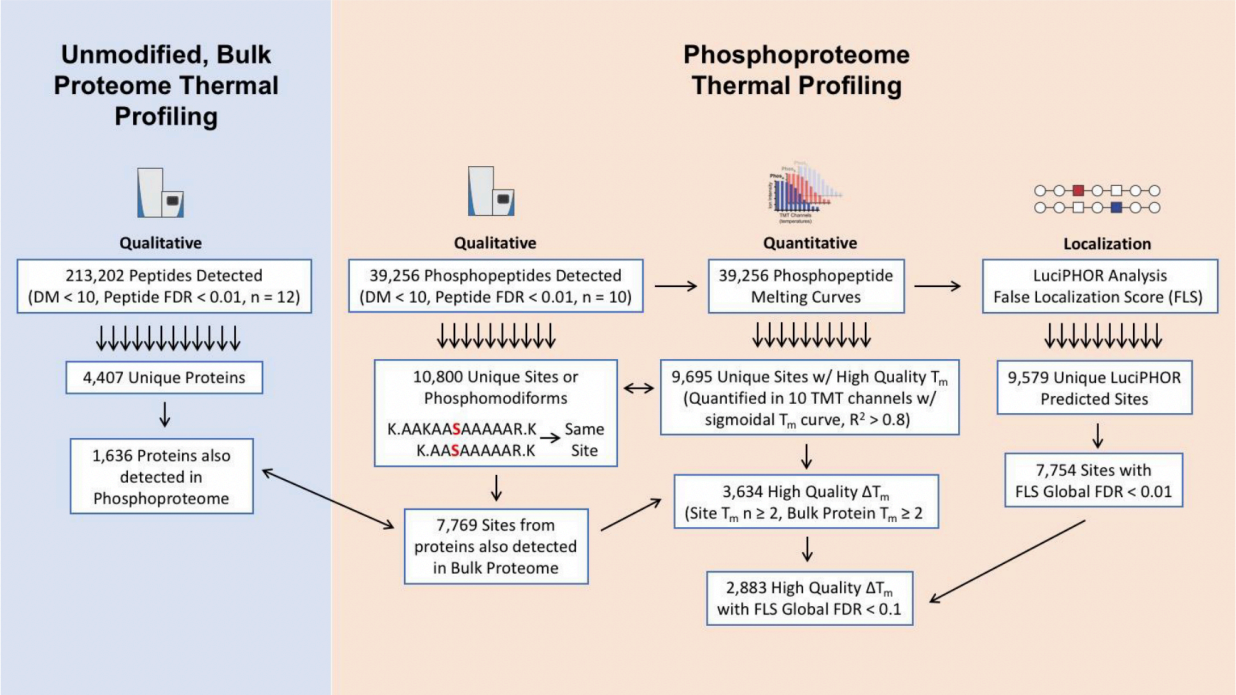

Figure 2.5 Schematic showing the total number of detected peptides and phosphopeptides in the unmodified, bulk proteome and phosphoproteome, and the data processing workflow. Stringent filters for qualitative, quantitative, and false localization analyses were used, which resulted in a total of 2,883 high quality ΔT_m's.

Global comparison of the unmodified, bulk protein pool with the phosphoproteome revealed nearly identical mean T_m values of 51.2 and 51.0°C, respectively (Figure 2.6B). Subsequent comparison of a given protein to its specific phosphomodiforms yielded a diverse database of ΔT_m values. The overall ΔT_m distribution was Gaussian in nature, and the majority of phosphorylated modiforms did not exhibit altered thermal stability compared to the corresponding unmodified protein. There were a sizeable number of phosphosites, however, associated with markedly different thermal stability compared to their unmodified, bulk protein pool in live cells.

Specifically, 719 of the phosphosites had statistically significant ΔT_m values relative to unmodified protein, defined by *p*-value < 0.05 calculated using 2-sided student's t-test (Figure 2.6C).

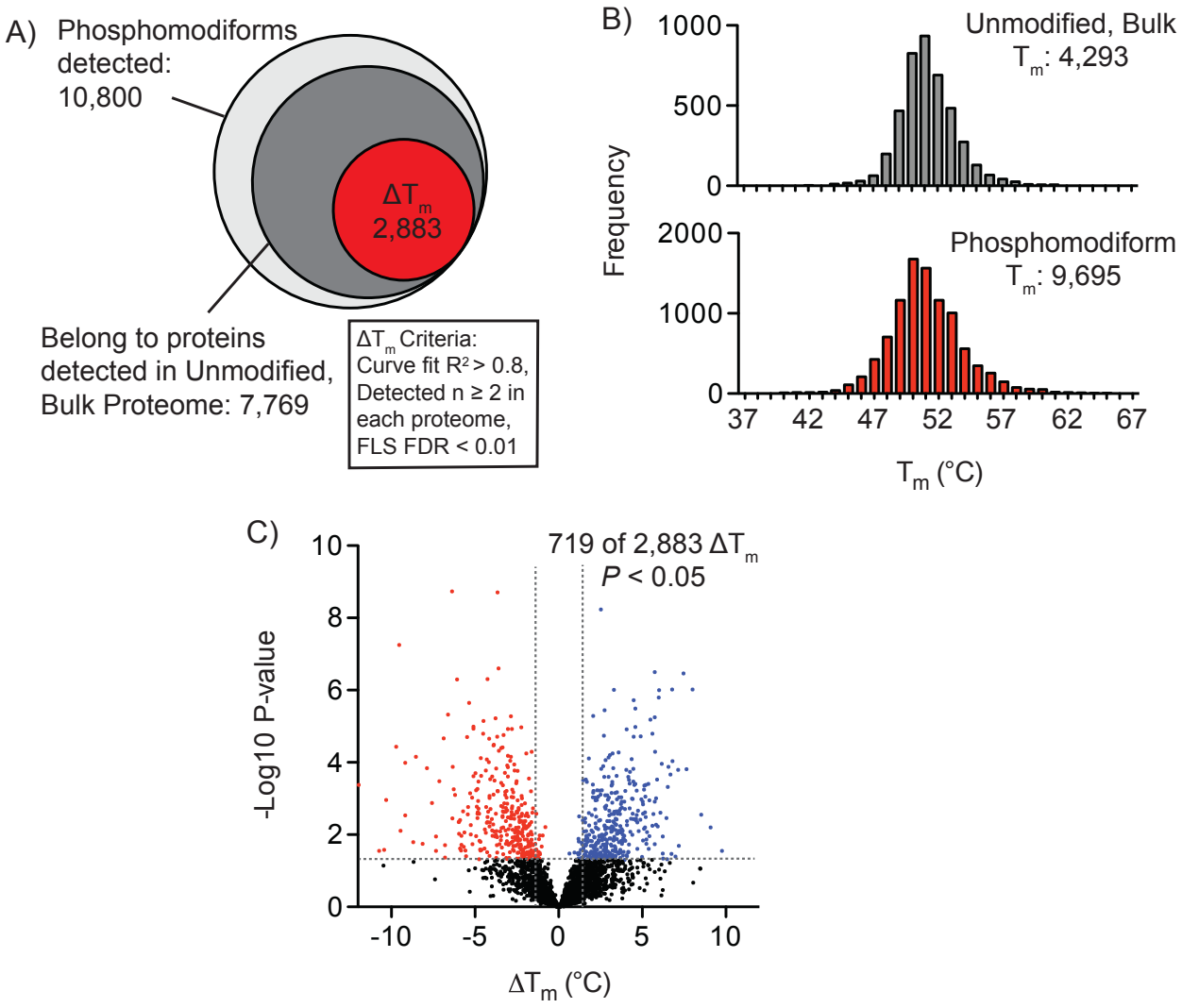

Figure 2.6 HTP identifies thousands of unmodified, bulk and phosphomodiform Tm's in HEK293T cells

A) The total number of unique phosphomodiforms detected in our dataset and criteria for identifying high-quality ΔT_m values. **B)** Global distribution of unmodified, bulk protein-level, and phosphomodiform T_m values. FLS FDR: localization site false discovery rate**. C)** Volcano plot of high-confidence ΔT_m values. Phosphomodiforms with significant shifts (P < 0.05, 2-sided t-test for within-protein comparisons) are shown in red (negative shift) and blue (positive shift).

It is important to note here that our analysis pipeline deliberately compares the thermal stability of specific phosphomodiforms (defined by unequivocal presence of a phosphorylation site) relative to the stability of the unmodified protein pool detected in cells (referred to as bulk, unmodified). In practice, this involves direct comparison of the mean T_m for a defined phosphomodiform to the mean T_m for the unmodified tryptic peptide pool for that protein in aggregate. We are not comparing each peptide to all other peptides one-by-one, as that represents a biochemically and statistically unique analysis. Therefore, each point on the volcano plot represents an internal comparison of two defined groups for each protein of interest, not comparisons among or between other populations (i.e., not a comparison of one T_m or ΔT_m to all of the other T_m values as is common in analyses presented on volcano plots) (Figure 2.7A). We considered applying more stringent multiple testing adjustment and observed the percentages of significant ΔT_m values go down uniformly: when we apply a population-level, multiple testing correction to generate *q*-values, the percentages of 'significant' ΔT_m values are reduced from 25% (*p*-value < 0.05) to 11% with adjusted *q*-value cutoff at < 0.05 and 4.1% with adjusted *q*-value cutoff at < 0.01 (Figure 2.7B).

2.2.3.2 Reproducibility of peptide-level and protein level T_m measurements

Protein level T_m measurements were highly stable across biological and technical replicates validating the reproducibility of cellular and proteome-level manipulations in the HTP workflow; bulk, unmodified protein T_m correlation R^2 between biological replicates is found to be 0.7 (Figure 2.8A). To understand how small of a phosphorylation-induced ΔT_m HTP can realistically capture, we looked into the how precise peptide-level T_m measurements are by examining the degrees of individual peptide-level T_m value deviations from the "aggregate" bulk protein T_m (Figure 2.8B).

Peptide-level T_m measurements across the proteome were found to be highly precise, with the median deviation for a given tryptic peptide T_m value relative to its aggregate protein-level T_m (i.e., the average T_m derived from many tryptic peptides in the same protein) being ~1.4ºC in four exemplary technical replicates, each comprised of >15,000 peptide and >3,000 composite protein melting curves (Figure 2.8C-F). These data indicate that even small shifts in thermal stability (e.g., ~1.5ºC) can be reliably detected with statistical significance.

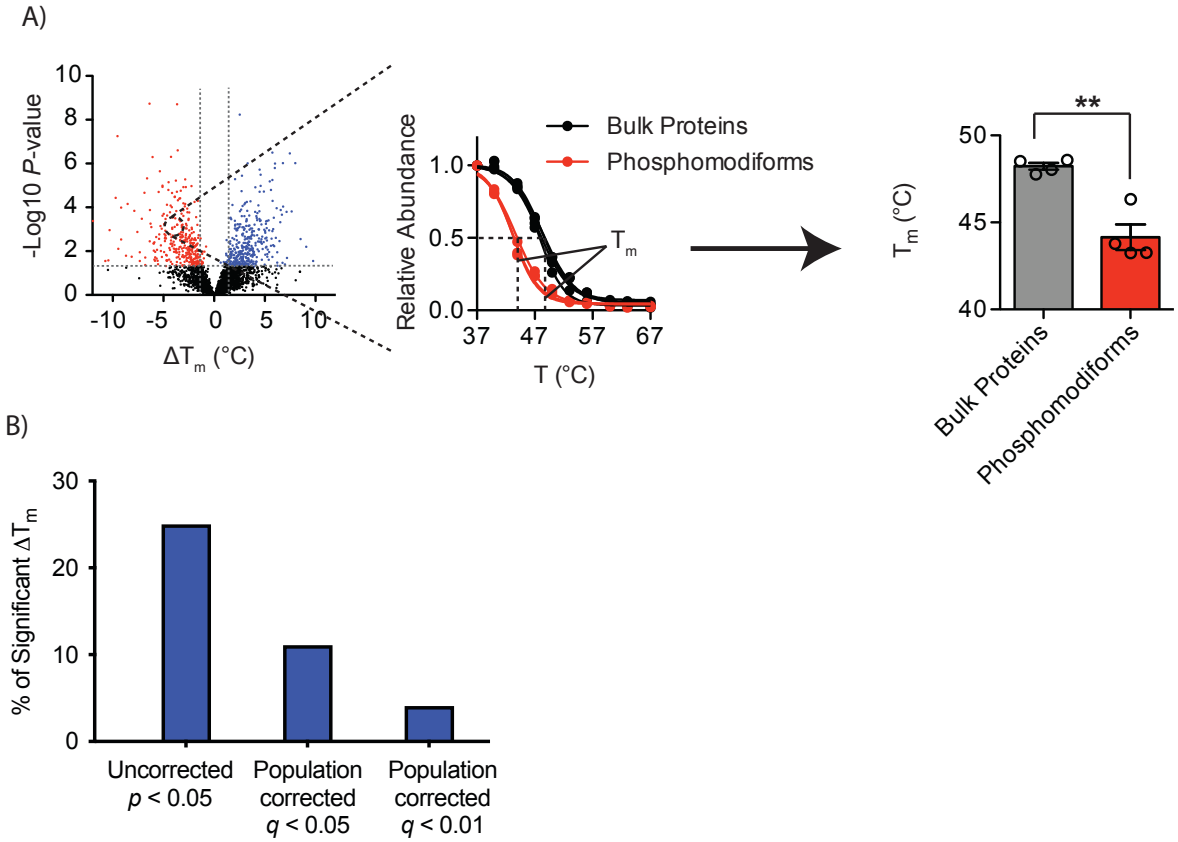

Figure 2.7 Statistical comparison strategy used to analyze ΔT_m values detected using HTP
A) Schematic of statistical comparisons used which compares the means and error surrounding bulk T_m and phosphomodiform T_m and assesses whether a site significantly alters protein stability. **B)** Comparing the percentages of significant ΔT_m values based on application of different statistical cutoffs. The percentages of significant ΔT_m values with p-value < 0.05 are 25% using analysis pipeline as described. When we apply a population-level, multiple testing correction to generate q-values, the percentages of 'significant' ΔT_m values are reduced. Specifically, application of q-value < 0.05 produces 11% and q-value < 0.01 produces 4.1% significantly shifted ΔT_m values.

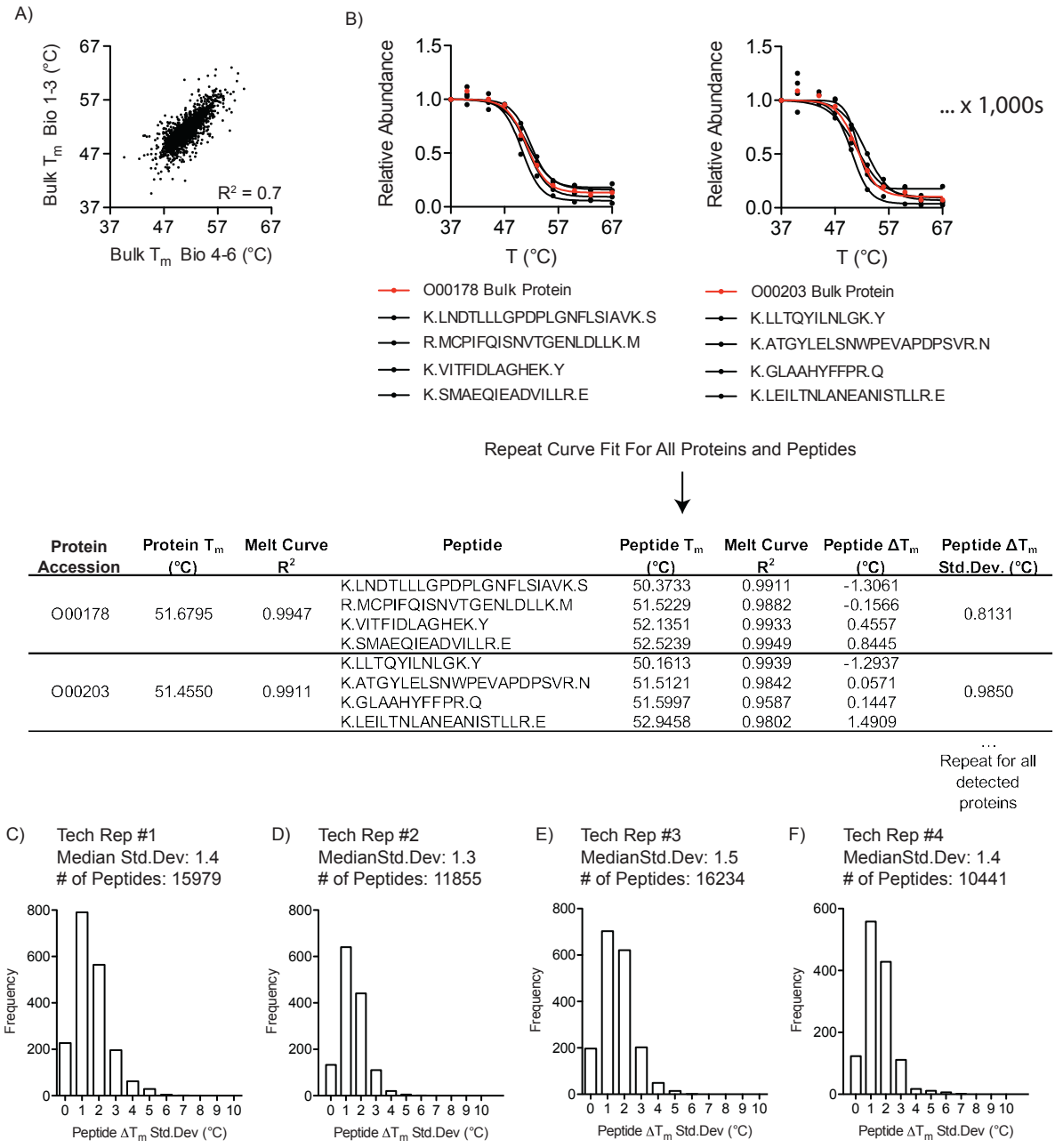

Figure 2.8 HTP dataset reproducibility and error analysis

A) Reproducibility of T_m values assessed with replicate experiments in bulk, unmodified proteome (linear regression). Only T_m values calculated from melting curves with $R^2 > 0.8$ are used for analysis in this paper. Data generated from n = 12 MS technical replicates from n = 6 independent biological replicates. **B)** Schematic that shows how peptide ΔT_m standard deviations are calculated for a typical MS technical run. Peptide ΔT_m standard deviations are the median deviation for a tryptic peptide T_m value relative to its aggregate protein-level T_m and this value demonstrates the "base error" of the HTP workflow. Reporter ion intensities of detected peptides were combined and converted to relative abundance to generate protein composite melt curve (red line) for each protein. **C-F)** The average deviation for any given tryptic peptide T_m value relative to its aggregate protein-level T_m is 1.4°C from four representative datasets.

2.2.4 Trends and functional correlations in phosphorylation-induced thermal stability shifts detected by HTP

We first queried our dataset to understand the types of thermal stability shift that can be detected by HTP from a site-specific view. Next, we investigated whether there are global biophysical, functional, and chemical trends imparted by the introduction of a small covalent modification on a protein. The intuitive idea that a locally charged PTM can increase steric bulk on a protein and alter inter- or intramolecular interactions as a result of local chemical environment has been generally assumed but never systematically tested.

2.2.4.1 HTP detects three types of site-specific thermal stability shifts

The resulting dataset demonstrated that three types of thermal stability shift that can be observed and detected with confidence by HTP. Many curves from parent protein-phosphomodiform pairs were typified by sigmoidal curves that overlay almost exactly, exemplified by the paired amphipathic helix protein Sin3A and its pS832 modiform, corresponding to negligible ΔT_m (Figure 2.9A) In contrast, numerous phosphomodiforms exhibited significantly different temperature-response curves compared to their unmodified, bulk protein counterpart, such as pY15 on cyclin-dependent kinase CDK1 ($\Delta T_m = 7.1 \pm 1.2°C$, $P = 7.1 \times 10^{-7}$; Figure 2.8B), and pT926 on kinesin-like protein KIF11 ($\Delta T_m = -5.4 \pm 1.3°C$, $P = 2.7 \times 10^{-3}$; Figure 2.8C). Interestingly, both of these phosphorylation sites are well-annotated to have critical roles in regulating the function of their target protein: the phosphorylation of Y15 has been shown to inactivate CDK1 through structural reorganization and autoinhibition[133, 134], while phosphorylation of T926 controls intermolecular association of KIF11 with the spindle apparatus to establish bipolarity during mitosis[135]. Instead, the direct interrogation of phosphomodiforms with amino-

acid resolution validates the ability to distinguish functional effects at different modification sites. Furthermore, the approach can provide a glimpse at co-occurrences of modification sites, as well as minor phosphomodiforms whose signal, and therefore function, would often be masked by other abundant forms of the same protein pool. For example, two distally located phosphorylation sites can be detected on glycolytic enzyme triosephosphate isomerase TPI with significantly different ΔT_m values: the pS58 phosphomodiform of TPI1 is associated with significant destabilization (ΔT_m = -3.6 ± 0.6°C, P = 7.4×10^{-6}), while pT210 has no effect on its stability (ΔT_m = -0.2 ± 0.7°C; Figure 2.8D). The difference in ΔT_m values may signify that these two modification sites do not occupy TPI at the same time as they are associated with distinctively different melting behaviours. In summary, these data confirm that changes in protein stability in response to protein phosphorylation on diverse proteins can be detected by thermal profiling of the modified proteome. Most importantly, we posit that "hotspot" modification sites identified in this screen – those significantly shifted relative to the global protein T_m and even other sites within the same protein – may represent sites with disproportionate effects on protein structure and function under specific cellular conditions.

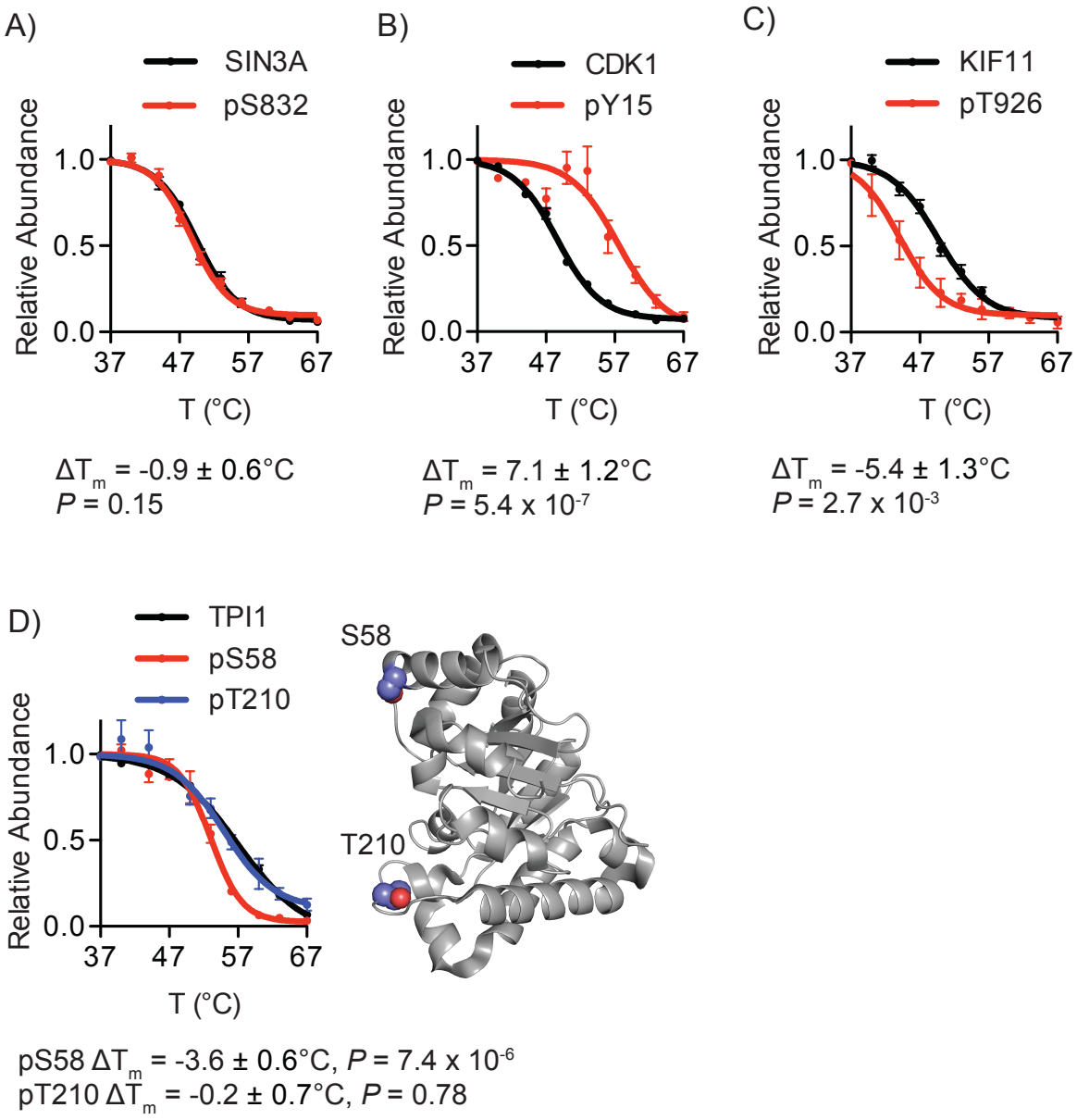

Figure 2.9 Types of T_m shifts detected by HTP

Representative T_m curves from phosphomodiforms that do not affect **(A),** decrease **(B),** and increase **(C),** thermal stability. Distinct phosphomodiforms of the same protein exhibit unique ΔT_m values, as exemplified in TPI1 **(D).** X-ray structure of human TPI1 is shown (grey) with detected phosphorylation sites shown as spheres. PDB accession: 1WYI. T_m curves and error bars in **(A-D)** correspond to mean and SEM. ΔT_m *P*-value computed from 2-sided t-test. All data points were generated from n = 12 total technical replicates from n = 6 independent biological replicates (bulk, unmodified proteome) and n = 10 total technical replicates from n = 5 independent biological replicates (phosphoproteome).

2.2.4.2 Perturbation of protein stability by phosphorylation is site-specific, not systematic

The idea that a covalent modification can change the chemical environment on a protein to alter inter- or intramolecular interactions has not been possible to query systematically because of the inability to test the effect of specific protein modifications on the stability of native proteins in parallel. With this global dataset in hand, we asked whether there were any systematic correlations between the types and contexts of protein phosphorylation sites and their effect on protein stability. Comparison of the ΔT_m values for phosphorylation on different residues (serine, threonine and tyrosine) across all detected phosphomodiforms did not reveal any general bias for stabilization or destabilization (Figure 2.10A). Likewise, comparison of modiforms harboring multiple phosphorylation sites within the same quantified tryptic peptide, and therefore in high proximity on the protein, did not reveal strong differences. Evaluation of general ΔT_m values among established kinase motifs also did not reveal significant correlations, indicating that the combination of phosphoamino acid and directly adjacent sequence context does not, at least on a global average in this dataset, impart a systematic stabilizing or destabilizing effect (Figure 2.10B). Detected phosphosites in our database showed enrichment in predicted loop- or coil secondary structural elements, which matches previous bioinformatic analysis of the PDB-deposited structures (Figure 2.10C)[136-140]. Phosphosites predicted to be in coil regions were correlated with increased stabilization relative to those that fell within helical or β-sheet secondary structures (Figure 2.10D). Additional computational predictions of solvent accessibility and local structure organization revealed a significant stabilizing trend for phosphosites located in exposed protein regions relative to those located in buried regions (Figure 2.10E-F). Taken together, these global measurements on native phosphomodiforms indicate that the effect of protein phosphorylation –

insofar as can be gleaned with this method - is governed by local context within proteins, as opposed to a systemic, biophysical "code" imparted by the phosphoamino acid or linear motif.

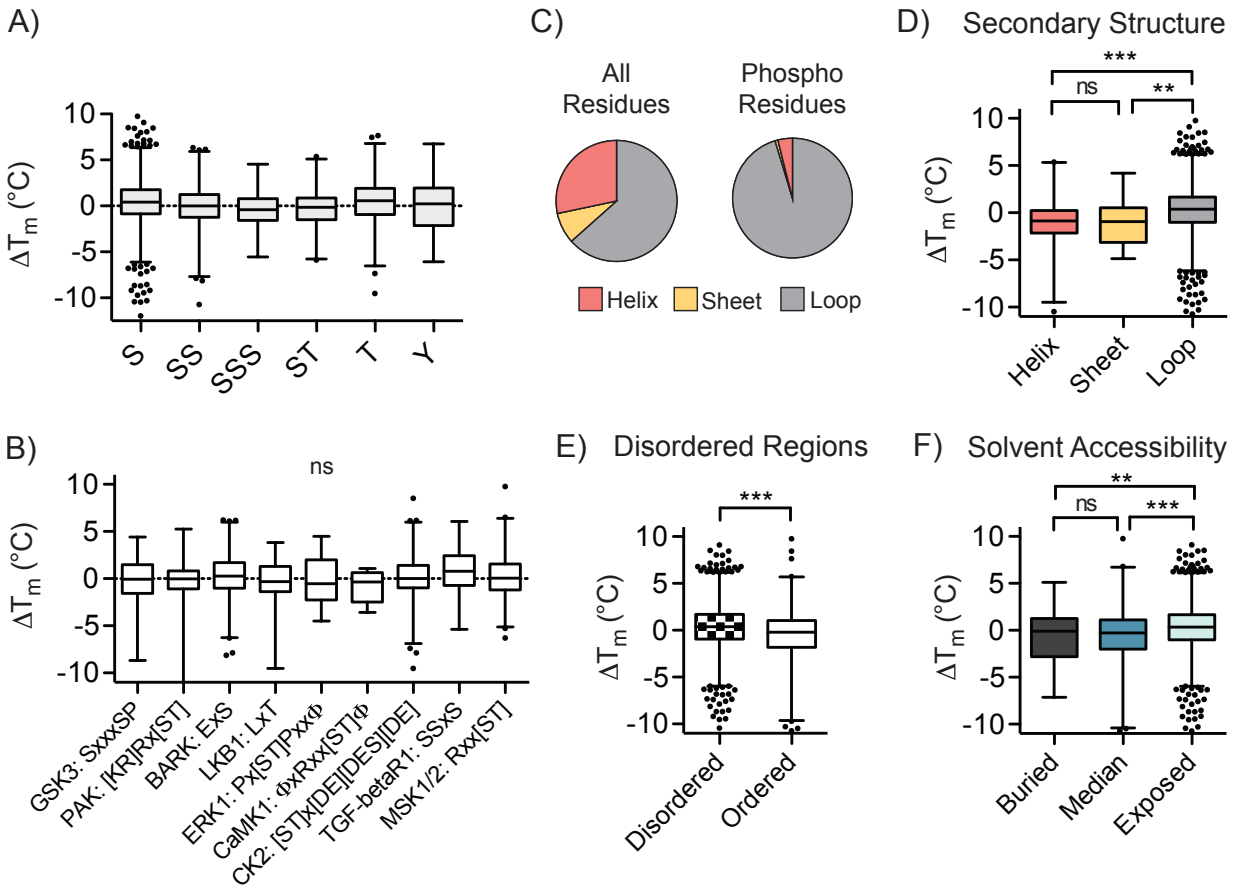

Figure 2.10 Global relationships between local phosphosite environment and altered phosphomodiform stability

A) Distribution of ΔT_m values for tryptic peptides containing indicated phosphoamino acids or coincidental combinations thereof ($P > 0.05$, One-way ANOVA, only groups with $n \geq 25$ are shown). **B)** Correlation between detected kinase substrate motifs and ΔT_m values for identified sites ($P > 0.05$, One-way ANOVA). **C)** Distribution of predicted secondary structure elements in which all amino acids in the human proteome (left) and detected phosphosites (right) are found. **D-F)** Comparisons of ΔT_m values and predicted secondary structure elements **(D)**, (***$P \leq 0.0001$, 2-sided t-test), ordered structural elements surrounding the phosphosite of interest **(E)**, (***$P \leq 0.0001$, 2-sided t-test), and solvent accessibility **(F)**, (***$P \leq 0.0001$, 2-sided t-test). Box plots (median, 1-99%) are shown in (A-B, D-F), with outliers shown as data points.

2.2.4.3 Functional significance and degrees of PTM-induced thermal perturbation

Next, we asked if the degrees of T_m shift have any correlation with the functional significance of that phosphorylation site. We explored this concept of "functional significance" by looking into number of annotations per detected phosphosites as we reasoned that the functionally important sites must have studied more closely. However, despite phosphorylation being the most commonly detected and well-studied of all PTMs, we were surprised to find that a vast majority of sites (92.5%) detected in our dataset has no annotation according to the PhosphositePlus online database[86]. As a result, there is little to no global correlation between the degrees of T_m shift and functional annotations (Figure 2.11). Besides well-characterized, functional sites in proteins like Ribosomal protein S6 kinase beta-1 (RPS6KB1), cyclin-dependent kinase 1 (CDK1), signal transducer and activator of transcription 1 and 3 (STAT1/STAT3) which stood out on this annotation plot, the vast majority of PTM sites are uncharacterized in the human proteome.

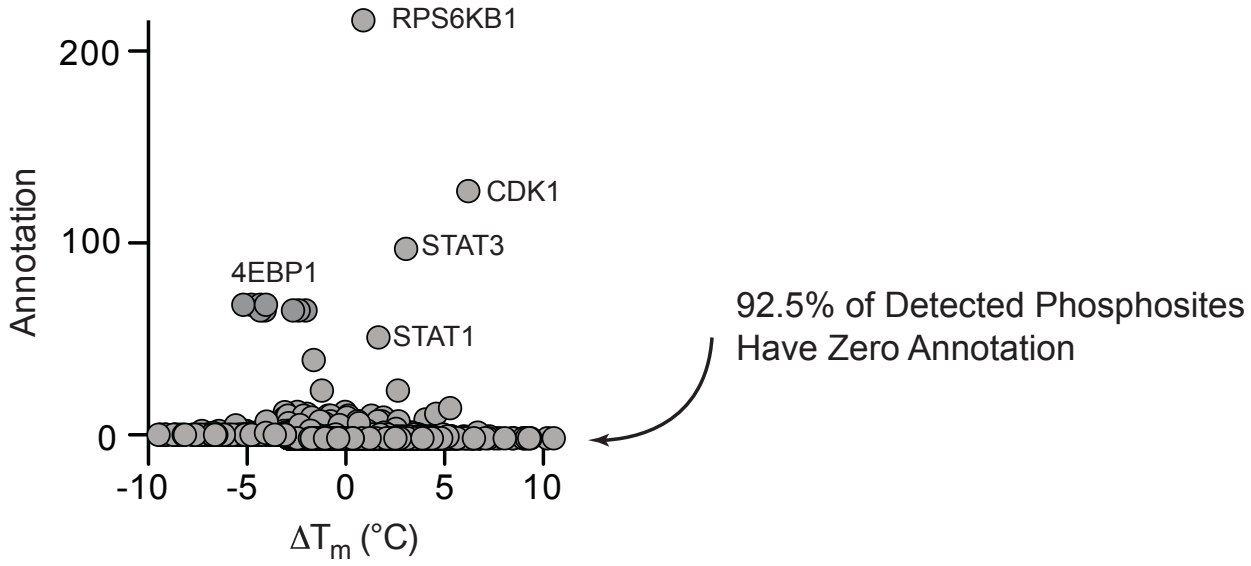

Figure 2.11 Plot of phosphosite ΔT_m and number of annotations in PhosphositePlus database Besides several well characterized proteins, 92.5% of detected phosphosites in our dataset have no annotations.

2.2.5 HTP detects known phosphorylation-mediated protein-protein interaction in 4EBP1-EIF4E

Our previous correlation analysis between the number of annotations and degrees of T_m shifts revealed a cluster of highly perturbed and well-annotated sites belonging to the protein 4-enhancer binding protein-1 (4EBP1) in the global annotation-ΔT_m plot (Figure 2.11A). The unmodified, bulk 4EBP1 pool exhibited a T_m of 53.8 ± 0.8°C by HTP, whereas the phosphomodiforms containing phosphorylation sites in the N-terminal region were significantly less stable, with T_m values depressed by 4.7 ± 1.0°C on average ($P = 4.4 \times 10^{-6}$) (Figure 2.12A). These sites were clustered in peptide stretch known to be responsible for direct binding to eukaryotic translation initiation factor 4E (EIF4E), which prevents ribosome binding and translation[141, 142]. Biochemical studies have established a phosphorylation cascade that requires T37, T46 and intervening phosphorylation sites as initiating modifications, which then lead to phosphorylation of Y54, which is necessary and sufficient to block 4EBP1 binding to EIF4E[143]. The thermal shifts resulting from phosphorylation around this site are consistent with steric and electronic perturbation of the peptide surrounding Y54 binding the hydrophobic ledge in EIF4E (Figure 2.12B), which further suggests that the shift in thermal stability observed for phosphorylated 4EBP1 modiforms is likely measuring the bound and unbound forms of this protein on a macromolecular complex, respectively. Specifically, 4EBP1 proteins lacking N-terminal modifications may demonstrate higher thermal stability due to involvement in the protein-protein interaction with EIF4E and likely other members of a macromolecular complex. This is contrasted by significantly reduced thermal stability of N-terminal phosphorylated modiforms that are incapable of engaging in this interaction and are either free in solution or engaged in an alternate set of interactions (Figure 2.12C). While the data provided by HTP global analysis cannot

distinguish these scenarios per se, this method is able to detect these previously annotated functional "hotspot" sites in a global, unbiased manner, and indicate that HTP can detect the consequences of altered intermolecular interactions, which would be impossible to identify in purely in vitro systems.

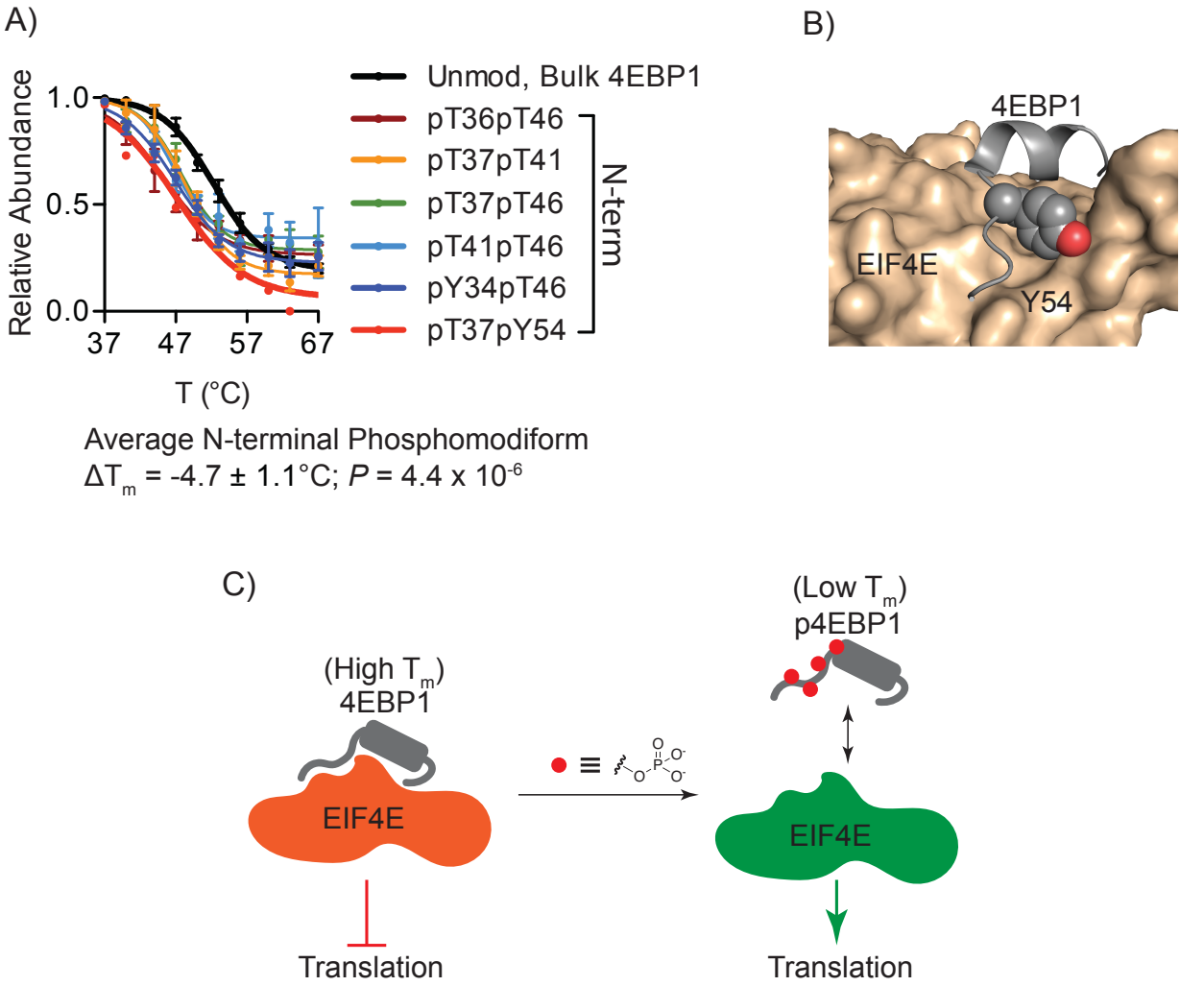

Figure 2.12 Effects of N-terminal phosphorylation sites on 4EBP1 protein stability and function

A) Representative HTP curves for the unmodified, bulk and N-terminal phosphomodiforms of 4EBP1. **B)** X-ray structure of the N-terminal helix-turn motif of 4EBP1 (grey) bound to EIF4E (tan), with the sidechain of Y54 in 4EBP1shown engaged with the hydrophobic ledge on EIF4E. The N-terminal phosphorylation sites upstream of T46 do not show electron density in this structure. **C)** Proposed model correlating high thermal stability of the non-phosphorylated 4EBP1 bound to an EIF4E macromolecular complex, transitioning to decreased stability of the unbound, phosphorylated modiform.

Furthermore, HTP detected several additional phosphorylation sites on 4EBP1 with significantly different melting behaviours to the N-terminal phosphomodiforms. Contrary to the thermally destabilizing N-terminal phosphorylation cascade, C-terminal phosphosite pS101 is associated with significantly increased stability (6.7 ± 1.3°C, P = 7.8 x 10^{-3}; Figure 2.13A). Phospho-S101 is known to be distal to the 4EBP1-eIF4E binding interface and not involved in the direct binding of these two proteins[144]. The large difference in ΔT_m values and melting behaviours may signify that these modification sites do not occupy 4EBP1 at the same time. The ability for HTP to detect unique phosphomodiforms from the same protein may therefore provide information on the co-occurrence and potential cooperation of modifications, as well as their mutual exclusivity, on the same protein molecules (Figure 2.13B).

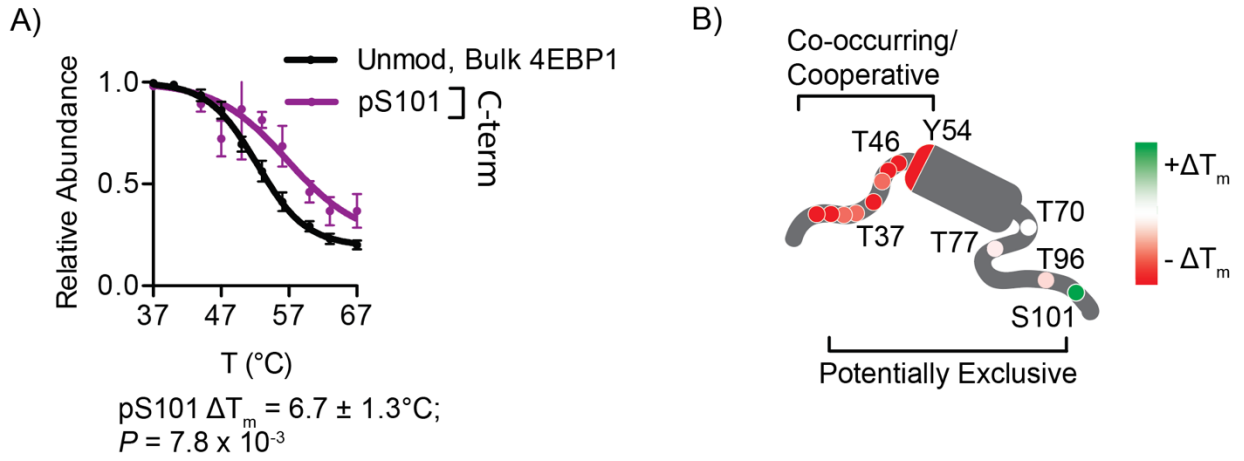

Figure 2.13 Effects of C-terminal phosphorylation site on 4EBP1 protein stability and function
A) HTP curve for C-terminal phosphomodiform S101. **B)** Schematic showing the location and relative ΔT_m values for all detected phosphomodiforms of 4EBP1, with annotation of potentially co-occurring and exclusive modification patterns based on thermal response.

2.2.6 HTP detects previously uncharacterized phosphorylation-mediated intramolecular changes in Vinculin

Learning from the previous example that HTP can detect protein-protein interaction mediated by site-specific phosphorylation, we hypothesized that PTM-induced shifts in thermal stability resulting from changes to protein biophysical properties can permit discovery of novel functional hotspot sites that would be otherwise unannotated in the sea of detected sites. To test this hypothesis, we first investigated an uncharacterized phosphorylation sites in cytoskeletal protein Vinculin that induced a significant thermal stability shift.

2.2.6.1 Significant thermal perturbation by pS721 in Vinculin

Vinculin exhibited typical unfolding in the unmodified, bulk protein population, which contrasted significantly with the pS721 phosphomodiform (ΔT_m = +7.6 ± 0.7°C, $P = 5.1 \times 10^{-6}$; Figure 2.14A). Analysis of published X-ray structures of Vinculin revealed that S721 lies at the N-terminus of an α-helix in the D4 domain, between the 'head' and 'tail' lobes of the protein[145] (Figure 2.14B). Residues within this hinge region have been implicated in conformational switching between an active open form, in which the two lobes are extended, and an inactive form, in which contacts between the tail, D1, and D4 domains stabilize a closed conformation and close proximity of the two nodes. The functional outcome of this switching is the regulation of Vinculin clustering, recruitment to focal adhesions and binding to F-actin stress fibers in the open conformation, and inability to form these "adhesome" complexes in the closed form, although the mechanisms governing vinculin activation remain incompletely understood.

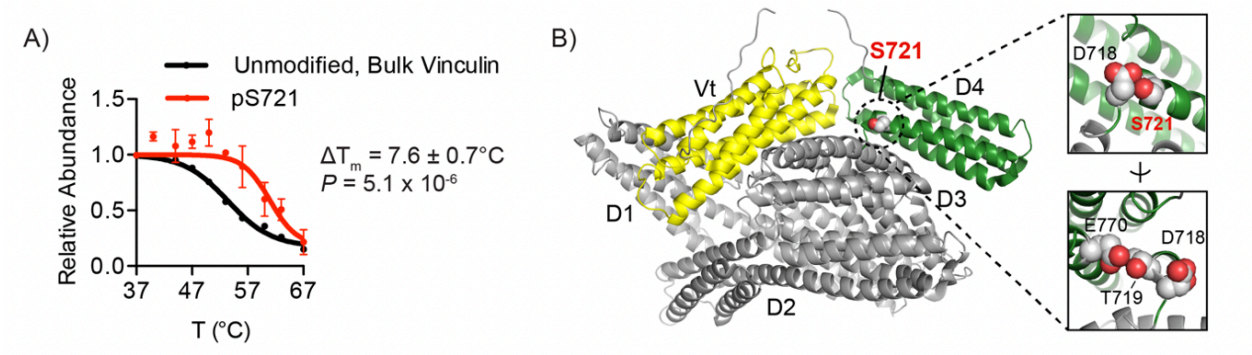

Figure 2.14 HTP detects significant phosphosite-induced thermal perturbation in Vinculin **A)** HTP curves for the unmodified, bulk and pS721 phosphomodiform of Vinculin. **B)** X-ray structure of human Vinculin with the uncharacterized phosphorylation site S721 shown as spheres packed in between the Vt (yellow) and D4 (green) domains. PDB accession: 1TR2. Inset images show close contacts between S721, D718, T219 and E770, which cap a buried helix in the D4 domain.

2.2.6.2 Phenotypic characterization of Vinculin phosphomimetics

Given the proximity of S721 to critical contacts within the D4 domain, we hypothesized that phosphorylation of this residue may result in altered intramolecular contacts to shift the equilibrium toward one state, and that altered thermal stability might be detecting this change or a functional consequence thereof. We interrogated wild-type, non-phosphorylatable (S721A) and phosphomimetic (S721D) Vinculin constructs for the formation of characteristic focal adhesions, the hallmark phenotype for proper function of this protein[146]. Immunofluorescent visualization of FLAG-tagged, wild-type Vinculin in HeLa cells showed high localization to distal membrane regions, which overlapped with phalloidin-stained F-Actin and a marker of focal adhesions, Paxillin (Figure 2.15A). The phosphorylation dead S721A protein exhibited a similar overall distribution, with the majority of this mutant co-localizing to a similar size and number of focal-adhesions per cell (Figure 2.15A-B). In contrast, localization of the phosphomimetic S721D mutant protein was significantly diminished, as measured by a 50% decrease (P = 0.002) in Vinculin/Paxillin co-localization compared to cells expressing S721A mutant (Figure 2.15C) as

well as a 58% reduction ($P < 0.0001$) in the average size Vinculin clusters (Figure 2.15D). These data are consistent with a model where phosphorylation at S721 is destabilizing the Vt-D4 contact in the closed form, and may stabilize an alternative, inactive conformation with diminished recruitment to focal adhesion complexes. Furthermore, we posit that the observed changes in thermal stability of the pS721 Vinculin phosphomodiform maybe capturing both a conformation change resulting from perturbed intramolecular contacts, as well as the stabilization of an alternate conformational state that is then capable of engaging in new intermolecular interactions. These data provide a proof-of-concept example of the effective prioritization and discovery of uncharacterized, but phenotypically relevant hotspot modification site by HTP.

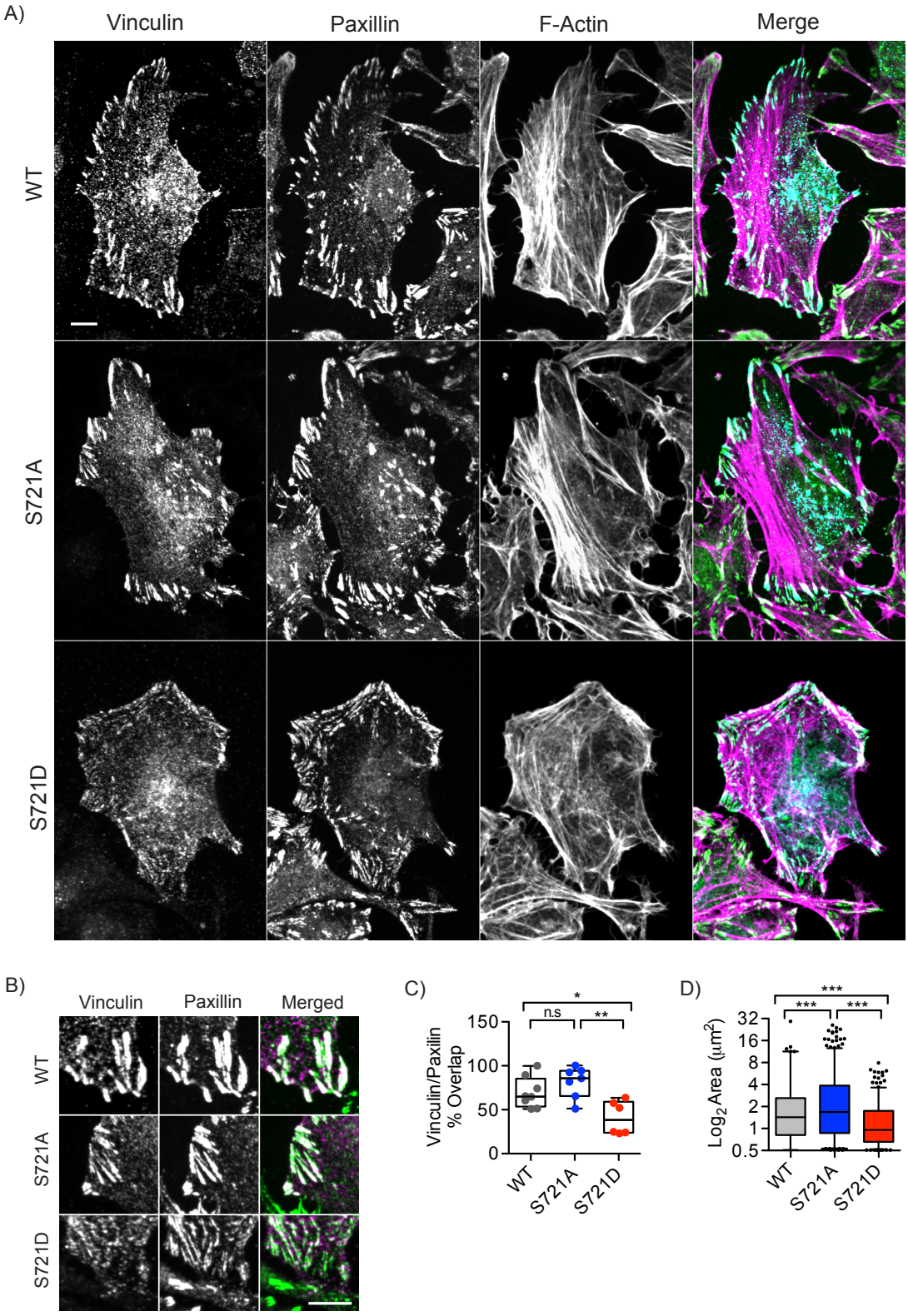

Figure 2.15 In situ characterization of Vinculin phosphomimetics

A) Confocal microscopy of wild-type, S721A and S721D FLAG-tagged vinculin constructs transiently expressed in HeLa cells. Merged, individual, and colocalization channels are

(**Figure 2.15 continued)** shown for Vinculin (cyan), Paxillin (green), and actin (purple). **B**) Representative focal adhesions at migrating fronts of the cell membrane are highlighted and show significant phenotypic differences between the three Vinculin constructs. **C**) Co-localization of Vinculin and Paxillin at Paxillin-positive focal adhesions, with each point represents percent colocalization in a quantified cell (n = 8, n = 7, and n = 6 cells, respectively). Each dot represents a distinct cell. **D**) Size of WT, S721A and S721D-Vinculin focal adhesions in n = 8, n = 7, and n = 6 cells, respectively. Each point represents a distinct focal adhesion. Scale bars in (C-D) = 10 μm. Data points shown in (A) represent the mean and SEM from $n \geq 2$ replicate measurements detected in $n = 6$ (bulk, unmodified) and $n = 5$ (phospho) biological replicates. ΔT_m *P*-value computed from 2-sided t-test. Experiment in (C-D) was independently repeated three times with similar results. Box plots (median, 1-99%) are shown in (E-F); $*P \leq 0.05$, $**P \leq 0.001$, $***P \leq 0.0001$, n.s. $P > 0.05$, One-way ANOVA.

2.2.7 HTP detects previously uncharacterized phosphorylation-mediated protein-metabolite interaction in GAPDH

Expanding on the potential to discover novel functional phosphorylation sites, we next interrogated one of the most well-characterized protein, glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase (GAPDH) which displayed significant altered thermal stability in response to phosphorylation.

2.2.7.1 Significant thermal perturbations by pS210 in GAPDH

The pS210 phosphomodiform of the GAPDH exhibited a significant destabilization ΔT_m of -6.3 ± 0.8°C relative to the unmodified, bulk GAPDH pool (Figure 2.16A). Due to high abundance of GAPDH in the proteome, we were also able to detect and measure the melting curve of GAPDH modiform that is unambiguously unphosphorylated at S210 through direct tryptic peptide comparison. GAPDH_pS210 modiform is significantly destabilized compared to the S210 unphosphorylated GAPDH modiform which has identical melting curve as the bulk, unmodified (Figure 2.16A). Other detected sites, such as pT154, were not associated with a significant ΔT_m

shift (Figure 2.16B). X-ray crystal structures show that S210 is located at the entrance of the D-glyceraldehyde 3-phosphate (GAP) substrate binding pocket, where it packs closely to L194 and by extension K195, which is involved in substrate binding (Figure 2.16C). Meanwhile, T154 is distal to the active site pocket which may explain its lack of impact on thermal stability of the enzyme. These results indicated that this previously observed yet uncharacterized site may affect GAPDH substrate binding and catalysis due to proximity to the GAP binding pocket.

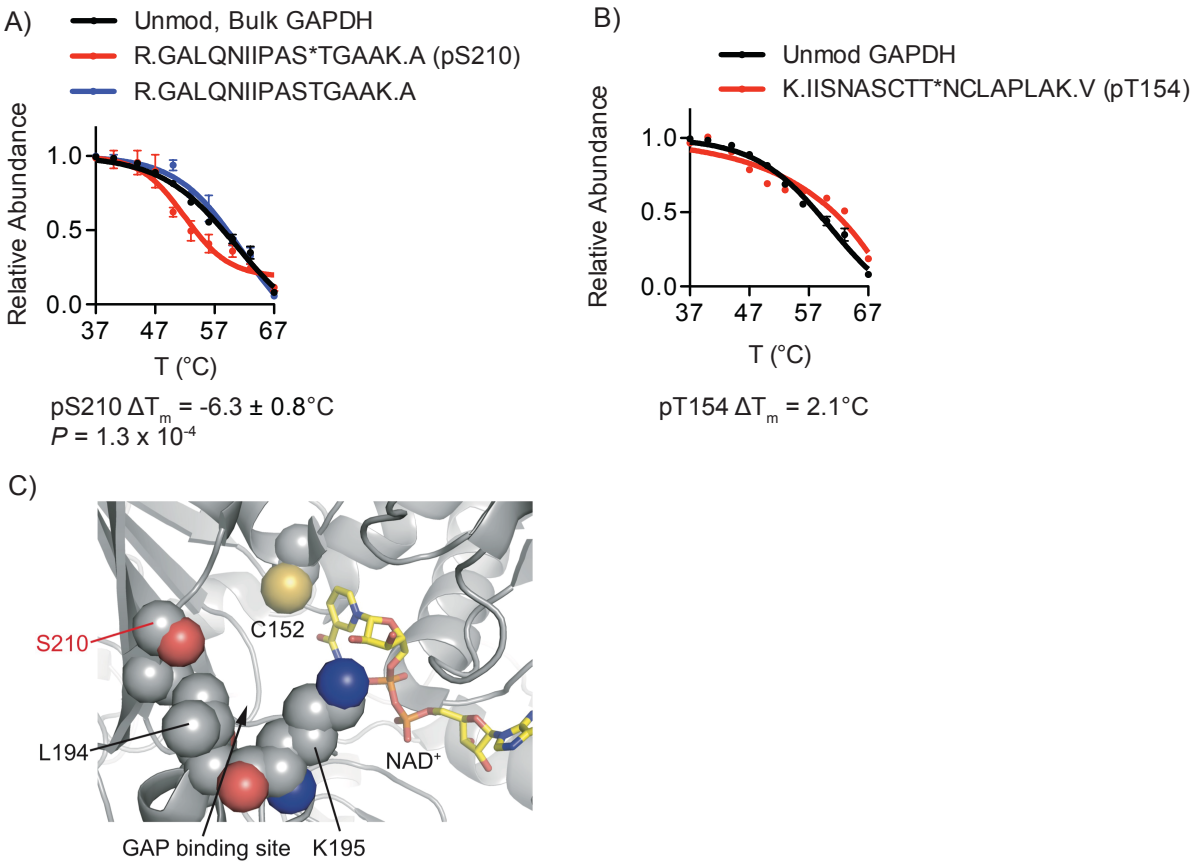

Figure 2.16 HTP detects phosphosite-induced thermal perturbations in GAPDH
A) Representative HTP curves for the bulk, unmodified and corresponding modiforms of GAPDH in HEK293T cells. **B)** Representative HTP curve of GAPDH pT154 phosphomodiform showing no significant ΔT_m relative to the bulk, unmodified GAPDH. **C)** X-ray structure of human GAPDH active site (grey) bound to NAD^+ (tan). The site of phosphorylation, S210, and adjacent residues lining the GAP substrate binding site are shown as spheres. The active site nucleophile, C152, demarcates the GAP binding tunnel. PDB accession: 1ZNQ.

2.2.7.2 Thermal stabilities of GAPDH modiforms under acute glucose-withdrawal

To further understand the functional relationship between pS210 and substrate binding in GAPDH, we performed HTP on cells under the limiting metabolic environment of acute glucose withdrawal. After HEK293T cells were incubated in glucose-free media for 30 minutes, cellular GAP metabolite levels were depleted by 4-fold (Figure 2.17A). Under these conditions, the thermal shift between the bulk, unmodified GAPDH pool and the pS210 phosphomodiform disappeared (Figure 2.17B). However, the thermal stability of the pS210 phosphomodiform was not affected in the presence of lower GAP levels (Figure 2.17C), which contrasted with the significant destabilization of the bulk, unmodified GAPDH protein pool under low-glucose cellular environment (Figure 2.17D). These data suggested that GAP availability and binding contribute to the thermal stability of GAPDH protein that is not phosphorylated at S210, and the negative shift in stability associated with the pS210 phosphomodiform may be capturing the reduced substrate occupancy of these proteins in cells.

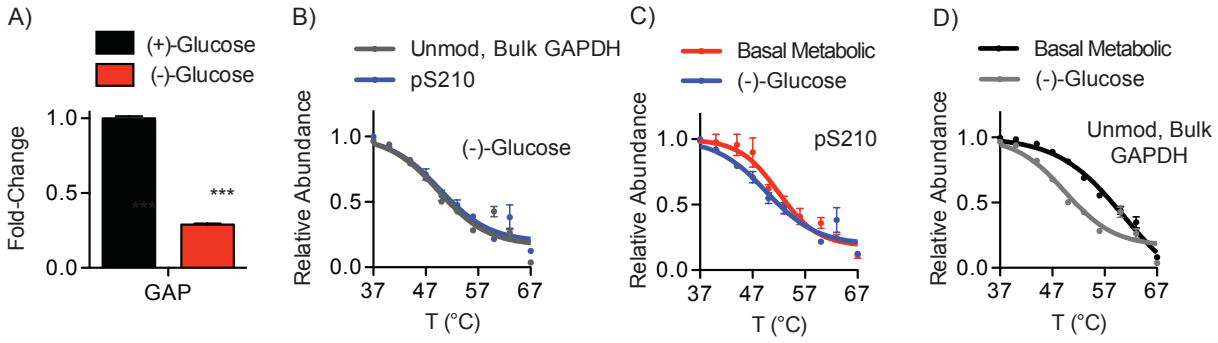

Figure 2.17 Thermal stability of GAPDH modiforms under acute glucose withdrawal
A) Relative intracellular GAP level after 30 minutes of acute glucose withdrawal (n = 2 biological replicates, four technical replicates each). **C-D)** HTP curve comparisons for unmodified and pS210 GAPDH after acute glucose restriction.

2.2.7.3 Phenotypic characterization of GAPDH phosphomimetics

To specifically test the effect of pS210 on GAP binding and catalytic activity of GAPDH, we interrogated the enzyme kinetics of wild-type GAPDH relative to mutants that are incapable of phosphorylation (S210A) or that mimic phosphorylation (S210D). Indeed, the reaction kinetics and K_m value for GAP was indistinguishable between the non-phosphorylatable S210A mutant and wild-type enzyme (Figure 2.18A-B). By contrast, the phosphomimetic S210D mutant remained active but exhibited decreased kinetics and a significantly increased K_m for GAP consistent with altered ability to bind GAP (Figure 2.18A-B). To test whether these effects are unique S210, we tested the enzymatic kinetics of a nearby detected site, T182, that is shown to be close in proximity to the GAP binding pocket (Figure 2.18C). Phosphomimetic mutants of T182 did not affect enzyme kinetics or K_m value for GAP (Figure 2.18 D-E).

Together, these data suggested a molecular model in which phosphorylation at S210 may be directly affecting the thermal stability of GAPDH by regulating GAP substrate binding. The high thermal stability of non-phosphorylated GAPDH may be directly correlated with the GAP-bound state of GAPDH in basal metabolic condition. Meanwhile, low thermal stabilities of pS210 in both basal metabolic and acute glucose withdrawal conditions may be directly correlated to low GAP occupancy of the enzyme (Figure 2.19).

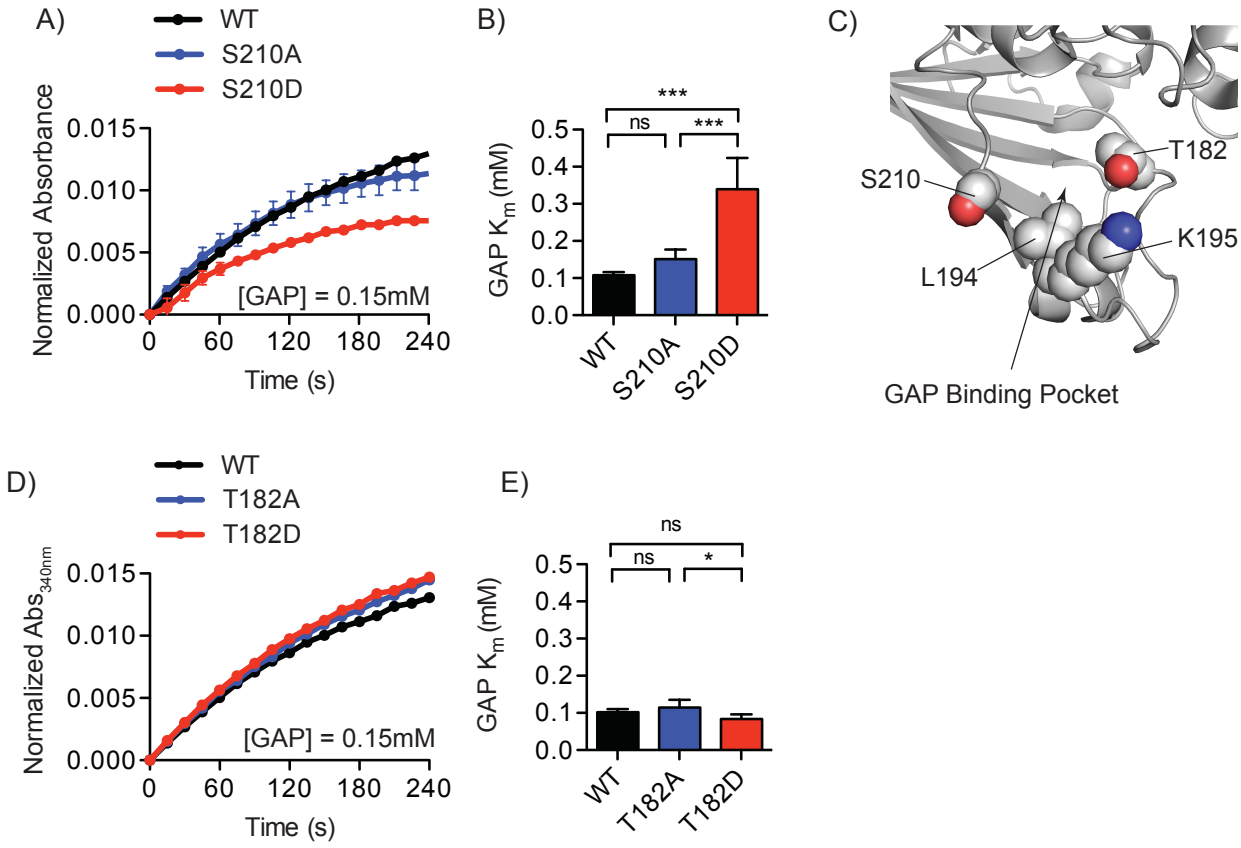

Figure 2.18 Phenotypic characterization of GAPDH phosphomimetics

A-B) Representative in vitro kinetic measurement of wild-type, S210A, and S210D mutant GAPDH **(A)** and Michaelis-Menten kinetic analysis **(B)** of wild-type, S210A and S210D mutant GAPDH confirmed a significant increase in K_M toward GAP for the phosphomimetic-containing enzyme (n = 3 biological replicates, $^{**}P \leq 0.001$, 2-sided t-test). **C)** Crystal structure showing close proximity of T182 to S210 and GAP binding pocket. PDB accession: 1ZNQ. **D)** Representative Michaelis-Menten kinetic measurement of wild-type, T182A, and T182D mutant GAPDH (n = 3 biological replicates). **E)** Graph of K_m toward GAP for wild-type (WT), T182A and T182D GAPDH enzymes (mean and SEM, two-sided t-test).

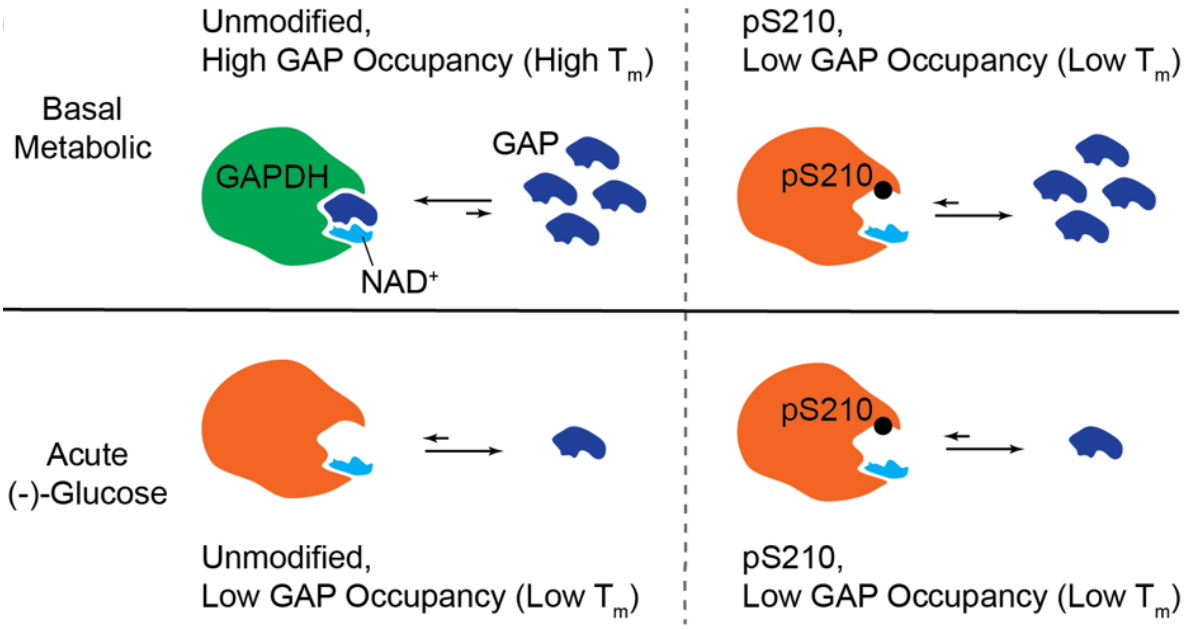

Figure 2.19 Proposed model of phosphoserine 210 function in GAPDH
High thermal stability of non-phosphorylated GAPDH may be directly correlated to the GAP-bound state of GAPDH in basal metabolic condition. Meanwhile, low thermal stabilities of pS210 in both basal metabolic and acute glucose withdrawal conditions may be directly correlated to low GAP occupancy of the enzyme.

2.3 Conclusion and Discussion

Protein post-translational modification represents an important mechanism for cells to regulate biological processes and rapidly respond to stimuli. Despite functional significance of these small covalent molecular moieties resulting in intense interest in PTM studies in the research community for the past decades, conventional biochemical approaches to determine the molecular mechanism of various PTMs have been labourious and low throughput. Recent technological advances such as mass spectrometry-based proteomics platforms have enabled high-throughput detection of thousands of novel modification sites in parallel. However, we still lack general methods with which to predict, or even prioritize, which modification sites are likely to be functional in the proteome.

Here we established a quantitative proteomic method to enable the unbiased global detection of protein stability changes in response to site-specific phosphorylation events on thousands of proteins (Figure 2.20). We hypothesized that phosphorylation-induced thermal stability shifts may serve as a readout on the biophysical and functional effect of a phosphosite on the protein of interest. We began by applying the HTP approach on live, steady state HEK293T cells and generated a proof-of-concept database consisting of 4,293 "bulk, unmodified" protein T_m's, 9,695 phosphomodiform T_m's, as well as 2,883 confident ΔT_m's. In contrast to traditional discovery-mode LC-MS/MS studies, the HTP method provides several novel insights into the functional effects of protein posttranslational modifications. First, it provides site-specific, quantitative relationships between the "unmodified, bulk" population of a protein with its distinct phosphomodiforms in the endogenous context of live cells. We defined the first pool as all unmodified peptides that map to the same protein to represent the "reference" protein pool. Similarly, we defined the pool of phosphomodiform as all modified peptides that contain the same

(combination of) modification site(s) to represent the pool of proteoforms that are unequivocally modified at a specific location. This approach enables the detection of modiforms even if they are only a minor member of the proteome. The resultant dataset showed that the vast majority of phosphorylation sites do not have an impact on the thermal stability of the protein of interest. However, a small but significant portion of sites can positively or negatively shift the thermal stability and these thermal shifts can be directly linked to a modification site thanks to the amino-acid resolution of LC-MS/MS proteomics detection. Global systematic trends analysis using this dataset also showed that the effect of protein phosphorylation on protein stability is entirely based on local chemical environments of the larger protein, instead of a biophysical barcode.

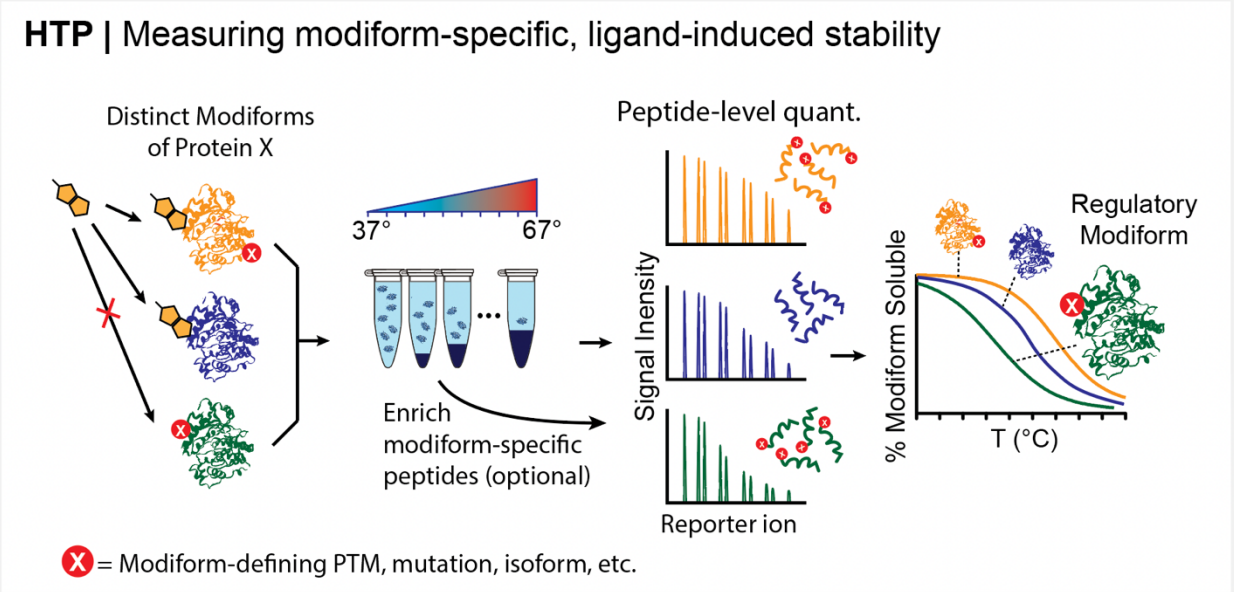

Figure 2.20 Simplified schematic of the HTP workflow

We next demonstrated that HTP can support the detection of both intrinsic alterations in protein structure, as well as the potential for context-specific intermolecular events such as protein-protein, and protein-metabolite interactions[128]. HTP rediscovered a well-known phosphorylation-mediated protein-protein interaction between protein 4-enhancer binding protein-1 (4EBP1) and eukaryotic translation initiation factor 4E (eIF4E). High thermal stability of unmodified 4EBP1

represented the bound form of the protein in the 4EBP1-eIF4E complex, and the low thermal stabilities of various N-terminally phosphorylated 4EBP1 modiforms represented the unbound form of the protein.

We also posit that the ability to perform HTP profiling under steady state and comparative "kinetic" conditions in response to small molecules or signaling events can be used to discover and characterize novel PTM-mediated interactions. We demonstrated this attribute of HTP through the identification of previously uncharacterized phosphorylation sites in extremely well-characterized proteins like GAPDH, as well as less studied cytoskeleton proteins like Vinculin. In Vinculin, we detected a novel, previously uncharacterized phosphosite that induced a functional and conformational shift to the protein. Vinculin phosphosite S721 resulted in significant thermal perturbation to the protein. Functional characterization of Vinculin phosphomimetics revealed a molecular model in which pS721 may be responsible for the conformational switch of Vinculin protein, resulting in an alternative, inactive conformation with diminished recruitment to focal adhesion complexes and increased thermal stability. In GAPDH, we revealed that previously uncharacterized phosphorylation at S210 may be directly impacting protein-metabolite interaction in GAPDH, which is demonstrated by low thermal stabilities of pS210 in both basal metabolic and acute glucose withdrawal conditions.

These results demonstrated that, beyond basic detection of phosphorylation sites, the biophysical information gained through HTP profiling of phosphomodiforms adds another functional layer to that measured by quantifying the abundance of modification sites in the proteome. These traditional studies can report on relative "writing" and "erasing" of modifications across the proteome but provide no information on the dynamic biophysical "reading" of these modifications. HTP profiling alone, or in concert with datasets focused on mapping protein-

protein[147] and metabolite-protein[128, 148] interactions will provide novel insights into the functional and interconnected roles of site-specific modification in the proteome.

It is important to consider the limitations in the interpretation of HTP-derived ΔT_m values in assuming functional impact on a protein of interest. First, this method specifically interrogates the inclusive pool of proteins that unambiguously contain a defining modification site – what we have described here as a "modiform." Thus, this method cannot determine the modification status of other sites that do not fall within this peptide on the same protein molecule[149]. This limitation is intrinsic to massively parallel detection by LC-MS/MS. We do not view this aspect of the approach as a limitation per se, because the primary goal of this approach is to correlate specific modification sites with putative changes to protein structure and function in cells, regardless of the inherent heterogeneity of the protein pool. As such, HTP may under-sample certain combinations of modifications, or miss them entirely, but it can unequivocally connect changes in apparent stability to the presence of site-specific, "hotspot" modifications. Additionally, the response of individual phosphomodiforms to thermal challenge under a defined cellular condition may be used to provide information about the co-occurrence or mutual exclusivity of specific modification sites on the same molecules in the larger protein pool, as we discussed here for TPI1, 4EBP1 and GAPDH. While beyond the scope of the current manuscript, we hope that focused bioinformatics analyses of this and future datasets could shed light on this aspect of the HTP method.

The major advance with this method is the ability to globally interrogate native modified proteins in relevant biological contexts to identify and prioritize sites for further study and provide mechanistic hypotheses for functional validation. In several cases presented here, the site of modification had been previously observed in numerous datasets, but had eluded dedicated study, which underscores the low-throughput nature of current PTM annotation. We expect the

phosphorylation dataset herein, as well as the extension of the HTP method to diverse organisms, cell types, and PTMs will greatly expand our current understanding of role of protein modifications in regulating protein structure, function and signal transduction.

2.4 Materials and Methods

Cell Culture

HeLa and HEK293T cell lines were propagated in RPMI 1640 with 2mM L-glutamine (HyClone) supplemented with 10% fetal bovine serum (Corning) and 1% penicillin-streptomycin (Thermo Fisher Scientific). All cell lines were grown at 37°C in a 5% CO2 humidified incubator.

HTP Proteomic Sample Preparation

HEK293 cells were grown to 90% confluency in 15 cm cell-culture treated plates (Denville). Cells were scraped from plates, briefly washed, and resuspended in PBS with EDTA-free protease inhibitor tablet (Roche) and phosphatase inhibitor cocktail 3 (Sigma-Aldrich) before being divided into ten equal aliquots. Each aliquot of live cells was exposed to a steady temperature between 37°C to 67°C for 3 minutes in parallel, incubated at 25°C for another 3 minutes, and lysed by rapid freeze-thawing. Insoluble proteins and cell debris were removed by centrifugation at 17,000g for 10 minutes. Small amount of supernatant from the 37°C aliquot was set aside to use for protein quantification by Bradford's assay. The remaining fraction of whole cell lysates was denatured in 8 M urea, followed by disulfide reduction with DTT (10 mM, 30 minutes, 65°C), alkylation (iodoacetamide, 15 mM, 30 min, room temperature, protected from light) and quenching (DTT, 5mM, 10 minutes, room temperature). The proteome solution was diluted 4-fold with ammonium bicarbonate solution (50 mM, pH 8.0), CaCl2 added (1 mM) and digested with sequencing grade trypsin (~1:100 enzyme/protein ratio; Thermo Pierce) at 37 ˚C while rotating overnight. Peptide digestion reactions were stopped by acidification to pH 2-3 with 1% formic acid, and peptides were then desalted with Sep-Pak tC18 cartridges (50 mg, Waters), dried under vacuum, resuspended with LC-MS grade water (Sigma-Aldrich) and then lyophilized. A small fraction (5%)

reserved for bulk unmodified proteome thermal profiling was split from each aliquot and lyophilized separately.

Phosphopeptide Enrichment and TMT Labeling

The major fraction of lyophilized peptides (95%) was re-dissolved in 28.5% lactic acid (Sigma-Aldrich), 57% acetonitrile (LC-MS grade, Sigma-Aldrich), 0.28% TFA (Sigma-Aldrich), and applied to TiO2 spin tips (88303, Thermo Pierce) for phosphopeptide enrichment according to manufacturer's protocol. Eluted phosphopeptides were acidified with formic acid (pH 2-3), desalted using ZipTip C18 tips (100 µL, Millipore), and lyophilized. The two sets of peptides (phospho-enriched and unenriched) were separately labeled with 10-plex isobaric tandem mass tags (90406, Thermo Scientific) according to manufacturer's protocol with slight modification. TMT reagents were reconstituted to 8 mg/mL in anhydrous acetonitrile (Sigma-Aldrich) and added to lyophilized peptides dissolved in 100 µL of 200 mM HEPES buffer, pH 8.0 (~8:1 reagent/peptide ratio). Labeling reaction was carried out at room temperature for 1 hour with gentle shaking and quenched with 5 µL of 5% hydroxylamine (Thermo Scientific). Labeled peptides were combined into a single pool per experiment, acidified with formic acid (pH 2-3), desalted using ZipTip C18 tips (100 µL, Millipore), and lyophilized. The final processed peptides were dissolved in LC-MS/MS Buffer A (H2O with 0.1% formic acid, LC-MS grade, Sigma-Aldrich) for LC-MS/MS analysis.

Proteomic LC-MS/MS and Data Analysis

LC-MS/MS experiments were performed with an Easy-nLC 1000 ultra-high pressure LC system (ThermoFisher) using a PepMap RSLC C18 column (75 µm x 50 cm; 3 µm, 100 Å, Thermo

Scientific) heated to 40˚C coupled to a Q Exactive HF orbitrap and Easy-Spray nanosource (ThermoFisher). TMT-labeled digested peptides in MS/MS Buffer A were injected onto the column and separated using the following gradient of buffer B (0.1% Formic acid acetonitrile) at 300 nL/min: 0-10% buffer B in 5 minutes, 10-40% buffer B in 240 minutes, 40-90% buffer B over 6 minutes, and hold at 90% for 20mins. MS/MS spectra were collected from 0 to 250 minutes using a data-dependent, top-10 ion setting with the following settings: full MS scans were acquired at a resolution of 120,000, scan range of 375-1500 m/z, maximum IT of 60 ms, AGC target of 1e6, and data collection in profile mode. MS2 scans was performed by HCD fragmentation with a resolution of 60,000, AGC target of 1e5, maximum IT of 60 ms, NCE of 30, and data type in centroid mode. Isolation window for precursor ions was set to 1.0 m/z with an underfill ratio of 0.5%. Peptides with charge state 1 and undefined were excluded and dynamic exclusion was set to 20 seconds. Furthermore, S-lens RF level was set to 60 with a spray voltage value of 2.60kV and ionization chamber temperature of 300˚C.

MS2 files were generated and searched using the ProLuCID algorithm in the Integrated Proteomics Pipeline (IP2) software platform. Human proteome data were searched using a concatenated target/decoy UniProt database (UniProt_Human_reviewed_04-10-2017.fasta). Basic searches were performed with the following search parameters: HCD fragmentation method; monoisotopic precursor ions; high resolution mode (3 isotopic peaks); precursor mass range 600-6,000 and initial fragment tolerance at 600 p.p.m.; enzyme cleavage specificity at C-terminal lysine and arginine residues with 3 missed cleavage sites permitted; static modification of +57.02146 on cysteine (carboxyamidomethylation), +229.1629 on N-terminal and lysine for TMT-10-plex tag; 4 total differential modification sites per peptide, including oxidized methionine (+15.9949), and

phosphorylation (+79.9663) on serine, threonine, and tyrosine (only for phospho-enriched samples); primary scoring type by XCorr and secondary by Zscore; minimum peptide length of six residues with a candidate peptide threshold of 500. A minimum of one peptide per protein and half-tryptic peptide specificity were required. Starting statistics were performed with a Δmass cutoff = 10 p.p.m. with modstat, and trypstat settings. False-discovery rates of peptide (sfp) were set to 1%. TMT quantification was peformed using the isobaric labeling 10-plex labeling algorithm, with a mass tolerance of 5.0 p.p.m. or less in cases where co-eluting peptide interfere. Reporter ions 126.127726, 127.124761, 127.131081, 128.128116, 128.134436, 129.131417, 129.13779, 130.134825, 130.141145, and 131.13838 were used for relative quantification. In general all quantified peptides have mass error within 3 p.p.m..

Modification Site Localization Analysis

Phosphorylation site localization analysis was performed using the LuciPHOR algorithm available on the Integrated Proteomics Pipeline (IP2). A global FDR score was calculated for every detected phosphopeptide from each MS run. Phosphosites, and therefore phosphomodiforms, were included in the dataset only if they satisfied a global FDR localization score of < 0.01.

Melting Curve Analyses

All detected phosphopeptides that map to the same phosphorylation sites were grouped together and given a new identifier (Gene_pSite). For example, 4 different tryptic peptides mapping to the same GAPDH phosphorylation site pS210 were often detected: R.DGRGALQNIIPAS(79.9963)TGAAKAVGK.V, R.DGRGALQNIIPAS(79.9963)TGAAK.A, R.GALQNIIPAS(79.9963)TGAAKAVGK.V, and R.GALQNIIPAS(79.9963)TGAAK.A. These

peptides were systematically re-labeled with the new custom identifier "GAPDH_pS210". TMT reporter ion intensities of phosphopeptides with the same identifier were combined for each temperature fraction from the same MS run. Fold change values were calculated using the lowest temperature fraction as the reference. Similarly, for the bulk unmodified proteome, TMT reporter ion intensities of all unmodified peptides mapped to the same protein were combined and fold change values relative to the lowest temperature condition were calculated. To generate unmodified protein and phosphosite-specific melt curves, relative fold changes as a function of temperature was fitted to the equation derived from the chemical denaturation theory using R. T_m values were calculated at which the sigmoidal curve crosses the 0.5-fold change level. Only T_m values calculated from melting curves with curve R2 > 0.8 are used in subsequent analyses. Shift in T_m values (e.g., ΔT_m) induced by phosphorylation were determined by subtracting the T_m of the unmodified protein from the T_m of the phosphomodiform:

ΔT_m = T_Phosphomodiform - Tm_Unmodified, bulk protein

ΔT_ms could only be calculated for phosphosites belonging to proteins that were also detected in the unmodified proteome. ΔT_ms were not determined for phosphosites that were detected in fewer than 2 MS technical replicates. T_m values of all unmodified proteins and phosphorylation sites, and their ΔT_m values and P-values are summarized in Data Tables.

Error Analyses

In our workflow, a composite protein-level T_m is calculated by averaging T_m values derived from all detected tryptic peptides of the same protein. To estimate the overall "base error" of our peptide-level T_m measurements, i.e., on average how far a peptide T_m deviates from its composite protein-level T_m, we took 4 representative datasets and calculated the median peptide ΔT_m standard

deviation. We individually curve fitted every detected tryptic peptide as well as their composite protein to calculate T_m values. Each dataset is comprised of between 10,000 to 16,000 peptide and >2,900 protein melt curves. We then subtracted the composite protein-level T_m from each peptide-level T_m to obtain peptide ΔT_ms and calculated the standard deviation of each set of peptide ΔT_m's. We plotted the distributions of the peptide ΔT_m s.d. of the 4 representative datasets which resulted in median peptide ΔT_m s.d. between 1.3 to 1.5°C. The average value of 1.4°C was used as vertical cutoff (aka smallest high confidence ΔT_m value) in the ΔT_m volcano plot.

In situ measurement of polar metabolome by LC-MS/MS

For acute glucose withdrawal studies, confluent 15 cm plates of HEK293T cells were switched from RPMI media containing 10 mM glucose into glucose-free RPMI media and grown for 30 minutes, whereas fresh, glucose-containing media was replaced for control cells. In both conditions, cells were scraped into ice-cold PBS and isolated by centrifugation at 1,400*g* at 4°C. Cell pellets were resuspended in 300 µl of an 80:20 mixture of MeOH/H_2O. Internal deuterated standards, 10 nmol d3-serine, were added to the extraction solution for quantification and sample normalization. LC-MS/MS analysis was then performed as described[29].

Western blot

To determine the effect of the 3-min heat pulses on global phosphorylation level, we exposed cells to the thermal profiling heat pulses as detailed above, but directly homogenized whole preoteome by addition of 4X Laemmli buffer containing 50 mM DTT. Samples were prepared for SDS-PAGE by heating to 95˚C for 5 minutes, cooled to room temperature, resolved on 10% SDS-page gel and transferred onto nitrocellulose membranes by standard western blotting methods. Membranes were

blocked in 2% BSA in TBS containing 0.1% tween-20 (TBST) and probed with primary and secondary antibodies. Antibodies used in this study include: anti-pS/T/Y (1:1000, ab15556, AbCam). Secondary anti-mouse (Licor) was used at 1:10,000 dilution in 2% BSA-containing TBST and incubated for 1hourprior to washing and imaging on a Licor infrared scanner. Densitometry measurements were performed with ImageJ software.

Vinculin Expression and Focal Adhesion Confocal Microscopy Assays

Full-length Myc-DDK-tagged-human vinculin (NM_003373, Origene) was used to generate mutants with PCR primers in Table 1 according to the QuikChange II Site-Directed Mutagenesis Kit manufacturer protocol (200523, Agilent). HeLa cells were grown to 70% confluency in 6-well plates before transfection of 500 ng plasmid with Lipofectamine 2000 (Invitrogen). After 4 hours, cells were reseeded onto round glass cover slides. After 24hourcells were washed once with PBS and fixed at 4°C for 10 min in 4% paraformaldehyde-containing PBS. Fixed cells were then washed three times with PBS, and blocked with 2% BSA with 0.05% Triton in PBS for 30 minutes at room temperature. Slides were stained with primary antibodies rabbit α-FLAG (1:500, Cell Signaling) and mouse α-Paxillin (1:100, Millipore) overnight at 4°C. Cells were washed three times with 0.05% Tween in TBST for 5 minutes and then stained with secondary antibodies Alexafluor488-conjugated goat-α-rabbit (1:200; Sigma-Aldrich), Alexa568-conjugated goat α-mouse (1:200; Sigma-Aldrich), and Alexa647-conjugated phalloidin (1:25, Invitrogen) for 30 minutes in 37°C. Finally, slides were washed three times with TBST and quickly dipped in DI water before mounting in Slow Fade Gold Antifade (ThermoFisher).

Images were taken using Metamorph (Molecular Devices) acquisition software on Nikon Eclipse Ti-PFS inverted microscope with a Yokogawa CSU-X1 confocal scanhead with Spectral Applied Research Laser Merge Module (491 nm, 561 nm, and 643 nm Lasers). Images were obtained using Andor Zyla sCMOS camera with a 60x Plan Apo TIRF 1.49NA objective (Nikon). Image processing and quantification was done in ImageJ. A single focal plane was acquired in which the lamella was in focus, as indicated by basal actin stress fibers stained with phalloidin and Paxillin. Paxillin and Vinculin channels were background subtracted. Paxillin channel was thresholded and made into a binary mask to calculate the area of focal adhesions (FAs) as indicated by Paxillin. The binary mask was then overlaid onto the vinculin channel for segmentation and area calculation (threshold at 0.5 μm2) of Vinculin incorporation into the FAs. For linescans analysis, fluorescence intensity line profiles over vinculin and paxillin channels were merged (raw images that were background subtracted) and done in ImageJ.

GAPDH Expression, Purification and Kinetic Assays

Full-length, human GAPDH (NM_002046, Origene) was used to generate GAPDH mutants with PCR primers in Table 1 according to the QuikChange II Site-Directed Mutagenesis Kit manufacturer protocol (200523, Agilent). Mutant GAPDH plasmids were verified by sequencing [CMV (forward), and XL39 (reverse)]. Wild type and GAPDH mutants were transiently expressed from a pCMV6 entry vector with a C-terminal Myc-DDK tag in confluent 10 cm plates of HEK293T cells by transfection of 2 μg plasmid with Lipofectamine 2000 (Invitrogen) according to manufacturer's protocol. After 24 hours, cells were harvested by scraping, pelleted by centrifugation, washed twice with PBS and lysed in PBS containing EDTA-free complete protease inhibitor (Roche), pH 8.0. After ultracentrifugation, lysates were added to Anti-Flag M2 affinity

resin (Sigma-Aldrich) and rotated overnight at 4˚C. Resins were washed 5x with PBS before eluting bound protein with 3X-FLAG peptide (100 μg/mL, 150 μL, twice, Sigma-Aldrich) in PBS for 1 hours. Relative protein levels were assessed by SDS-PAGE and enzymes were used without further purification.

Michaelis-Menten kinetic assays to assess the catalytic activities of GAPDH proteins were performed on a Tecan Infinite M200 plate reader. GAPDH proteins were supplemented with constant NAD+ (0.278 mM) and serial three-fold dilutions of D/L-glyceraldehyde-3-phosphate (1.36 to 0.0056 mM) and NADH production (340nm absorbance) was monitored. Enzyme velocity at different GAP concentrations was measured in the first 4 minutes and Michaelis constants (KM) were generated by applying a Michaelis-Menten curve fit to the plot of enzyme velocity as a function of [GAP] in Prism 5 software.

Kinase Substrate Motif Analysis

Motif analysis was performed by comparing peptide sequence containing the target phosphorylation site to commonly described kinase/phosphatase substrate consensus motifs listed in the Human Protein Reference Database (http://www.hprd.org/serine_motifs). Nine common kinase motifs were chosen for this analysis.

Secondary Structure, Solvent Accessibility, and Protein Disorder Prediction

Secondary structure, solvent accessibility, and protein disorder were predicted using RaptorX (http://raptorx.uchicago.edu)43 Protein FASTA sequences were uploaded to RaptorX Structure Prediction portal as described. To calculate the secondary structure percentage, the number of

residues predicted to be alpha-helix, beta-sheet, or loop was divided by the total number of residues of all detected proteins. The secondary structure per centage of the phosphosites was calculated by dividing the number of sites predicted to be alpha-helix, beta-sheet, or loop over the total number of detected phosphosites. Secondary structure elements, solvent accessibility and disorderness of detected phosphosites were plotted against their ΔT_m values.

Phosphosite Annotation Analysis

The number of annotations with site-specific biochemical studies was extracted from the PhosphoSite database (https://www.phosphosite.org) and plotted against ΔT_m values.

2.5 Tables

Protein	Mutation	Primer (Forward)	Primer (Reverse)
GAPDH	S210A	5'- gcagcgccagtagcggcagggatgatg- 3’	5'-catcatccctgccgctactggcgctgc-3’
	S210D	5'- cttggcagcgccagtatcggcagggatga tgttc-3’	5'- gaacatcatccctgccgatactggcgctgccaag- 3’
	T182A	5'- cttctgggtggcagcgatggcatggactg -3'	5'-cagtccatgccatcgctgccacccagaag-3'
	T182D	5'- ccacagtccatgccatcgatgccacccag aagactg-3'	5'- cagtcttctgggtggcatcgatggcatggactgtg g-3'
Vinculin	S721A	5'- ctgaagcatccaacagagctttggtatcaa tggcttcg-3'	5'- cgaagccattgataccaaagctctgttggatgcttc ag-3’
	S721D	5'- gaagcatccaacagatctttggtatcaatg gcttcgtccacc-3’	5'- ggtggacgaagccattgataccaaagatctgttgg atgcttc-3’

Table 2.1 Primers used for site-directed mutagenesis of GAPDH and vinculin

Primer	Sequence
CMV	5'-cgcaaatgggcggtaggcgtg-3'
XL39	5’-attaggacaaggctggtggg-3’
Vinculin middle	5'-tggtctcggctgctcgtatcttac-3'

Table 2.2 Sequencing primers used to verify GAPDH and vinculin mutant plasmids

CHAPTER 3

Diverse Hotspot Thermal Profiling methods detect phosphorylation-dependent changes in protein stability[*]

We are preparing for the publication of the work described in this chapter.

3.1 Introduction

In the previous Chapter, we established a quantitative proteomic workflow titled Hotspot Thermal Profiling (HTP) that enables the unbiased global detection of protein stability changes in response to site-specific modifications in live cells[149]. The resulting proof-of-concept dataset demonstrated 1) that peptide-level thermal stability information can be obtained through HTP which allows physical interrogation of protein modiforms – in this case, phosphomodiforms and their unmodified, bulk proteoforms – in live cells a high-throughput, global, and unbiased manner using LC-MS/MS-based quantitative proteomics; 2) HTP can measure and detect site-specific PTMs that significantly affect proteoform T_m or solubility and these "hotspot" sites likely represent functionally relevant PTMs that perturb protein structure and function; 3) uncharacterized modifications that impact inter- and intramolecular interactions can be discovered and prioritized for subsequent functional studies. Specifically, our first HTP dataset detected phosphorylation-

[*] The author's contribution to the studies presented in this chapter: The author conducted experiments described in this chapter that led to the development of the HTP variants method. Benjamin D. Stein and David Wu contributed to the development the LFE-HTP method, performed cell-based and mass spectrometry experiments. The author designed and performed cell-based experiments, mass spectrometry experiments and analyzed data.

induced thermal stability shifts of thousands protein modiforms in parallel in live cells, and supported the discovery of novel, previously uncharacterized intra- and intermolecular interactions mediated by phosphorylations in glycolytic protein GAPDH and cytoskeleton protein Vinculin.

Our primary goal for developing the original HTP workflow was to demonstrate proof-of-concept and applicability of the method. There are two important and practical benefits to the original workflow as developed (Figure 2.2): significant reagent cost- and instrument usage time-savings. First, the TMT isobaric labeling reagents are the one of the most expensive reagents used in the workflow, costing up to $5,000 per reagent kit. Therefore, there is a significant cost-saving to isobarically labeling the enriched phosphoproteome at the very end. Rather than labeling the bulk proteome before enrichment just for 90% of it to be discarded after the phospho-enrichment step, tagging the final phospho-enriched proteome requires almost 10 times less TMT reagents upfront. Secondly, we performed 1D proteomics analysis, meaning that one single bulk injection was run for each biological experiment. This translates to about 6 hours of total instrument running time, allowing more biological experiments to be performed and boosting statistical confidence of the final dataset. These two practical advantages gave rise to the first iteration of the HTP workflow and allowed the proof-of-concept dataset to be produced in a reasonable cost and amount of time.

We have since examined some areas that can be improved upon to further strengthen the depth of proteome coverage and precision of the method. We mentioned previously that TMT labeling enables parallel multiplexing of up to 16 independent quantitative measurements[81, 82, 150]. In the original HTP workflow, the labeling and multiplexing step is delayed until the very end which may contribute to additional sample handling variability. The second area of improvement is proteome coverage. Since our original publication, several off-line fractionation approaches and products became more widely used. Chief among them are high pH reverse phase fractionation

columns that demonstrated deep proteome coverage across a variety of samples and proteomes[151-153]. The last area of improvement has to do with a phenomenon known as ratio suppression or interference contamination of TMT reporter tags (Figure 3.1)[154-158]. As described in the previous chapter, TMT isobaric tags have the same nominal mass but different versions of mass reporters[83]. After identical peptides derived from different temperature fractions are labeled with different versions of the tags, they appear indistinguishable in the full mass spectra scan (MS1). At the MS2 step where peptide ions are isolated and fragmented, each differentially labeled peptide produces a unique reporter ion that correspond to the version of TMT tag that the peptide was labeled with. In theory, the signal intensity of the reporter ion should reflect the relative amount of its precursor ion. But in reality, the signal intensity can often be distorted due to coisolation and cofragmentation of nearby interfering ions that are near-isobaric. It has been reported in a series of publications that the accuracy and precision of quantitative ratios from TMT experiments can be affected by the interference phenomenon resulting in a ratio suppressive effect[154-158]. The contaminating isobaric ions are isolated and fragment together with the target reporter ions, and their signal intensities are erroneously attributed to those of the target ions, resulting in the ratio differences between species that are less prominent than what they should be in actuality (Figure 3.1).

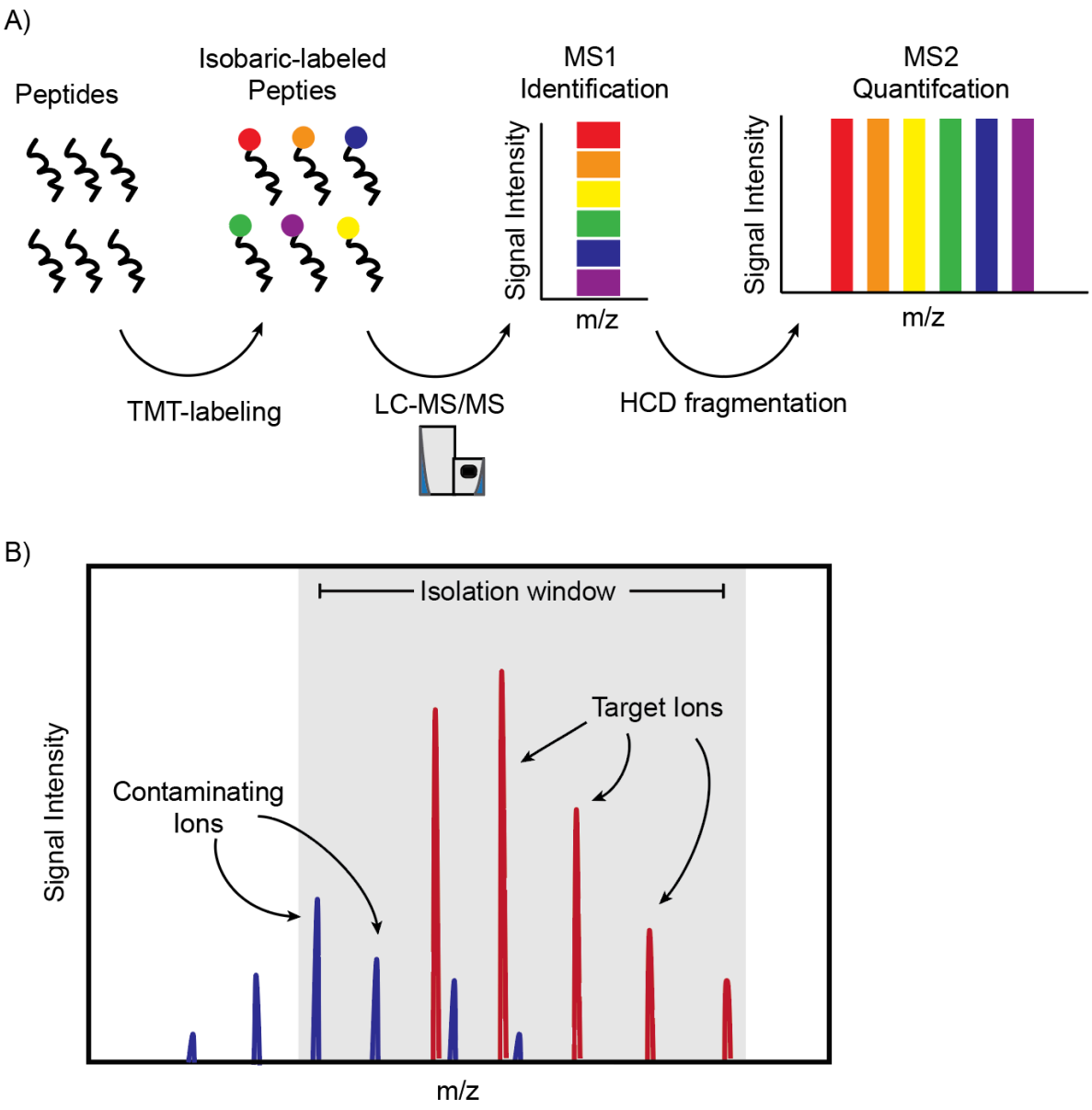

Figure 3.1 Schematic showing interference contamination from TMT reporter tags using MS2 quantification

A) After tryptic peptides are TMT-labeled, they undergo full-spectrum MS1 mode before target ions are isolated and further fragmented in MS2 mode for quantification. **B)** Contaminating ions are isolated together with the target ion inside the same isolation window, interfering with the accurate quantification of the target ion.

Several strategies have been proposed to address this interference phenomenon. The first is to narrow ion isolation window using mass spectrometers that has a higher isolation efficiency at narrower mass windows[159]. It has been reported that narrowing the precursor-ion isolation window from 2 to 0.5 on a LTQ Orbitrap Velos mass spectrometer can reduce ratio suppression effect by up to 56%[158]. However, using an isolation window that is too narrow to boost quantification accuracy increases the risk of losing spectral identifications, which accounted for up to a quarter of peptide identifications loss in some cases. The second strategy to minimize isobaric ion contamination is to reduce the complexity of the sample by extensive offline and multidimensional LC fractionation, such as acidic pH reverse phase, strong cation exchange (SCX), and high pH reverse-phase [159]. It has been reported that SCX fractionation can reduce the ratio suppression effect by up to 14%[158]. High pH reverse phase (also referred to as basic-reverse phase or bRP later) fractionation columns have not been tested on reducing TMT ratio suppressive effect, but they have been shown to generate deep proteome coverage across a variety of samples and proteomes[151-153]. The last strategy is to perform an additional isolation and fragmentation event known as a MS3 scan[158, 160, 161]. In a MS3-experiment, after a peptide ion is isolated and fragmented in the MS2 step for identification, a selected fragment ion is isolated and fragmented again to produce an MS3 spectrum used for quantification. Only the most high-intensity target fragment ion belonging to the correct reporter tag is selected for MS3, thus near completely eliminating the interference effect. MS3 methods such as MultiNotch MS3 method have been shown to enable more accurate quantifications, at the loss of spectral identifications of a small reduction (13%) of peptide identifications[158].

In this Chapter, we proposed several variants of the original HTP workflow with the intentions of improving precision in sample processing, the depth of (phospho)proteome coverage, and LC-MS/MS quantitative measurement precision.

3.2 Results

3.2.1 "Label-Enrich" Hotspot Thermal Profiling (LE-HTP) workflow

The first HTP variant, termed "Label-Enrich" Hotspot Thermal Profiling (LE-HTP), is similar to the original HTP (also referred to as "Enrich-Label" or EL-HTP later) with two major differences: 1) the TMT labeling step happens before phosphoenrichment, and 2) one single TiO_2 enrichment is performed on the multiplex sample instead of ten separate TiO_2 enrichments on the non-multiplexed sample (Figure 3.2). The rationale behind switching the order of these two steps is to take full advantage of the parallel multiplexing capability that isobaric labeling offers. We posit that combining the temperature fractions as early as possible should reduce sample processing variability.

The start of the LE-HTP workflow is the same as that of the original HTP: equivalent aliquots of live cells are exposed to a 3-minute heat pulse of temperatures ranging from 37 to 67°C, followed by cooling to room temperature, rapid cell lysis in the presence of protease and phosphatase inhibitors, and centrifugation to remove unfolded proteins. The soluble proteome remaining in the supernatant of each temperature-matched aliquot is then proteolyzed into tryptic peptides. Starting from here, LE-HTP workflow differs from the original HTP in that all peptides from each temperature aliquot is labeled with specific tandem mass tag (TMT) isotope channels and combined into one single pool. After multiplexing, 5% of the combined pool of TMT-labeled

sample is set aside for protein-level thermal profiling of the "unmodified, bulk" protein pool. The remaining 95% of the TMT-labeled multiplexed sample is enriched in one single TiO_2 enrichment step to generate the phosphoproteome pool. The two respective pools undergo 1D LC-MS/MS proteomic analysis and quantitation of relative peptide levels across the temperature range. Finally, these peptide-level measurements are analyzed en masse with peptide identification, relative quantification and automated curve fitting algorithms to determine peptide-specific melting points (T_m) for tens of thousands of specific peptides in the same sample.

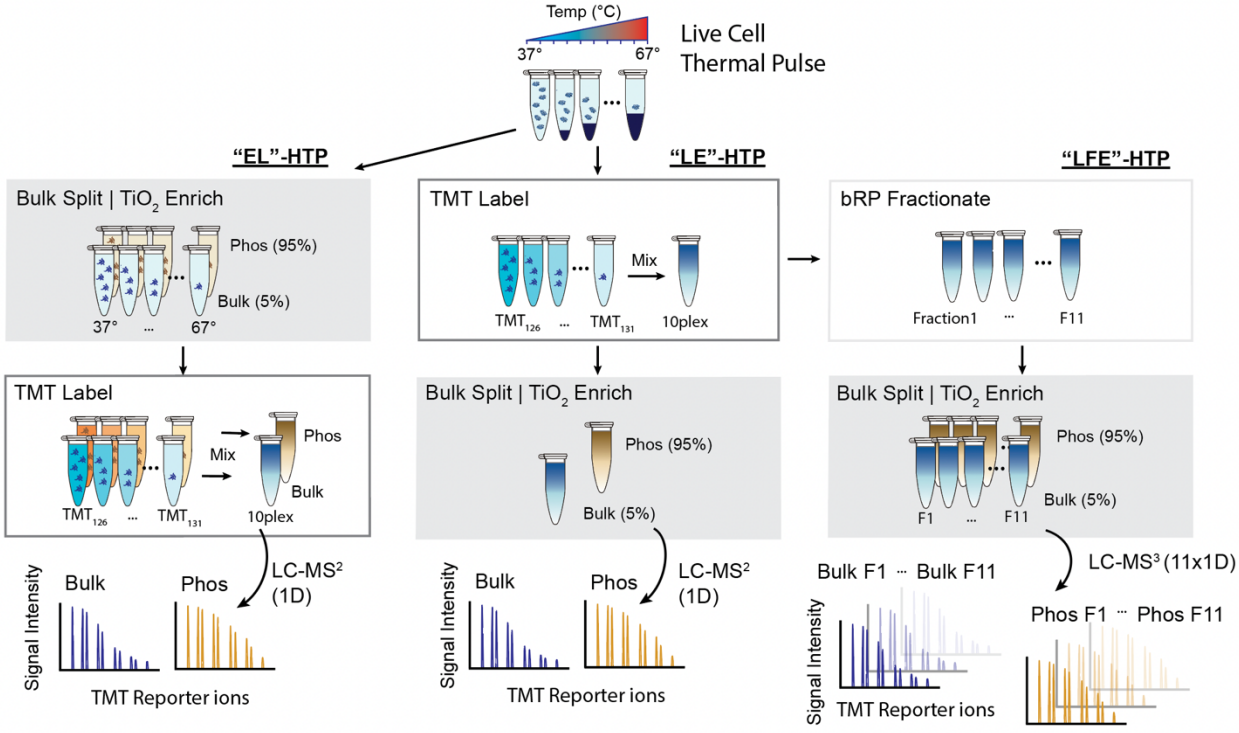

Figure 3.2 Comparison of new variant HTP workflows
Proteomic workflow schematics of the three HTP variants: EL (enrich-label), LE (label-enrich), and LFE (label-fractionate-enrich).

3.2.2 "Label-Fractionate-Enrich" Hotspot Thermal Profiling (LFE-HTP) workflow

The second HTP variant, termed "Label-Fractionate-Enrich" Hotspot Thermal Profiling (LFE-HTP), differs from the original HTP in several ways (Figure 3.2). First, the TMT labeling step happens before phosphoenrichment, which is similar to the LE-HTP workflow described previously. Second, the multiplexed TMT sample is separated by offline high pH reverse phase (bRP) columns to generate 11 fractions. This fractionation step not only helps reduce sample complexity by minimizing the effects of quantitative TMT ratio suppression, but also significantly expands the depth of (phospho)proteome coverage. Lastly, the LFE-HTP method performs 11 LC-MS/MS runs aided by the use of Multi-Notch MS3 acquisition mode to reduce TMT ratio compression and improve precision in TMT quantification of phosphopeptides.

The start of the LFE-HTP workflow is the same as the original HTP and LE-HTP described in the previous section. After the live cell heat-pulses, rapid cell lysis, isolation of soluble proteome, and digestion into tryptic peptides, all peptides from the same temperature aliquot are labeled with specific tandem mass tag (TMT) isotope channels and combined into one single pool. After multiplexing, the combined pool of TMT-labeled sample undergoes offline bRP fractionation into 11 different fractions. From each bRP fraction, 5% is reserved for protein-level thermal profiling of the "unmodified, bulk" protein pool, and the remaining 95% from each bRP fraction undergoes separate phosphopeptide enrichment by TiO_2 resin. Each series fractionated of two proteome pools undergo 11 LC-MS/MS proteomic analyses and multi-notch MS3 acquisition quantitation of relative peptide levels across the temperature range. Finally, these peptide-level measurements are analyzed en masse with peptide identification, relative quantification and automated curve fitting algorithms to determine peptide-specific melting points (T_m) for tens of thousands of specific peptides in the same sample.

3.2.3 Global thermal stability profiles comparison between methods

Using LE-HTP workflow, we detected 4,076 proteins in HEK293T cells across all LC-MS/MS runs (comprised of n = 4 biological replicates, each biological replicate of the unmodified, bulk proteome or the phosphoproteome underwent n = 1 technical LC-MS/MS injections) (Figure 3.3C). Statistical filtering at the peptide identification level identified 3,476 unique unmodified, bulk protein, and 4,186 unique phosphomodiforms from 1,544 unique proteins after grouping distinct phosphopeptides that map to a common site according to description in the previous chapter. Of those 4,186 unique phosphomodiforms, 3,108 of them belong to proteins also detected in unmodified bulk proteome. The application of stringent automated T_m curve fitting cutoffs ($R^2 > 0.8$) (and site-localization filter specifically for the phosphoproteome) resulted in a database of site-specific T_m values for 3,131 unmodified, bulk proteins, and 3,719 distinct phosphomodiforms in the steady state HEK293T proteome (Figure 3.3D). The LE-HTP dataset was overall smaller than that of the original HTP because of fewer biological replicate experiments (Figure 3.3A-B).

Using LFE-HTP workflow, we detected significantly more proteins and phosphosite-specific T_m compared to LE-HTP and the original EL-HTP dataset. This was expected of the fractionation method used to reduce the complexity of the sample which allows for a deeper phospho- and global proteome coverage. Specifically, we detected 6,613 unique proteins in HEK293T cells across all LC-MS/MS runs. The dataset is comprised of n = 4 biological replicates, and each biological replicate underwent 11 LC-MS/MS injections (one per bRP fraction) for the unmodified, bulk proteome series, and another 11 injections for phosphoproteome series. Statistical filtering at the peptide identification level identified 5,175 unique unmodified, bulk protein, and 15,621 unique phosphomodiforms (Figure 3.3E). Of those 15,621 unique phosphomodiforms, 12,637 of them belong to proteins also detected in unmodified bulk proteome.

The application of stringent automated T_m curve fitting cutoffs ($R^2 > 0.8$) (and site-localization filter specifically for the phosproteome) resulted in a database of site-specific T_m values for 4,626 unmodified, bulk proteins, and 14,143 distinct phosphomodiforms (Figure 3.3F).

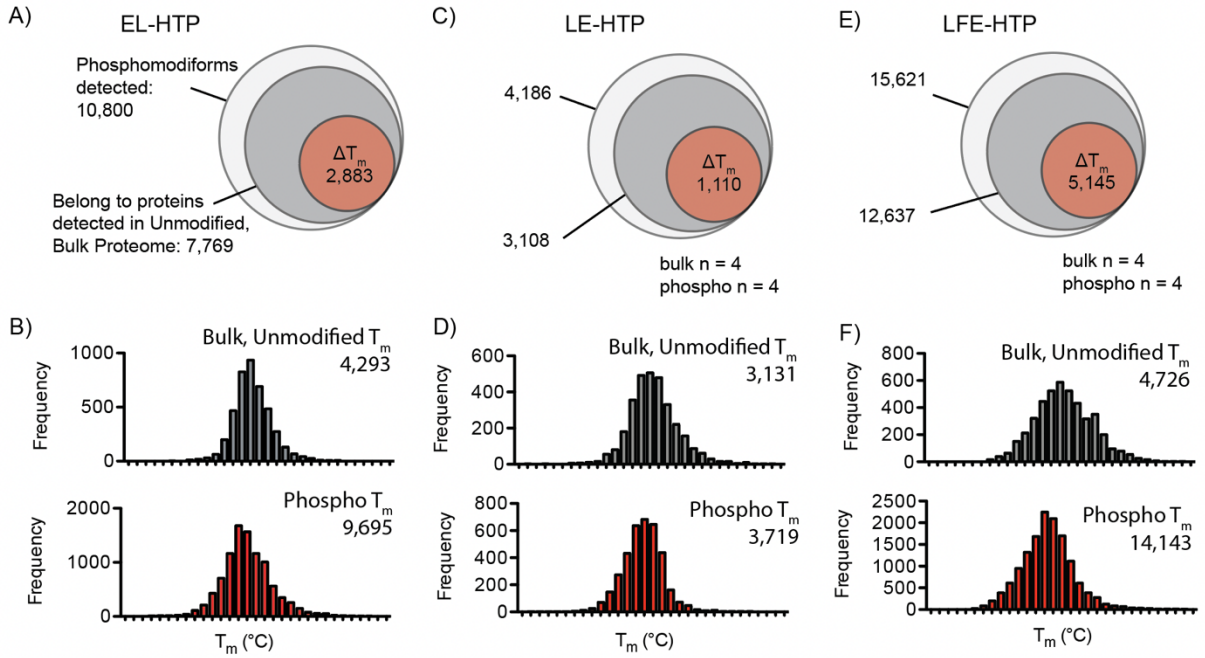

Figure 3.3 Global T_m values measured by three HTP variant workflows
The total number of unique phosphomodiforms and high-quality ΔT_m values detected in EL-HTP **(A-B)**, LE-HTP **(C-D)**, and LFE-HTP **(E-F).** Global T_m distribution of unmodified, bulk protein-level, and phosphomodiform T_m values detected in EL-HTP **(A-B)**, LE-HTP **(C-D)**, and LFE-HTP **(E-F).**

Global comparisons of the LE-HTP of unmodified, bulk protein pool with the phosphoproteome showed that median T_m values of 52.4 and 50.8°C, respectively, which were nearly identical to those of the original HTP dataset. In LFE-HTP, they were 52.5°C for bulk, unmodified proteome and 50.4°C for the phospho pool. The overall ΔT_m distributions were all Gaussian in nature. Filtering of only high-confidence T_m values detected twice or more in both

phospho- and unmodified, bulk proteomes, yielded 1,110 high-quality ΔT_m values for LE-HTP (Figure 3.3C), and 5,145 ΔT_m values for LFE-HTP (Figure 3.3E). Specifically, within the segment of ΔT_m values, ~25% and ~26% of the phosphosites had statistically significant ΔT_m values relative to unmodified protein in LE-HTP and LFE-HTP respectively, which was defined by *p*-value < 0.05 calculated using 2-sided student's t-test. These percentages were nearly identical to that observed in our original HTP dataset.

3.2.4 Data reproducibility comparison between methods

Compared to our published method and dataset, LFE-HTP variant method resulted in significantly higher precision as various areas of improvements have been addressed. The LFE-HTP dataset has a bulk T_m biological replicate correlation R^2 of 0.78 (Figure 3.4A), and a phospho-T_m biological replicate correlation R^2 of 0.72 (Figure 3.4B). This significant reproducibility improvement is the result of earlier multiplexing of the temperature fractions before phosphoenrichment to help reduce sample processing variability (Figure 3.2). Furthermore, the LFE-HTP dataset has an even better precision due to increased fractionation resulting in 22 (vs 2 in LE- and EL-HTP) MS runs per biological replicate which significantly boosted the depth of the proteome as well as Multi-Notch MS3 quantification which eliminated TMT ratio suppression. Compared to our published EL-HTP dataset, LE-HTP variant method resulted in moderate improvement in phospho-T_m reproducibility correlation (R^2 = 0.5, Figure 3.4D). The reproducibility improvement is the result of the multiplexing step taking place before further sample processing as well as only one TiO_2 enrichment step compared to ten separate enrichments in EL-HTP. LE-HTP variant resulted in no improvement in bulk-T_m reproducibility correlation,

which is not unexpected as the bulk-T_m sample preparation protocol is essentially the same between EL- and LE-HTP.

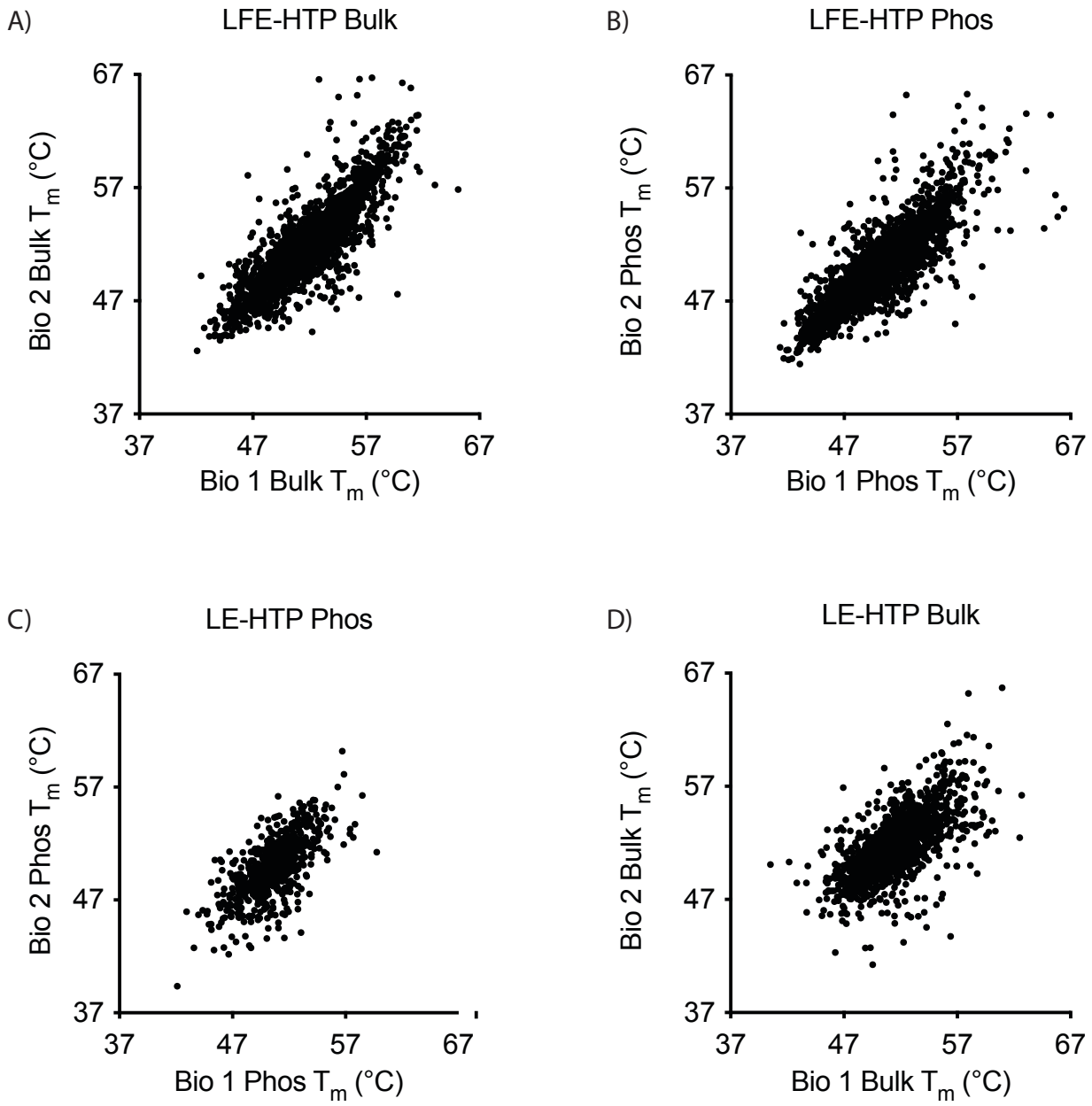

Figure 3.4 Representative T_m correlation plots in LFE- and LE-HTP datasets
A-B) LFE-HTP representative bulk, unmodified proteome **(A)** and phospho **(B)**T_m correlation plots demonstrate significant improvements in T_m measurement reproducibility. **C-D)** LE-HTP representative bulk, unmodified proteome **(B)** showed similar T_m reproducibility to the original HTP dataset, meanwhile the representative phosphoproteome T_m correlation plot shows improvements in reproducibility.

3.2.5 Global trends

Global relationships between phosphosite environment and stability were similar to that reported in our original HTP dataset. We first compared the ΔT_m values for phosphorylation on different residues (serine, threonine and tyrosine) across all detected phosphomodiforms, including modiforms that harbor multiple phosphosites within the same quantified tryptic peptide. The screen did not reveal any general bias for stabilization or destabilization in neither LE- nor LFE-HTP datasets (Figure 3.5A-C). Next, we performed, for LE-HTP and LFE-HTP datasets, structural elements (secondary structure, disordered regions, and solvent accessibility) analysis on proteins detected in the original HTP dataset (Figure 3.5D-L). We saw similar trends to those of our original HTP. These analyzes and comparisons reconfirmed that the effect of phosphorylation on the thermal stability is governed by local context within proteins as opposed to a systemic, biophysical "code" imparted by the sequence or identity of phosphoamino acids. This observation was reproducibly detected by different HTP variant methods.

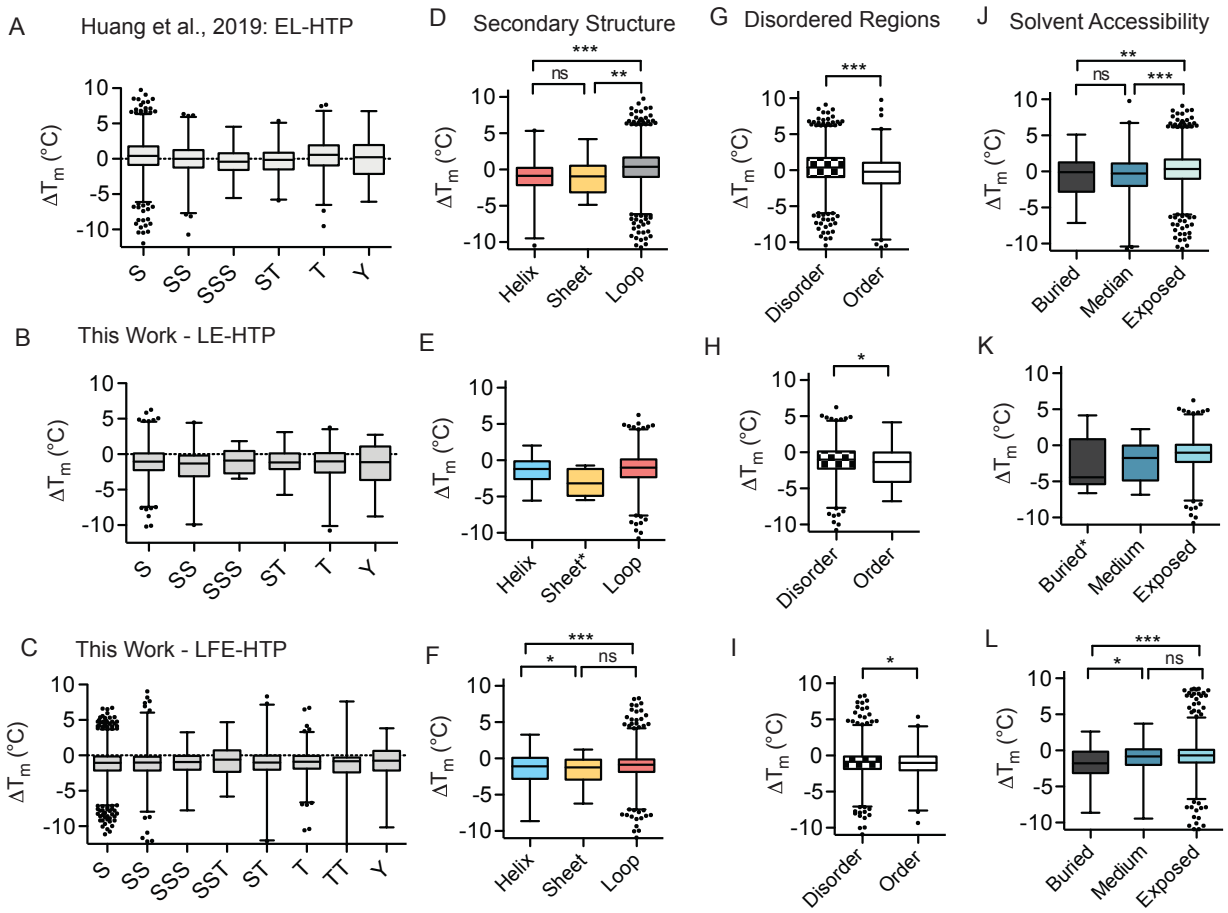

Figure 3.5 Comparing global relationships between local phosphosite environment and altered phosphomodiform stability of the 3 HTP variants

A-C) Distributions of ΔT_m values for tryptic peptides containing indicated phosphoamino acids or coincidental combinations thereof for Huang et al 2019 EL-HTP (A), LE-HTP (B), and LFE-HTP (C). **D-L**) Comparisons of ΔT_m values and predicted secondary structure elements (D-F), ordered structural elements surrounding the phosphosite of interest (G-I) and solvent accessibility (J-L) of the three datasets respectively. For secondary structure (E-F), disordered regions (H-H) and solvent accessibility (K-L) analyses performed for LE-HTP and LFE-HTP datasets, only proteins detected in the original Huang et al. publication[149] are included in the analyses. For E) and K), the number of values obtained for "Sheet*" and "Buried*" are too few (<15) to allow for meaningful statistical analyses. For D-L), ***$P \leq 0.0001$, **$P \leq 0.001$, *$P \leq 0.05$, two-sided t-test. Box plots (median, 1–99%) are shown in A-L), with outliers shown as data points.

3.2.6 Site-specific thermal stability profiles comparison between methods

In agreement with our original published dataset, we observed replication of previous hotspot phosphosite ΔT_m values in our two new datasets. SIN3A_pS832, which did not affect proteoform stability in our original dataset, likewise showed overlaid curves in the LE-HTP and LFE-HTP datasets (Figure 3.6A). Other phosphosites highlighted in our first dataset, including pT926 in KIF11 (Figure 3.6B) and pS58 in TPI1 (Figure 3.6C), showed highly similar and statistically significant ΔT_m values across our three HTP workflows and datasets. Consistent shifts were observed in all datasets for pS58 in TPI1, which led to modest statistical significance in our LE dataset ($P = 0.03$). We also observed similar changes in stability for phosphosites in 4EBP1 (Figure 3.6D) and GAPDH (Figure 3.6E). Because the depth of coverage in LE-HTP was low, KIF11_pT926 and GAPDH_pS210 were only detected once. Comparisons of datasets show consistent reproductions of both the fraction of phosphosites that were significantly shifted relative to corresponding bulk, unmodified proteoforms, as well as specific ΔT_m values for highlighted proteins. Finally, in line with the main assertions of our previous study, we also observed many other phosphorylation sites that significantly impact protein thermal stability across these new datasets and in diverse protein families.

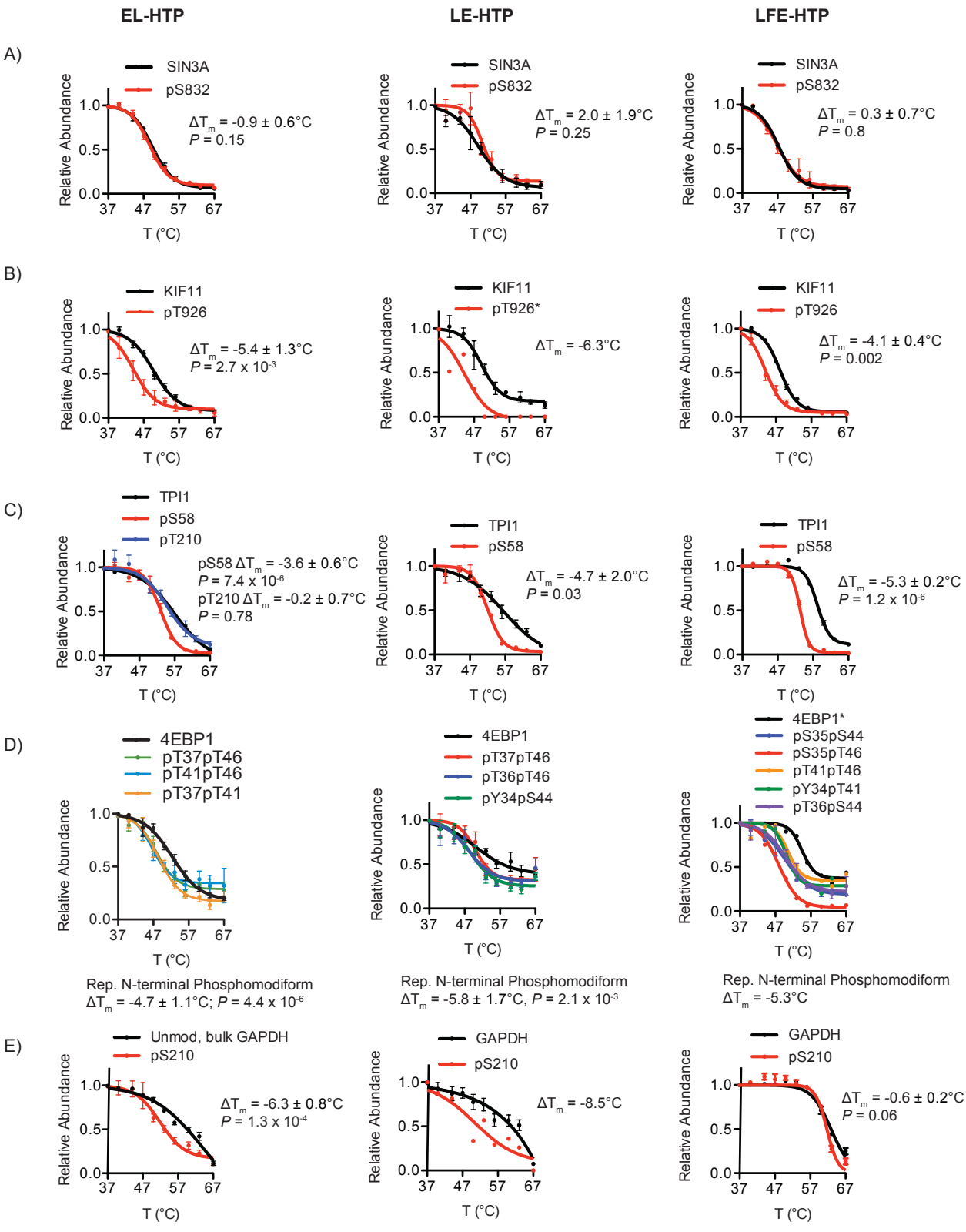

Figure 3.6 Site-specific thermal stability profiles comparison between methods
A-E) Representative T_m curves of bulk, unmodified protein (black) and indicated phosphomodiforms (red) from SIN3A **(A)**, KIF11 **(B)**, TPI1 **(C)**, 4EBP1 **(D)**, and GAPDH **(E)**.

(**Figure 3.6 continued)** Left, middle, and right column belong to melting curves generated from original EL-HTP, new LE-HTP and LFE-HTP datasets respective. Only curves with $R^2 > 0.8$ are plotted, including cases where only one curve for specific bulk or phosphosites were detected and passed this criterion (denoted by an *). Curves and error bars in **A-E)** correspond to mean and SEM; ΔT_m *P*-values calculated with two-sided *t*-test.

3.2.7 LFE-HTP detects uncharacterized phosphorylation sites in ALDOA

Among novel phosphosites that significantly impacted protein thermal stability detected in the new datasets were several sites belonging to Aldolase A (ALDOA) that caught our attention. Significant negative thermal stability shifts were detected in pS46 (ΔT_m = -7.6 ± 0.5°C, P = 9.7 x 10^{-5}), pS360 (ΔT_m = -9.0 ± 0.4°C, P = 2.0 x 10^{-5}), and pY364 ΔT_m = (-10.1 ± 1.0°C, P = 9.2 x 10^{-4}) (Figure 3.7A). These three phosphosites had been all been detected in the past according to PhosphositePlus database but not none of them had been functionally characterized. Two conformations of ALDOA had been reported in the literature with different positionings of the C-terminal tail where residues Y360 and Y364 are located. In the monomeric conformation, the C-terminal residue pY364 is located deep inside the enzyme active site while the rest of the flexible C-term tail wraps around the protein (PDB: 1ALD, Figure 3.7B)[162]. Meanwhile, in the homotetramer conformation, the C-terminal tail is much more open and exposed, with the tail no longer tugged inside the active site but located next to the tetramer binding interface (PDB: 5KY6, Figure 3.7C). It has been reported homotetramer ALDOA is catalytically less active than monomer[163] and tetramerization of the enzyme promotes its association with actin filament of the cytoskeleton[164]. We posit that the significant decrease in ALDOA thermal stability is due Y364 and S360 phosphorylations interfering with the substrate binding ability of the protein. Phosphorylation on Y364 may also affect subcellular localization of ALDOA by promoting the tetramerization of the enzyme and recruitment to the cytoskeleton. Although tetrameric ALDOA

is reportedly very thermally stable *in vitro*[165], association with the cytoskeleton in the cell makes a protein less soluble, and increases its chance to be co-aggregated and co-precipitated with the rest of the cytoskeleton during thermal denaturation process. It is interesting to note pS46 is located near the entrance of the substrate binding pocket, but it is unlike pS364 which is densely packed inside. The smaller ΔT_m value of pS46 may be reflective of its lesser influence on regulating enzyme-substrate interaction. Phenotypical characterizations could be used to uncover the full extent of these three phosphorylation sites' regulations on ALDOA activity and conformation.

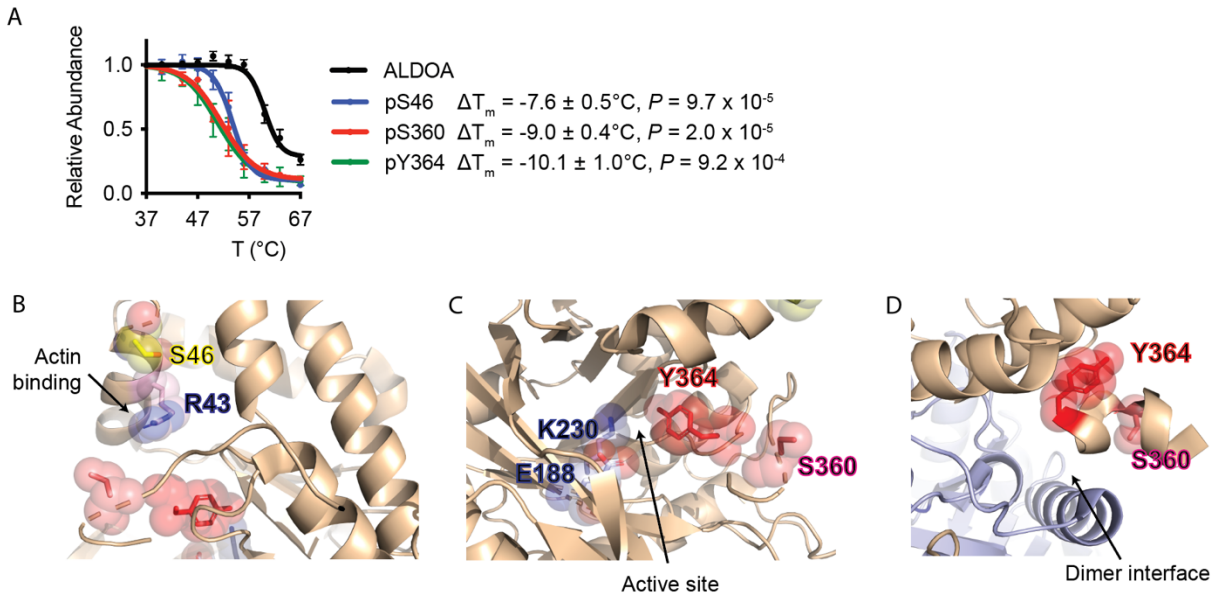

Figure 3.7 Phosphorylation induced thermal stability shifts and structural correlation in aldolase A

A) Phosphosites pS46, pS360, and pY364 induced significant and similar negative thermal stability shifts on aldolase A. **B-C**) In the monomeric conformation, S46 (yellow) is adjacent to R43 (blue) which has been shown to be involved in ALDOA recruitment to F-actin (B), meanwhile, C-terminal Y364 (red) and S360 (pink) are packed deep inside the enzyme active sites (blue) (C) (PDB: 1ALD). **D**) In the tetramer structure Y364 (red) and S360 (pink) are far away from the active site (blue) but packed next to the dimer interface (PDB: 5KY6). Curves and error bars in A) correspond to Tm mean and SEM; Only curves with R2 > 0.8 are plotted. ΔTm P-value is calculated using 2-sided student's t-test.

3.2.8 Different data analysis pipelines and statistical treatments, but not upfront biochemical workflow, can affect the significance threshold of ΔT_m values

Following our report of the EL-HTP method, several other reports have used the approach in various forms to measure the effect of protein modifications or mutations on thermal stability[166]. Among these, a method similar to our LFE-HTP approach was reported to also employ up-front TMT labeling and fractionation[167]. Despite significant differences in the subsequent biochemical workflow and data analysis pipeline used in this workflow, the researchers reported that their methodology yielded substantially fewer significantly T_m shifted phosphorylation sites (~3%) relative to our original report using the EL-HTP method (24%) (Figure 3.8). In light of the fact that our analogous LFE-HTP dataset also reported similar percentage (26%) to our EL-HTP report, we were intrigued by the conclusion of Potel et al. and wanted to explore reasons behind the difference.

We noticed the first major difference between the two methodologies is in the statistical comparison. Our HTP data analysis pipeline deliberately compares the thermal stability of specific phosphomodiforms (defined by unequivocal presence of a phosphorylation site) relative to the stability of the unmodified protein pool detected in cells (referred to as bulk, unmodified). In practice, this involves direct comparison of the mean T_m for a defined phosphomodiform to the mean T_m for the unmodified tryptic peptide pool for that protein in aggregate (Figure 3.9). We are not comparing each peptide to all other peptides one-by-one, as that represents a biochemically and statistically unique analysis. Therefore, each point on the volcano plot represents an internal comparison of two defined groups for each protein of interest, not comparisons among or between other populations (i.e., not a comparison of one T_m or ΔT_m to all of the other T_m values as is common in analyses presented on volcano plots). The statistical test applied for each modified

proteoform/bulk protein pool pair is therefore no different than that applied to compare the relative abundance change of a phosphosite/total protein pair between two conditions[168], where multiple hypothesis testing correction is not applied. Meanwhile, the Potel et al. method employed an alternative and more stringent statistical comparison pipeline that involves comparing each phosphopeptide T_m to the T_m of every tryptic peptide from that protein individually. This represents a different set of biochemical comparisons and empirically imposes more stringent statistical bar to detect significant shifts where multiple hypothesis testing correction is required.

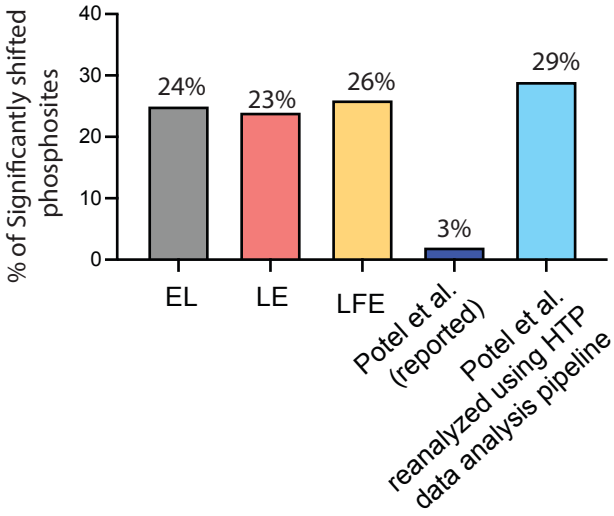

Figure 3.8 Percentage of significantly shifted phosphosites reported in different HTP datasets
Datasets generated using the three HTP variants (EL-, LE-, and LFE-) all report a similar percentage of significantly shifted T_m values, which stand in contrast with reported percentage by Potel et al.

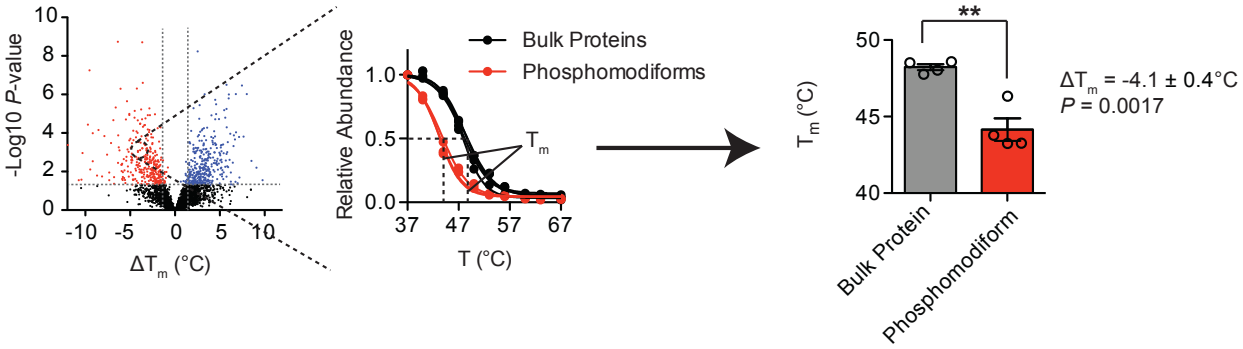

Figure 3.9 Schematic of statistical comparisons used in our HTP methods
Our data analysis pipeline compares the means and error surrounding bulk Tm and phosphomodiform T_m and assesses whether a site significantly alters protein stability. A two-side t-test is used to compare the T_m values.

Nonetheless, this approach and resulting dataset provided an appropriate comparator to the EL-HTP and LFE-HTP methods. We were particularly intrigued by the conclusion that their methodology yielded only ~3% of all detected phosphorylation sites with a significantly shifted T_m value (Figure 3.8). To understand this discrepancy, we directly analyzed their raw data using our published data analysis pipeline and statistical treatment, which reproduced nearly identical T_m curves and ΔT_m values for the phosphomodiforms LYN_pY397, LMNA_pT19pS22, and CARHSP1_pS41 highlighted in their dataset (Figure 3.10A). Therefore, despite differences in the overall analysis pipelines and statistical treatment of the resulting data (i.e., comparing a phosphopeptide to an aggregate protein-level T_m, or comparing a phosphopeptide to each unmodified peptide one-by-one), their analysis pipeline objectively reproduces the curves and shifts in thermal stability generated by our published analysis method, and vice versa.

Next, since both analysis pipelines produce essentially identical shifts on a site-by-site and protein-by-protein basis, we performed a global reanalysis of their data and compared to our published profile. This objective reanalysis confirmed that their data harbors just as many significantly shifted phosphorylation sites as our original dataset. In particular, their alternative

HTP workflow yields a dataset in which ~29% of detected phosphorylation sites have a ΔT_m with *P*-value < 0.05; this value was 23% in our original study, and 26% in our analogous LFE-HTP dataset (Figure 3.8, Figure 3.10B).

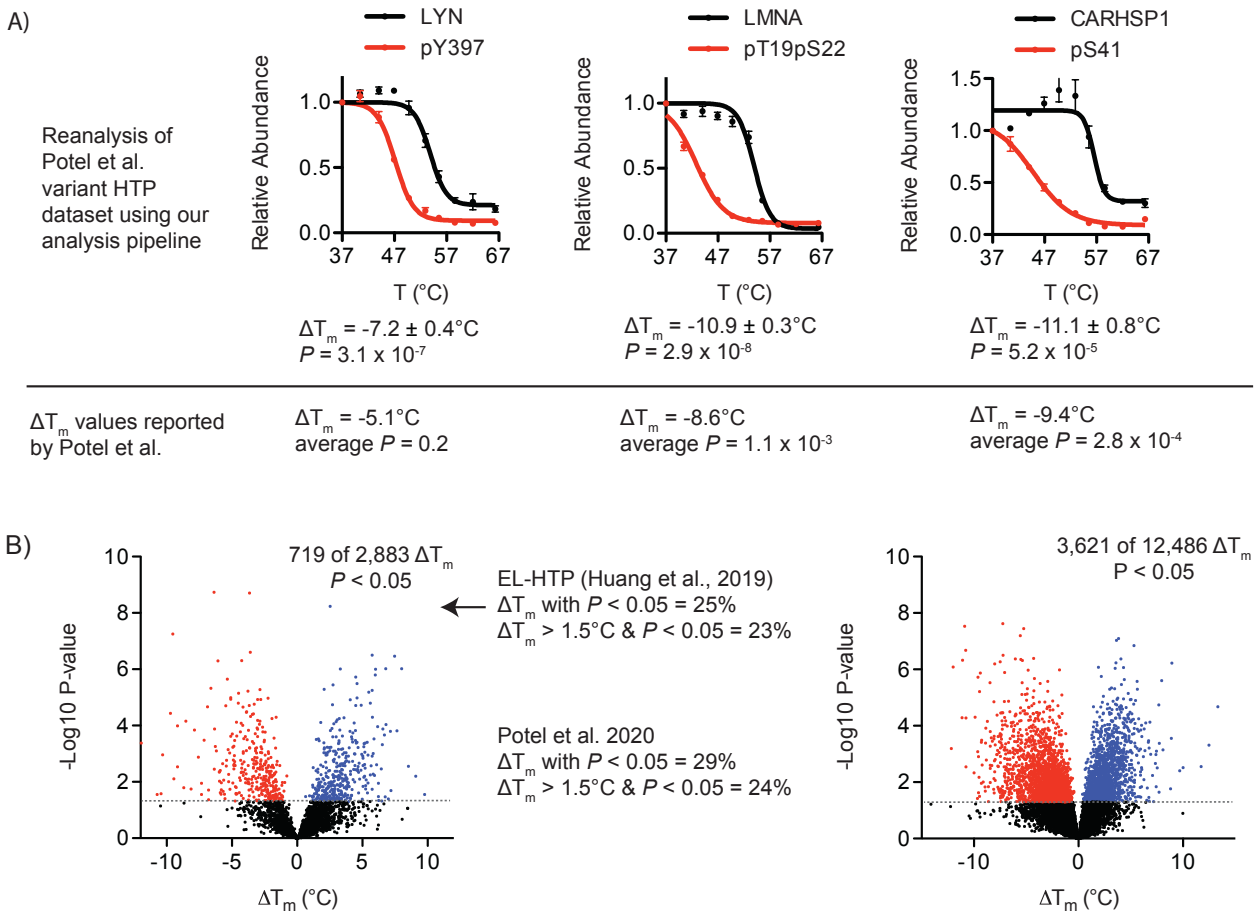

Figure 3.10 Comparing data analysis pipelines
A) Reanalysis of dataset produced by Potel et al. with our published data analysis pipeline directly reproduced the melting curves highlighted in their commentary and yielded identical ΔT_m values for specific reported modiforms. Curves in (A) are directly reproduced from Potel et al.[1] for comparison. **B)** Global ΔT_m volcano plots from Huang et al., 2019 (left) and Potel et al., commentary (right). The Potel et al. dataset produces contains a near-identical (24%) percentage of significantly shifted phosphomodiforms relative to our published dataset (23%) when data are analyzed head-to-head with our published analysis pipeline).

Diving deeper into their dataset, we also find that ΔT_m's for many of the anecdotal 'hotspot' phosphorylation sites detected in both our EL- and LFE-HTP datasets are reproduced in their dataset. While not presented in their analysis, their HTP workflow detects significant shifts in the stability of phosphomodiforms of TPI1_pS58, CDK1_pY15, 4EBP1 (several sites in the N-terminal tail) and GADPH_S210 (Figure 3.11A). The last two are notable as the observed shifts are not captured by their computational workflow due to arbitrary cutoffs imposed on curve shape and plateaus. Additionally, we found that pT926 in KIF11, which was highlighted in our first EL-HTP dataset and showed highly similar and statistically significant ΔT_m values in other HTP workflows and datasets such as LFE-HTP, was not considered significantly shifted (P = 0.4 by their analysis) by the Potel et al. analysis, despite a near identical ΔT_m as reported in our datasets (Figure 3.11B). Furthermore, we found that the application of a more stringent multiple testing adjustment brings the perceived significant hits down in their dataset and ours, but they come down uniformly (Figure 3.11). This comparison is useful as it is important to weigh the relative merits of including more or less strict significance filters for hypothesis-generating experiments, and it is unsurprising that application of divergent significance comparisons to a dataset after the fact will yield different results.

Taken together, these comparative data analyses demonstrated that the primary driver of difference in "percentage of significantly shifted phosphosite T_m" is driven by the use of a divergent downstream statistical processing pipeline. Different data analysis pipelines and statistical treatments – not upfront biochemical workflow – can affect the significance threshold of ΔT_m values.

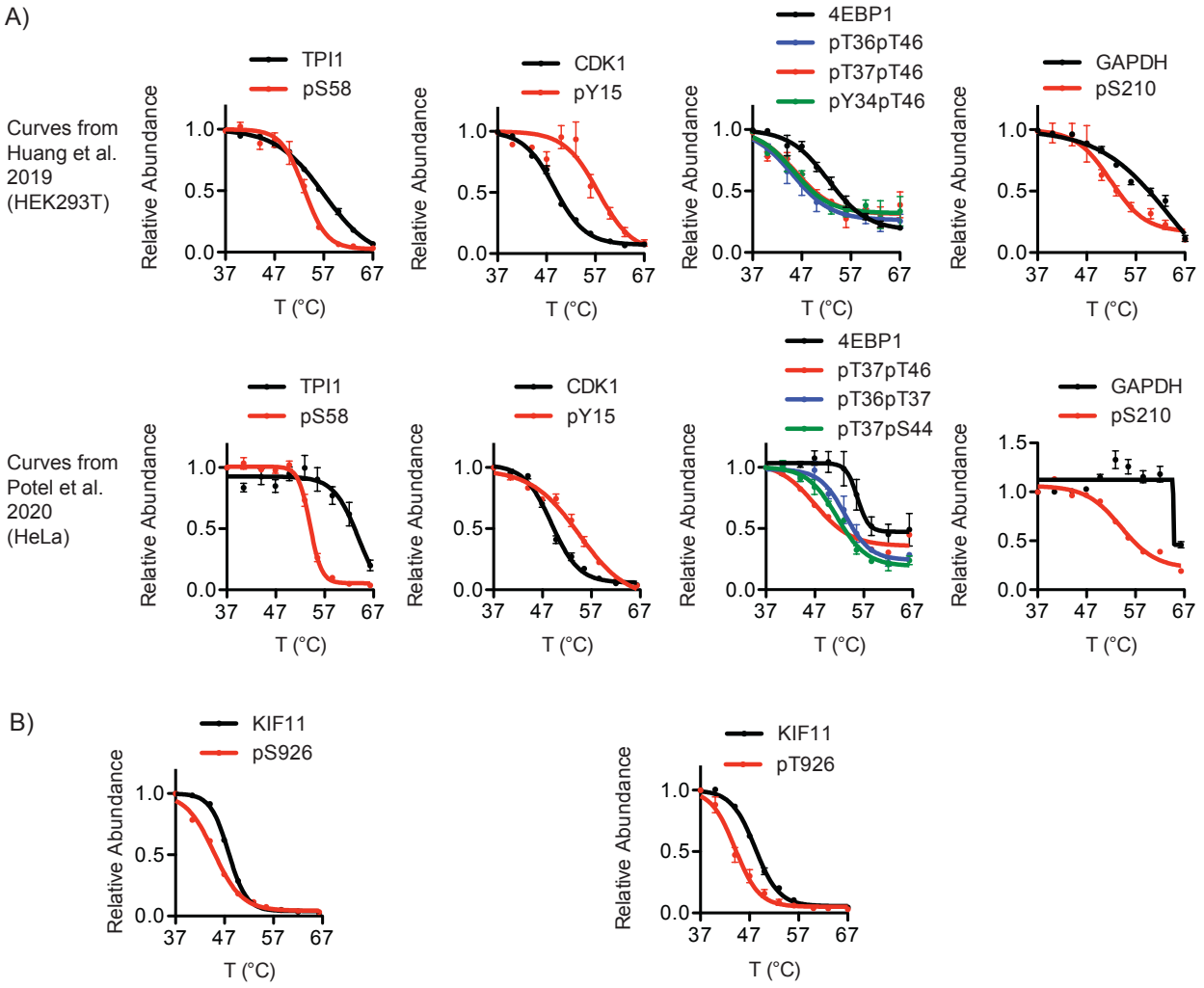

Figure 3.11 Comparing of anecdotal phosphosites measured by different data analysis pipelines

A) Representative melting curves with significant shifts in thermal stability reported in our original publication including TPI1 (modiform pS58), CDK1 (modiform pY15), 4EBP1 (several modiforms in the N-terminal tail) and GADPH (pS210) were reproduced in the Potel et al. dataset. **B**) Potel et al. reported a near identical ΔT_m in pS926 in KIF11 as in our LFE-HTP dataset, but is considered statistically insignificant using their statistical analysis pipeline.

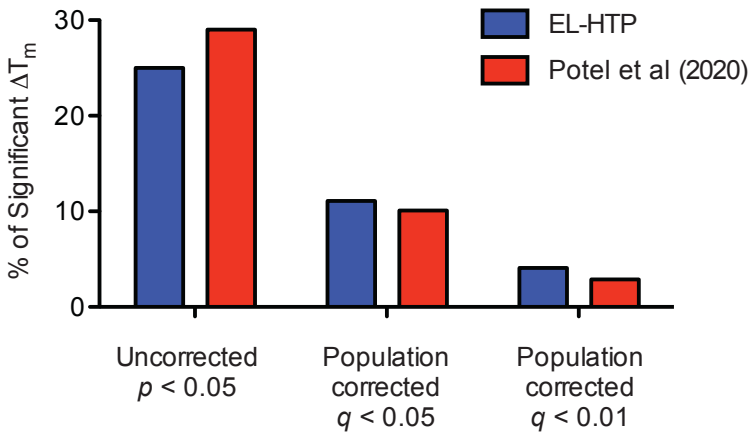

Figure 3.12 Comparing the percentages of significant ΔT_m values based on application of different statistical cutoffs

The percentages of significant ΔT_m values with *P*-value < 0.05 are 25% (Huang) and 29% (Potel) of the two datasets using our published analysis pipeline. When we apply a population-level, multiple testing correction to generate *q*-values, the percentages of 'significant' ΔT_m values are reduced. Specifically, application thresholds values of *q*-value < 0.05 reduced the percentage of significant ΔT_m to 11% (Huang) and 10% (Potel). Meanwhile, *q*-value < 0.01 reduced the percentages further to 4.1% (Huang) and 2.9% (Potel) significantly shifted ΔT_m values.

3.3 Conclusion and Discussion

We explored several modifications to the HTP workflow that could affect the depth of (phospho)proteome coverage, precision in sample processing, LC-MS/MS measurement precision, and data interpretation. We hypothesized that separate TMT-labeling of bulk and phosphopeptide fractions, use of TiO_2 spin-tips and LC-MS/MS acquisition with a single chromatographic separation could collectively contribute to lower measurement precision in our original workflow. The resulting datasets generated from the two novel HTP variant workflows, termed LE-HTP and LFE-HTP, demonstrated varying degrees of improvements to the original HTP workflow. Notably, phospho-T_m measurements have a higher reproducibility under LE-HTP as a result of performing the TMT-labeling/multiplexing step before phosphoenrichment, as well as only having to perform one single TiO_2 enrichment instead of ten separate enrichments on ten different temperature samples. Meanwhile, LFE-HTP workflow produced a significantly deeper coverage of the proteomes as well as overall greater T_m measurements reproducibility in both the bulk, unmodified- and phosphoproteomes. This is due to deep-fractionation of the proteome, resulting in 11 separate LC-MS/MS injections as well as MultiNotch MS3 quantification that helped to eliminate TMT ration suppression effects[153, 158, 160, 161]. We also discovered significant thermal perturbations from three previously uncharacterized phosphosites on ALDOA. We suggested a molecular model on how these phosphorylation sites may impact the substrate-binding ability of ALDOA as well as the enzyme conformation and subcellular localization.

The resulting datasets generated from the two novel HTP variant workflows support the notion that more cost- and resource-intensive protocols for deeper profiling should result in higher precision within runs. It is interesting to note that anecdotal proteins and phosphomodiforms detected in the original HTP dataset demonstrated similar melting profiles in the two novel HTP

variant datasets. Despite lower precision in the original HTP workflow, highly abundant proteins can still be reliably detected and generate highly reproducible T_m curves regardless of which method was used. Depending on what the goal of the experiment is, researchers may consider whether more resource-intensive workflows such as LE- and LFE-HTP justify the experimental data gains. Nevertheless, all HTP variant workflows demonstrate that they can reliably capture peptide-level thermal stability information to enable biophysical interrogation of proteoforms. All HTP variant method can measure and detect site-specific PTMs that significantly affect proteoform T_m or solubility and discover novel uncharacterized modifications that impact biophysical properties of proteins.

3.4 Materials and Methods

Cell Culture

HEK293T cells were propagated in RPMI 1640 with 2mM L-glutamine (HyClone) supplemented with 10% fetal bovine serum (Corning) and 1% penicillin-streptomycin (Thermo Fisher Scientific) and grown at 37°C in a 5% CO_2, humidified incubator.

LE-HTP Cell Harvesting and Proteomic Sample Preparation

LE (Label-Enrich) HTP proteomic samples were prepared as reported previously with several modifications[1]. Briefly, HEK293T cells were grown to 90% confluency in 15 cm cell-culture treated plates (Denville). Cells were scraped from plates, briefly washed, and resuspended in PBS with EDTA-free protease inhibitor tablet (Roche) and phosphatase inhibitor cocktail 3 (Sigma-Aldrich) before being divided into ten equal aliquots. Each aliquot of live cells was exposed to a steady temperature between 37°C to 67°C (37, 40, 44, 47, 50, 53, 56, 60, 63, and 67°C) for 3 minutes in parallel, incubated at 25°C for another 3 minutes, and lysed by three cycles of rapid freeze-thawing. Insoluble proteins and cell debris were removed by centrifugation at 17,000 *x g* for 10 minutes. A small amount of supernatant from the 37°C aliquot was set aside to use for protein quantification by Bradford's assay. The remaining fraction of the soluble fraction was denatured in 8 M urea, followed by disulfide reduction with DTT (10 mM, 30 minutes, 65°C), alkylation (iodoacetamide, 15 mM, 30 min, room temperature, protected from light) and quenching (DTT, 5 mM, 10 minutes, room temperature). The proteome solution was diluted 4-fold with ammonium bicarbonate solution (50 mM, pH 8.0), $CaCl_2$ added (1 mM) and digested with sequencing grade trypsin (~1:100 enzyme/protein ratio; Thermo Pierce) at 37°C while rotating overnight. Peptide digestion reactions were stopped by acidification to pH 2-3

with 1% formic acid, and peptides were then desalted with Sep-Pak tC18 cartridges (50 mg, Waters), dried under vacuum, resuspended with LC-MS grade water (Sigma-Aldrich) and then lyophilized.

LFE-HTP Cell Harvesting and Proteomic Sample Preparation

For LFE (Label-Fractionate-Enrich) HTP samples, cells were lifted using TrypLE Express and neutralized following 5-minute incubation using complete media (DMEM + 10% FBS penicillin/streptomycin) and centrifuged at 1100 r.p.m. for 4 minutes. The cell pellet was reconstituted in 10 mL PBS containing protease and phosphatase inhibitors (Roche) and centrifuged again at 1100 RPM for 4 minutes. Following centrifugation, the cell pellet was resuspended in 1 mL PBS with inhibitors and distributed into a thin-wall PCR tubes at 100 µL of cell suspension in each tube. Thermal denaturation was performed as previously described[1], and the resulting cellular suspension was transferred to clean 1.5 mL microcentrifuge tubes and PCR tubes were additionally rinsed with 30 µL of PBS with inhibitors to ensure complete transfer of cellular suspension. Cellular suspension was next snap frozen in liquid nitrogen for 1 minute followed by thawing and re-equilibration back to room temperature. This freeze-thaw cycle was repeated 2 additional times and the soluble fraction of each lysate was generated by centrifugation at 20,000 *x g* for 30 minutes at 4°C. Supernatants were transferred to clean 1.5 mL microcentifuge tubes, and protein was quantified in the supernatant for temperatures 37°C and 41°C by micro BCA assay (Thermo Pierce). Following quantification, the average of the two lowest temperatures was taken and the volume equivalent to 30 µg of protein in the lowest temperature was moved from each temperature fraction into a clean 1.5 mL tube. Following distribution of protein, each tube was brought to a final volume of 300 µL by addition of PBS

with inhibitors, followed by precipitation with trichloroacetic acid (TCA) to a final concentration of 25%, vigorously vortexed and incubated on ice overnight. TCA precipitates were centrifuged at 20,000 *x g* for 30 minutes at 4°C, washed twice in 500 µL of ice-cold acetone, and centrifuged at 20,000 *x g* for 10 minutes after each wash. Following precipitation and washes, pellets were allowed to completely dry at room temperature. Dry pellets were re-suspended in 100 µL of 100 mM TEAB, 0.5% SDS and reduced with 9.5 mM tris-carboxyethyl phosphine (TCEP) for 60 minutes at 55°C. Following reduction of disulfide bonds with TCEP, the denatured protein mix was centrifuged at 20,000 *x g* for 5 minutes then alkylated with 4.5 mM iodoacetamide (IAA) for 30 minutes in the dark at room temperature. After reduction and alkylation of disulfide bonds, the denatured protein mixture was precipitated out of solution by addition of 600 µL of ice-cold acetone and placed in the -20°C freezer overnight. The following day precipitated proteins were centrifuged at 8,000 *x g* for 10 minutes to pellet precipitated protein. Following centrifugation supernatant was decanted off and pellets were allowed to air-dry at room temperature. Once dry, protein pellets were reconstituted in 100 µL 100 mM TEAB and $CaCl_2$ was supplemented to a final concentration of 1 mM, 1 mg of sequencing grade Trypsin (Promega) was added, and reactions were placed in the dark on a thermal mixer (Eppendorf) set to 37°C and shaking at 800 r.p.m. for 16 hours. The next day, digested samples were centrifuged at 20,000 *x g* for 10 minutes and proceeded to TMT labeling of digested samples.

LE-HTP Phosphopeptide Enrichment and TMT Labeling

For LE-HTP (Labeled-Enriched) experiments, lyophilized peptides were labeled with 10-plex isobaric tandem mass tags (90406, Thermo Scientific) according to manufacturer's protocol with slight modification. TMT reagents were reconstituted to 16 mg/mL in anhydrous acetonitrile

(Sigma-Aldrich-) and added to lyophilized peptides dissolved in 50 µL of 50 mM HEPES buffer, pH 8.0 (~4:1 reagent/peptide ratio). Labeling reaction was carried out at room temperature for 1 hour with gentle shaking and quenched with 5 µL of 5% hydroxylamine (Thermo Scientific). Labeled peptides were combined into a single pool per experiment, acidified with formic acid (pH 2-3), desalted using ZipTip C18 tips (100 µL, Millipore), and lyophilized. A small fraction (5%) reserved for bulk unmodified proteome thermal profiling was split from the labeled pool and lyophilized separately. The major fraction of lyophilized TMT-labeled peptides (95%) was re-dissolved in 40% acetonitrile, 6% TFA before phosphopeptide enrichment with Titansphere TiO_2 beads (GL Sciences). Titansphere TiO_2 beads were reconstituted in buffer containing 80% acetonitrile, 6% TFA, and 2,5-dihydroxybenzoic acid (20 mg/mL) and rotated for 15 min in 25°C. Equal amount of beads slurry (~10:1 beads-to-peptide ratio based on concentration of peptides in 37°C aliquot) was added to each temperature aliquot of reconstituted peptides and rotated for 20 minutes at 25°C. Beads were then washed twice times with higher percentages of acetonitrile (10% and 40%) in 6% TFA and supernatant was removed by centrifugation at 500 *x g* for 2 min. Washed beads were then added to self-packed stage tip with C8 SPE (Sigma-Aldrich) and washed once more with 60% acetonitrile in 6% TFA. Phosphopeptides were first eluted with 5% NH_4OH, then 10% NH_4OH, 25% acetonitrile, and dried with speedvac. Dried phosphopeptides were reconstituted in 5% acetonitrile, 1% TFA, desalted with self-packed stage tip with C18 SPE (Sigma-Aldrich), and dried with speedvac once more. The two sets of peptides (phospho-enriched and unenriched) were separately dissolved in LC-MS/MS Buffer A (H_2O with 0.1% formic acid, LC-MS grade, Sigma-Aldrich) for LC-MS/MS analysis.

LFE-HTP TMT Labeling, Fractionation, and Phosphopeptide Enrichment

For LFE-HTP (Label-Fractionate-Enrich) experiments, TMT labeling was performed generally as above in LE-HTP experiments. Briefly, each TMT tag was re-suspended in 164 µL anhydrous acetonitrile with intermittent vortexing for 10 minutes. Following resuspension, 41 µL was added to corresponding temperatures (TMT-126 = 37°C; four separate aliquots of each for subsequent desalting and fractionation) and labeling reaction was allowed to proceed for 1 hour at room temperature. Reactions were quenched by addition of 8 µL of 5% hydroxylamine in 100 mM TEAB and incubated for 15 minutes. Labeled temperature fractions were pooled, desalted on 1cc/50 mg C18 SepPAK columns (Waters # WAT054955) on a vacuum manifold and desalted peptides were dried down in a speedvac. Dried peptides were reconstituted in 300 µL of 0.1% TFA in H_2O, high-pH reverse phase spin-columns (Pierce/Thermo-Fisher) were equilibrated, and samples fractionated per manufacturer's instructions into 8 fractions, 2 washes and a flow-through fraction (11 total). Separate samples from the same fractions were then combined and dried. Peptide fractions were reconstituted in 200 µL of 5% acetonitrile, 0.1% TFA in water, and 10 µL was removed for bulk HTP analysis. The remaining fractionated labeled peptides dried and re-dissolved in 40% acetonitrile, 6% TFA in water before phosphopeptide enrichment with Titansphere TiO_2 beads (GL Sciences) according to protocol described above for LE-HTP experiments. Each high-pH reverse-phase fraction was enriched separately with equal amounts of beads (~10:1 beads-to-peptide ratio). The final processed phosphopeptides were reconstituted in 5% acetonitrile, 0.1% TFA in water for LC-MS/MS analysis.

LC-MS/MS and Data Acquisition (QExactive HF)

LC-MS/MS experiments were performed with an Easy-nLC 1000 ultra-high pressure LC system (ThermoFisher) using a PepMap RSLC C18 column (75 μm x 50 cm; 3 μm, 100 Å, Thermo Scientific) heated to 40˚C, and coupled to a Q Exactive HF orbitrap and Easy-Spray nanosource (ThermoFisher). TMT-labeled digested peptides in MS/MS Buffer A were injected onto the column and separated using the following gradient of buffer B (0.1% formic acid, acetonitrile) at 300 nL/min: 0-10% buffer B in 5 minutes, 10-40% buffer B in 240 minutes, 40-90% buffer B over 6 minutes, and hold at 90% for 20 minutes. MS/MS spectra were collected from 0 to 250 minutes using a data-dependent, top-10 ion setting with the following settings: full MS scans were acquired at a resolution of 120,000, scan range of 375-1500 m/z, maximum IT of 60 ms, AGC target of $1e^6$, and data collection in profile mode. MS2 scans were performed by HCD fragmentation with a resolution of 60,000, AGC target of $1e^5$, maximum IT of 60 ms, NCE of 30, and data type in centroid mode. Isolation window for precursor ions was set to 1.0 *m*/z with an underfill ratio of 0.5%. Peptides with charge state 1 and undefined were excluded and dynamic exclusion was set to 20 seconds. Furthermore, S-lens RF level was set to 60 with a spray voltage value of 2.60 kV and ionization chamber temperature of 300˚C. Raw files were parsed into MS1 and MS2 spectra using RawConverter.

LC-MS/MS Analysis and Data Acquisition (Orbitrap Fusion Lumos)

High-pH reverse-phase fractions were run on a 4-hour instrument method with an effective linear gradient of 180 minutes from 5% to 25% mobile phase B with the following mobile phases: A: 0.1% formic acid in H_2O, B: 80% acetonitrile/0.1% formic acid in water on a 50 cm Acclaim PepMap RSLC C18 column (#164942) operated by a Dionex ultimate 3000 RSLC nano pump

with column heating at 50°C connected to an Orbitrap Fusion Lumos. Briefly, the instrument method was a data-dependent analysis and cycle time set to 3 seconds total. Each cycle consisted of one full-scan mass spectrum (400-1500 m/z) at a resolution of 120,000, RF Lens: 60%, maximum injection time of 100 ms followed by data-dependent MS/MS spectra with precursor selection determined by the following parameters: AGC Target of 4.0 e^5, maximum injection time of 100 ms, monoisotopic peak determination: peptide, charge state inclusion: 2-7, dynamic exclusion 10 sec with an intensity threshold filter: 5.0 e^3. Data-dependent MS/MS spectra were generated by isolating in the quadrupole with an isolation window of 0.4 *m*/z with CID activation and corresponding collision energy of 35%, CID activation time of 10 ms, activation Q of 0.25, detector type Ion Trap in Turbo mode, AGC target of 1.0 e^4 and maximum injection time of 120 ms. Data-dependent MS3 was done in synchronous precursor selection mode (SPS, multinotch MS3) with the following settings: Precursor selection Range; Mass Range 400-1200, Precursor Ion Exclusion Properties *m*/z Low: 18 High: 5, Isobaric Tag Loss Exclusion Properties: TMT. Number of SPS precursors was set to 10 and data-dependent MS3 was detected in the Orbitrap (60,000 resolution, scan range 120-500) with an isolation window of 2 *m*/z HCD activation type with collision energy of 55%, AGC target of 1.2 e^5 and a maximum injection time of 150 ms. Raw files were parsed into MS1 and MS2 spectra using RawConverter.

Data Analysis

Both data generated from QExactive HF and Orbitrap Fusion Lumos were searched using the ProLuCID algorithm in the Integrated Proteomics Pipeline (IP2) software platform. Human proteome data were searched using a concatenated target/decoy UniProt database. Basic searches were performed with the following search parameters: HCD fragmentation method;

monoisotopic precursor ions; high resolution mode (3 isotopic peaks); precursor mass range 600-6,000 and initial fragment tolerance at 600 p.p.m.; enzyme cleavage specificity at C-terminal lysine and arginine residues with 3 missed cleavage sites permitted; static modification of +57.02146 on cysteine (carboxyamidomethylation), +229.1629 on N-terminal and lysine for TMT-10-plex tag; 4 total differential modification sites per peptide, including oxidized methionine (+15.9949), and phosphorylation (+79.9663) on serine, threonine, and tyrosine (only for phospho-enriched samples); primary scoring type by XCorr and secondary by Zscore; minimum peptide length of six residues with a candidate peptide threshold of 500. A minimum of one peptide per protein and half-tryptic peptide specificity were required. Non-unique peptides were included in search. Starting statistics were performed with a Δmass cutoff = 10 p.p.m. with modstat, and trypstat settings. False-discovery rates of peptide (sfp) were set to 1%. TMT quantification was performed using the isobaric labeling 10-plex labeling algorithm, with a mass tolerance of 5.0 p.p.m. or less. Reporter ions 126.127726, 127.124761, 127.131081, 128.128116, 128.134436, 129.131417, 129.13779, 130.134825, 130.141145, and 131.13838 were used for relative quantification.

Statistical and Melting Curve Analysis

The LE-HTP dataset is comprised of n = 4 independent experiments derived from materials from independent cell cultures. Each independent experiment corresponded to n = 2 unique mass spectrometry measurements (n = 1 for bulk proteome, n = 1 for phospho-enriched proteome). The LFE-HTP dataset is comprised of n = 4 independent experiments derived from materials from independent cell cultures. Each independent experiment corresponded to n = 22 unique mass spectrometry measurements (n = 11 for fractionated bulk proteome, n = 11 for fractionated

phospho-enriched proteome). For melting curve analysis, all detected phosphopeptides that map to the same phosphorylation sites were grouped together and given a new identifier (Gene_pSite). For example, 4 different tryptic peptides mapping to the same GAPDH phosphorylation site pS210 were often detected: R.DGRGALQNIIPAS(79.9963)TGAAKAVGK.V, R.DGRGALQNIIPAS(79.9963)TGAAK.A, R.GALQNIIPAS(79.9963)TGAAKAVGK.V, and R.GALQNIIPAS(79.9963)TGAAK.A. These peptides were systematically re-labeled with the new custom identifier "GAPDH_pS210". TMT reporter ion intensities of phosphopeptides with the same identifier were combined for each temperature fraction from the same MS run. Fold change values were calculated using the lowest temperature fraction as the reference. Similarly, for the bulk unmodified proteome, TMT reporter ion intensities of all unmodified peptides mapped to the same protein were combined and fold-change values relative to the lowest temperature condition were calculated. To generate unmodified protein and phosphosite-specific melting curves, relative fold-changes as a function of temperature was fitted to the equation derived from the chemical denaturation theory using the R package and used in the original HTP manuscript[121121121121121121121121121121121121121121121121121121121121121121121 118]. T_m values were calculated at which the sigmoidal curve crosses the 0.5 fold-change level (which is not necessarily the inflection point, as discussed in the main text). Only T_m values calculated from melting curves with curve $R^2 > 0.8$ were used in subsequent analyses. Shift in T_m values (e.g., ΔT_m) induced by phosphorylation were determined by subtracting the T_m of the bulk, unmodified protein from the T_m of the phosphomodiform:

$$\Delta T_m = T_m\_\text{Phosphomodiform} - T_m\_\text{Unmodified, bulk protein}$$

ΔT_ms could only be calculated for phosphosites belonging to proteins that were also detected in the unmodified proteome. T_m values calculated by the R script of all unmodified proteins and phosphorylation sites are summarized in Supplementary Tables. For proteins highlighted in Figure 3.6, normalized relative abundance ratios were plotted and fitted with variable-slope (four-parameter) curve fit function in Prism5 (GraphPad). For proteins and phosphomodiforms highlighted in which the R script was unable to generate a curve fit, the variable-slope (four-parameter) curve fit function Prism5 was used to generate a T_m value for ΔT_m calculation. To calculate p-value for ΔT_m values, two-sided student's t-test was used to compare the mean $T_m \pm$ SEM for a defined phosphomodiform of interest to the mean $T_m \pm$ SEM for the bulk, unmodified protein pool, both derived from independent measurements from unique MS runs.

CHAPTER 4

Application of thermal profiling methods to interrogate dynamic cell states *

4.1 Introduction

4.1.1 Deregulation of glucose metabolism in diseases

The long road to elucidating the molecular mechanism of glycolysis first began in the 1850s when French scientist Louis Pasteur investigated why wine sometimes failed to ferment into alcohol[169]. He described the oxygen-independent anaerobic fermentation process of glucose as the Pasteur effect. Over the next century, many collaborative efforts went into characterizing this ancient metabolic pathway that converts glucose to pyruvate via a series of intermediate metabolites[170-172]. With generating two ATP and two NADH molecules in the process, glycolysis plays an essential role in central carbon metabolism by supplying both free energy and biosynthetic precursors for the cell.

Deregulation of the glycolytic pathway has been linked to a number of diseases, including aging-related diseases such as Alzheimer's[173-176] , diabetes[177, 178], and cancers[179-181]. In brain of patient with Alzheimer's, glucose metabolism is dramatically decreased due to damages to glycolytic proteins resulting in impairment to many downstream processes of glycolysis[173]. One study suggests that reduced glycolytic flux results in NADH deficiency and mitochondrial hypo-

* The author's contribution to the studies presented in this chapter: The author conducted all experiments described in this chapter. The author designed and performed functional biochemical experiments, cell-based experiments, mass spectrometry experiments and analyzed data.

metabolism, impairing synaptic functions and leading to neuronal death[174, 182]. Accumulation of glycolysis intermediate DHAP causes spontaneous decomposition into reactive glycolytic metabolite methylglyoxal (MGx) which is thought to be responsible for much age-related formation of advance glycation end-products (AGEs)[183]. In type 2 diabetes, multi-omics studies have associated glycolysis deregulation with reduced mitochondrial metabolism as the main driver of diabetes progression [184]. The progressive failure of pancreatic β-cells to secrete adequate level of insulin which is used stimulate cellular-uptake of glucose leads to intracellular deregulation of glycolytic flux. In cancer, proliferating tumour cells overwhelmingly rely on glycolysis for energy production even in the presence of sufficient oxygen to support mitochondrial oxidative phosphorylation to produce ATP, in a phenomenon known as Warburg effect[180, 181]. Cancer cells prefer to "ferment" most glucose to lactate and recent studies have discovered that in non-small lung cancer cells, lactate can be used as a fuel source in vivo[179].

Glycolysis also has an important role in maintaining cellular homeostasis by directly communicating and regulating other pathways in the cell. It has been recently reported that glucose metabolism directly influences the KEAP1-NRF2 pathway which serves as a central regulator of cell stress[21]. Reactive glycolytic metabolite MGx functions as a signaling molecule to induce a non-enzymatic covalent modification of KEAP1, subsequently resulting in activation of NRF2 transcriptional program of antioxidant genes[57-60].

Although glycolysis is the most well-known and well-studied pathway in nutrient metabolism, much remains unknown about how glycolysis mediates crosstalk between pathways and what role it has in a wide variety of biological functions.

4.1.2 Proteome thermal stability under low-glucose cellular environment

In Chapter 2, we touched upon the potential to perform proteome-wide interrogation of the effect of glycolytic flux on the thermal stability of the proteome. At first, we applied the HTP approach to investigate proteome thermal stability of cells under a steady-state environment without additional stimuli. While trying to understand the phosphorylation-induced thermal stability shifts in GAPDH and its phosphomodiform GAPDH_pS210, we performed HTP in live cells under an acute glucose withdrawal environment and discovered a novel phosphosite-mediated protein-metabolite interaction in GAPDH (Figure 2.17 – 2.19). Specifically, phosphorylation on S210 reduces the enzymatic activity of GAPDH by inhibiting GAP substrate-binding to the active site. As a proof-of-concept, we generated a small HTP profile of cells under low-glucose metabolic stress to show melting curves of the two GAPDH modiforms (bulk, unmodified- and phosphomodiform) under two metabolic conditions (steady-state and glucose-starved).

The ability to interrogate thermal stability of the proteome under different metabolic conditions presents another dimension of information to help further deconvolute the functional relevance of a modification site or protein to the stimuli introduced. In our pilot study investigating the functional significance of GAPDH_pS210, we observed that the GAPDH_pS210's negative T_m shift is only present under steady-state environment and disappears when cells are deprived of glucose for 30 minutes. This two-dimensional, two-state T_m shift information directly linked GAPDH_pS210 to glucose metabolite level in the cell and provided deep mechanistic insights into how glycolytic flux and phosphorylation regulate the activity of a major glycolytic enzyme. The data provided clues to the molecular mechanism of phosphoserine S210 before conventional and

labourious biochemistry efforts were carried out to exhaustively characterize the functional role of the PTM.

We envision the application of deep thermal profiling of non-steady state proteomes such as low-glucose cellular environments would help with functional discoveries of signaling pathways and targets. However, this kind of dynamic thermal profiling involving the generation of several massively parallel datasets across multiple cellular states presented new challenges to the workflow and data collection. First, the accuracy of HTP data is dependent on a confident T_m calculation. The curve fitting algorithm biases towards smooth sigmoidal melting curve that generally comes from proteins with a classic two-state unfolding model. This strict prerequisite leaves out portions of the proteome that do not follow the bimodal modal[185, 186]. Proteins that have intermediate conformation(s), proteins that have high thermal resistance and thus do not melt within the temperature range tested, and proteins that change in solubility with temperature would be missed because of poorly fitted melting curves. Secondly, the melt curve generating method require a high sample input because each temperature point must be independently quantified. This could restrict the utility of HTP in interrogating sample types such as clinical tissue samples that are limited in quantity.

An alternate thermal stability approach has been proposed by Gaetani et al.[187] to address the curve-fitting and sample limiting issues. Termed Proteome Integral Solubility Alteration (PISA), the assay measures area-under-the-curve of a protein melting profile and calculates a single S_m (signal) value per protein to describe its unfolding behaviour. Using area-under-the-curve as a qualitative and quantitative measurement of protein thermal stability allows the PISA method to bypass sigmoidal curve-fitting issue which increases the risk of excluding potentially mechanistically important proteins that do not follow a bimodal unfolding mechanism.

The PISA assay has been used to profile 9 different drugs in human lung carcinoma A549 cells and rediscovered many known protein targets of the drugs[187]. Furthermore, the selection of a subset heating temperatures closer to the medium range of human protein T_m[124] has been shown to improve the sensitivity and sampling handling throughput of the PISA assay[188].

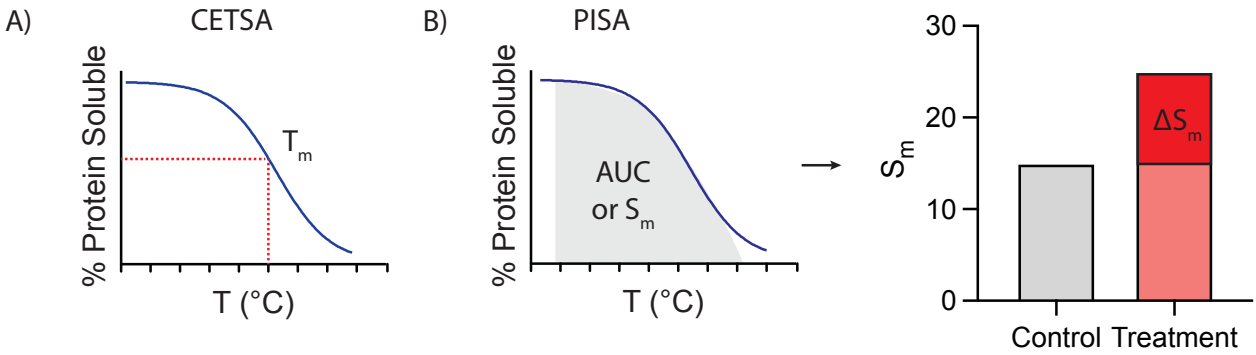

Figure 4.1 Schematic showing the data quantification difference between CETSA and PISA assays
A) In a classic CETSA assay, protein melting follows a sigmoidal curve. Protein melting temperature T_m is measured at which half of proteins population remain soluble. **B)** The PISA assay measures the total area-under-curve of the protein melting curve. A new term S_m is assigned to describe the melting behaviour. The difference between S_m values (ΔS_m) of different treatment groups is used to measure the thermal stability shifts.

In this Chapter, we sought to explore the application of the original HTP workflow in dynamic cellular states (acute glucose starvation and steady state) with the intension of significantly increasing the depth of proteome coverage to further our understanding of the proteome thermal stability impact of glycolytic flux. Then, we apply a modified PISA assay in a kinetic series of time-dependent acute glucose starvation to profile the dynamic thermal stability changes of the proteome.

4.2 Results

4.2.1 Deep HTP profiling of proteome under acute glucose withdrawal

To generate a deep thermal stability profile of the proteome under acute glucose withdrawal, we performed an updated HTP workflow specifically looking at the bulk, unmodified proteome. We incubated HEK293T cells in either normal growth media (steady-state), or glucose-free media (acute glucose withdrawal) for 30 minutes to deplete cellular glucose level before live cell heat-pulses, rapid cell lysis, isolation of soluble proteome, and preoteolyzation into tryptic peptides (Figure 4.2). All peptides from the same temperature aliquot are labeled with specific tandem mass tag (TMT) isotope channels and combined into one single pool. After multiplexing, the combined pool of TMT-labeled sample undergoes offline bRP fractionation into 8 different fractions. Each series fractionated of two proteome pools undergo 8 LC-MS/MS proteomic analyses and MS2-based acquisition quantification of relative peptide levels across the temperature range. Peptide-level measurements are analyzed en masse with peptide identification, relative quantification and automated curve fitting algorithms to determine peptide-specific melting points (T_m) for tens of thousands of specific peptides in the same sample.

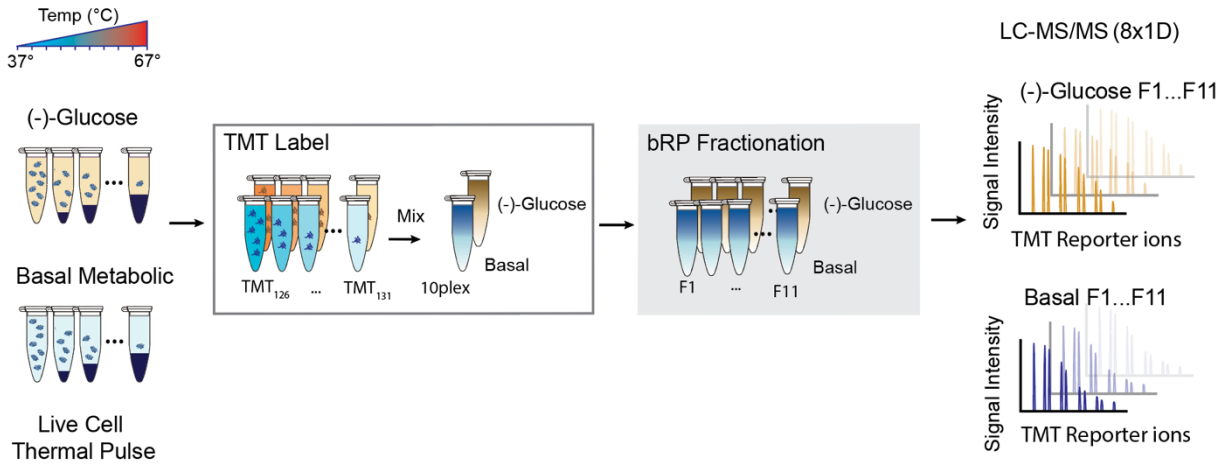

Figure 4.2 Schematic of showing thermal proteome profiling workflow with deep bRP fractionation

We detected 6,251 proteins in HEK293T cells across all LC-MS/MS runs. The dataset is comprised of n = 3 biological replicates from either steady-state or acute glucose withdrawal proteome. Each biological replicate underwent 8 LC-MS/MS injections (one per bRP fraction) for the unmodified, bulk proteome series. Statistical filtering at the peptide identification level identified a total 5,658 and 4,979 unique proteins in the basal metabolic and (-)-glucose proteomes respectively (Figure 4.3A). More than 63% (3,931) of them are common in both proteomes. The application of stringent automated T_m curve fitting cutoffs ($R^2 > 0.8$) resulted in a database of site-specific T_m values for 5,163 and 4,515 proteins under basal metabolic and acute (-)-glucose conditions respective (Figure 4.3B). The T_m distributions are Gaussian and the average T_m values between two proteomes are nearly identical at 50.9°C and 51.5°C for basal metabolic and (-)-glucose respectively. The volcano plot shows a small enrichment of positive ΔT_m values. Surprisingly, there are very few proteins with significantly thermal perturbation when compared to the difference between phosphomodiform and bulk, unmodified proteome.

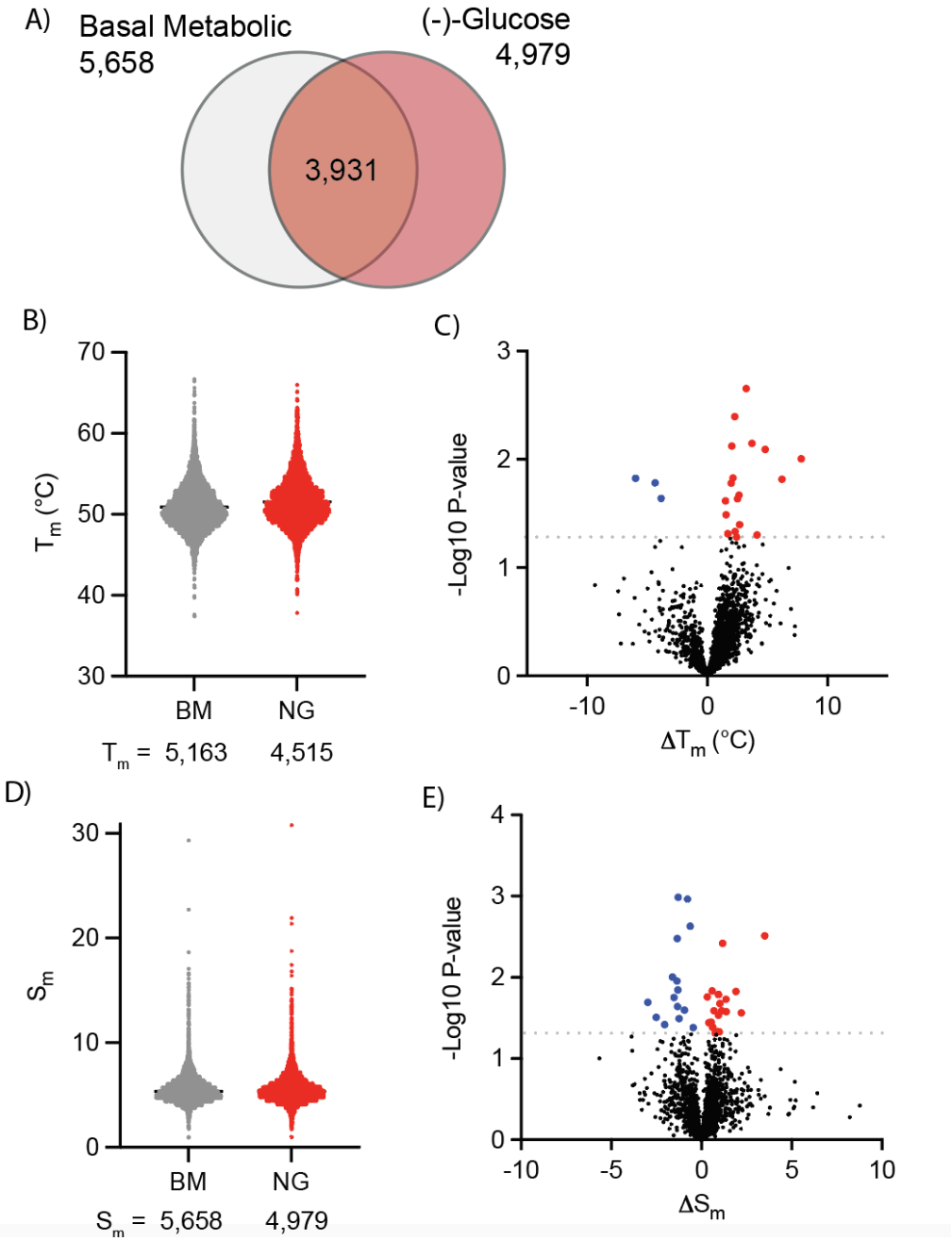

Figure 4.3 Deep global thermal proteome profiles of proteome under acute glucose withdrawal perturbation

A) The total number of unique proteins detected in the datasets. **B)** Global T_m value distribution of the basal metabolic (BM) and acute glucose withdrawal (NG) datasets**. C)** Volcano plot of high-confidence ΔT_m values. Proteins with significant shifts ($P < 0.05$, 2-sided t-test for protein-to-protein comparisons) are shown in red (negative shift) and blue (positive shift). **D**) Global S_m value distribution of the basal metabolic (BM) and acute glucose withdrawal (NG) datasets. S_m values are generated from T_m curves by calculating the area-under-curve of the protein melting curves. **E**) Volcano plot of high-confidence ΔS_m values. Proteins with significant shifts ($P < 0.05$, 2-sided t-test for protein-to-protein comparisons) are shown in red (negative shift) and blue (positive shift).

Next, we converted melting curves to area-under-curve S_m values in in the style of a PISA dataset. Signal intensities of all TMT channels are combined to generate a single S_m value for each protein. The S_m value distributions are mainly Gaussian in nature, but there is a slight enrichment in the high S_m value range (Figure 4.3D). This is not surprising as AUC quantification allows proteins that do not readily melt, and therefore have a low confidence sigmoidal fit, to still be captured. As expected, every detected protein was quantified using AUC style quantifications. The volcano plot of ΔS_m shows an even distribution (Figure 4.4E).

Despite small in number, a small subset of proteins was significantly thermally perturbed under acute glucose withdrawal condition. These include potassium channel tetramization BTB/POZ domain-containing protein (KCTD12) which showed a large T_m of 7.8℃ ($P = 0.001$) and an equally large converted ΔS_m of 1.5 ($P = 0.02$) (Figure 4.4A). Enzyme glyoxalase I (GLO1) showed a large T_m of 4.8℃ ($P = 0.008$) but a small converted ΔS_m of 0.26 ($P = 0.8$) (Figure 4.4B). In the opposite thermal stability shift is protein junction plakoglobin (JUP) which has negative ΔT_m and ΔS_m values (Figure 4.4C).

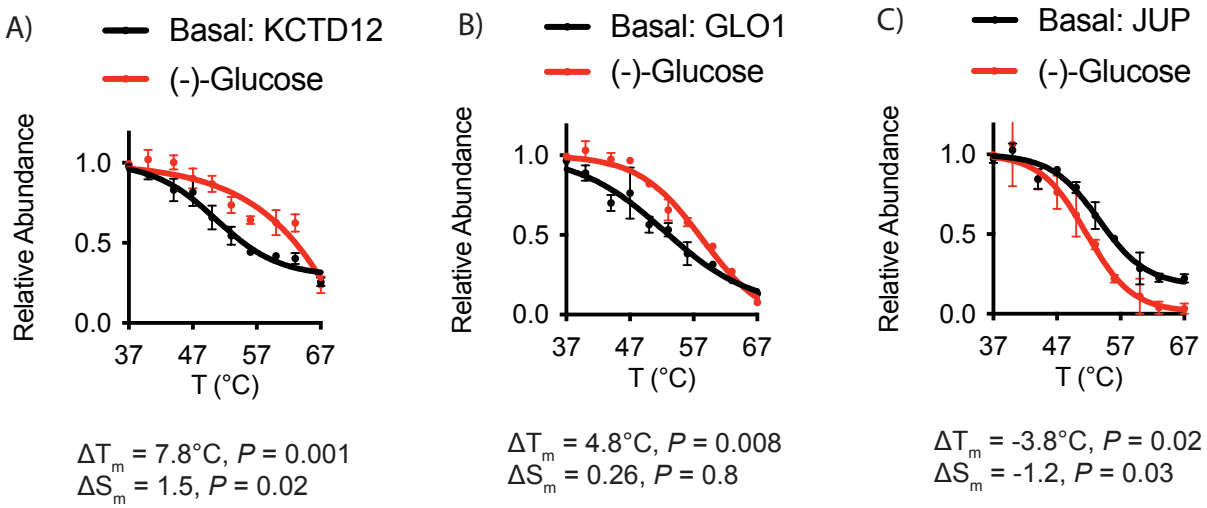

Figure 4.4 Representative protein melting curves
Representative protein melting curves of KCTD12 **(A)**, GLO1 **(B)** and JUP **(C)** are shown. Their ΔT_m and converted ΔS_m values are show below.

4.2.2 Deep dynamic PISA profiling of proteome under acute glucose withdrawal

To generate a deep dynamic PISA profile of the proteome under acute glucose withdrawal, we performed an updated PISA workflow adapted from Gaetani et al to specifically look at the bulk, unmodified proteome. We incubated heavy and light SILAC HEK293T cells in normal growth media (basal metabolic or steady state condition) and in glucose-free media (acute glucose withdrawal) respectively for 30, 60, and 120 minutes before live cell heat-pulses, rapid cell lysis, isolation of soluble proteome (Figure 4.5). Corresponding SILAC cells from the same incubation time point are multiplexed, proteolyzed, and undergo offline bRP fractionation into 8 different fractions. Each timepoint series undergoes 8 LC-MS/MS proteomic analyses. Peptide-level measurements are analyzed en masse with peptide identification, and relative quantification of heavy and light peptide ratios to determine protein-specific area-under-the-curve values for tens of thousands of proteins in the same sample.

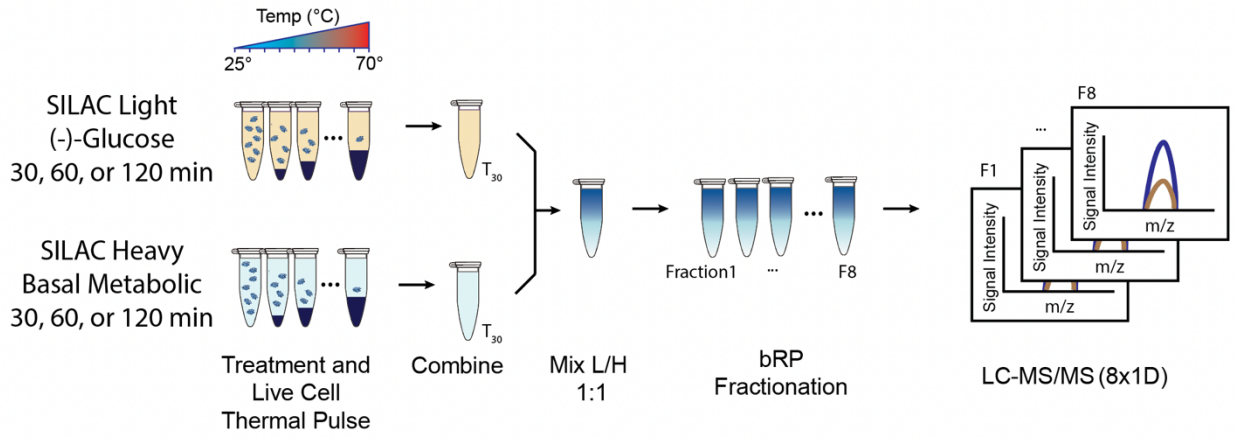

Figure 4.5 Schematic of modified PISA assay using SILAC cells for dynamic profiling of acute glucose withdrawal

We detected 6,429 proteins in HEK293T cells across all time points and all LC-MS/MS runs (Figure 4.6A). The dataset is comprised of n = 3 biological replicates for each time-point and each biological replicate underwent 8 separate LC-MS/MS injections (one per bRP fraction). Because we used a SILAC generate this data, we defined the AUC S_m value as the L/H ratio, in which light cells were incubated with glucose-free media and heavy cells were incubated with regular growth media. For peptides that are detected only in either the light or heavy proteome (also known as singletons), a log2(L/H) value of +/- 6 is manually assigned. The distribution of L/H ratios are evenly distributed, with a small but global negative thermal shift in both T30 and T60 timepoints (Figure 4.6B) meanwhile T120 time-point had a small but positive global thermal shift. Volcano plots also reveal a similar enrichment of negatively shifted proteins after 30 and 60 minutes of glucose-free media incubation (Figure 4.6 C-D) and a small positively shifted proteins after in the 120-minute timepoint (Figure 4.6 E)

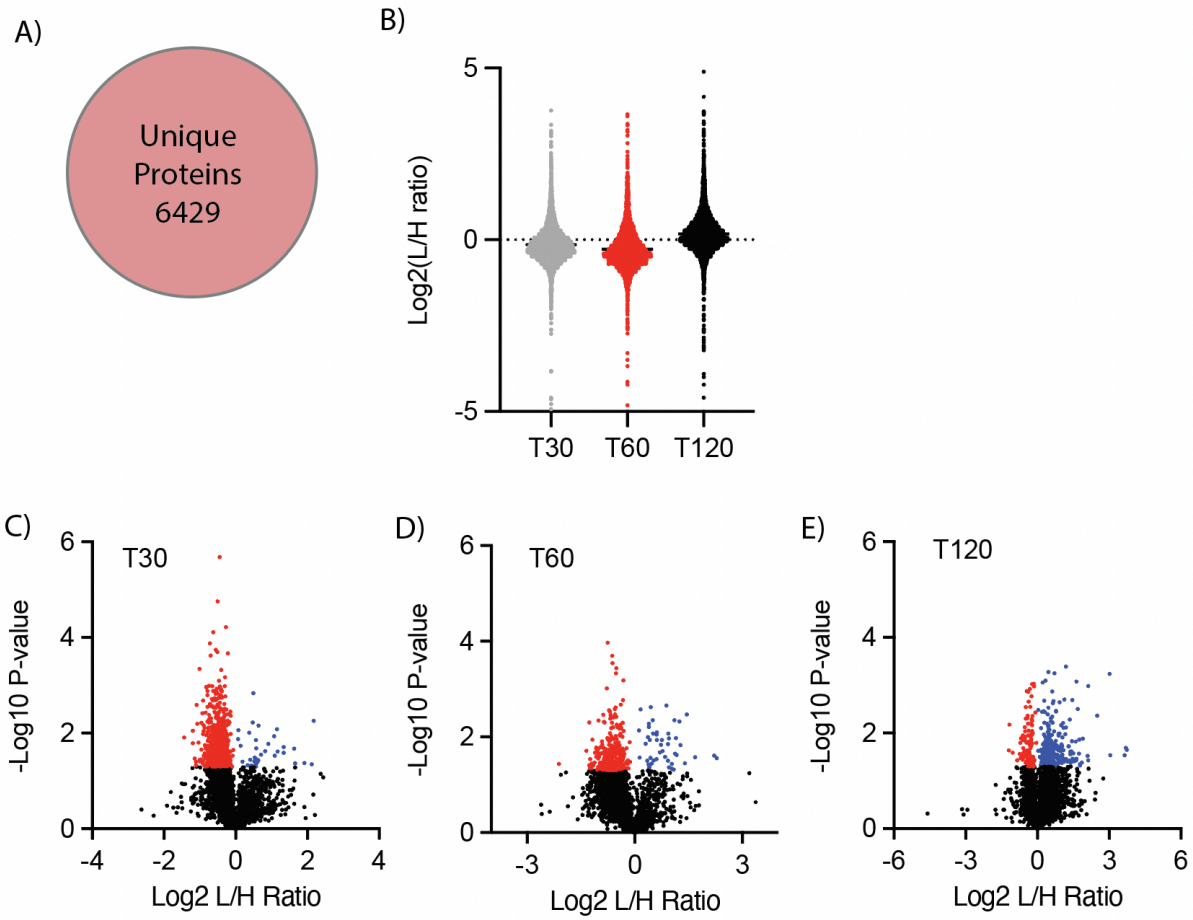

Figure 4.6 Deep global PISA thermal stability profiles of cells under acute glucose withdrawal perturbation
(A) Total number of unique proteins detected across the timepoints. **B**) The global distribution of L/H values for the three incubation timepoints of HEK293T cells in glucose-free media: 30-, 60- and 120-min. **D-E)** Volcano plots of high-confidence L/H values in the 30-min (**C**), 60-min (**D**), and 120-min (**E**) time points. Proteins with significant shifts ($P < 0.05$, 2-sided t-test for protein-to-protein comparisons) are shown in red (negative shift) and blue (positive shift).

To understand whether there is a systematic effect of glucose-withdrawal on the thermal stability of the entire proteome, we looked at proteins that are known to interact with glucose metabolites. Glycolytic enzyme GAPDH was previously reported to demonstrate significant negative T_m shift after cells were incubated with glucose-free media for 30 minutes[149]. Indeed, this effect on GAPDH is observed again (Figure 4.7A). The magnitude of negative thermal shift

increased after 60 minutes and appeared to be slightly stabilizing after 120 minutes of cells-incubation in glucose-free media. This profile is also observed in glycolytic enzymes GPI, ALDOA, TPI1, PGK1, and ENO1 (Figure 4.7B-F). However, this profile is different in HK1 in which the enzyme remained consistently destabilized throughout the three timepoints (Figure 4.7G).

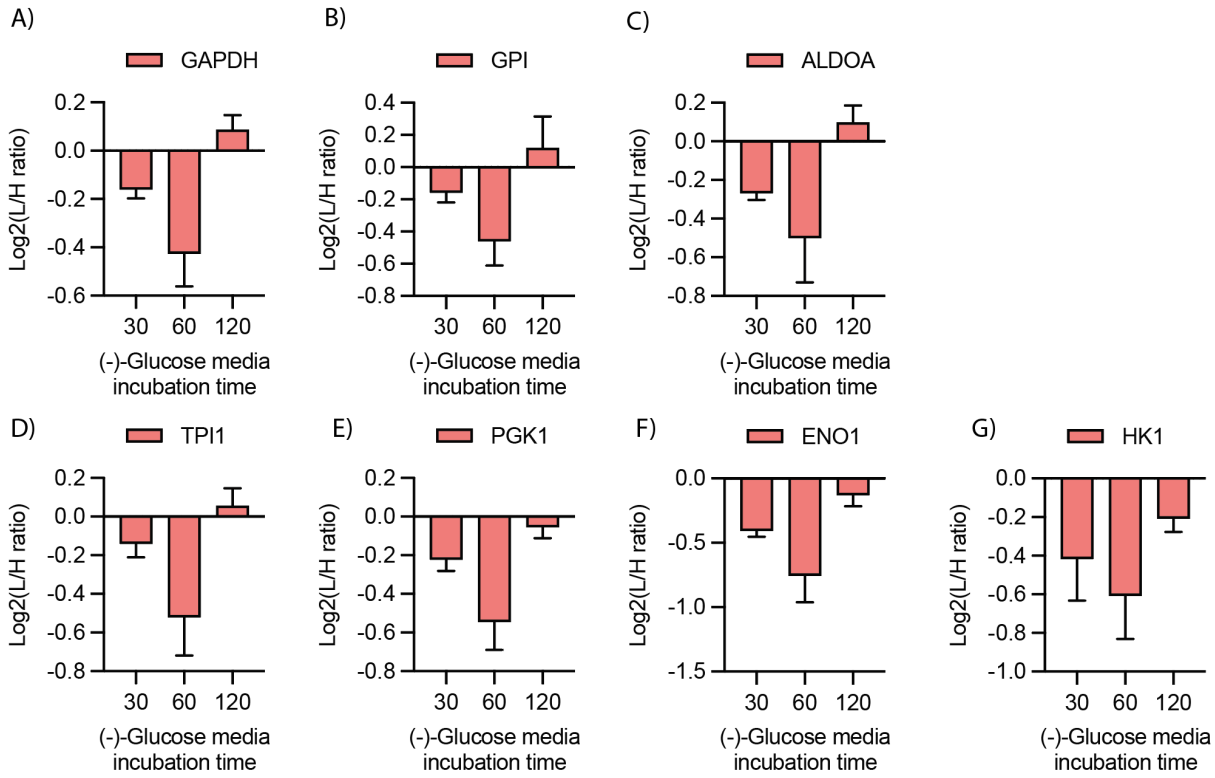

Figure 4.7 Glucose-deficiency induced stability shifts in glycolytic enzymes
The thermal stability shifts in glycolytic enzymes after HEK293T cells underwent 30-, 60-, and 120-min incubation in glucose-free media are shown: glyceraldehyde 3-phosphate dehydrogenase (**A**), glucose-6-phosphate isomerase (**B**), aldolase A (**C**), triosephosphate isomerase 1 (**D**), phosphoglycerate kinase 1 (**E**), alpha-enolase 1 (**F**), and hexokinase 1 (**G**).

4.3 Conclusion and Discussion

We applied a modified HTP workflow to generate a deep proteome dataset of protein thermal stabilities under acute glucose withdrawal. We were able to detect significantly more proteins compared to the first generation of HTP dataset thanks to offline fractionation approaches we employed. This dataset revealed surprisingly few proteins with significant thermal perturbation but helped provide a proof-of-concept that T_m melting curve data can be converted to a PISA style data output to capture proteins that would normally be missed in using stringent curve fitting strategies.

We next developed a modified version of the original PISA method[187]. Instead of using TMT labeling, we performed the experiment in SILAC cells to investigate the thermal stabilities of proteomes after 30, 60, and 120 minutes of incubation with glucose-free media. The resultant dataset revealed a small global negative shift in the first two timepoints but not in the last. Looking at specific proteins that are known to interact with glucose metabolites, we observed that only a subset of glycolytic enzymes is affected by the acute glucose-deficiency. These data may suggest that the thermal stability response to acute glycose flux is mainly protein specific.

The SILAC-PISA approach has a great advantage over the classic thermal shift assay which best captures proteins that undergo two-state unfolding model but may miss or overlook proteins that do not display prototypical melting behaviours with a sigmoidal protein melting curve. Meanwhile, the PISA assay does not require sigmoidal curve fitting, since the method takes into account the entire melting curve and uses total area under the protein melting curve as a signal readout. PISA can help capture proteins that have atypical melting behaviour that would otherwise be missed by the more stringent sigmoidal curve fit. We believe that SILAC-PISA offers a greater multiplexing capability that could enable profiling different conditions more easily.

4.3.1 Future directions

The kinetic glucose-withdrawal SILAC-PISA dataset revealed surprising shifts in thermal stabilities in the proteome. We propose several future directions to investigate the underlying mechanisms in regulating the global proteome as well as protein-specific thermal stability shifts induced by the glucose-deficient environment in cancer cells. A destabilizing trend in the global proteome was observed after 30 and 60 minutes of glucose-free media incubation. Digging deeper into the dataset, we also measured thermal destabilization of many glycolytic proteins, with maximal negative shifts detected in the 60-minute time point. This destabilizing trend was expected of glycolytic proteins as they are known to interact with glucose metabolites. For example, we previously demonstrated that low substrate occupancy could thermally destabilize GAPDH (Chapter 2.2.7). However, thermal destabilization was nearly abolished in the 120-minute time point and even trended marginally positive for several glycolytic enzymes. Among them, the only enzyme that remained negatively shifted was hexokinase-1 (Figure 4.7G). Hexokinase is the first enzyme in glycolysis and a rate-limiting enzyme for glycolysis[189]. Overexpression of hexokinase was found to be associated with aerobic glycolysis and to help maintain the high glycolytic phenotype in multiple tumours[190]. Given the importance of the enzyme as a metabolic target in cancer treatment, an immediate next step is to validate and understand how glucose-deficiency progressively and consistently thermally destabilizes hexokinase in a time-dependent manner. To determine whether protein abundance may also have a role in altering the thermal stability of glucose-dependent proteins, we propose to perform relative quantification experiments of the proteome to provide insights on the reversal of thermal stability trend after 120-minute of glucose-deplete environment in the cell.

Against the consistently destabilization trend in hexokinase, several downstream glycolytic enzymes, including GAPDH, TPI1, ALDOA and GPI1, were able to seemingly recover from the destabilization effect after 120 minutes of glucose-withdrawal treatments. It was recent reported that glycolytic enzymes coalesce in non-membrane-bound granules known as G bodies under stress[191, 192]. This liquid-liquid phase-separation event is an adaptive stress response to concentrate glycolytic enzymes and glucose metabolites to increase glycolytic output. We hypothesize that G bodies formation may play a role in the widespread thermal stabilization of glycolytic enzymes after 120 minutes of glucose-deficiency through interactions with chaperone proteins in the non-membrane-bound granules. To validate this hypothesis, we propose to perform microscopy experiments with GFP-tagged glycolytic proteins to observe the formation of G bodies under a kinetic series of glucose-depleted environments. To understand how phase separation affects the stabilities of proteins, we also suggest comparing and contrasting the thermal stabilities of cytosolic and phase-separated glycolytic enzymes by isolating G bodies and performing thermal shift assays using intact G bodies. A purification strategy for G bodies were recently developed by Jin et al.[191]. Quantitative proteomics experiment can also be performed to measure the abundance of phase-separated glycolytic proteins to demonstrate time-dependent recruitment of proteins to G bodies under metabolic stress.

The next category of future directions is in fine-tuning the SILAC-PISA workflow. The proof-of-concept SILAC-PISA dataset interrogating thermal stabilities of the global proteome after a kinetic series of glucose-withdrawal perturbation demonstrated the feasibility of the method to screen through different conditions in a hypothesis generating experiment. However, despite the more straight-forward sample handling and higher multiplexing capabilities, the magnitude of thermal shifts captured in the SILAC-PISA glucose-withdrawal dataset was relatively small. We

wonder if improvements can be made to the current PISA workflow to enhance sensitivity. The first area of improvement is in the selection of temperature ranges used to thermally perturb the proteome. In an attempt to match the conditions used in HTP workflow, we used a wide range of temperatures between 25℃ and 70℃ in the health pulse step. It was recently suggested by Li et al. that an optimized selection of temperature ranges could improve sensitivity and dynamic range of PISA measurments[188]. A subset of temperature ranges between 45-55℃ where the majority of proteome unfold was reported to capture the biggest thermal shifts. It was also shown that fewer temperature points could be used as long as they stay within the dynamic range of melting temperatures. Using fewer temperature points also provide the added benefit of requiring less sample input and which could be hugely beneficial when working with limited sample such as primary patient tissues or sensitive cell lines. The next step in profiling the thermal stability effects of glucose withdrawal using PISA is to fine tune the workflow to maximize dynamic range of PISA ratio measurements.

The SILAC-PISA workflow we developed in this Chapter does not allow direct thermal stability comparison between bulk, unmodified proteins and phosphoproteoforms. This is due to the fact that PISA compares abundance differences between two of the same species of modiforms, meanwhile phosphomodiforms and unmodified proteoforms are different species that have different abundance level in the cell. In CETSA, the T_m value of a protein is self-normalizing since the melting curve is constructed using relative fold-change compared to the lowest temperature time fraction. This means even if the abundance level of the PTM modiform and unmodified proteoform is different, it is still possible to compare and contrast the two since the calculated T_m value is derived from relative fold change that is internally normalized within each protein or modiform. For a PISA-style method to work for protein modiform comparisons, an internal

normalizing parameter must be incorporated so that the PISA ratios between PTM modiform and the unmodified protein can be directly compared. We propose an altered SILAC-PISA workflow in which a set of heavy cells acting as the internal reference is kept at isobaric temperature, while a matched set of light SILAC cells is treated to a 3-minute heat pulse (Figure 4.8). The two sets of SILAC cells then undergo rapid cell lysis, isolation of soluble proteome, combined into one SILAC 2-plex sample. After multiplexing, 5% of the combined SILAC pool is set aside for protein-level thermal profiling of the "unmodified, bulk" protein pool. The remaining 95% of SILAC multiplexed sample is enriched in one single TiO_2 enrichment step to generate the phosphoproteome pool. The two respective pools undergo 1D LC-MS/MS proteomic analysis or fractionated further before quantitative proteomics analysis. TMT labeling can also be performed on each SILAC 2-plex sample if multiple timepoints or treatment conditions are performed. A SILAC-TMT experiment requires significant amount of bioinformatics computation power in order to search and quantify these complex peptides that are multiply modified (i.e., first metabolically labeled with heavy amino acids and then isobarically labeled with TMT reagents), therefore the method might not be doable in every research lab. However, the multiplexing capability of PISA-based approach holds huge potential for thermal profiling of protein modiforms in a high-throughput manner.

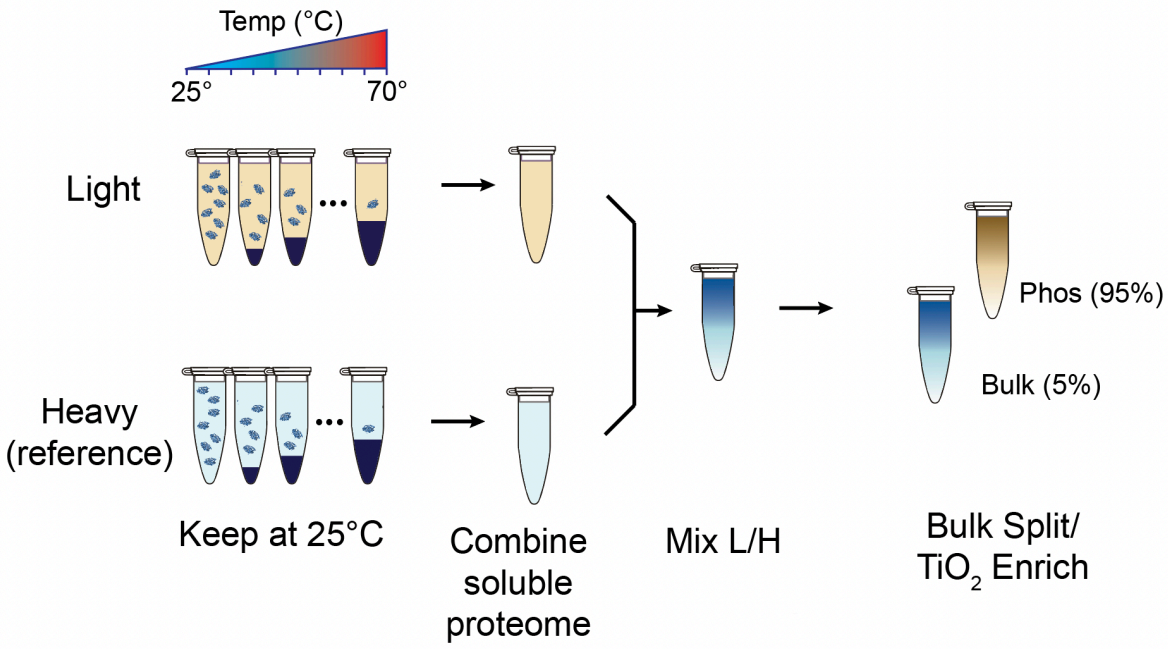

Figure 4.8 Workflow schematic for the proposed PTM-PISA method
A set of heavy SILAC cells which acts the internal reference is kept at isobaric temperature, while a matched set of light SILAC cells is treated to a 3-minute heat pulse, before they both undergo rapid cell lysis, isolation of soluble proteome, and combined into one SILAC 2-plex sample. After SILAC multiplexing, 5% of the combined SILAC pool is set aside for protein-level thermal profiling of the "unmodified, bulk" protein pool, while the remaining 95% is enriched to generate the phosphoproteome pool.

4.4 Materials and Methods

Cell Culture

HEK293T cell lines were propagated in RPMI 1640 with 2mM L-glutamine (HyClone) supplemented with 10% fetal bovine serum (Corning) and 1% penicillin-streptomycin (Thermo Fisher Scientific). SILAC HEK293T cells were propagated in SILAC RPMI1640 with 2mM L-glutamine (Thermo) supplemented with 10% fetal bovine serum (Corning), and 1% penicillin-streptomycin (Thermo Fisher Scientific). "Light" and "Heavy media were supplemented with natural lysine and arginine (0.1 mg/mL), and ^{13}C-, ^{15}N- labeled lysine and arginine (0.1 mg/mL) respectively. All cell lines were grown at 37°C in a 5% CO2 humidified incubator.

Thermal proteome profiling Sample Preparation

HEK293 cells were grown to 90% confluency in 10 cm cell-culture treated plates (Denville). Cells were then incubated with 5 mL of glucose-free RPMI media (Hyclone) for 30 minutes before they are scraped from plates, briefly washed, and resuspended in PBS with EDTA-free protease inhibitor tablet (Roche) and phosphatase inhibitor cocktail 3 (Sigma-Aldrich) before being divided into ten equal aliquots. Each aliquot of live cells was exposed to a steady temperature between 37°C to 67°C for 3 minutes in parallel, incubated at 25°C for another 3 minutes, and lysed by rapid freeze-thawing. Insoluble proteins and cell debris were removed by centrifugation at 17,000*g* for 10 minutes. Small amount of supernatants from the 37°C aliquot was set aside to use for protein quantification by Bradford's assay. The remaining fraction of whole cell lysates was denatured in 8 M urea, followed by disulfide reduction with DTT (10 mM, 30 minutes, 65°C), alkylation (iodoacetamide, 15 mM, 30 minutes, room temperature, protected from light) and quenching (DTT, 5mM, 10 minutes, room temperature). The proteome solution was diluted 4-fold with ammonium

bicarbonate solution (50 mM, pH 8.0), $CaCl_2$ added (1 mM) and digested with sequencing grade trypsin (~1:100 enzyme/protein ratio; Thermo Pierce) at 37 ˚C while rotating overnight. Peptide digestion reactions were stopped by acidification to pH 2-3 with 1% formic acid, and peptides were then desalted with Sep-Pak tC18 cartridges (50 mg, Waters), dried under vacuum, resuspended with LC-MS grade water (Sigma-Aldrich) and then lyophilized.

PISA Proteomic Sample Preparation

SILAC HEK293 cells were grown to 90% confluency in 10 cm cell-culture treated plates (Denville). Light SILAC cells are incubated with 5 mL of glucose-free RPMI media (Hyclone) for 30, 60, or 120 minutes. At the same time, fresh normal growth media is added to heavy SILAC cells to act as the basal metabolic control. Cells were scraped from plates, briefly washed, and resuspended in PBS with EDTA-free protease inhibitor tablet (Roche) and phosphatase inhibitor cocktail 3 (Sigma-Aldrich) before being divided into eight equal aliquots. Each aliquot of live cells was exposed to a steady temperature between 25°C to 70°C for 3 minutes in parallel, incubated at 25°C for another 3 minutes, and lysed by rapid freeze-thawing. Insoluble proteins and cell debris were removed by centrifugation at 17,000*g* for 10 minutes. The isolated supernatant from different temperature fractions were combined, then small amount of combined supernatant was used for protein quantification by Bradford's assay before light and heavy fractions are mixed at 1:1 ratio. The combined SILAC lysates was denatured in 8 M urea, followed by disulfide reduction with DTT (10 mM, 30 minutes, 65°C), alkylation (iodoacetamide, 15 mM, 30 min, room temperature, protected from light) and quenching (DTT, 5mM, 10 minutes, room temperature). The proteome solution was diluted 4-fold with ammonium bicarbonate solution (50 mM, pH 8.0), $CaCl_2$ added (1 mM) and digested with sequencing grade trypsin (~1:100 enzyme/protein ratio; Thermo Pierce)

at 37 ˚C while rotating overnight. Peptide digestion reactions were stopped by acidification to pH 2-3 with 1% formic acid, and peptides were then desalted with Sep-Pak tC18 cartridges (50 mg, Waters), dried under vacuum, resuspended with LC-MS grade water (Sigma-Aldrich) and then lyophilized.

TMT Labeling

Lyophilized peptides were labeled with 10-plex isobaric tandem mass tags (90406, Thermo Scientific) according to manufacturer's protocol with slight modification. TMT reagents were reconstituted to 8 mg/mL in anhydrous acetonitrile (Sigma-Aldrich) and added to lyophilized peptides dissolved in 100 µL of 200 mM HEPES buffer, pH 8.0 (~8:1 reagent/peptide ratio). Labeling reaction was carried out at room temperature for 1 hour with gentle shaking and quenched with 5 µL of 5% hydroxylamine (Thermo Scientific). Labeled peptides were combined into a single pool per experiment, acidified with formic acid (pH 2-3), desalted using ZipTip C18 tips (100 µL, Millipore), and lyophilized. The final processed peptides were dissolved in LC-MS/MS Buffer A (H_2O with 0.1% formic acid, LC-MS grade, Sigma-Aldrich) for LC-MS/MS analysis.

High-pH Reverse Phase Fractionation

High-pH reverse phase fractionation was performed on both SILAC PISA sample and TMT-labeled TPP samples. Dried peptides were reconstituted in 300 µL of 0.1% TFA in H_2O, high-pH reverse phase spin-columns (Pierce/ThermoFisher) were equilibrated, and samples fractionated per manufacturer's instructions into 8 fractions, 2 washes and a flow-through fraction (11 total). Only 8 fractions were used for LC-MS/MS analysis

Proteomic LC-MS/MS and Data Analysis

LC-MS/MS experiments were performed with an Easy-nLC 1000 ultra-high pressure LC system (ThermoFisher) using a PepMap RSLC C18 column (75 μm x 50 cm; 3 μm, 100 Å, Thermo Scientific) heated to 40˚C coupled to a Q Exactive HF orbitrap and Easy-Spray nanosource (ThermoFisher). TMT-labeled digested peptides in MS/MS Buffer A were injected onto the column and separated using the following gradient of buffer B (0.1% formic acid in acetonitrile) at 300 nL/min: 0-10% buffer B in 5 minutes, 10-40% buffer B in 240 minutes, 40-90% buffer B over 6 minutes, and hold at 90% for 20 minutes. MS/MS spectra were collected from 0 to 250 minutes using a data-dependent, top-10 ion setting with the following settings: full MS scans were acquired at a resolution of 120,000, scan range of 375-1500 m/z, maximum IT of 60 ms, AGC target of 1e6, and data collection in profile mode. MS2 scans was performed by HCD fragmentation with a resolution of 60,000, AGC target of 1e5, maximum IT of 60 ms, NCE of 30, and data type in centroid mode. Isolation window for precursor ions was set to 1.0 m/z with an underfill ratio of 0.5%. Peptides with charge state 1 and undefined were excluded and dynamic exclusion was set to 20 seconds. Furthermore, S-lens RF level was set to 60 with a spray voltage value of 2.60kV and ionization chamber temperature of 300˚C.

MS2 files were generated and searched using the ProLuCID algorithm in the Integrated Proteomics Pipeline (IP2) software platform. Human proteome data were searched using a concatenated target/decoy UniProt database (UniProt_Human_reviewed_04-10-2017.fasta). Basic searches were performed with the following search parameters: HCD fragmentation method; monoisotopic precursor ions; high resolution mode (3 isotopic peaks); precursor mass range 600-6,000 and initial fragment tolerance at 600 p.p.m.; enzyme cleavage specificity at C-terminal lysine and arginine

residues with 3 missed cleavage sites permitted; static modification of +57.02146 on cysteine (carboxyamidomethylation), +229.1629 on N-terminal and lysine for TMT-10-plex tag; 4 total differential modification sites per peptide, including oxidized methionine (+15.9949), and phosphorylation (+79.9663) on serine, threonine, and tyrosine (only for phospho-enriched samples); primary scoring type by XCorr and secondary by Zscore; minimum peptide length of six residues with a candidate peptide threshold of 500. A minimum of one peptide per protein and half-tryptic peptide specificity were required. Starting statistics were performed with a Δmass cutoff = 10 p.p.m. with modstat, and trypstat settings. False-discovery rates of peptide (sfp) were set to 1%. TMT quantification was performed using the isobaric labeling 10-plex labeling algorithm, with a mass tolerance of 5.0 p.p.m. or less in cases where co-eluting peptide interfere. Reporter ions 126.127726, 127.124761, 127.131081, 128.128116, 128.134436, 129.131417, 129.13779, 130.134825, 130.141145, and 131.13838 were used for relative quantification. In general all quantified peptides have mass error within 3 p.p.m..

Melting Curve Analyses

For the bulk proteome, TMT reporter ion intensities of all unmodified peptides mapped to the same protein were combined and fold change values relative to the lowest temperature condition were calculated. To generate protein melt curves, relative fold changes as a function of temperature was fitted to the equation derived from the chemical denaturation theory using R. T_m values were calculated at which the sigmoidal curve crosses the 0.5 fold change level. Only T_m values calculated from melting curves with curve $R^2 > 0.8$ were used in subsequent analyses. Shift in T_m values (e.g. ΔT_m) induced by phosphorylation were determined by subtracting the T_m of the

unmodified protein from the T_m of the phosphomodiform: $\Delta T_m = T_{m\_}$(-)-glucose - $T_{m_}$basal. ΔT_ms could only be calculated for proteins found in both basal metabolic and (-)-glucose conditions.

CHAPTER 5

Summary and Perspective

5.1 Summary

Protein post-translational modification represents an important mechanism for cells to regulate biological processes and rapidly respond to stimuli. Through electronic and steric effects, PTMs impact the chemical and biophysical properties of proteins to alter their intra- and intermolecular interactions. Technological advances in the field of LC-MS/MS-based quantitative proteomics have enabled massively parallel detection of PTMs in the past few decades. Yet a central question remains at the core of PTM studies – how these small changes are altering protein function and leading to phenotypic responses in the cell. Conventional biochemical approaches for PTM characterization are usually confined to individual proteins and isolated pathways, and therefore are labourious and low throughput in nature. With the ever-expanding gap between the number of novel PTM sites detected by LC-MS/MS approaches and the number of sites that have been functionally validated, we lack general methods with which to predict, or even prioritize, which modification sites are likely to be functional in the proteome.

In Chapter 2 of this thesis, I described the development of a quantitative proteomic method, termed "Hotspot Thermal Profiling" (HTP)[149], to enable the unbiased, global, and high-throughput measurements of protein stability changes in response to site-specific phosphorylation events in live cells (Figure 5.1A). We hypothesized that 1) phosphorylation-induced thermal stability shifts may serve as a readout on the biophysical and functional effect of a phosphosite on the protein of interest, and 2) this information can help identify modification marks with functional significance.

Using HTP, we generated steady-state thermal stability profiles of thousands of bulk, unmodified proteoforms and phosphomodiforms. This dataset provided site-specific information about the biophysical properties impacted by phosphorylation. HTP rediscovered a well-known phosphorylation-mediated protein-protein interaction between translation initiation factor proteins 4EBP1 and eIF4E, in which a large negative ΔT_m indicated the unbound form of hyperphosphorylated-4EBP1. HTP also enabled the discovery two previously uncharacterized functional PTMs. In cytoskeleton protein Vinculin, we revealed that phospho-S721 may impact the intramolecular interactions in Vinculin, resulting in a conformational shift and phenotypical change. In glycolytic enzyme GAPDH, we revealed S210 may impact protein-metabolite interactions, resulting in altered substrate occupancy of the protein.

Next, in Chapter 3, we discussed modifications to the original HTP workflow to increase the depth of (phospho)proteome coverage, precision in sample processing, LC-MS/MS measurement precision, and data interpretation. We termed these method variants "Label-Enrich"- and "Label-Fractionate-Enrich"-HTP. The resulting datasets generated from the two HTP variant workflows support the notion that more cost- and resource-intensive protocols for deeper profiling should result in higher precision within runs. In particular, the LFE-HTP method resulted in significantly increased proteome coverage and T_m measurements reproducibility due to the addition of a fractionation step and multi-notch MS3 acquisition resulting in 22 LC-MS/MS runs. In addition, both datasets revealed that, via anecdotal proteins and phosphomodiforms detected in the original HTP dataset, highly abundant proteins could be still be reliably detected with highly reproducible T_m curves regardless of which method was used. We also explored how different biochemical processing and statistical analyses may affect the global T_m profiles. We discovered

that different data analysis pipelines and statistical treatments – not upfront biochemical workflow – can affect the significance threshold of ΔT_m values.

Lastly, in Chapter 4 we explored the application of thermal profiling strategies to profile proteomes under different metabolic states. Using HTP workflow modified with a fractionation step to reduce sample complexity and increase proteome coverage, we measured thousands of protein T_m's under acute glucose-withdrawal. The dataset also showed that T_m melting curves data can be converted to an area-under-curve data output to capture proteins that would normally be missed in using stringent curve fitting strategies. We next developed a modified version of the original PISA method[187] that we called SILAC-PISA to investigate the thermal stabilities of proteomes under acute-glucose withdrawal in a kinetic series (Figure 5.1B). The resultant dataset revealed the time-dependent shifts in thermal stabilities of thousands of proteins, including in glycolytic enzymes known to interact with glucose metabolites. The SILAC-PISA demonstrated a more straightforward data processing workflow as well as the higher multiplexing capabilities of the method.

We envision that the extension of HTP methods to diverse organisms, cell types, PTMs, and kinetic states will greatly expand our current understanding of the role of protein modifications in regulating protein structure, function and signal transduction.

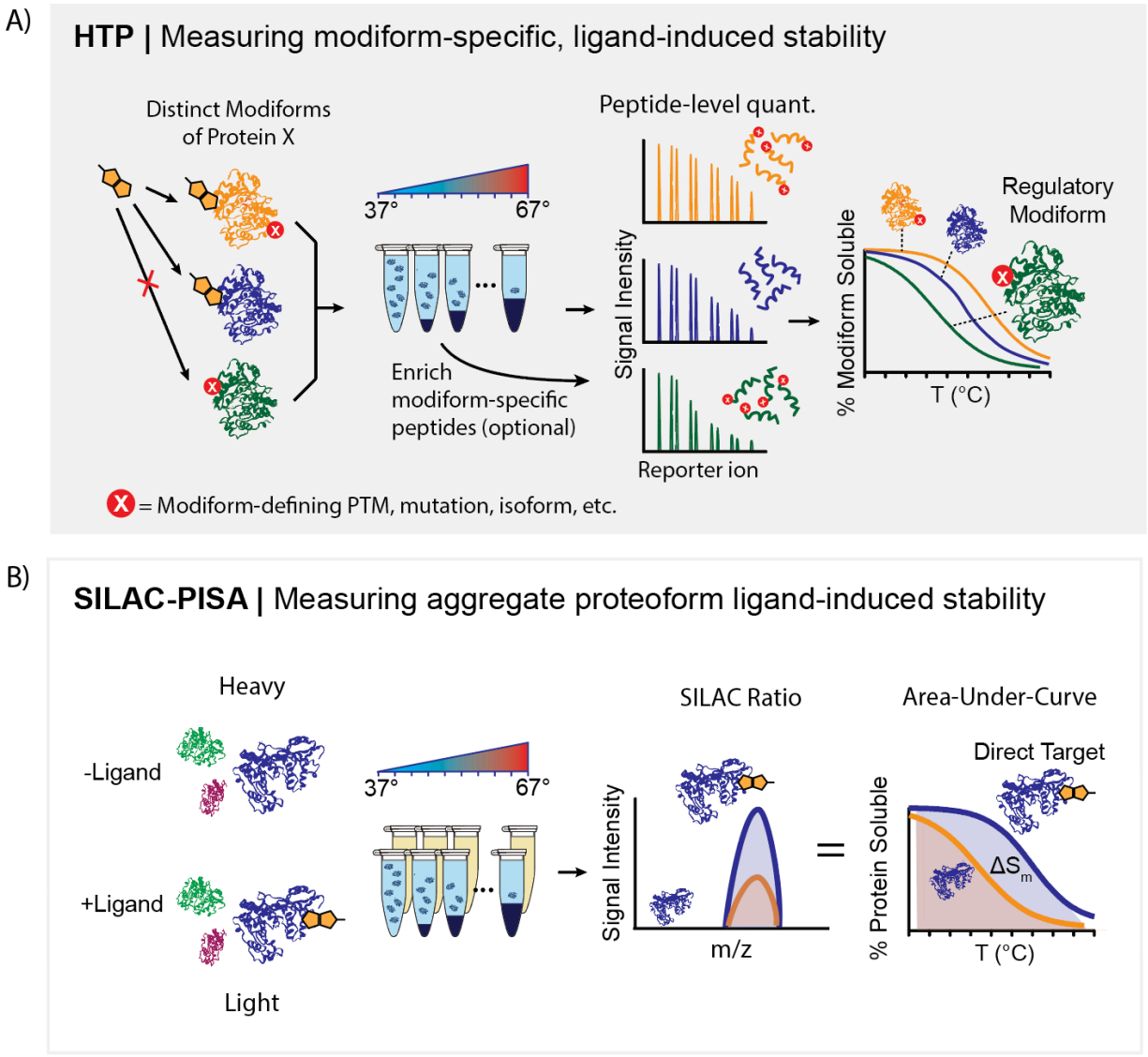

Figure 5.1 Schematics of thermal stabilities profiling approaches described in this thesis
HTP enables the measurement of modiform-specific thermal stability perturbations in melting temperature shifts **(A)**, meanwhile SILAC-PISA (**B**) collapses the T_m information into a one-dimensional SILAC ratio value to inform on the shifts in protein thermal stability.

5.2 Perspectives and Future Directions: integrating interaction profiling methods to elucidate PTM-mediated signaling mechanisms

LC-MS/MS-based proteomics strategies ushered in a new era of massively paralleled detection and discovery of proteome information. This unbiased, global view provides a comprehensive examination of biomolecular interactions in the proteome in response to stimuli[193]. The functional readouts can be as simple as abundance changes of proteins and PTMs, which is used in more traditional and straightforward protein identification and quantitative proteomics studies, and as complex as biophysical property changes such as shifts in protein melting temperature or solubility. In this section, we will discuss future directions in integrating different interaction profiling methods for the unbiased discovery of proteins interactions, functional PTMs, and pathways, to provide a unique and dynamic view into the mechanistic inner workings of cellular signaling networks for both basic and translational biological studies.

5.2.1 Orthogonal strategies to complement proteome thermal stability profiling of PTMs

Thermal stability-based assays measure protein melting temperature or solubility shifts to report on complex biophysical property changes to report on biomolecular interactions[117, 119, 150]. In cellular thermal shift-based assays (CETSA), changes in protein melting temperature T_m are used to indicate functional perturbations induced by protein-biomolecular interactions. While these thermal stability shift-based strategies have been successfully employed to interrogate protein-drug[120, 194], protein-protein[147], protein-metabolite[129], and intramolecular interactions[149], there are caveats to the technique and thus there is great value in applying orthogonal assays in conjunction to best capture protein stability shifts information in the cell.

The first caveat with the classic thermal shift assay is that the method best captures proteins that undergo two-state unfolding model and therefore may miss or overlook proteins that do not display prototypical melting behaviours with a sigmoidal protein melting curve. Orthogonal approaches that do not rely on sigmoidal curve fitting could be used in parallel to complement CETSA-style proteome thermal stability screenings. One of such approach is the proteome integral stability alteration (PISA) assay which is based on a similar premise that proteins will unfold and aggregate upon heating[187]. The advantage that PISA assay has over CETSA is that it does not require sigmoidal curve fitting, since the assay takes into account the entire melting curve and uses total area under the protein melting curve (AUC) as a signal readout. PISA can help capture proteins that have atypical melting behaviour that would otherwise be missed by the more stringent sigmoidal curve fit. This advantage has been demonstrated in Chapter 4 in which the modified PISA assay developed, called SILAC-PISA, was able to capture more S_m than T_m values in the same experiment (Figure 4.3). Because of multiplexing capability of the assay, PISA can be used as a higher throughput technique to screen through kinetic conditions to provide targets identification before the "higher resolution" approach of CETSA is used to further characterize protein melting behaviour.

A second category of orthogonal strategies to complement thermal shift profiling approaches consists of methods that are capable of disrupting protein stability without heat. In principle, any method that affect the molecular attribute of protein structure or chemical state could be used to detect altered protein interactions in response to biomolecular interactions. Similar to heat-induced protein unfolding, chemical denaturants can disrupt proteome stability and have been mainly used to interrogate ligand-induced protein stabilization *in vitro*[195-198]. Unlike thermal stability shift approaches which cannot extract any thermodynamic parameters beyond the T_m

value, chemical denaturation has the added bonus of enabling thermodynamic parameters (e.g., ΔG_f, $\Delta\Delta G_f$, and K_d values) to be derived since the relationship between the chemical denaturant-induced unfolding properties of a protein and it's folding free energy is well established[199]. One of such approach termed chemical denaturation and protein precipitation (CPP) uses increasing concentrations of guanidium chloride (GdmCl) to denature the proteome, followed by removal of precipitated proteins and quantitative proteomic analysis to generate protein denaturation curves to find ligand-induced shifts in protein's folding equilibrium[196]. Similarly, the use of organic solvent as a chemical denaturant has also been developed[197]. Because of similarity in sample processing between chemical and heat denaturation approaches, we posit that chemical denaturation methods can be applied alongside thermal profiling techniques such as HTP to interrogate and validate PTM-induced stability shifts in proteins.

The last category of orthogonal strategies consists of methods that take advantage of increased protease resistance of protein surfaces that are engaged in biomolecular interactions[147, 200-202]. Drug affinity response sensitivity (DARTS)[201, 202] and limited proteolysis-mass spectrometry (LiP-MS)[147] both measures the kinetics of proteolytic cleavage in whole proteome, and compare between conditions to identify proteins that are likely engaged in differential interactions (Figure 5.2). DARTS has been successfully used to identify a novel protein target for the metabolite α-ketoglutarate[203]. Meanwhile, LiP-MS expands upon the fundamental principles of DARTS and utilizes LC-MS/MS to enable mass detection of biomolecular-induced protein stabilizations[148]. In LiP-MS, the nonspecific protease Proteinase K is incubated with cell lysate after treatment with a small molecule of interest. The binding of ligands can locally alter the protease accessibility of its target proteins and protects the binding interfaces from non-specific protease digestion. A secondary proteolysis step using trypsin is performed which digests binding

interface regions into peptide sequences with recognizable terminal residues. Tryptic peptide abundances are then measured by LC-MS/MS which yields structural fingerprints for protein surfaces that interact with the ligand of interest. LiP-MS was used to identify more than 1,000 metabolite-protein interactions in *E. coli* and more than 7,000 putative binding sites involved in those interactions[148]. We posit that limited proteolysis strategies can be applied to the discovery of functional protein modifications because the presence of PTM on a daughter peptide may represent an PTM-induced binding event that limited protease accessibility to the target protein. Utilizing LiP-MS alongside HTP could generate both protein stabilities and structural elements data in PTM studies.

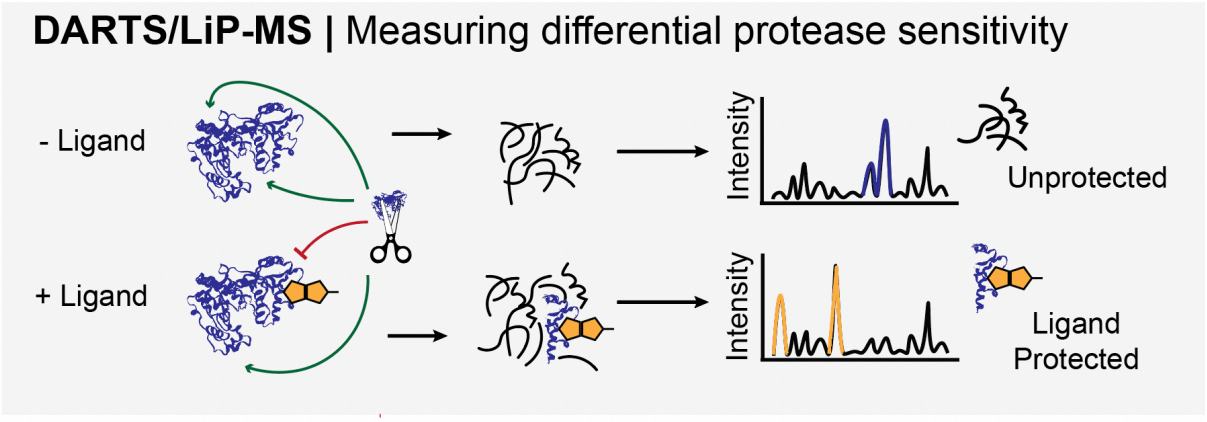

Figure 5.2 Schematics of limited-proteolysis profiling approaches
Ligand binding in lysates leads to stabilization or protection of specific protein surfaces against proteolysis. Quantitative LC-MS/MS detection and peptide-level mapping can identify proteins and specific sites that are protected, and thus candidate ligand binding sites. (DARTS, drug affinity response sensitivity; LiP-MS, limited proteolysis-mass spectrometry). Figure is adapted from [193].

5.2.2 Proximity profiling strategies to elucidate local protein networks

Global profiling approaches are a powerful first-stop shop in hypothesis-generating experiments as they provide an unbiased view of network interactions in the proteome. These so-called "top-down" profiling approaches we describe in this Chapter are assays such as MS-CETSA that can observe both direct and indirect interactions in the proteome in a massively parallel manner. Despite the large scale of information that can be generated using these global profiling methods, it is very difficult to differentiate direct and indirect targets of biomolecular interactions and these datasets do not provide detailed mechanistic insight into the local protein networks. We posit that the next step in omic-based profiling experiments, after the first "top-down" hypothesis-generating dataset is completed, is to perform "bottom-up" approaches in parallel to enable a comprehensive interrogation of targeted molecular interactions of interest[193].

We describe a "bottom-up" approach as methods that zoom in on a more localized view of interaction networks that originates from a protein or PTM site of interest[193]. Early examples of such "bottom-up" approaches are intracellular proximity labeling strategies such as BioID[204, 205] and TurboID[206] (Figure 5.3). Both employ an engineered biotin ligase fused to a protein of interest which generates a biotinoyl-50-AMP product that can covalently label surface amines on proximal proteins. An updated variant is the APEX system[207-209] which utilizes an engineered fusion enzyme comprised of horseradish peroxidase and ascorbic acid peroxidase to convert an exogenous biotin phenol probe into a reactive phenoxyl radical in the presence of heme and H_2O_2 to label proximal proteins at a distance shorter than BioID and Turbo ID (Figure 5.3). Lastly, a more recent example of proximity protein labeling that provides an even shorter diffusion distance is the light-activated photoproximity protein interaction (PhotoPPI) profiling method[210]. In PhotoPPI, cells expressing the SNAP-Tag-fusion protein-of-interest are treated with the photoproximity probe. Upon light

excitation, dual photocleavage events happen which release a reactive carbene from the probe to covalently label proximal proteins (Figure 5.3). This modular system enables the identification of direct protein-binding partners and interrogation of local interaction networks from the view of the protein of interest and provides one of shortest diffusion distance of labeling methods. These three technologies offer different experimental properties, and we posit that once a protein or PTM target of interest is identified from a "top-down" hypothesis generating approach such as HTP, proximity labeling strategies that offer varied intracellular distances could be applied in parallel to provide a mechanistic insight from the protein level and differentiate between direct and indirect interactions.

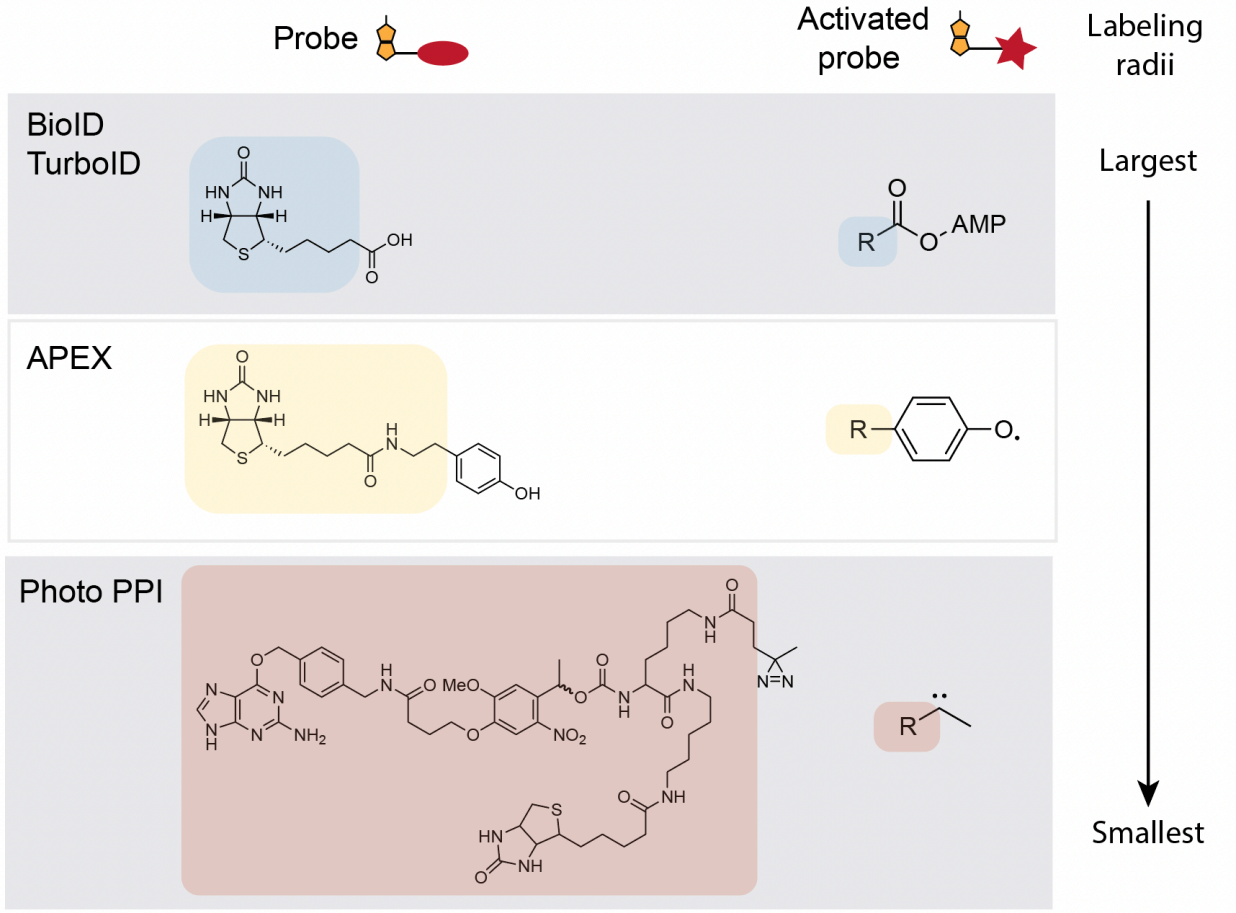

Figure 5.3 Schematics of "bottom-up" proximity profiling methods
Overview of intracellular proximity methods BioID/TurboID, APEX, and Photo PPI that rely on production of a diffusible reactive probe. Figure is adapted from [193].

Lastly, we propose that integrating the two layers of profiling approaches – "top-down" and "bottom-up" – in a kinetic series of experiments could enable tracking of network cascades in response to signaling events. For example, kinetic monitoring of protein or phosphomodiform T_m shifts using HTP could enable proteome-wide tracking of altered and progressively changing biophysical interactions that originate at sites of direct molecular interactions and subsequently flow outward within protein-protein and other protein-biomolecule interaction networks[193] (Figure 5.4). Differentiating between direct and indirection interactions is a current bottleneck in the analysis of global "top-down" datasets. We suggest that "bottom-up" approaches such as proximity labeling methods should be used after identification of candidates with progressively altering responses in "top-down" experiments to help differential between direct and secondary targets as well as to validate interactions. Integrating these multi-omic interaction profiling techniques could significantly enhance the mechanistic understanding of proteins and pathways that are impacted by biological signals.

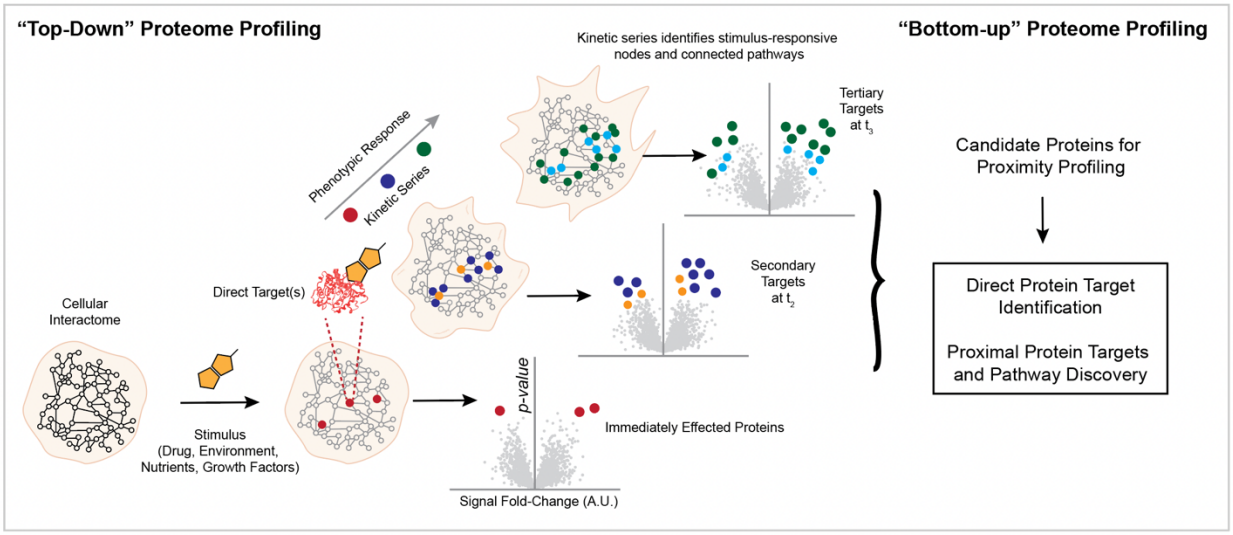

Figure 5.4 Schematic depiction of integrating interaction profiling strategies in a kinetic series to discover protein targets and signaling networks

The direct (red) and secondary (blue) targets of the stimulus (in this case a bioactive small molecule) would theoretically be the most significantly altered proteins in "top-down" datasets in early timepoints, as depicted in hypothetical volcano plots. Subsequent timepoints can capture signal propagation as interactions ripple outwards through affected protein networks and pathways. "Bottom-up" proteome profiling approaches can then be applied in parallel to validate direct protein targets and discover proximal pathways. Figure is adapted from [193].

5.2.3 Top-down proteomics to interrogate thermal stability of intact proteoforms

Top-down proteomics (not to be confused with "top-down" profiling techniques described in the previous section) is an emerging field that directly analyzes intact proteins using mass spectrometry[211-214]. It offers a unique type of proteoform identity information that the more commonly used bottom-up proteomics methods cannot provide. In bottom-up proteomics, proteins are digested by protease to generate short peptide segments which are then directly analyzed by LC-MS/MS to infer protein identification (Figure 5.5). Because peptides from different proteoforms are mixed and matched in a bottom-up experiment, it is impossible to reconstruct the complete proteoform from these peptide measurements, and information about the unique combination and co-occurrence of PTMs on a protein is lost. On the other hand, top-down proteomics analyze intact proteins that have not been digested, which allows for the direct detection of specific proteoforms including the location and identity of PTMs[215] (Figure 5.5).

Top-down proteomics is a still developing field with many analytical challenges. Low abundance of many proteoforms, low signal-to-noise ratios, low proteome coverage and resolution due to mass spectrometry instrumentation limitations, and challenging data analysis pipeline due to the complexity of intact proteoform MS data are current bottlenecks in this field. Recent developments with better fractionation and separation techniques[216, 217] have been combined to reduce the complexity of samples delivered to the mass spectrometer to enhance detection of low abundance proteoforms. Hardware innovations on mass spectrometer side such as the Field Asymmetric Ion Mobility Spectrometry (FAIMS)[218] that enables multidimensional on-line fractionation has been used in top-down proteomics studies to improve sign-to-noise ratio and proteoform coverage[219-221]. A recent study has identified more than 1,000 unique proteoforms in $CD3^+$ T cells using a combination of technical improvements[219]. We previously discussed the

possibility of inferring co-occurring PTM information based on the degrees of thermal shifts from our HTP datasets in Chapter 2. With top-down proteomics, information about the combination and co-occurrence of PTMs can be easily determined as the method directly measures whole proteins. Although top-down proteomics is still at its infancy with many analytical challenges, we envision that our upfront HTP sample processing steps could be modified to suit a top-down proteomics workflow derive PTM-induced thermal stability shifts of intact proteoforms. Although the proteome coverage of top-down techniques is still limited compared to what the more sophisticated and robust bottom-up technologies can offer, thermal profiling of intact proteins can provide novel mechanistic insights about what PTMs and proteoforms have functional relevance in cellular biological processes.

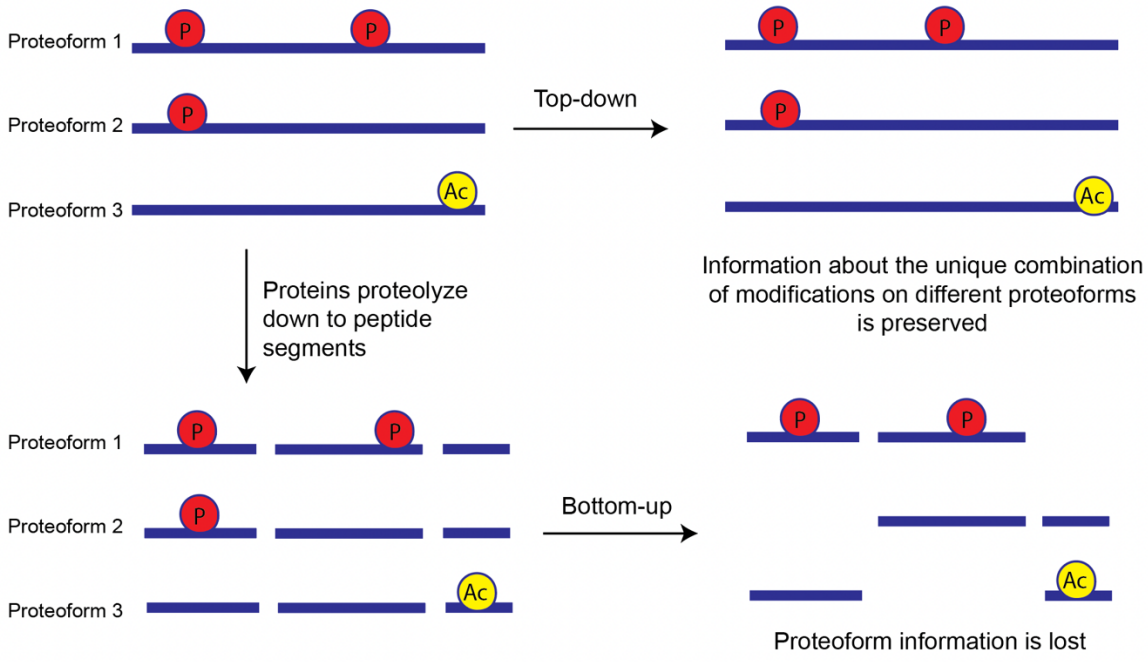

Figure 5.5 Differences between top-down and bottom-up proteomics methods
In bottom-up proteomics, proteins are proteyolyze down to peptide segments before analysis by LC-MS/MS. Information about which proteoform a peptide came from is lost. Meanwhile, in top-down proteomics, intact proteins are analyzed by LC-MS/MS and information about the unique combination of modifications on different proteoforms is preserved.

5.2.4 Integrating kinetic profiling strategies to derive mechanism of a novel PTM

PTMs are typically enzymatically installed on proteins, such as phosphorylation by kinases, ubiquitination by ubiquinases, and methylation by methylases. In Chapter 1, we introduced a novel class of enzyme independent PTMs which have been discovered in recent years called reactive metabolites modifications (rmPTM)[20, 21, 222]. These nonenzymatic modifications occur via spontaneous chemical reactions between reactive metabolites and nucleophilic residue in proteins. However, the scope and broad functional significance of nonenzymatic PTMs remain poorly understood.

One particular rmPTM of interest is the 3-phosphoglyceryl-lysine (pgK)[20]. This nonenzymatic modification originates from the reaction of lysine with primary glycolytic metabolite, 1,3-bisphosphate-glycerate (1,3-BPG), which is the end-product and substrate of glycolytic enzymes glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and phosphoglycerate kinase 1 (PGK1), respectively. The pgK modification has been detected in more than 200 proteins of diverse classes[20], but for most proteins, pgK is only detected on a single conserved lysine. The fact that the pgK modification is not indiscriminately modifying any surface lysine of nearby proteins underscores the high level of specificity and potential functional significance of this PTM. It has been shown that pgK modifications can change the structure and functions of proteins within glycolysis[20]. Also, as the level of 1,3-BPG changes due to metabolic flux, pgK modifies and accumulates on nearby glycolytic proteins to create a potential feedback loop, resulting in carbon flow rewiring to alternate biosynthetic pathways. Other central glycolytic metabolites have also been observed to impact metabolic pathways such as the Keap1-Nrf2 pathway[21]. Since glycolysis is a ubiquitous and ancient pathway as well as the fact that pgK modifications have been detected in non-glycolytic proteins, one has to ask what important biological function this modification

provides and if there are other pathways that also respond to this intrinsically reactive and highly conserved metabolite.

In order to begin to provide answers to these overarching questions, we previously coupled pgK-specific enrichment and isotopic labeling to profile pgKylation in *S. elongatus*. Using quantitative proteomics, we profiled the level pgKylation over a 24-hour period of light and dark cycles. Cyanobacteria *S. elongatus* contains the simplest circadian rhythm clock system which is defined as a self-sustained cycle that allows organisms to prepare for regular environmental changes such as light/dark or temperature changes occurring within 24-hour periods[223]. Circadian rhythms are widely observed in many different species of organisms, from plants to animals to bacteria, and are believed to have evolved at least 2.3 billion years ago[224]. Although cyanobacteria *S. elongatus* consists of a small proteome of ~2,400 proteins, more than 369 unique pgK sites were detected in our preliminary dataset (Figure 5.5A). The number of pgK found in the simple organism stands in stark contrast to the <150 pgK sites detected in human[20] and 280 phosphorylation sites detected in the bacteria[223]. Additionally, we tracked the levels of pgKylation over a 24-hour period of light and dark cycles and detected a subset of pgK sites follow a similar pattern attuned to the light and dark cycles (Figure 5.5B).

Given that the regulatory mechanism of the central circadian rhythm is not well defined, we wonder if the pgK modifications play role in these widely observed and evolutionarily conserved pathways. We propose that a kinetic series of *in vitro* HTP experiments can be performed on *S. elongatus* over a 24-hour period to elucidate the effect of pgKylation on proteome function and stability. Cell lysates are extracted from different timepoints can undergo HTP sample processing and analysis. We posit that the HTP platform can help discover functional pgK sites in the cyanobacteria to elucidate molecular mechanisms of pgK and their correlation to important

pathways such as circadian rhythm in the organism. We also envision that this approach can be applied to other classes of PTMs of interest in any organisms.

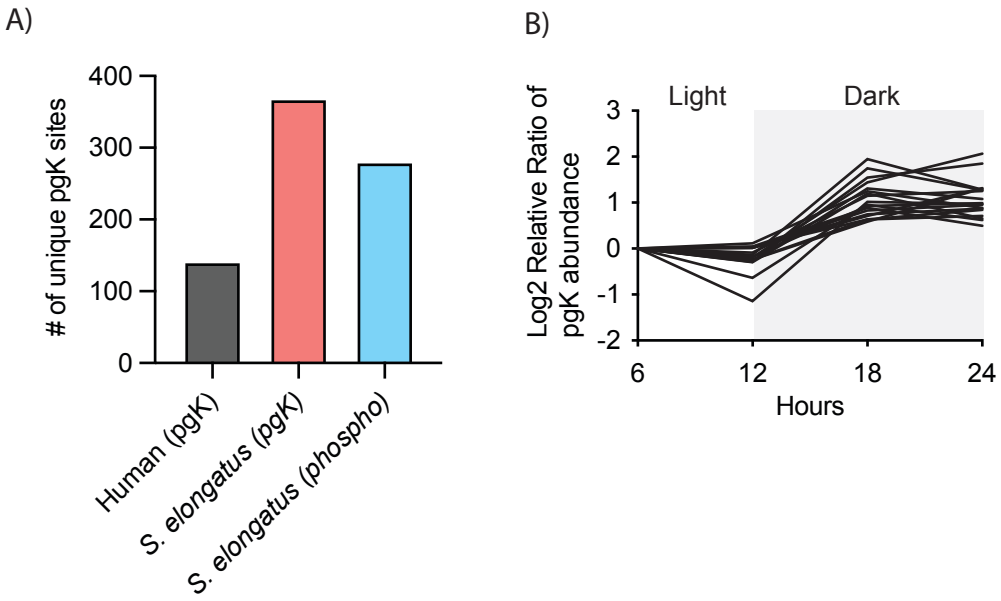

Figure 5.6 Global kinetic quantitative profiling of pgK sites in *S. elongatus*
A) Number of unique pgK modification sites detected in human and *S. elongatus* compared to the number of phosphorylation sites detected in *S. elongatus.* **B)** Quantitative profiling of pgK modification sites in *S. elongatus* over a 24-hour light and dark cycle shows that a subset of pgK sites have elevated abundance level during the dark periods.

List of References

1. Conibear, A.C., *Deciphering protein post-translational modifications using chemical biology tools.* Nature Reviews Chemistry, 2020. **4**(12): p. 674-695.
2. Aebersold, R., et al., *How many human proteoforms are there?* Nat Chem Biol, 2018. **14**(3): p. 206-214.
3. Su, M., et al., *Small proteins: untapped area of potential biological importance.* Front Genet, 2013. **4**: p. 286.
4. Walsh, C.T., S. Garneau-Tsodikova, and G.J. Gatto, Jr., *Protein posttranslational modifications: the chemistry of proteome diversifications.* Angew Chem Int Ed Engl, 2005. **44**(45): p. 7342-72.
5. Dominguez, R., *Actin filament nucleation and elongation factors--structure-function relationships.* Crit Rev Biochem Mol Biol, 2009. **44**(6): p. 351-66.
6. Dominguez, R. and K.C. Holmes, *Actin structure and function.* Annu Rev Biophys, 2011. **40**: p. 169-86.
7. Denhardt, D.T., *Signal-transducing protein phosphorylation cascades mediated by Ras/Rho proteins in the mammalian cell: the potential for multiplex signalling.* Biochem J, 1996. **318 (Pt 3)**: p. 729-47.
8. Magnaghi-Jaulin, L., S. Ait-Si-Ali, and A. Harel-Bellan, *Histone acetylation in signal transduction by growth regulatory signals.* Semin Cell Dev Biol, 1999. **10**(2): p. 197-203.
9. Johnson, C.A. and B.M. Turner, *Histone deacetylases: complex transducers of nuclear signals.* Semin Cell Dev Biol, 1999. **10**(2): p. 179-88.
10. Kim, J.W. and C.V. Dang, *Multifaceted roles of glycolytic enzymes.* Trends Biochem Sci, 2005. **30**(3): p. 142-50.
11. Yu, X. and S. Li, *Non-metabolic functions of glycolytic enzymes in tumorigenesis.* Oncogene, 2017. **36**(19): p. 2629-2636.
12. Aken, B.L., et al., *Ensembl 2017.* Nucleic Acids Res, 2017. **45**(D1): p. D635-D642.
13. Gaudet, P., et al., *The neXtProt knowledgebase on human proteins: 2017 update.* Nucleic Acids Res, 2017. **45**(D1): p. D177-D182.
14. Uhlen, M., et al., *Proteomics. Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.
15. Ponomarenko, E.A., et al., *The Size of the Human Proteome: The Width and Depth.* Int J Anal Chem, 2016. **2016**: p. 7436849.
16. Greenberg, D.S. and H. Soreq, *Alternative Splicing*, in *Brenner's Encyclopedia of Genetics (Second Edition)*, S. Maloy and K. Hughes, Editors. 2013, Academic Press: San Diego. p. 97-98.
17. Pan, Q., et al., *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing.* Nat Genet, 2008. **40**(12): p. 1413-5.
18. Wang, E.T., et al., *Alternative isoform regulation in human tissue transcriptomes.* Nature, 2008. **456**(7221): p. 470-6.
19. UniProt, C., *UniProt: the universal protein knowledgebase in 2021.* Nucleic Acids Res, 2021. **49**(D1): p. D480-D489.
20. Moellering, R.E. and B.F. Cravatt, *Functional lysine modification by an intrinsically reactive primary glycolytic metabolite.* Science, 2013. **341**(6145): p. 549-53.

21. Bollong, M.J., et al., *A metabolite-derived protein modification integrates glycolysis with KEAP1-NRF2 signalling.* Nature, 2018. **562**(7728): p. 600-604.
22. Keenan, E.K., D.K. Zachman, and M.D. Hirschey, *Discovering the landscape of protein modifications.* Mol Cell, 2021.
23. Levene, P. and C. Alsberg, *The cleavage products of vitellin.* Journal of Biological Chemistry, 1906. **2**(1): p. 127-133.
24. Carr, S.M., et al., *Post-translational control of transcription factors: methylation ranks highly.* FEBS J, 2015. **282**(23): p. 4450-65.
25. Song, L. and Z.Q. Luo, *Post-translational regulation of ubiquitin signaling.* J Cell Biol, 2019. **218**(6): p. 1776-1786.
26. Prabakaran, S., et al., *Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding.* Wiley Interdiscip Rev Syst Biol Med, 2012. **4**(6): p. 565-83.
27. Cheng, H.C., et al., *Regulation and function of protein kinases and phosphatases.* Enzyme Res, 2011. **2011**: p. 794089.
28. Beranek, V., et al., *Genetically Encoded Protein Phosphorylation in Mammalian Cells.* Cell Chem Biol, 2018. **25**(9): p. 1067-1074 e5.
29. Chang, J.W., et al., *Chemoproteomic Profiling of Phosphoaspartate Modifications in Prokaryotes.* Angew Chem Int Ed Engl, 2018. **57**(48): p. 15712-15716.
30. Fischer, E.H. and E.G. Krebs, *Conversion of phosphorylase b to phosphorylase a in muscle extracts.* J Biol Chem, 1955. **216**(1): p. 121-32.
31. Krebs, E.G. and E.H. Fischer, *The phosphorylase b to a converting enzyme of rabbit skeletal muscle.* Biochim Biophys Acta, 1956. **20**(1): p. 150-7.
32. Shieh, S.Y., et al., *DNA damage-induced phosphorylation of p53 alleviates inhibition by MDM2.* Cell, 1997. **91**(3): p. 325-34.
33. Hafner, A., et al., *The multiple mechanisms that regulate p53 activity and cell fate.* Nature Reviews Molecular Cell Biology, 2019. **20**(4): p. 199-210.
34. Aylon, Y. and M. Oren, *The Paradox of p53: What, How, and Why?* Cold Spring Harb Perspect Med, 2016. **6**(10).
35. Ardito, F., et al., *The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review).* Int J Mol Med, 2017. **40**(2): p. 271-280.
36. Druker, B.J., *Imatinib as a paradigm of targeted therapies.* Advances in cancer research, 2004. **91**(1): p. 1-30.
37. Marcucci, G., D. Perrotti, and M.A. Caligiuri, *Understanding the molecular basis of imatinib mesylate therapy in chronic myelogenous leukemia and the related mechanisms of resistance. Commentary re: A. N. Mohamed et al., The effect of imatinib mesylate on patients with Philadelphia chromosome-positive chronic myeloid leukemia with secondary chromosomal aberrations. Clin. Cancer Res., 9: 1333-1337, 2003.* Clin Cancer Res, 2003. **9**(4): p. 1248-52.
38. Wu, P., T.E. Nielsen, and M.H. Clausen, *FDA-approved small-molecule kinase inhibitors.* Trends Pharmacol Sci, 2015. **36**(7): p. 422-39.
39. Chrencik, J.E., et al., *Structural and thermodynamic characterization of the TYK2 and JAK3 kinase domains in complex with CP-690550 and CMP-6.* J Mol Biol, 2010. **400**(3): p. 413-33.
40. Pagel, O., et al., *Current strategies and findings in clinically relevant post-translational modification-specific proteomics.* Expert Rev Proteomics, 2015. **12**(3): p. 235-53.

41. Narita, T., B.T. Weinert, and C. Choudhary, *Functions and mechanisms of non-histone protein acetylation.* Nature Reviews Molecular Cell Biology, 2019. **20**(3): p. 156-174.
42. Choudhary, C. and M. Mann, *Sequencing of the First Draft of the Human Acetylome.* Clinical Chemistry, 2020. **66**(6): p. 852-853.
43. Verdin, E. and M. Ott, *50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond.* Nat Rev Mol Cell Biol, 2015. **16**(4): p. 258-64.
44. Kim, S.C., et al., *Substrate and functional diversity of lysine acetylation revealed by a proteomics survey.* Mol Cell, 2006. **23**(4): p. 607-18.
45. Kovacs, J.J., et al., *HDAC6 regulates Hsp90 acetylation and chaperone-dependent activation of glucocorticoid receptor.* Mol Cell, 2005. **18**(5): p. 601-7.
46. Cohen, T.J., et al., *An acetylation switch controls TDP-43 function and aggregation propensity.* Nat Commun, 2015. **6**: p. 5845.
47. Bubna, A.K., *Vorinostat-An Overview.* Indian J Dermatol, 2015. **60**(4): p. 419.
48. Gross, S., et al., *Targeting cancer with kinase inhibitors.* J Clin Invest, 2015. **125**(5): p. 1780-9.
49. Deng, L., et al., *The role of ubiquitination in tumorigenesis and targeted drug discovery.* Signal Transduct Target Ther, 2020. **5**(1): p. 11.
50. Hochstrasser, M., *Ubiquitin-dependent protein degradation.* Annu Rev Genet, 1996. **30**: p. 405-39.
51. Takahashi, T.S., et al., *Structural insights into two distinct binding modules for Lys63-linked polyubiquitin chains in RNF168.* Nat Commun, 2018. **9**(1): p. 170.
52. Harmel, R. and D. Fiedler, *Features and regulation of non-enzymatic post-translational modifications.* Nat Chem Biol, 2018. **14**(3): p. 244-252.
53. Wagner, G.R., et al., *A Class of Reactive Acyl-CoA Species Reveals the Non-enzymatic Origins of Protein Acylation.* Cell Metab, 2017. **25**(4): p. 823-837 e8.
54. Hirschey, M.D. and Y. Zhao, *Metabolic Regulation by Lysine Malonylation, Succinylation, and Glutarylation.* Mol Cell Proteomics, 2015. **14**(9): p. 2308-15.
55. Tan, M., et al., *Lysine glutarylation is a protein posttranslational modification regulated by SIRT5.* Cell Metab, 2014. **19**(4): p. 605-17.
56. Xie, L., et al., *Proteome-wide Lysine Glutarylation Profiling of the Mycobacterium tuberculosis H37Rv.* J Proteome Res, 2016. **15**(4): p. 1379-85.
57. Baird, L., et al., *Regulatory flexibility in the Nrf2-mediated stress response is conferred by conformational cycling of the Keap1-Nrf2 protein complex.* Proc Natl Acad Sci U S A, 2013. **110**(38): p. 15259-64.
58. Kobayashi, M. and M. Yamamoto, *Nrf2-Keap1 regulation of cellular defense mechanisms against electrophiles and reactive oxygen species.* Adv Enzyme Regul, 2006. **46**: p. 113-40.
59. Lo, S.C., et al., *Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling.* EMBO J, 2006. **25**(15): p. 3605-17.
60. Nguyen, T., P.J. Sherratt, and C.B. Pickett, *Regulatory mechanisms controlling gene expression mediated by the antioxidant response element.* Annu Rev Pharmacol Toxicol, 2003. **43**: p. 233-60.
61. Olsen, J.V. and M. Mann, *Status of large-scale analysis of post-translational modifications by mass spectrometry.* Mol Cell Proteomics, 2013. **12**(12): p. 3444-52.
62. Barnouin, K., *Special issue in quantitative mass spectrometric proteomics.* Amino Acids, 2012. **43**(3): p. 1005-7.

63. Ho, C.S., et al., *Electrospray ionisation mass spectrometry: principles and clinical applications.* Clin Biochem Rev, 2003. **24**(1): p. 3-12.
64. Whitehouse, C.M., et al., *Electrospray interface for liquid chromatographs and mass spectrometers.* Anal Chem, 1985. **57**(3): p. 675-9.
65. Miladi, M., B. Harper, and T. Solouki, *Evidence for sequence scrambling in collision-induced dissociation of y-type fragment ions.* J Am Soc Mass Spectrom, 2013. **24**(11): p. 1755-66.
66. Washburn, M.P., D. Wolters, and J.R. Yates, 3rd, *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nat Biotechnol, 2001. **19**(3): p. 242-7.
67. Wolters, D.A., M.P. Washburn, and J.R. Yates, 3rd, *An automated multidimensional protein identification technology for shotgun proteomics.* Anal Chem, 2001. **73**(23): p. 5683-90.
68. Kalli, A., et al., *Evaluation and optimization of mass spectrometric settings during data-dependent acquisition mode: focus on LTQ-Orbitrap mass analyzers.* J Proteome Res, 2013. **12**(7): p. 3071-86.
69. Aebersold, R. and M. Mann, *Mass-spectrometric exploration of proteome structure and function.* Nature, 2016. **537**(7620): p. 347-55.
70. Wu, C.C. and M.J. MacCoss, *Shotgun proteomics: tools for the analysis of complex biological systems.* Curr Opin Mol Ther, 2002. **4**(3): p. 242-50.
71. Bateman, N.W., et al., *Maximizing peptide identification events in proteomic workflows using data-dependent acquisition (DDA).* Mol Cell Proteomics, 2014. **13**(1): p. 329-38.
72. Zhao, Y. and O.N. Jensen, *Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques.* Proteomics, 2009. **9**(20): p. 4632-41.
73. Nishida, Y., et al., *SIRT5 Regulates both Cytosolic and Mitochondrial Protein Malonylation with Glycolysis as a Major Target.* Mol Cell, 2015. **59**(2): p. 321-32.
74. Pinkse, M.W., et al., *Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns.* Anal Chem, 2004. **76**(14): p. 3935-43.
75. Ke, M., et al., *Identification, Quantification, and Site Localization of Protein Posttranslational Modifications via Mass Spectrometry-Based Proteomics.* Adv Exp Med Biol, 2016. **919**: p. 345-382.
76. Vocadlo, D.J., et al., *A chemical approach for identifying O-GlcNAc-modified proteins in cells.* Proc Natl Acad Sci U S A, 2003. **100**(16): p. 9116-21.
77. Olsen, J.V., et al., *Global, in vivo, and site-specific phosphorylation dynamics in signaling networks.* Cell, 2006. **127**(3): p. 635-48.
78. Matsuoka, S., et al., *ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage.* Science, 2007. **316**(5828): p. 1160-6.
79. Olsen, J.V., et al., *Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis.* Sci Signal, 2010. **3**(104): p. ra3.
80. Deng, J., H. Erdjument-Bromage, and T.A. Neubert, *Quantitative Comparison of Proteomes Using SILAC.* Current Protocols in Protein Science, 2019. **95**(1): p. e74.
81. Li, J., et al., *TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples.* Nat Methods, 2020. **17**(4): p. 399-404.

82. Thompson, A., et al., *TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set.* Anal Chem, 2019. **91**(24): p. 15941-15950.
83. Thompson, A., et al., *Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.* Anal Chem, 2003. **75**(8): p. 1895-904.
84. Rose, C.M., et al., *Highly Multiplexed Quantitative Mass Spectrometry Analysis of Ubiquitylomes.* Cell Syst, 2016. **3**(4): p. 395-403 e4.
85. Aggarwal, S., et al., *Post-translational Modification Crosstalk and Hotspots in Sirtuin Interactors Implicated in Cardiovascular Diseases.* Front Genet, 2020. **11**: p. 356.
86. PhosphositePlus.
87. Wesseling, H., et al., *Tau PTM Profiles Identify Patient Heterogeneity and Stages of Alzheimer's Disease.* Cell, 2020. **183**(6): p. 1699-1713 e13.
88. Graff, J. and L.H. Tsai, *Histone acetylation: molecular mnemonics on the chromatin.* Nat Rev Neurosci, 2013. **14**(2): p. 97-111.
89. Shahbazian, M.D. and M. Grunstein, *Functions of site-specific histone acetylation and deacetylation.* Annu Rev Biochem, 2007. **76**: p. 75-100.
90. Perez-Mejias, G., et al., *Exploring protein phosphorylation by combining computational approaches and biochemical methods.* Comput Struct Biotechnol J, 2020. **18**: p. 1852-1863.
91. Hunter, T., *Why nature chose phosphate to modify proteins.* Philos Trans R Soc Lond B Biol Sci, 2012. **367**(1602): p. 2513-6.
92. Elam, W.A., et al., *Phosphomimetic S3D cofilin binds but only weakly severs actin filaments.* J Biol Chem, 2017. **292**(48): p. 19565-19579.
93. Lamper, A.M., et al., *A phosphorylation-regulated eIF3d translation switch mediates cellular adaptation to metabolic stress.* Science, 2020. **370**(6518): p. 853-856.
94. He, W., L. Wei, and Q. Zou, *Research progress in protein posttranslational modification site prediction.* Brief Funct Genomics, 2018. **18**(4): p. 220-229.
95. Torres, M.P., H. Dewhurst, and N. Sundararaman, *Proteome-wide Structural Analysis of PTM Hotspots Reveals Regulatory Elements Predicted to Impact Biological Function and Disease.* Mol Cell Proteomics, 2016. **15**(11): p. 3513-3528.
96. Su, M.-G., et al., *Investigation and identification of functional post-translational modification sites associated with drug binding and protein-protein interactions.* BMC Systems Biology, 2017. **11**(7): p. 132.
97. O'Sullivan, C., Tompson, F.W.,, *LX.—Invertase: a contribution to the history of an enzyme or unorganised ferment. Journal of the Chemical Society, Transactions 57, 834–931.. doi:10.1039/ct8905700834.* 1890.
98. Brandts, J.F. and L.N. Lin, *Study of strong to ultratight protein interactions using differential scanning calorimetry.* Biochemistry, 1990. **29**(29): p. 6927-40.
99. Crothers, D.M., *Statistical thermodynamics of nucleic acid melting transitions with coupled binding equilibria.* Biopolymers, 1971. **10**(11): p. 2147-60.
100. Matulis, D., et al., *Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor.* Biochemistry, 2005. **44**(13): p. 5258-66.
101. Niesen, F.H., H. Berglund, and M. Vedadi, *The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability.* Nat Protoc, 2007. **2**(9): p. 2212-21.

102. Schwarz, F.P., *Interaction of cytidine 3'-monophosphate and uridine 3'-monophosphate with ribonuclease a at the denaturation temperature.* Biochemistry, 1988. **27**(22): p. 8429-36.
103. Pace, C.N. and T. McGrath, *Substrate stabilization of lysozyme to thermal and guanidine hydrochloride denaturation.* J Biol Chem, 1980. **255**(9): p. 3862-5.
104. Bryngelson, J.D., et al., *Funnels, pathways, and the energy landscape of protein folding: a synthesis.* Proteins, 1995. **21**(3): p. 167-95.
105. Honig, B., *Protein folding: from the levinthal paradox to structure prediction.* J Mol Biol, 1999. **293**(2): p. 283-93.
106. Baker, D., *A surprising simplicity to protein folding.* Nature, 2000. **405**(6782): p. 39-42.
107. Rader, A.J., et al., *Protein unfolding: rigidity lost.* Proc Natl Acad Sci U S A, 2002. **99**(6): p. 3540-5.
108. Ericsson, U.B., et al., *Thermofluor-based high-throughput stability optimization of proteins for structural studies.* Anal Biochem, 2006. **357**(2): p. 289-98.
109. Pantoliano, M.W., et al., *High-density miniaturized thermal shift assays as a general strategy for drug discovery.* J Biomol Screen, 2001. **6**(6): p. 429-40.
110. Senisterra, G.A., et al., *Screening for ligands using a generic and high-throughput light-scattering-based assay.* J Biomol Screen, 2006. **11**(8): p. 940-8.
111. Vedadi, M., et al., *Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination.* Proc Natl Acad Sci U S A, 2006. **103**(43): p. 15835-40.
112. Garnett, M.J., et al., *Systematic identification of genomic markers of drug sensitivity in cancer cells.* Nature, 2012. **483**(7391): p. 570-5.
113. Fedorov, O., et al., *A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases.* Proc Natl Acad Sci U S A, 2007. **104**(51): p. 20523-8.
114. Wahlberg, E., et al., *Family-wide chemical profiling and structural analysis of PARP and tankyrase inhibitors.* Nat Biotechnol, 2012. **30**(3): p. 283-8.
115. Kopec, J. and G. Schneider, *Comparison of fluorescence and light scattering based methods to assess formation and stability of protein-protein complexes.* J Struct Biol, 2011. **175**(2): p. 216-23.
116. Dahlroth, S.L., et al., *Crystal structure of the shutoff and exonuclease protein from the oncogenic Kaposi's sarcoma-associated herpesvirus.* FEBS J, 2009. **276**(22): p. 6636-45.
117. Martinez Molina, D., et al., *Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay.* Science, 2013. **341**(6141): p. 84-7.
118. Jafari, R., et al., *The cellular thermal shift assay for evaluating drug target interactions in cells.* Nat Protoc, 2014. **9**(9): p. 2100-22.
119. Martinez Molina, D. and P. Nordlund, *The Cellular Thermal Shift Assay: A Novel Biophysical Assay for In Situ Drug Target Engagement and Mechanistic Biomarker Studies.* Annu Rev Pharmacol Toxicol, 2016. **56**: p. 141-61.
120. Almqvist, H., et al., *CETSA screening identifies known and novel thymidylate synthase inhibitors and slow intracellular activation of 5-fluorouracil.* Nat Commun, 2016. **7**: p. 11040.
121. Savitski, M.M., et al., *Tracking cancer drugs in living cells by thermal profiling of the proteome.* Science, 2014. **346**(6205): p. 1255784.
122. Reckzeh, E.S., et al., *Target Engagement of Small Molecules: Thermal Profiling Approaches on Different Levels.* Methods Mol Biol, 2019. **1888**: p. 73-98.

123. Kitagawa, M., et al., *Dual blockade of the lipid kinase PIP4Ks and mitotic pathways leads to cancer-selective lethality.* Nat Commun, 2017. **8**(1): p. 2200.
124. Miettinen, T.P., et al., *Thermal proteome profiling of breast cancer cells reveals proteasomal activation by CDK4/6 inhibitor palbociclib.* EMBO J, 2018. **37**(10).
125. Warpman Berglund, U., et al., *Validation and development of MTH1 inhibitors for treatment of cancer.* Ann Oncol, 2016. **27**(12): p. 2275-2283.
126. Leuenberger, P., et al., *Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability.* Science, 2017. **355**(6327).
127. Dziekan, J.M., et al., *Identifying purine nucleoside phosphorylase as the target of quinine using cellular thermal shift assay.* Sci Transl Med, 2019. **11**(473).
128. Huber, K.V., et al., *Proteome-wide drug and metabolite interaction mapping by thermal-stability profiling.* Nat Methods, 2015. **12**(11): p. 1055-7.
129. Sridharan, S., et al., *Proteome-wide solubility and thermal stability profiling reveals distinct regulatory roles for ATP.* Nat Commun, 2019. **10**(1): p. 1155.
130. LeDuc, R.D., et al., *ProForma: A Standard Proteoform Notation.* J Proteome Res, 2018. **17**(3): p. 1321-1325.
131. Smith, L.M., N.L. Kelleher, and P. Consortium for Top Down, *Proteoform: a single term describing protein complexity.* Nat Methods, 2013. **10**(3): p. 186-7.
132. Smith, L.M., et al., *A five-level classification system for proteoform identifications.* Nat Methods, 2019. **16**(10): p. 939-940.
133. McGowan, C.H. and P. Russell, *Human Wee1 kinase inhibits cell division by phosphorylating p34cdc2 exclusively on Tyr15.* EMBO J, 1993. **12**(1): p. 75-85.
134. Watanabe, N., M. Broome, and T. Hunter, *Regulation of the human WEE1Hu CDK tyrosine 15-kinase during the cell cycle.* EMBO J, 1995. **14**(9): p. 1878-91.
135. Blangy, A., et al., *Phosphorylation by p34cdc2 regulates spindle association of human Eg5, a kinesin-related motor essential for bipolar spindle formation in vivo.* Cell, 1995. **83**(7): p. 1159-69.
136. Gnad, F., J. Gunawardena, and M. Mann, *PHOSIDA 2011: the posttranslational modification database.* Nucleic Acids Res, 2011. **39**(Database issue): p. D253-60.
137. Gnad, F., et al., *PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites.* Genome Biol, 2007. **8**(11): p. R250.
138. Kallberg, M., et al., *Template-based protein structure modeling using the RaptorX web server.* Nat Protoc, 2012. **7**(8): p. 1511-22.
139. Wang, S., et al., *RaptorX-Property: a web server for protein structure property prediction.* Nucleic Acids Res, 2016. **44**(W1): p. W430-5.
140. Jimenez, J.L., et al., *A systematic comparative and structural analysis of protein phosphorylation sites based on the mtcPTM database.* Genome Biol, 2007. **8**(5): p. R90.
141. Beretta, L., et al., *Rapamycin blocks the phosphorylation of 4E-BP1 and inhibits cap-dependent initiation of translation.* EMBO J, 1996. **15**(3): p. 658-64.
142. Sekiyama, N., et al., *Molecular mechanism of the dual activity of 4EGI-1: Dissociating eIF4G from eIF4E but stabilizing the binding of unphosphorylated 4E-BP1.* Proc Natl Acad Sci U S A, 2015. **112**(30): p. E4036-45.
143. Gingras, A.C., et al., *Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism.* Genes Dev, 1999. **13**(11): p. 1422-37.

144. Wang, X., et al., *The C terminus of initiation factor 4E-binding protein 1 contains multiple regulatory features that influence its function and phosphorylation.* Mol Cell Biol, 2003. **23**(5): p. 1546-57.
145. Borgon, R.A., et al., *Crystal structure of human vinculin.* Structure, 2004. **12**(7): p. 1189-97.
146. Chorev, D.S., et al., *Conformational states during vinculin unlocking differentially regulate focal adhesion properties.* Sci Rep, 2018. **8**(1): p. 2693.
147. Tan, C.S.H., et al., *Thermal proximity coaggregation for system-wide profiling of protein complex dynamics in cells.* Science, 2018. **359**(6380): p. 1170-1177.
148. Piazza, I., et al., *A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication.* Cell, 2018. **172**(1-2): p. 358-372 e23.
149. Huang, J.X., et al., *High throughput discovery of functional protein modifications by Hotspot Thermal Profiling.* Nat Methods, 2019. **16**(9): p. 894-901.
150. Zinn, N., et al., *Improved Proteomics-Based Drug Mechanism-of-Action Studies Using 16-Plex Isobaric Mass Tags.* J Proteome Res, 2021. **20**(3): p. 1792-1801.
151. Sze, Y.H., et al., *High-pH reversed-phase fractionated neural retina proteome of normal growing C57BL/6 mouse.* Sci Data, 2021. **8**(1): p. 27.
152. Stryinski, R., et al., *Proteome profiling of L3 and L4 Anisakis simplex development stages by TMT-based quantitative proteomics.* J Proteomics, 2019. **201**: p. 1-11.
153. Huang, F.K., et al., *Deep Coverage of Global Protein Expression and Phosphorylation in Breast Tumor Cell Lines Using TMT 10-plex Isobaric Labeling.* J Proteome Res, 2017. **16**(3): p. 1121-1132.
154. Shirran, S.L. and C.H. Botting, *A comparison of the accuracy of iTRAQ quantification by nLC-ESI MSMS and nLC-MALDI MSMS methods.* J Proteomics, 2010. **73**(7): p. 1391-403.
155. Bantscheff, M., et al., *Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer.* Mol Cell Proteomics, 2008. **7**(9): p. 1702-13.
156. Karp, N.A., et al., *Addressing accuracy and precision issues in iTRAQ quantitation.* Mol Cell Proteomics, 2010. **9**(9): p. 1885-97.
157. Ow, S.Y., et al., *iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly".* J Proteome Res, 2009. **8**(11): p. 5347-55.
158. Ting, L., et al., *MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics.* Nat Methods, 2011. **8**(11): p. 937-40.
159. Bai, B., et al., *Deep Profiling of Proteome and Phosphoproteome by Isobaric Labeling, Extensive Liquid Chromatography, and Mass Spectrometry.* Methods Enzymol, 2017. **585**: p. 377-395.
160. McAlister, G.C., et al., *MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes.* Anal Chem, 2014. **86**(14): p. 7150-8.
161. Navarrete-Perea, J., et al., *Streamlined Tandem Mass Tag (SL-TMT) Protocol: An Efficient Strategy for Quantitative (Phospho)proteome Profiling Using Tandem Mass Tag-Synchronous Precursor Selection-MS3.* J Proteome Res, 2018. **17**(6): p. 2226-2236.
162. Gamblin, S.J., et al., *Activity and specificity of human aldolases.* J Mol Biol, 1991. **219**(4): p. 573-6.

163. Beernink, P.T. and D.R. Tolan, *Disruption of the aldolase A tetramer into catalytically active monomers.* Proceedings of the National Academy of Sciences, 1996. **93**(11): p. 5374-5379.
164. Starnes, G.L., et al., *Aldolase Is Essential for Energy Production and Bridging Adhesin-Actin Cytoskeletal Interactions during Parasite Invasion of Host Cells.* Cell Host & Microbe, 2009. **5**(4): p. 353-364.
165. Tolan, D.R., et al., *Thermodynamic analysis of the dissociation of the aldolase tetramer substituted at one or both of the subunit interfaces.* Biol Chem, 2003. **384**(10-11): p. 1463-71.
166. Justice, S.A.P., et al., *Mutant thermal proteome profiling for characterization of missense protein variants and their associated phenotypes within the proteome.* Journal of Biological Chemistry, 2020. **295**(48): p. 16219-16238.
167. Potel, C.M., et al., *Impact of phosphorylation on thermal stability of proteins.* bioRxiv, 2020: p. 2020.01.14.903849.
168. Wu, R., et al., *Correct interpretation of comprehensive phosphorylation dynamics requires normalization by protein expression changes.* Mol Cell Proteomics, 2011. **10**(8): p. M111 009654.
169. Pasteur, L., *Le Memoire sur la fermentation alcoolique.* Annales de Chimie et de Physique, 1860. **58**: p. 323-426.
170. Buchner, E., *Alkoholische Gärung ohne Hefezellen.* Berichte der deutschen chemischen Gesellschaft, 1897. **30**: p. 117-124.
171. Barnett, J.A., *A history of research on yeasts 5: the fermentation pathway.* Yeast, 2003. **20**(6): p. 509-43.
172. Kresge, N., R.D. Simoni, and R.L. Hill, *Otto Fritz Meyerhof and the elucidation of the glycolytic pathway.* J Biol Chem, 2005. **280**(4): p. e3.
173. Butterfield, D.A. and B. Halliwell, *Oxidative stress, dysfunctional glucose metabolism and Alzheimer disease.* Nat Rev Neurosci, 2019. **20**(3): p. 148-160.
174. Theurey, P., et al., *Systems biology identifies preserved integrity but impaired metabolism of mitochondria due to a glycolytic defect in Alzheimer's disease neurons.* Aging Cell, 2019. **18**(3): p. e12924.
175. Priebe, S., et al., *Extension of life span by impaired glucose metabolism in Caenorhabditis elegans is accompanied by structural rearrangements of the transcriptomic network.* PLoS One, 2013. **8**(10): p. e77776.
176. Missios, P., et al., *Glucose substitution prolongs maintenance of energy homeostasis and lifespan of telomere dysfunctional mice.* Nat Commun, 2014. **5**: p. 4924.
177. DeFronzo, R.A., et al., *Effects of insulin on peripheral and splanchnic glucose metabolism in noninsulin-dependent (type II) diabetes mellitus.* J Clin Invest, 1985. **76**(1): p. 149-55.
178. Rossetti, L. and A. Giaccari, *Relative contribution of glycogen synthesis and glycolysis to insulin-mediated glucose uptake. A dose-response euglycemic clamp study in normal and diabetic rats.* J Clin Invest, 1990. **85**(6): p. 1785-92.
179. Faubert, B., et al., *Lactate Metabolism in Human Lung Tumors.* Cell, 2017. **171**(2): p. 358-371 e9.
180. Vander Heiden, M.G., L.C. Cantley, and C.B. Thompson, *Understanding the Warburg effect: the metabolic requirements of cell proliferation.* Science, 2009. **324**(5930): p. 1029-33.

181. Warburg, O., F. Wind, and E. Negelein, *The Metabolism of Tumors in the Body.* J Gen Physiol, 1927. **8**(6): p. 519-30.
182. Hipkiss, A.R., *Aging, Alzheimer's Disease and Dysfunctional Glycolysis; Similar Effects of Too Much and Too Little.* Aging Dis, 2019. **10**(6): p. 1328-1331.
183. Koike, S., et al., *Age-related alteration in the distribution of methylglyoxal and its metabolic enzymes in the mouse brain.* Brain Res Bull, 2019. **144**: p. 164-170.
184. Haythorne, E., et al., *Diabetes causes marked inhibition of mitochondrial metabolism in pancreatic beta-cells.* Nat Commun, 2019. **10**(1): p. 2474.
185. El-Baba, T.J., et al., *Melting Proteins: Evidence for Multiple Stable Structures upon Thermal Denaturation of Native Ubiquitin from Ion Mobility Spectrometry-Mass Spectrometry Measurements.* J Am Chem Soc, 2017. **139**(18): p. 6306-6309.
186. Biltonen, R.L. and E. Freire, *Thermodynamic characterization of conformational states of biological macromolecules using differential scanning calorimetry.* CRC Crit Rev Biochem, 1978. **5**(2): p. 85-124.
187. Gaetani, M., et al., *Proteome Integral Solubility Alteration: A High-Throughput Proteomics Assay for Target Deconvolution.* J Proteome Res, 2019. **18**(11): p. 4027-4037.
188. Li, J., et al., *Selection of Heating Temperatures Improves the Sensitivity of the Proteome Integral Solubility Alteration Assay.* J Proteome Res, 2020. **19**(5): p. 2159-2166.
189. Valle-Mendiola, A. and I. Soto-Cruz, *Energy Metabolism in Cancer: The Roles of STAT3 and STAT5 in the Regulation of Metabolism-Related Genes.* Cancers, 2020. **12**(1): p. 124.
190. Vander Heiden, M.G., *Targeting cancer metabolism: a therapeutic window opens.* Nat Rev Drug Discov, 2011. **10**(9): p. 671-84.
191. Jin, M., et al., *Glycolytic Enzymes Coalesce in G Bodies under Hypoxic Stress.* Cell Rep, 2017. **20**(4): p. 895-908.
192. Fuller, G.G., et al., *RNA promotes phase separation of glycolysis enzymes into yeast G bodies in hypoxia.* Elife, 2020. **9**.
193. Huang, J.X., J.S. Coukos, and R.E. Moellering, *Interaction profiling methods to map protein and pathway targets of bioactive ligands.* Curr Opin Chem Biol, 2020. **54**: p. 76-84.
194. Martinez, N.J., et al., *A widely-applicable high-throughput cellular thermal shift assay (CETSA) using split Nano Luciferase.* Sci Rep, 2018. **8**(1): p. 9472.
195. West, G.M., et al., *Quantitative proteomics approach for identifying protein-drug interactions in complex mixtures using protein stability measurements.* Proc Natl Acad Sci U S A, 2010. **107**(20): p. 9078-82.
196. Meng, H., R. Ma, and M.C. Fitzgerald, *Chemical Denaturation and Protein Precipitation Approach for Discovery and Quantitation of Protein–Drug Interactions.* Analytical Chemistry, 2018. **90**(15): p. 9249-9255.
197. Zhang, X., et al., *Solvent-Induced Protein Precipitation for Drug Target Discovery on the Proteomic Scale.* Anal Chem, 2020. **92**(1): p. 1363-1371.
198. Strickland, E.C., et al., *Thermodynamic analysis of protein-ligand binding interactions in complex biological mixtures using the stability of proteins from rates of oxidation.* Nat Protoc, 2013. **8**(1): p. 148-61.
199. Pace, C., *[24] Evaluating contribution of hydrogen bonding and hydrophobic bonding to protein folding.* Methods in enzymology, 1995. **259**: p. 538-554.

200. Trindade, R.V., et al., *Pulse Proteolysis and Precipitation for Target Identification.* J Proteome Res, 2016. **15**(7): p. 2236-45.
201. Pai, M.Y., et al., *Drug affinity responsive target stability (DARTS) for small-molecule target identification.* Methods Mol Biol, 2015. **1263**: p. 287-98.
202. Lomenick, B., et al., *Target identification using drug affinity responsive target stability (DARTS).* Proc Natl Acad Sci U S A, 2009. **106**(51): p. 21984-9.
203. Chin, R.M., et al., *The metabolite alpha-ketoglutarate extends lifespan by inhibiting ATP synthase and TOR.* Nature, 2014. **510**(7505): p. 397-401.
204. Roux, K.J., et al., *A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells.* Journal of Cell Biology, 2012. **196**(6): p. 801-810.
205. Gupta, Gagan D., et al., *A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface.* Cell, 2015. **163**(6): p. 1484-1499.
206. Branon, T.C., et al., *Efficient proximity labeling in living cells and organisms with TurboID.* Nat Biotechnol, 2018. **36**(9): p. 880-887.
207. Martell, J.D., et al., *Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy.* Nat Biotechnol, 2012. **30**(11): p. 1143-8.
208. Rhee, H.W., et al., *Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging.* Science, 2013. **339**(6125): p. 1328-1331.
209. Lee, S.Y., et al., *APEX Fingerprinting Reveals the Subcellular Localization of Proteins of Interest.* Cell Rep, 2016. **15**(8): p. 1837-47.
210. McCutcheon, D.C., et al., *Photoproximity Profiling of Protein-Protein Interactions in Cells.* J Am Chem Soc, 2020. **142**(1): p. 146-153.
211. Cai, W., et al., *Top-down Proteomics: Technology Advancements and Applications to Heart Diseases.* Expert Rev Proteomics, 2016. **13**(8): p. 717-30.
212. Gregorich, Z.R. and Y. Ge, *Top-down proteomics in health and disease: challenges and opportunities.* Proteomics, 2014. **14**(10): p. 1195-210.
213. Schaffer, L.V., et al., *Identification and Quantification of Proteoforms by Mass Spectrometry.* Proteomics, 2019. **19**(10): p. e1800361.
214. Siuti, N. and N.L. Kelleher, *Decoding protein modifications using top-down mass spectrometry.* Nat Methods, 2007. **4**(10): p. 817-21.
215. Cupp-Sutton, K.A. and S. Wu, *High-throughput quantitative top-down proteomics.* Mol Omics, 2020. **16**(2): p. 91-99.
216. Riley, N.M., et al., *Enhanced dissociation of intact proteins with high capacity electron transfer dissociation.* Journal of The American Society for Mass Spectrometry, 2015. **27**(3): p. 520-531.
217. Shen, Y., et al., *High-resolution ultrahigh-pressure long column reversed-phase liquid chromatography for top-down proteomics.* Journal of Chromatography A, 2017. **1498**: p. 99-110.
218. Cooper, H.J., *To What Extent is FAIMS Beneficial in the Analysis of Proteins?* J Am Soc Mass Spectrom, 2016. **27**(4): p. 566-77.
219. Gerbasi, V.R., et al., *Deeper Protein Identification Using Field Asymmetric Ion Mobility Spectrometry in Top-Down Proteomics.* Anal Chem, 2021. **93**(16): p. 6323-6328.
220. Hale, O.J., et al., *High-Field Asymmetric Waveform Ion Mobility Spectrometry and Native Mass Spectrometry: Analysis of Intact Protein Assemblies and Protein Complexes.* Anal Chem, 2020. **92**(10): p. 6811-6816.

221. Griffiths, R.L., et al., *Comprehensive LESA Mass Spectrometry Imaging of Intact Proteins by Integration of Cylindrical FAIMS.* Anal Chem, 2020. **92**(4): p. 2885-2890.
222. Gaffney, D.O., et al., *Non-enzymatic Lysine Lactoylation of Glycolytic Enzymes.* Cell Chem Biol, 2020. **27**(2): p. 206-213 e6.
223. Yang, M.K., et al., *Global phosphoproteomic analysis reveals diverse functions of serine/threonine/tyrosine phosphorylation in the model cyanobacterium Synechococcus sp. strain PCC 7002.* J Proteome Res, 2013. **12**(4): p. 1909-23.
224. Young, M.W. and S.A. Kay, *Time zones: a comparative genetics of circadian clocks.* Nature Reviews Genetics, 2001. **2**(9): p. 702-715.