UNIVERSITY OF CALIFORNIA
RIVERSIDE

Cellular Protein Client Recovery of Human HSP40s

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Maureen R. Montoya

December 2020

Dissertation Committee:
Dr. Joseph Genereux, Chairperson
Dr. Ryan Julian
Dr. Jason Cheng

The Dissertation of Maureen R. Montoya is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

Throughout my graduate studies here at the University of California, Riverside, I have received much support from my research advisor, peers, and family.

I would like to thank my research advisor, Dr. Joey Genereux, for supporting and guiding me through research, presentations, and writing. His feedback and expertise were invaluable to my research progress and development as a scientist. I am grateful for his patience and encouragement that have helped me to improve and overcome difficulties in my research project.

I would like to thank my lab colleagues, Dr. Liangyong Mei, Guy Quanrud, Dr. Khanh Nguyen, Mateo Espinoza, Ziqi Lyu, and Melody Sycks. Each of them has helped me with practice presentations, troubleshooting lab protocols, and emotional support through stressful times.

I also want to thank Gwen Gonzalez for her kindness, friendship, and for always being available to talk through research problems with me.

I want to thank Dr. Kevin Simpson for being a good teaching mentor. Some of the quarters I have taught general chemistry were better because I had his support and guidance.

Finally, I would like to thank my family, especially my mother, who has always believed in me and supported my dreams. Without my family I could not have gotten through these difficult and stressful five years.

Copyright Acknowledgement

ABSTRACT OF THE DISSERTATION

Cellular Protein Client Recovery of Human HSP40s

by

Maureen R. Montoya

Doctor of Philosophy, Analytical Chemistry
University of California, Riverside, December 2020
Dr. Joseph Genereux, Chairperson

The integrity of the proteome must be maintained to ensure normal protein function and prevent cellular toxicity. Proteostasis factors scan the proteome to promote protein folding, trafficking, disassembly, and degradation. In all forms of life and cellular environments, the Hsp70/Hsp40 chaperone machinery rescues misfolded proteins, with over 40 human Hsp40s responsible for identifying and recruiting misfolded protein substrates to Hsp70. This Hsp40 diversity could be responsible for matching client recognition to Hsp70 functional diversity. To evaluate the client diversity of class B Hsp40s, we applied tandem mass tag-affinity purification-mass spectrometry (TMT-AP-MS) to characterize the interactor profiles of human Hsp40s. To increase interactor recovery yield, we used mutations that inactivate handoff of clients to Hsp70s. We found > 400 high-confidence interactors of DNAJB8 in HEK293T cells by using crosslinking with high stringency washing conditions to decrease false positives, TMT labeling to allow head-to-head comparison across different samples, and identification of specific interactors based on integrated TMT intensity correlation between prey and bait levels.

With bait correlation coupled with TMT-AP-MS we could assess the substrate profiles of different Hsp40 members. DNAJB1 and DNAJB8 are two Hsp40s that differ in size, structure, and functionality, that are ideal candidates for probing differential substrate binding across human class B Hsp40s. We first evaluated the proteome-wide effect of crosslinking and J-domain activity on interactor recovery for each Hsp40. After finding crosslinking to be unnecessary and J-domain activity to be neutral for recovery of interactors, we profiled substrates of DNAJB1^{WT} and DNAJB8^{WT}. Only 19 interactors, largely Hsp70 chaperones, were identified for DNAJB1^{WT}. Approximately half of the 249 interactors for DNAJB8^{WT} were shared with DNAJB8^{H31Q} clients. However, heat shock, which induces protein misfolding, increases global client binding to DNAJB8^{H31Q} but not to DNAJB8^{WT}, suggesting that the J-domain mutation is necessary to maintain increased client binding under stress conditions. We demonstrate an effective technique that could probe client interactions for Hsp40 co-chaperones, which could lead to insight into how proteostasis machineries subdivide their large substrate pool between different fates and minimize the gap in understanding of the role of Hsp40s in Hsp70 functional diversity.

Table of Contents

List of Tables

Chapter 2

Chapter 3

List of Figures

Chapter 1

Chapter 2

Chapter 1: Introduction

1.1 Protein Folding and Proteostasis

Every cell has a system for maintaining protein homeostasis, including protein synthesis, protein assembly/disassembly, preservation of protein stability, and response to stimuli[1]. External stresses, such as environmental toxins, heat, or oxidation, can cause imbalances (increased production of protein or decreased assembly of properly folded protein) in functional protein levels in the cell. These stresses can lead proteins into an unfolded or misfolded state by changing the cellular environment, causing mutations, or damage to DNA that can render proteins non-functional[2]. The overabundance of misfolded protein can lead to their aggregation, which can cause disease in the organism. For example, accumulation of misfolded protein (gain-of-function disorder)[3] can lead to the formation of aggregates unable to be degraded, which can overwhelm the cell and its functions, such as happens with alpha-synuclein in Parkinson's patients[4]. Conversely, if reversibly misfolded proteins are directed into degradation pathways, the levels of functional protein (loss-of-function disorders)[3] will also decrease, as seen in cystic fibrosis[5]. To maintain a healthy state, there must be a balance of protein synthesis, folding, trafficking, recycling, and degradation (**Figure 1.1**), and this balance must be maintained in the presence of stress. The system within the cell that maintains protein homeostasis is the proteostasis network (PN)[2]. The PN is comprised of protein factors that promote cellular maintenance and restoration of proteostasis for newly synthesized proteins. While nascent peptide is translated by the ribosome, it is already forming secondary

and tertiary contacts. These intermediate conformations can contribute to folding into its native functional form, to misfolding, or to aggregation, the two latter of which are non-functional (**Figure 1.2**)[6]. Protein factors (depicted as colored arrows in **Figure 1.2**), which guide nascent proteins to return to their folding intermediate, illustrate a key role of the proteostasis network: to restore proteins to their native (functional) state to preserve proteostasis. It is essential for proteins to be in a state that will allow proper folding, because once a protein has become misfolded or aggregated it is at risk of being degraded without performing its job.

It is more difficult for proteins to escape misfolded or aggregated states because of the energy required to change out of these conformations than it is for proteins to reach native conformations from an unfolded state[7,8]. Taking a closer look at the relative energies these conformations may reach, the unfolded state is high-energy, and will quickly fall to a lower-energy state (**Figure 1.3**). If a protein falls to the low-energy aggregated state, it may be unable to obtain the energy to return to a high-energy (intermediate) conformation, so the protein can attempt to refold into its native state.

One group of protein factors in the proteostasis network that are responsible for promoting substrate folding are molecular chaperones[9]. Chaperones with foldase activity can allow misfolded proteins to return to an intermediate state, so they can try to fold again into the native state or assist aggregated proteins in returning to a misfolded or intermediate state (**Figure 1.4**). Other chaperones that can act as holdases, such as small heat shock proteins, can bind partially unfolded proteins and hold onto them to prevent formation of aggregates[10].

One of the most well-studied chaperone systems is discussed in the next section.

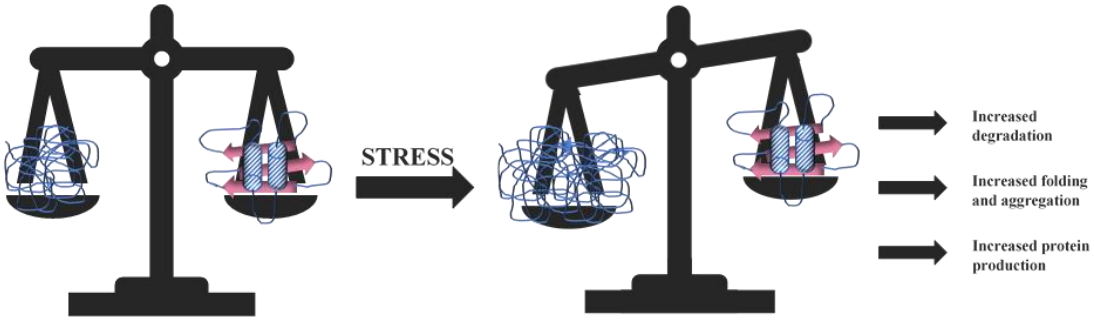


Figure 1.1 Proteostasis will be disrupted when affected by stress to the cell, which could lead to gain-of-function or loss-of-function disorders from improper protein folding[3]. The arrows on the right refer to the consequences of the cellular stress response.

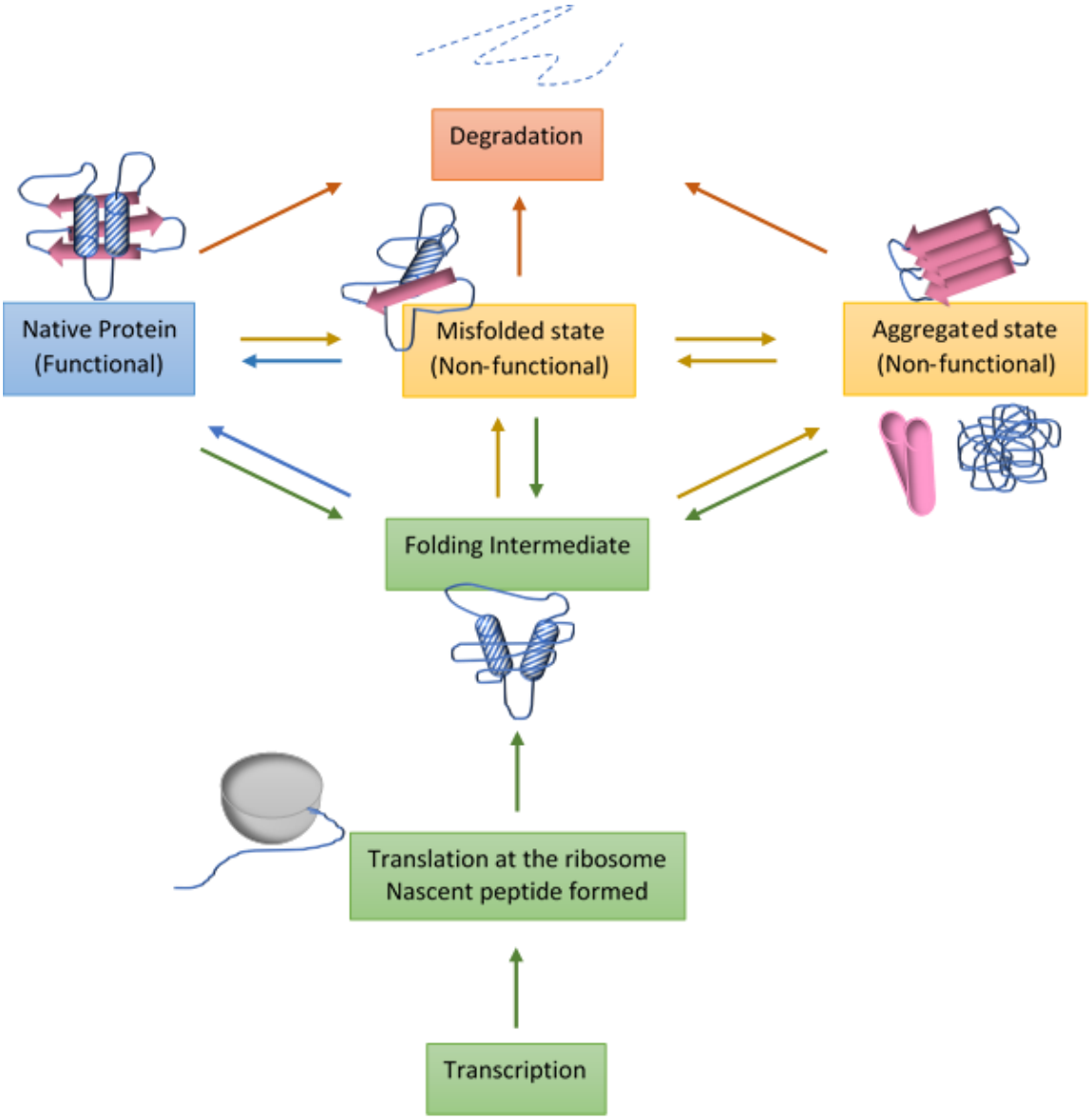


Figure 1.2 The possible fates of newly synthesized proteins regulated by protein factors of the proteostasis network. Image adapted from Kim et al[6].

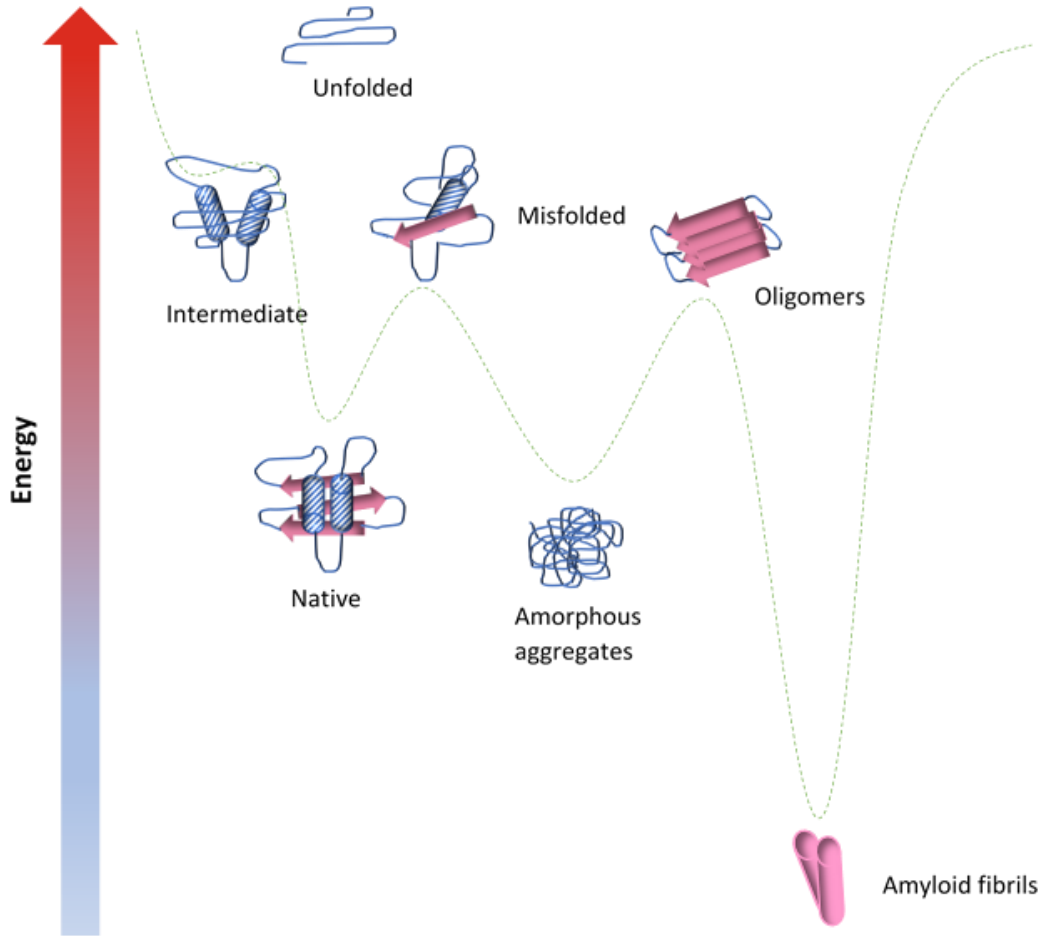


Figure 1.3 The protein folding energy landscape. The native form is the lowest energy conformation that is functional. Aggregates are low-energy as well but are unlikely to escape these conformations because of the energy barrier it would take to reach the intermediate state[6-8]. Image adapted from Jahn et al[8].

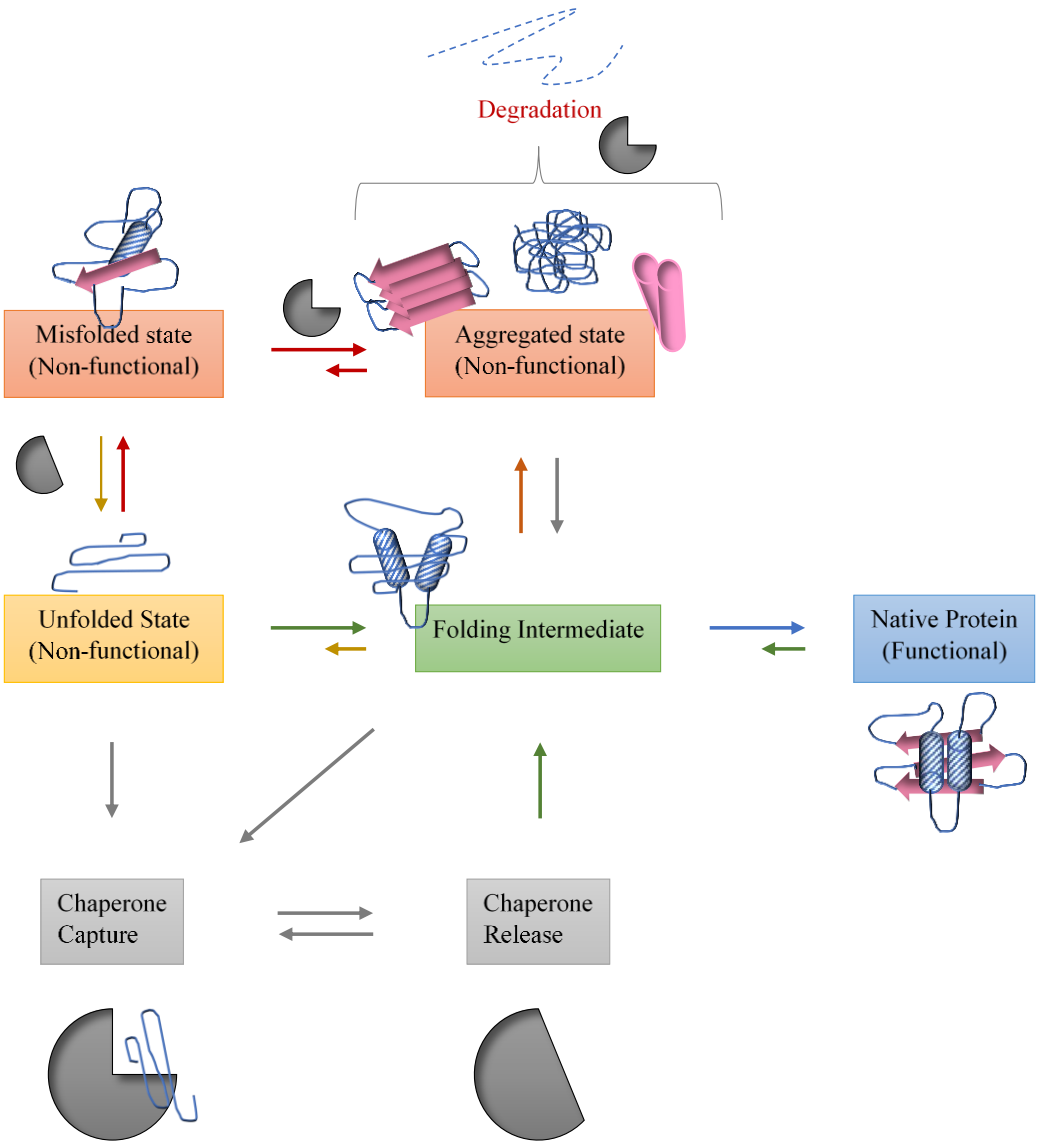


Figure 1.4 Chaperones can intervene in the folding pathway and allow unfolded proteins to return to an intermediate conformational state. Chaperones can also prevent intermediates from reaching an unfolded/misfolded state or native proteins from leaving the functional state. Image adapted from Kim et al[6].

1.2 The Role of Heat Shock Proteins in Proteostasis

Heat shock protein 70 (HSP70) is a molecular chaperone that is highly conserved across all forms of life. Its main function is to aid in the folding of misfolded proteins. This is not its only function. HSP70 assists in protein translocation and disaggregation[11,12]. Here, we focus on its main function, which relates to its structure and role in its chaperoning cycle.

HSP70 consists of three main subunits: the nucleotide binding domain (NBD), substrate binding domain (SBD), and linker region (**Figure 1.5A**). The NBD contains an ATP binding pocket that allows HSP70 to enter an open state when ATP is bound to that pocket. It is this open state that permits a misfolded protein to bind HSP70. Heat shock protein 40 (HSP40) is a co-chaperone that assists HSP70 in the refolding of proteins. HSP40 recruits the misfolded (substrate) protein and brings this substrate to the SBD of HSP70. HSP40 stimulates HSP70 ATPase activity in the NBD. ATP hydrolysis promotes substrate unfolding and transfer to HSP70. The ATP present inside the ATP binding pocket of HSP70 is then converted to ADP and HSP70 closes and clamps onto the unfolded substrate. The cycle is complete when a nucleotide exchange factor promotes release of the unfolded substrate from HSP70, providing the substrate an opportunity to fold to its native state[11,13]. In short, the HSP70 cycle (**Figure 1.5B**) intervenes in the folding pathway of misfolded substrates to unfold these substrates and increase their likelihood of forming into their native state.

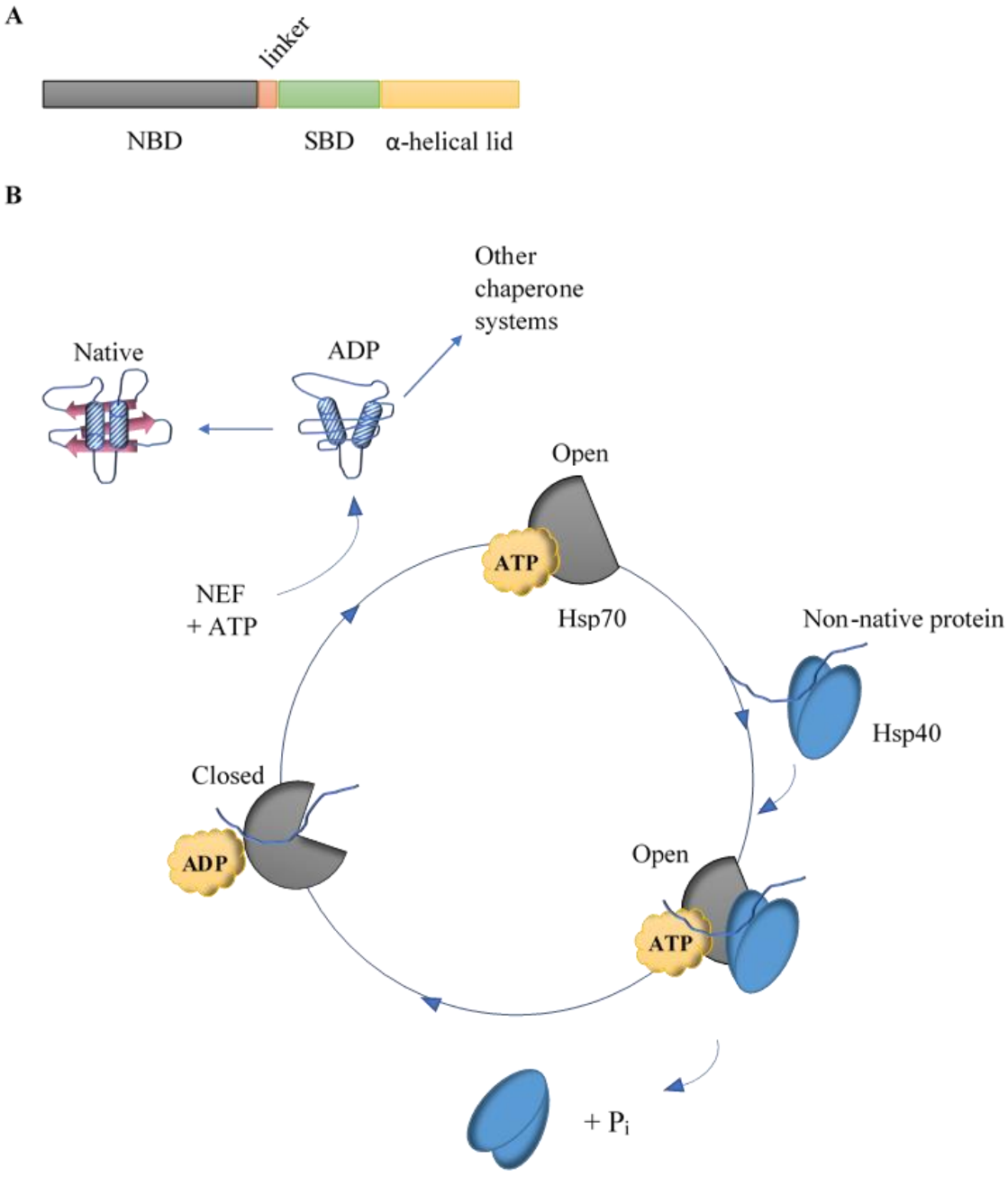


Figure 1.5 Chaperones can intervene in the folding pathway and allow unfolded proteins to return to an intermediate conformational state. Chaperones can also prevent intermediates from reaching an unfolded/misfolded state or native proteins from leaving the functional state[6,13]. Image adapted from Kim et al[6].

1.3 HSP40 Family of Co-Chaperones (J-Domain Proteins)

Intervention of HSP70 is crucial for protein folding. As described in the previous section, the HSP70 cycle is dependent on a key factor, the co-chaperone HSP40. HSP40s, which were named by the molecular weight (40 kDa) of the first observed member DnaJ in *E. coli*, are more diverse than their HSP70 interaction partner. While HSP70 is ubiquitous in the cell and there 13 HSP70s found in humans, there are over 40 different members of the HSP40 family[13]. Since HSP40s recruit misfolded substrates to HSP70, it has become increasingly important to study the binding properties and substrates of HSP40s to expand the knowledge of protein folding. There are three classes of HSP40s (**Figure 1.6**): A, B, and C, named by their domain structure and classified by their structural similarity to bacterial DnaJ. All classes are characterized by the highly conserved J-domain[14,15], with a conserved tripeptide, histidine-proline-aspartate (HPD). The HPD motif is in a loop between the helices I and II of the J-domain and is critical for the stimulation of Hsp70 ATPase activity[15]. Class A HSP40s consist of an N-terminal J-domain, a glycine/phenylalanine-rich domain, a ß-barrel domain containing a zinc finger-like region, and a ß-barrel dimerization domain. Class B HSP40s also contain an N-terminal J-domain, glycine/phenylalanine-rich domain, and contain double ß-barrels, but lack a zinc finger-like region and one or no substrate interaction domain. Class C are those HSP40s that have a J-domain anywhere in its structure and do not fit into class A or B. Recently, HSP40s (which are not always 40 kDa) have been reclassified as J-domain proteins (JDPs) because the J-domain is conserved across all HSP40s[16].

Although, the J-domain is conserved across all JDPs, that does not mean that all JDPs have the same functionality[17]. On the contrary, some JDPs do not require the J-domain to perform crucial functions like aggregation prevention. Additionally, different JDPs can have more than one function or substrate binding capability[18,19]. Some JDPs can function solely as a stimulus for ATPase activity of HSP70 without substrate binding. These JDPs can be free-moving or membrane-bound, as with yeast Hlj1 which sits at the ER membrane and engages Hsp70 from that location[20]. JDPs can simply bind substrates to assist in translocation across membranes, like human DNAJC19 which aids substrates in translocation from cytosol to the mitochondrial matrix, without facilitating refolding[21]. Other JDPs, like the canonical HSP70 model, bind substrates and stimulate ATP hydrolysis (human DNAJB1).

There have been a few recent studies on the client binding of JDPs. Jiang et al. has reported on client recognition of a *Thermus thermophilus* class B Hsp40 (ttHsp40)[19]. In this study, titration of physiological substrates alkaline phosphatase (PhoA) and maltose-binding protein (MBP) to isotopically labeled ttHsp40 show by NMR structural analysis that the two unfolded substrates interact with the two ß-barrel domains of ttHsp40 referred to as client binding domains 1 and 2 (CBD1, CBD2). Fifteen binding sites of PhoA were observed: seven interacted strongly with ttHsp40, while the other eight interacted weakly. Further NMR studies between ttHsp40 and ttHsp70 show there is a specific interaction between the CBD2 of ttHsp40 and the C-tail of ttHsp70. Similarly, the CBD1 of human class B Hsp40 DNAJB1 interacts with the C-tail of human Hsp70, and CBD1 of yeast class A Hsp40 Ydj1 interacts with the

C-tail of SsaI. The C-tail and client interaction with ttHsp40 appear to overlap, suggesting that clients may have to compete with ttHsp70 for binding to ttHsp40. Additionally, the C-tail and CBD must interact for client refolding, as removal of C-tail residues in ttHsp70 significantly decreased refolding by ttHsp70/ttHsp40 of chemically denatured luciferase. This reliance of Hsp40 on Hsp70 interaction for client refolding has been linked to an EEVD motif in the Hsp70 C-terminal[22]. Both studies suggest that for Hsp40s that rely on Hsp70 for client refolding, Hsp70 binding to Hsp40 can regulate client binding and release. Whereas Hsp70 may regulate client binding for some JDPs, one study characterizing DNAJB8 structure by NMR shows that for this JDP, interdomain contact in DNAJB8 oligomeric structure regulates Hsp70 binding[23]. This contact between the J-domain and C-terminal domain of DNAJB8 oligomers engages the J-domain, making it unavailable to Hsp70, but when a client is bound to DNAJB8, the oligomeric structure is disrupted, and the J-domain is available to Hsp70. The variation in composition of client binding domains and the number of binding sites available to a JDP, together with internal regulation of client versus Hsp70 binding, could result in distinct client selectivity among JDPs and account for their different functions.

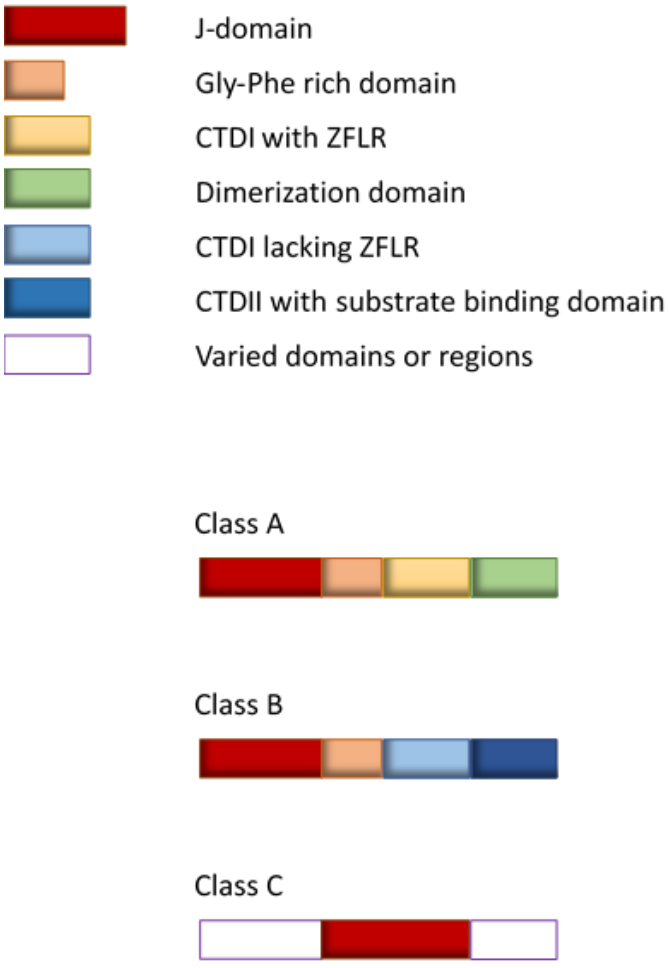


Figure 1.6 J-domain proteins can be divided into three classes: A, B, and C. All classes have a highly conserved J-domain. Class A and B have a G/F-rich region C-terminal to the J-domain. Class C can have the J-domain anywhere in its structure[13,15].

1.4 JDPs and Disease

Besides enabling folding processes, several functions of JDPs have been linked to disease. Neurodegenerative disorders like Alzheimer's, Huntington's, and Parkinson's disease result from the aggregation of proteins associated with the disease (amyloid-ß, polyglutamine, and α-synuclein/parkin, respectively)[24-26]. Expanded polyglutamine (poly Q) repeats in the huntingtin gene leads to protein aggregation[25]. One study exploring whether HSPs (Hsp70s, Hsp110s, Hsp40s) could suppress poly Q aggregation in cells co-transfected with poly Q-containing huntingtin protein, demonstrated that only members of the JDP family, DNAJB6 and DNAJB8, could prevent aggregation of poly Q proteins[27]. This finding was confirmed by another study in cells on DNAJB6 and DNAJB8 for suppression of poly Q peptide aggregation[28]. The authors also demonstrated that DNAJB1 did not provide the same aggregation prevention of poly Q peptides. The anti-aggregation activity of these JDPs is shown to be Hsp70-independent, as removal of the J-domain did not prevent DNAJB6 or DNAJB8 from suppressing aggregation. The C-terminal domain, specifically the SSF-SST serine-rich region, however, was crucial for anti-aggregation activity. This serine-rich region has been studied further by NMR structural analysis with respect to DNAJB6 and amyloid ß (Aß) proteins[29]. Based on two studies, DNAJB6 is an efficient suppressor of amyloid fibril formation[29], dependent on the serine-rich region and its own oligomeric structure, and DNAJB6 can inhibit primary nucleation rate in Aß by reacting with large aggregated species rather than monomeric Aß in solution[30]. Aside from poly Q and amyloid ß, α-

synuclein and a parkin mutant C289G have been associated with JDP members[31-33]. A characteristic of Parkinson's disease is the formation of Lewy bodies that are rich in α-synuclein (α-syn) proteins[34-36]. α-syn proteins can misfold and aggregate causing degeneration of neurons. CRISPR-Cas9 knockout of DNAJB6 in HEK293T cells expressing α-syn, had an increase in α-syn. When DNAJB6 was reestablished in the cells, α-syn aggregation was reduced back to normal levels (matching the parent cells), but only with an intact J-domain (Hsp70-dependent)[31]. It is interesting that DNAJB6 requires interaction with Hsp70 to suppress α-syn aggregation, but not for suppression of poly Q aggregation. This demonstrates that JDPs can have more than one function and substrate binding capability. Another study on α-syn aggregation with JDPs, shows that DNAJB1 does not suppress α-syn aggregation since DNAJB1 knockout cells had no effect on α-syn[32]. Several mutations in different genes (over 20) are at risk of mutation causing Parkinson's disorder[37]. One such mutation is parkin RING1 C289G. This loss-of-function mutant has lower solubility and is prone to aggregation. Overexpression of DNAJB2a (isoform), DNAJB6b (short isoform), and DNAJB8 efficiently reduces parkin C289G aggregation in an Hsp70-dependent manner[33]. Suppression of parkin C289G aggregation also did not require the serine-rich (SSF-SST) region that was necessary for poly Q aggregation. This suggests that different disease-prone proteins (substrates of JDPs) challenge cellular proteostasis in different ways. Additionally, DNAJB6 and DNAJB8 can handle disease-prone substrates by different mechanisms. Although DNAJB2a, DNAJB6b, and DNAJB8, can suppress aggregation of disease-prone substrates, it is significant that other JDPs,

like DNAJA1 and DNAJB1 cannot. Knockout of DNAJA1 in cells, decreases aggregation of poly Q, while knockout of DNAJB1 has no effect[38]. Evidently, different JDPs, even those from the same class (both DNAJB1 and DNAJB6 are class B), have unique substrate pools. Substrate binding among the various JDPs can be elucidated by exploring their protein-protein interactions.

1.5 Methods for Determining Protein-Protein Interactions

Protein-protein interactions (PPIs) are studied intensively, yet the techniques available struggle with differentiating true interactions versus non-specific interactions. A few techniques to characterize PPIs are yeast two-hybrid assays[39], chemical crosslinking[40], LUminescence-based Mammalian InteRactome (LUMIER)[41], and affinity purification-mass spectrometry (AP-MS)[42].

The yeast two-hybrid assay was developed for high-throughput screening of PPIs[39]. In this assay, a pair of proteins to be tested for interaction are expressed in yeast. One protein is fused to a DNA-binding domain upstream of reporter genes, and the other protein is fused to a transcription activation domain. The two strains are mated, and interaction is revealed by the formation of a colony on media that requires an activated reporter to have growth. Although this assay allows for high-throughput screening, it suffers from poor coverage from missed interactions[43]. Even large data sets of the same species have little overlap of identified interactions[44-46].

Another technique, chemical crosslinking, was developed to study three-dimensional protein and protein complex structures[40]. This approach uses a chemical crosslinking reagent to covalently link functional groups of one protein to another nearby protein.

The functional group that is targeted and its tagging distance depends on the crosslinking reagent[40]. One common reagent is dithiobis succinimidyl propionate, DSP, that reacts with primary amines. Chemical crosslinking is often combined with mass spectrometry techniques. A disadvantage to crosslinking is that it can be difficult to determine structural features of the peptide pairs that are physically crosslinked to each other, but this problem can be solved with isotope-labeled crosslinkers[47].

LUMIER is a protein tagging and immunoprecipitation technique where the protein of interest (POI) is fused to Renilla luciferase enzyme (RL) and is co-expressed in mammalian cells with its substrates that are individually Flag-tagged[41]. An RL enzymatic assay is then performed on immunoprecipitates (with antibody against Flag). The eluted immunoprecipitates with the RL-tagged POI and its Flag-tagged prey can then be detected by light emission from the enzyme.

LUMIER and other affinity purification techniques have issues with false negatives and false positives. Contaminants can falsely be identified as interactors because they are abundant, such as housekeeping proteins or Hsps[42]. Conversely, the conditions of the AP experiment could fail to preserve interactions and lead to true interactors not being identified[42]. These issues can be overcome by choosing a set of conditions that preserve PPIs, reducing non-specific binding, comparing to lists of interactors or non-interactors in the literature, and combining AP techniques with quantitative analysis. Recently, one common way in which protein interactors are identified is through Affinity Purification-Mass Spectrometry.

1.6 Affinity Purification-Mass Spectrometry

This approach usually places an affinity tag onto a protein-of-interest (POI) so that protein can be isolated using immunobeads specific to the tag. When the protein is overexpressed in cells, the cells can be lysed, then incubation with the immunobeads allows the POI and any other proteins bound to it, to be co-immunoprecipitated. The proteins are eluted off the beads and samples are then purified with only the proteins that attached to the beads, predominantly the POI (bait) and its interactors (prey)[42]. An experiment can be set up so that many conditions can be tested at once, using quantitative labeling approaches such as stable isotope labeling using amino acids in cell culture (SILAC)[48,49], isobaric tags for relative and absolute quantitation (iTRAQ)[50,51], or tandem mass tags (TMT)[52,53].

In the SILAC labeling approach, cells are cultured in isotopically labeled media containing lysine and arginine[48]. Cells grown in the isobaric amino acid-supplemented media will stably express these isotopes, labeling endogenous proteins with the isotopes. Two experimental conditions can be explored and compared head-to-head: one set of conditions performed in cells grown in the isotopically labeled media (heavy), with another set of conditions performed in cells without amino acid labeled media (light)[48]. After cells from these two conditions are harvested and lysed, lysates are combined in an equal ratio, prepared for mass spectrometry analysis, and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). A pair of peptides can now be differentiated between the two conditions (light and heavy) and quantified by the relative MS signal intensity[48]. SILAC has the advantage of

reducing sample handling error by labeling before affinity purification but is limited to testing a maximum of three different conditions[54,42].

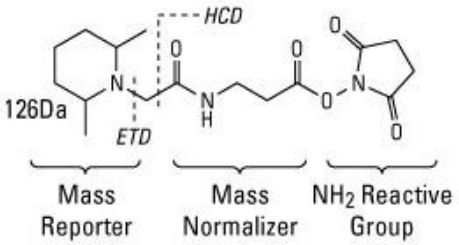
iTRAQ is an isobaric labeling strategy that can test up to eight different conditions but obtains higher identification with 4-plex experiments[55]. An iTRAQ isobaric tag contains a mass reporter group, mass normalization group, and an amine-reactive NHS group[50]. Each tag co-elutes in the chromatographic separation of an LC-MS/MS run. Then, HCD or CID fragmentation breaks the covalent bond between the mass reporter and mass normalizer groups. The reporter ions from the four conditions then have different masses that can be quantified by their relative protein abundance[50].

Tandem mass tags (TMTs) are reagents that label proteins to be quantified by LC-MS/MS. Typically, proteins are digested by trypsin and so are split into smaller subunits called peptides. TMTs can label the N-terminus of each protein or peptide fragment. One advantage of using TMTs for protein quantification in a proteomics experiment, is that you can label up to ten different samples for a single MS run. However, resolution of every single tag is only possible in mass analyzers with resolving power greater than 60,000.

TMTs, like iTRAQ, consist of three regions: the mass reporter group, the mass normalization group, and the amine reactive group (**Figure 1.7A**)[52]. Upon incubation with the TMT, the N-terminus of each peptide in one sample mixture reacts with the carboxylate group and removes the amine reactive group of the TMT. Then each peptide for that sample is labeled with the mass reporter and mass normalizer portion of the TMT. Up to ten different samples can be labeled each with a different TMT

label (126-131) and are mixed before MS analysis. Each uniquely labeled set of peptides will maintain the same molecular weight upon MS1 acquisition (**Figure 1.7B**). Once this peak is selected for isolation to acquire MS2, the peptides are cleaved by HCD between the mass reporter and mass normalizer in the TMT label. Then each sample is differentiated by one another by a single-low mass reporter ion (**Figure 1.7B,** right panel)[56]. Each peptide can then be quantified. This labeling strategy combined with the AP platform has the advantage of allowing quantification and comparison of several different samples in one run. However, it also has the potential for co-isolation of contaminating near-isobaric ions with target ions, which can interfere with detection of small protein expression changes in complex protein mixtures[57].

A



B

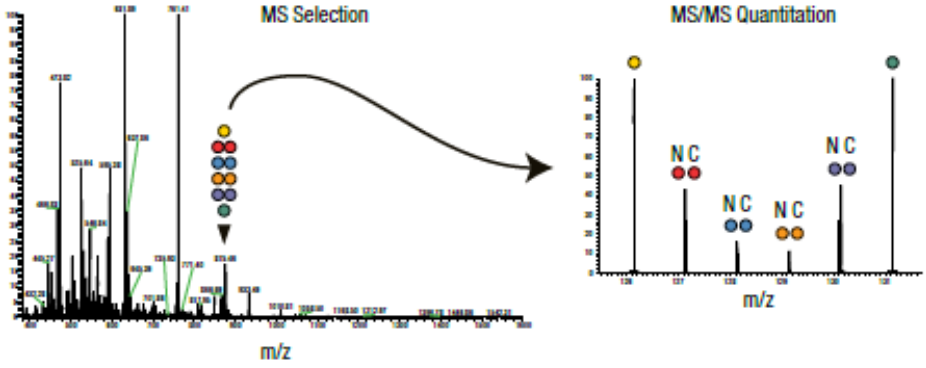


Figure 1.7 Image from Thermo Scientific TMT 10-plex labeling reagent manual[56]. A) TMT structure and identity of the mass reporter, mass normalizer, and NH_2 reactive group. B) MS^1 of ten TMTs eluting at same m/z, followed by separation in MS^2 after cleavage.

1.7 Scope of the Dissertation

This work will focus on three main ideas: (1) AP-MS can target potential clients of J-domain proteins (JDPs), (2) different JDPs within the same class may have separate client pools from one another, and (3) the variety of JDPs and their respective clients can give insight into the functional diversity of the Hsp70 cycle machinery.

In chapter 2, the TMT-AP-MS is established as a platform for identifying clients of DNAJB8.

In chapter 3, the same TMT-AP-MS platform can be used to determine the client binding capability of two class B JDPs: DNAJB1 and DNAJB8.

In chapter 4, the diverse client pool of class B JDPs (DNAJB2a, DNAJB4, DNAJB6b) are discussed. Comparisons are made about DNAJB6 and DNAJB8, and how they differ from other subfamily members like DNAJB1. The implications of the findings in chapters 2-4 are discussed as well as the future directions for the project.

1.7 References

(1) Balch, W. E.; Morimoto, R. I.; Dillin, A.; Kelly, J. W. *Science* **2008**, *319* (5865), 916–919.

(2) Morimoto, R. *Cold Spring Harb Symp Quant Biol.* **2011**, 76, 91-99.

(3) Narayan, P.; Ehsani, S.; Lindquist, S. *Nat. Chem. Biol.* **2014**, *10* (11), 911–920.

(4) Lakso, M.; Vartiainen, S.; Moilanen, A.M.; et al. *J. Neurochem*. **2003**, 86, 165–172.

(5) Ren, H. Y.; Grove, D. E.; De La Rosa, O.; et al. *Mol. Biol. Cell* **2013**, 24, 3016–3024.

(6) Kim, Y. E.; Hipp, M. S.; Bracher, A.; Hayer-Hartl, M.; Ulrich Hartl, F. *Annu Review Biochem* **2013**, 82, 323-355.

(7) Hartl, F., Bracher, A.; Hayer-Hartl, M. *Nature* **2011**, 475, 324–332.

(8) Jahn, T. R.; Radford, S. E. *FEBS J*. **2005**, *272* (23), 5962-5970.

(9) Nollen, E. A.; Morimoto, R. I. *Journal of Cell Science* **2002**, 115, 2809-2816.

(10) Mymrikov, E. V.; Daake, M.; Richter, B.; Haslbeck, M.; Buchner, J. *J Biol Chem*. **2017**, *292* (2), 672-684.

(11) Szabo, A.; Langer, H.; Schroder, H.; Flanagan, J.; Bukau, B.; Hartl, F. U. *PNAS*. **1994**. *91* (22): 10345-10349.

(12) Meng, W.; Clerico, E. M.; McArthur, N.; Gierasch, L. M. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (47), 11970–11975.

(13) Kampinga, H. H.; Craig, E. A. *Nat. Rev. Mol. Cell Biol.* **2010**, *11* (8), 579–592.

(14) Hennessy, F.; Cheetham, M. E.; Dirr, H. W.; Blatch, G. L. *Cell Stress Chaperones* **2000**, *5* (4), 347–358.

(15) Cheetham, M. E.; Caplan, A. J. *Cell Stress Chaperones* **1998**, *3*, (1), 28-36.

(16) Kampinga, H. H.; Andreasson, C.; Barducci, A.; et al. *Cell Stress Chaperones* **2019**, *24* (1), 7-15.

(17) Craig, E. A.; Marszalek, J. *Trends in Biochem Sci.* **2017**, *41* (5): 355-368.

(18) Ajit Tamadaddi, C.; Sahi, C. *Cell Stress Chaperones* **2016**, *21* (4), 563–570.

(19) Jiang, Y.; Rossi, P.; Kalodimos, C. G. *Science (80).* **2019**, *365* (6459), 1313–1319.

(20) Nakatsukasa, K.; Huyer, G.; Michaelis, S.; Brodsky, J. L. *Cell* **2008**, 132, 101–112.

(21) Chacinska, A.; Koehler, C. M.; Milenkovic, D.; Lithgow, T.; Pfanner, N. *Cell* **2009**, 138, 628–644.

(22) Yu, H. Y.; Ziegelhoffer, T.; Craig, E. A. *FEBS Lett*. **2015**, *589* (19 Pt B), 2825-2830.

(23) Ryder, B. D.; Matlahov, I.; Bali, S.; Vaquer-Alicea, J.; van der Wel, P. C. A. *BioRxiv*. **2020**.

(24) Murphy, M. P.; LeVine, H. 3rd. **2010**, *19* (1), 311-323.

(25) Stoyas, C. A.; La Spada, A. R. *Handb. Clin. Neurol*. **2018**, 147, 143–170.

(26) Hansen, C.; Li, J. Y. *Trends Mol Med* **2012**, 18, 248–255.

(27) Hageman, J.; Rujano, M. A.; van Waarde, M. A., et al. *Mol. Cell* **2010**, *37* (3), 355–369.

(28) Gillis, J.; Schipper-Krom, S.; Juenemann, K.; et al. *J. Biol. Chem.* **2013**, *288* (24), 17225–17237.

(29) Söderberg, C. A. G.; Månsson, C.; Bernfur, K.; et al. *Sci. Rep.* **2018**, *8* (1), 1–15.

(30) Österlund, N.; Lundquist, M.; Ilag, L. L.; Gräslund, A.; Emanuelsson, C. **2020**, No. 18.

(31) Aprile, F. A.; Källstig, E.; Limorenko, G.; Vendruscolo, M.; Ron, D.; Hansen, C. *Sci Rep* **2017**, 7, 9039.

(32) Deshayes, N.; Arkan, S.; Hansen, C. *Int J Mol Sci*. **2019**, *20* (18), 4495.

(33) Kakkar, V.; Kuiper, E. F. E.; Pandey, A.; Braakman, I.; Kampinga, H. H. *Sci. Rep.* **2016**, *6* (September), 1–12.

(34) Spillantini, M. G.; Schmidt, M. L.; Lee, V. M.; Trojanowski, J. Q.; Jakes, R.; Goedert, M. *Nature* **1997**, 388, 839–840.

(35) Goedert, M.; Spillantini, M. G.; Del Tredici, K.; Braak, H. *Nat Rev Neurol* **2013**, 9, 13–24.

(36) Waxman, E. A.; Giasson, B. I. *Biochim Biophys Acta* **2009**, 1792, 616–624.

(37) Hasegawa, T.; Yoshida, S.; Sugeno, N.; Kobayashi, J.; Aoki, M. *Front. Neurosci.* **2018**, *11* (JAN), 1–9.

(38) Rodríguez-González, C.; Lin, S.; Arkan, S.; Hansen, C. *Sci Rep* **2020**, 10, 8130.

(39) Fields, S.; Song, O. *Nature* **1989**, 340, 245-246.

(40) Sinz, A. *Mass Spectrom Rev*. **2006**, *25* (4), 663-682.

(41) Barrios-Rodiles, M.; Ellis, J. D.; Blencowe, B. J.; Wrana, J. L. *Methods Mol Biol*. **2017**, 1550, 137-148.

(42) Gingras, A. C.; Gstaiger, M.; Raught, B.; Aebersold, R. *Nat Rev Mol Cell Biol*. **2007**, *8* (8), 645-654.

(43) Parrish, J. R.; Gulyas, K. D.; Finley, R. L., Jr. *Curr Opin Biotechnol*. **2006**, *17* (4), 387-393.

(44) Giot, L.; Bader, J. S.; Brouwer, C.; et al. *Science* **2003**, 302, 1727-1736.

(45) Stanyon, C. A.; Liu, G.; Mangiola, B. A.; et al. *Genome Biol* **2004**, 5, R96.

(46) Formstecher, E.; Aresta, S.; Collura, V.; et al. *Genome Res* **2005**, 15, 376-384.

(47) Muller, D. R.; et al. *Anal Chem*. **2001**, 73, 1927–1934.

(48) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; et al. *Mol Cell Proteomics* **2002**, 1, 376-86.

(49) Wang, X.; Huang, L. *Methods Mol Biol*. **2014**, 1188, 191-205.

(50) Wiese, S.; Reidegeld, K. A.; Meyer, H. E.; Warscheid, B. *Proteomics* **2007**, *7* (3), 340-350.

(51) Haura, E. B.; Müller, A.; Breitwieser, F. P.; et al. *J Proteome Res*. **2011**, *10* (1), 182-190.

(52) Thompson, A.; Schäfer, J.; Kuhn, K.; et al. *Anal. Chem.* **2003**, *75*, 1895-1904.

(53) Mei, L.; Montoya, M. R.; Quanrud, G. M.; et al. *J Proteome Res*. **2020**, *19* (4), 1565-1573.

(54) Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. *Anal Bioanal Chem*. **2007**, 389, 1017-31.

(55) Pichler, P.; Köcher, T.; Holzmann, J.; et al. *Anal Chem*. **2010**, *82* (15), 6549-6558.

(56) Thermo Scientific, "TMT10 plex Mass Tag Labeling Kits and Reagents," 90406 datasheet, **2015**.

(57) Bantscheff, M.; Boesche, M.; Eberhard, D.; Matthieson, T.; Sweetman, G.; Kuster, B. *Mol Cell Proteomics* **2008**, *7* (9), 1702-1713.

Chapter 2: Bait Correlation Improves Interactor Identification by Tandem Mass Tag-Affinity Purification-Mass Spectrometry

2.1 Introduction

Protein-protein interactions (PPIs) allow the assembly of protein complexes, mediate signaling pathways, and define the client distribution of enzymes involved in post-translational modifications[1–5]. Identifying and characterizing PPIs provides valuable molecular insights into cell function and physiology. However, the large dynamic range of PPI strength and abundance challenges the discrimination of bona fide from artefactual interactions. This challenge is central to the interpretation of most PPI datasets, including those acquired by the most prevalent techniques: yeast two-hybrid[6–8], sandwich assays such as LUMIER[39,10], microarrays[11], chemical crosslinking[12,13], and Affinity Purification coupled with Mass Spectrometry (AP-MS)[14–19]. In the latter approach, a "bait" protein is isolated, and its co-isolating "prey" interactors are identified by MS. Because many proteins associate with beads independent of bait, this approach is prone to false positives (Type I errors). To decrease false positives, more stringent wash buffers can be used, but this in turn increases false negatives (Type II errors). Alternatively, false positives can be minimized by carefully filtering out potential prey that are known to strongly associate with beads, or by comparison to controls obtained under similar conditions[20,21]. Large datasets comprised of dozens to thousands of AP-MS runs reveal the identities and intensities of non-specific binding proteins associated with an individual AP-MS platform, which can then be compared to each individual AP-MS experiment[21–23]. This was most dramatically demonstrated in the BioPlex, which determined over 70,000

interactions from over 5,000 proteins (and counting) using the Comparative Proteomics Analysis (ComPASS) methodology[24,25]. For smaller datasets, however, while aggressive statistical filters ensure that only high-quality interactors are reported, they can also lead to few identified prey [26,27].

TMT-AP-MS allows direct quantitative comparisons of prey recovery across multiple replicates in a single run, simplifying evaluation of whether potential interactors are preferentially recovered with the bait[22,28]. However, variation in bait levels between conditions, particularly when using transient transfection, leads to variability in recovered prey levels. For experiments comparing interaction networks for a single bait between cellular conditions, this variability has been controlled by normalizing prey levels to bait levels[28,29]; this approach is not available when differentiating real vs. artefactual interactors, as minimal bait levels are present under mock transfection conditions. Several studies have been successful at identifying protein complex composition by global correlation analysis following native protein chromatography. In this approach, proteins are chromatographically separated under conditions that preserve native complexes, with the expectation that proteins that co-fractionate are likely to interact. A similar approach is to perform a large number of immunoprecipitations, and then to determine which proteins co-IP reproducibly; in this case correlations are made between each prey:prey combination rather than solely between prey and bait. [30–33]. This approach relies upon the tendency of interacting proteins to maintain stoichiometry. However, both approaches require large scale experiments and probe global interaction maps, while researchers frequently want to identify interactors of a specific bait protein with the fewest number of

replicates. We thus consider whether using a correlation-based analysis can assist in identifying prey following individual TMT-AP-MS runs for a single bait. Herein, we report that evaluating potential interactors based on their Pearson's correlation with bait levels decreases Type II errors without increasing Type I errors.

2.2 Materials and Methods

2.2.1 Materials

Buffer components and other biochemical reagents were all purchased from Fisher, VWR, or Millipore Sigma. Nanopure water and sterilized consumables were used for all biochemical experiments.

DNAJB8 was amplified from pcDNA5/FRT/TO V5 DNAJB8[34] and inserted into the pFLAG.CMV2 vector by PIPE cloning[35]. The H31Q mutation was introduced into DNAJB8 using site-directed mutagenesis. 14-3-3ζ was amplified from cDNA derived from HepG2 (ATCC) using TRIzol (Fisher) and inserted into the pFLAG.CMV2 vector by PIPE cloning. eGFP.pDEST30 has been reported[36]. All constructs were subjected to analytical digest and sequenced (Retrogen) to confirm identity. Primers were purchased from IDT, and all cloning enzymes purchased from New England Biolabs. Primer sequences are shown in **Table 2.1**.

Table 2.1. Primers used for Molecular Cloning		
Primer Name		**Sequence**
DNAJB8 PIPES Vector	Fwd	5’-GAC AGC AAG TAG GCG AAT TCA TCG ATA GAT CTG-3’
DNAJB8 PIPES Vector	Rev	5’-CTT CGT AGT AGT TAG CCA TAA GCT TGT CGT CAT CGT C-3’
DNAJB8 PIPES Insert	Fwd	5’-GAC GAT GAC GAC AAG CTT ATG GCT AAC TAC TAC GAA G-3’
DNAJB8 PIPES Insert	Rev	5’-CAG ATC TAT CGA TGA ATT CGC CTA CTT GCT GTC-3’
DNAJB8 H31Q SDM	Fwd	5’- CTT CGT TGG CAG CCC GAC AAG AAC CCT GAC AAT AAG-3’
DNAJB8 H31Q SDM	Rev	5’- GTT CTT GTC GGG CTG CCA ACG AAG GGC-3’
14-3-3z PIPES Vector	Fwd	5’-GGA GAA GGA GGG GAA AAT TAA GCG AAT TCA TCG ATA GAT CTG-3’
14-3-3z PIPES Vector	Rev	5’-GAA CCA GCT CAT TTT TAT CCA TAA GCT TGT CGT CAT CGT C-3’
14-3-3z PIPES Insert	Fwd	5’-GAC GAT GAC GAC AAG CTT ATG GAT AAA AAT GAG CTG GTT C-3’
14-3-3z PIPES Insert	Rev	5’-CAG ATC TAT CGA TGA ATT CGC TTA ATT TTC CCC TCC TTC TCC-3’

HEK293T cells (ATCC) were cultured in DMEM (Corning) supplemented with 10% fetal bovine serum (FBS; Seradigm), 2 mM L-Glutamine (Corning), and penicillin (100 IU/mL)-streptomycin (100 μg/mL; Corning) and used within 30 passages. Cells were checked monthly for mycoplasma contamination by PCR assay. Plasmid DNA was introduced into cells by the method of calcium phosphate transfection. Transfection efficiency >80% was confirmed in all experiments by GFP positive control.

2.2.2 Simulations in Mathematica

Calculations were performed in Mathematica. Each parameter is pulled from a truncated normal distribution with mean and variance parameters as described in **Table 2.2**. ϵ_{ij} represents sampling variation. Each protein is assumed to have a non-specific background intensity, μ^{ns} independent of bait levels. Under this simple model, the levels of a non-interacting protein *i* in the TMT channel *j* is taken as:

$$TMT_{ij}^{ni} = \mu_i^{ni,ns}\epsilon_{ij}^{ni,ns} \qquad (1)$$

where $\mu_i^{ni,ns}$ is the mean intensity of the *i* non-interacting protein. The levels of the bait protein consist of a non-specific and specific term:

$$TMT_j^{bait} = \mu^{bait,ns}\epsilon_j^{bait,ns} + \delta\mu_j^{bait}\epsilon_j^{bait} \qquad (2)$$

where $\delta = 1$ in the three "bait" channels and $\delta = 0$ in the three "mock" channels. Prey intensities also consist of a specific and non-specific term:

$$TMT_{kj}^{prey} = \mu_k^{prey,ns}\epsilon_{kj}^{prey,ns} + \delta R_k\mu_j^{bait}\epsilon_{kj}^{prey} \qquad (3)$$

where R_k represents the mean ratio of the k protein to bait. For each simulation, 100 replicates were ran, with each replicate including three "bait" channels and three "mock" channels. The Mathematica code is included in **Appendix**.

TMT runs were simulated by populating reporter ion ratios for three types of protein: bait, non-specific interactors (ni), and specific interactors (prey). Background signals for each protein, bait levels in an individual replicate, the ratio of each prey to the bait, and replicate-level error for all measurements were drawn from truncated normal distributions. For each parameter set, 100 TMT 6-plex samples were populated with 3 channels including "bait" and 3 channels serving as "mock"s, with no bait expressed. Area under the curve was calculated by integration of a 100-point parametrization of the Receiver Operating Characteristic curve, with prey taken as true-positives and non-specific interactors taken as false positives. Mathematica code may be found at https://github.com/josephgenereux/ROC-Simulations-in-Mei-et-al.

Table 2.2. Parameters for simulations of TMT-AP-MS data.			
Parameter	**Symbol**	**Mean (Value in Figure 2.1)**	**Variance (Value in Figure 2.1)**
Non-specific Non-interactor Levels	$\mu_i^{ni,bg}$	NonSpecificNoninteractMean (0.3)	NonSpecificNoninteractStandard Deviation (0.6)
Non-specific Non-interactor Levels Sampling Variation	$\epsilon_{ij}^{ni,bg}$	1	NonSpecificNoninteract Channel Standard Deviation (0.2)
Non-specific Prey Levels	$\mu_k^{prey,bg}$	NonSpecificPreyMean (0.1)	NonSpecificPrey Standard Deviation (0.6)
Non-specific Prey Levels Sampling Variation	$\epsilon_{kj}^{prey,bg}$	1	NonSpecificPrey Channel Standard Deviation (0.2)
Prey:Bait Ratio	R_k	RatioMean (0.3)	Ratio Standard Deviation (0.3)
Prey:Bait Ratio Sampling Variation	ϵ_{kj}^{prey}	1	Ratio Channel Standard Deviation (0.2)
Non-specific Bait Levels	$\mu^{bait,bg}$	NonSpecificBaitMean (0.1)	NonSpecificBaitStandard Deviation (0.01)
Non-specific Bait Levels Sampling Variation	$\epsilon_j^{bait,bg}$	1	NonSpecificBait Channel Standard Deviation (0.2)
Bait Levels	μ_j^{bait}	1	BaitLevelsStandardDeviation (indicated)
Bait Levels Sampling Variation	ϵ_j^{bait}	1	BaitLevels Channel Standard Deviation (0.2)

2.2.3 Immunoprecipitation

Cells were harvested from confluent 10 cm dishes at 36 to 48 h post-transfection by scraping in TBS buffer with 1 mM EDTA. High stringency lysis was performed in RIPA buffer (150 mM NaCl, 50 mM Tris pH 7.5, 1% Triton X 100, 0.5% sodium deoxycholate, 0.1 % SDS). Low stringency lysis and washes was performed with 0.1% Triton X-100 in TBS (10 mM Tris pH 7.5, 150 mM NaCl). For conditions involving dithiobis succinimidyl propionate (DSP) crosslinking, cells were incubated in 1 mM DSP/1%DMSO/PBS for 30 min. with rotating at ambient temperature, and then quenched by addition of Tris pH 8.0 (to 90 mM) and rotating for 15 min at ambient temperature. Cells were lysed for 30 min on ice in lysis buffer supplemented with fresh protease inhibitors (Roche). Lysate was separated from cell debris centrifugation at 21,100 x g for 15 min at 4 °C. Protein was quantified by Bradford assay (Bio-rad). Lysates were pre-cleared with 15 uL sepharose-4B beads (Millipore Sigma) for 30 min at 4 °C, then centrifuged at 1,500 × g for 1 min to pellet beads, denatured by boiling for 10 min at 100 °C in 20% SDS, followed by immunoprecipitation (at 0.1% SDS) with 15 µL M2 anti-FLAG Magnetic Beads (Millipore Sigma) and overnight rotation at 4 °C. The denaturation step was excluded for low-stringency immunopurifications. Immunodepletion was validated by immunoblot. Beads were washed the next day four times with lysis buffer, 10 min each wash rotating at ambient temperature. Bound proteins were eluted from the beads by boiling for 5 min at 100 °C in 30 µL of Laemmli concentrate (120 mM Tris pH 6.8, 60% glycerol, 12% SDS, brilliant phenol blue to

color). About 17% of eluates were reserved for silver stain analysis, while the remainder was prepared for mass spectrometry.

2.2.4 Silver Stain

Eluates were boiled for 5 min at 100 °C with 0.17 M DTT and separated by SDS-PAGE. Gels were fixed overnight in 30% ethanol/10% acetic acid or for a few hours with 50% methanol/12% acetic acid. Gels were washed in 35% ethanol three times for 20 min each, sensitized for 2 min (0.02% $Na_2S_2O_3$ in H_2O), washed three times for 1 min each in H_2O, and stained for 30 min to overnight in Ag staining solution (0.2% $AgNO_3$, 0.076% formalin). After two 5 min rinses in H_2O, gels were developed with 6% $NaCO_3$/0.05% formalin/0.0004% $Na_2S_2O_3$. Development was stopped with 5% acetic acid. Gels were imaged on a white-light transilluminator (UVP).

2.2.5 TMT-MuDPIT Analysis of Interactomes

Immunoprecipitates were prepared for TMT-AP-MS according to standard protocols[28,37]. Only MS grade organic solvents were used during sample preparation. Buffer A is 0.1% formic acid, 5% acetonitrile in water. Buffer B is 0.1% formic acid, 80% acetonitrile in water. Buffer C is 500 mM ammonium acetate in Buffer A. Proteins in eluates were precipitated by methanol/chloroform precipitation and pellets air-dried, followed by resuspension in 1% Rapigest (Waters) in water. Resuspended protein solutions were then brought to 50 µL in 100 mM HEPES, pH 8.0, reduced with10 mM TCEP (Millipore Sigma) for 30 min at 37 °C, alkylated with 5 mM iodoacetamide (Millipore Sigma) for 30 min in the dark and at ambient temperature, and digested overnight at 37 °C with agitation (600 rpm) in the presence of 0.5 µg sequencing grade trypsin (Thermo Fisher

Scientific). TMT isotopic labels (Pierce) were resuspended (100 µg/80 µL acetonitrile) and 40 µL of label was added to each 100 µL sample of digested peptides[38]. Several runs were searched for the presence of unlabeled peptides by allowing TMT labeling as a dynamic modification, and results indicated an unlabeled fraction similar to the false discovery rate. Samples were labeled for 1 hour at ambient temperature, followed by quenching with 0.4% ammonium bicarbonate at ambient temperature for 1 hour. Samples were pooled, acidified, centrifuged for 30 min at 21,100 x g to remove any insoluble debris, and then dried by centrifugal evaporation to 10 µL. Solutions were then brought to 200 µL in Buffer A, incubated at 37 °C for 1 hour, and centrifuged for 30 min. at 21,100 x g to complete elimination of Rapigest. Samples were analyzed using two dimensional LC/MS/MS on an LTQ Orbitrap Velos hybrid mass spectrometer (Thermo) interfaced with an Easy-nLC 1000 (Thermo) according to standard MuDPIT protocols[4]. Triphasic loading columns were prepared by polymerizing a Kasil 1624 frit into a 250 µm diameter fused silica capillary (Agilent) and then packing with 2.5 cm reversed-phase 5 µm Aqua C18 resin (Phenomenex), 2.5 cm 5 µm strong cation exchange resin (Partisphere, GE Healthcare) and again with 2.5 cm reversed-phase 5 µm Aqua C18 resin. Analytical columns were prepared by pulling 100 µm diameter fused silica columns (Agilent) with a P-2000 laser tip puller (Sutter Instrument Co., Novato, CA), followed by packing with at least 15 cm reversed-phase 3 µm Aqua C18 resin (Phenomenex). Analysis was performed using a six-cycle chromatographic run, with progressively increasing ammonium acetate salt bumps injected prior to each cycle (0% C, 25% C, 50% C, 75% C, 100% C, 90% C+10% B; balance of each buffer A), followed

by acetonitrile gradient (5 min from 1% B to 7% B, 60 min to 55% B, 15 min to 100% B, 5 min at 100% B, 5 min to 1% B; 300 nL/min flow rate). Eluted peptides were ionized by electrospray (3.0 kV) and scanned from 110 to 2000 m/z in the Orbitrap with resolution 30000 in data dependent acquisition mode. The top ten peaks with charge states of 2+, 3+, or 4+ from each full scan were fragmented by HCD using a stepped collision energy of 36%, 42%, and 48%, a 100 ms activation time, and a resolution of 7500. Dynamic exclusion parameters were 1 repeat count, 30 ms repeat duration, 500 exclusion list size, 120 s exclusion duration, and 2.00 Da exclusion width. MS/MS spectra were extracted using RAW Xtractor (version 1.1.0.19, available at fields.scripps.edu) without deisotoping and searched using ProLuCID[39,40] against a Uniprot human proteome database (05/05/2016 release) containing 20245 human sequences (longest entry for each protein) concatenated with reverse sequences for each entry as the decoy set, plus 200 select known contaminants (e.g albumen, porcine trypsin, etc.). ProLuCID searches allowed for static modification of cysteine residues (57.02146 Da, acetylation) N-termini, and lysine residues (229.1629 Da, TMT-tagging), half tryptic peptidolysis specificity, and mass tolerance of 20 ppm for precursor mass and 20 ppm for product ion masses**.** Spectra matches were assembled and filtered by DTASelect2 (version 2.0.27)[41]. The stringency of spectral matching was chosen such that <1% of identified peptides were from the decoy database. Decoy proteins, contaminants, and keratins were filtered from the final protein list. Quantitation in Census[42] was performed by averaging TMT reporter ion intensities for all spectra associated with an individual peptide and deconvolution of isotopic impurity as reported in the lot analysis supplied by Thermo Fisher. Only unique

peptides were used for quantification. Empty TMT channels were substituted with a "1", which is well below the threshold value for TMT quantification.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[43] partner repository with the dataset identifier PXD016613 and 10.6019/PXD016613.

2.2.6 Parallel Reaction Monitoring

Human DNAJB8 peptides appropriate for PRM (not including H31) were chosen using the Picky online interface[44,45] (**Table 2.3**). Lysates were prepared for PRM according to standard protocols[28,37]. Lysates were desalted by chloroform/methanol precipitation. Protein pellets were solubilized in 8 M urea in 50 mM Tris pH 8, reduced with 10 mM TCEP in 50 mM Tris pH 8 for 30 min at ambient temperature, alkylated with 5 mM iodoacetamide for 30 min in the dark at ambient temperature, and diluted 3-fold in 50 mM Tris pH 8 to lower the urea concentration to 2 M. Trypsin digestion was performed with 100:1 sequencing grade trypsin 20 h at 37 ºC, 600 rpm.

Digests (20 µg) were injected onto a homemade C18 trapping column for desalting and then separated on a reversed-phase analytical column prior to electrospray ionization. Peptides were analyzed according to a scheduled isolation method using an LTQ Orbitrap Velos Pro. Chromatograms were integrated in Xcalibur. An internal reference peptide[45] (VFFAEDVGSNK) was spiked into each sample and used for normalization. Peptides were scanned over scheduled 5 min windows centered around their average retention time, and isolated with a 2.0 m/z isolation window. Peptides were fragmented with CID with a normalized collision energy of 35, activation Q of 0.25 and activation time of 10

ms. MS2 were acquired in the Orbitrap at 7500 resolution and saved in profile mode. Peptide separations by LC-MS proceeded between Buffer A (5% acetonitrile:95 % water: 0.1% formic acid) and Buffer B (80% acetonitrile: 20% water: 0.1% formic acid) over a 100 minute gradient with the following segments: 1-5 minutes: 1-6% Buffer B. 5-75 minutes: 6-36% Buffer B. 75-80 minutes: 36-100% Buffer B. 80-85 minutes: 100% Buffer B. 85-90 minutes: 100-1% Buffer B. 90-100 minutes: 1% Buffer B.

Table 2.3. Transitions for PRM Analysis of ^{Flag}DNAJB8^{H31Q}.			
Sequence	**Charge State**	**Precursor m/z**	**Retention Time (min.)**
SVMSSTEMINGHK	2+	710.8316	35.04
NPEDIFR	2+	445.722	44.95
LVSEAYEVLSDSK	2+	720.3669	26.19
AGGGASTPYHSPFDTGYTFR	2+	696.9852	38.19

2.2.7 Statistical Methods

Ordinary t-statistics for each protein were determined from $t = (\bar{x} - \bar{y})\sqrt{\frac{n}{s_x^2 + s_y^2}}$ where $\bar{x}$ and $\bar{y}$ are the sample means for TMT reporter ion intensities, s_x and s_y are the respective sample standard deviations, and n is the number of measurements. p-values were then inferred based on the two-tailed Student's t-distribution. Raw TMT intensities ratios were analyzed with the LIMMA package in R to generate moderated t-statistics and p-values[47,48]. This data package transforms TMT reporter ion intensities to a logarithmic scale due to the assumption that fold changes are constant between conditions[49], and then moderates the standard errors for each protein against a global estimated standard error. This moderated standard error is then used to generate a t-statistic and p-values in the standard manner.

Because the Pearson's coefficient of non-correlated data (the null hypothesis) is normally distributed[50], a simple t-statistic can be directly calculated from: $t = R\sqrt{\frac{n-2}{1-R^2}}$ (for derivation see ref 51), where $n - 2$ is the degrees of freedom and R is the correlation coefficient, $r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$. This t-statistic is equivalent to the ratio of the measured slope from the linear fit divided by its standard error. p-values are then inferred by comparing the t-statistic to the two-tailed Student's t-distribution with $n - 2$ degrees of freedom.

q-values (q_{BH}) were determined from p-values using the Benjamini-Hochberg methodology, $q_k = p_k \frac{N}{k}$, where k is the rank for each protein, arranged as increasing p-value, and N is the total number of proteins examined (representing the number of

hypotheses)[52]. These q-values are then adjusted to preserve monotonicity by replacing each q_i with $\min\{q_j|j>i\}$, a transformation that has been shown to offer greater power without sacrificing false discovery rate control[53]. q-values were originally developed to allow construction of a set $\{k \mid q_k < Q\}$ such that the False Discovery Rate of {k} is below Q. However, individual q-values still provide a measure of the local false discovery rate even in the absence of a set value of Q[54].

Box and whisker plots are presented with lines marking median values, X marking average values, boxes from the first to third quartiles, whiskers extending to minimum and maximum values (excluding outliers), and outliers defined at points greater than 1.5-fold the interquartile range beyond the first and third quartiles. These outliers are shown explicitly. Coefficients of Variation are calculated as the standard deviation divided by the mean.

2.3 Results and Discussion

We constructed a simple model (See **Methods 2.2.2**) for the recovery of prey proteins (defined as proteins whose specific recovery is linearly dependent on the bait) and non-interacting proteins (those whose recovery is independent of the bait). Simulations indicate that both Pearson's and Student's t-statistics perform well when bait level variation is low (**Figure 2.1**). However, as bait level variance between replicates increases, Student's derived t-statistics are less able to distinguish prey from non-specific interactors, as compared to Pearson's correlation-derived t-statistics. This finding was robust against a wide range of parameters sets (**Figure 2.2**). Hence, we considered that

perhaps Pearson-derived t-statistics would enable better discrimination of true and false interactors from TMT-AP-MS data.

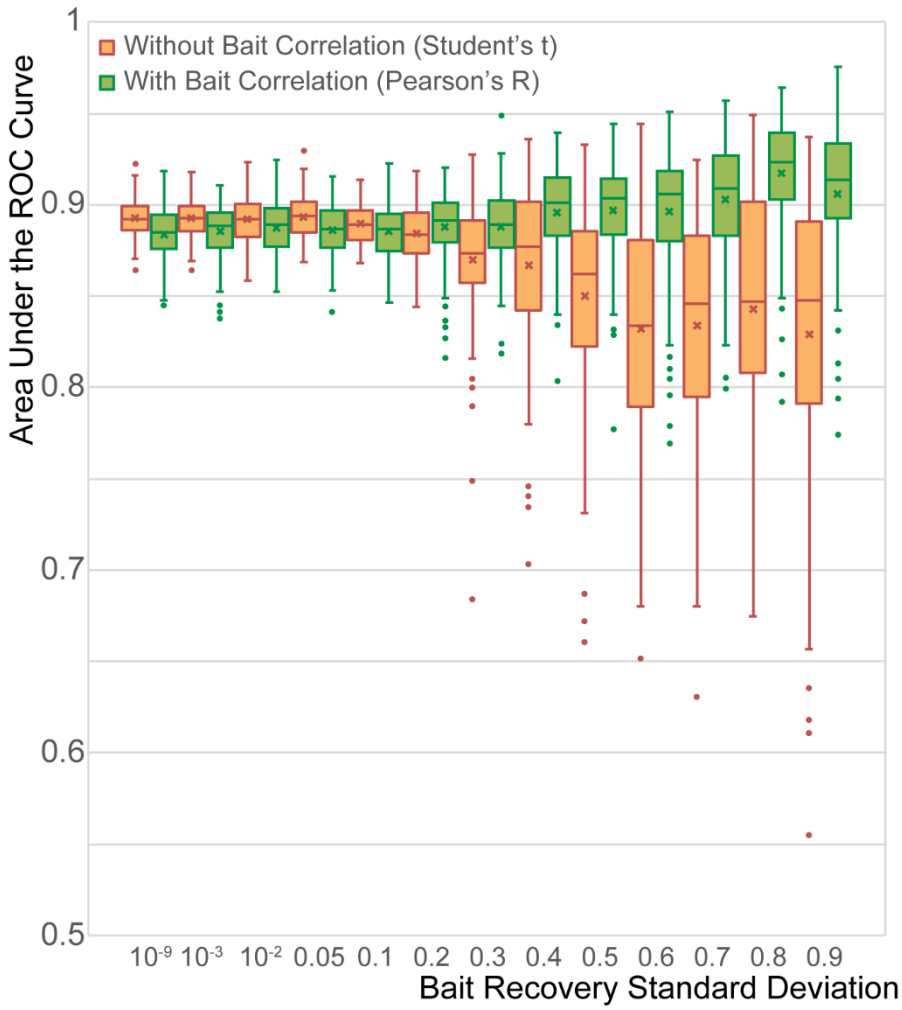


Figure 2.1

Figure 2.1. Box and whisker plots of interactor identification accuracy, as determined using t-statistics derived from either Student's t test or Pearson's correlation, as indicated. Areas Under the Curve (AUCs) for Receiver Operating Characteristic (ROC) Curves, which reflect interactor accuracy, were generated from simulations (100 replicates) of 500 non-specific (non-interactor) and 500 specific (prey) proteins recovered from bait immunoprecipitation under conditions of bait level variation. For each simulation, we generate a ROC curve. This curve compares *sensitivity* (true positive rate) to *1 – specificity* (false positive rate), based on assignment of non-interactors and prey before the simulation[55]. The area under this curve represents the accuracy with which non-specific and specific interactors are distinguished. Areas are averaged across the 100 replicate simulations. The abscissa represents the standard deviation of the distribution from which individual bait levels are drawn. Parameters are provided in **Table 2.1**. At highly reproducible bait recovery levels (low standard deviation), the bait correlation approach (green) slightly underperforms a traditional t-statistic (orange), while as the variability of bait recovery increases, t-statistics derived from Pearson's correlations outperform Student's t-statistics.

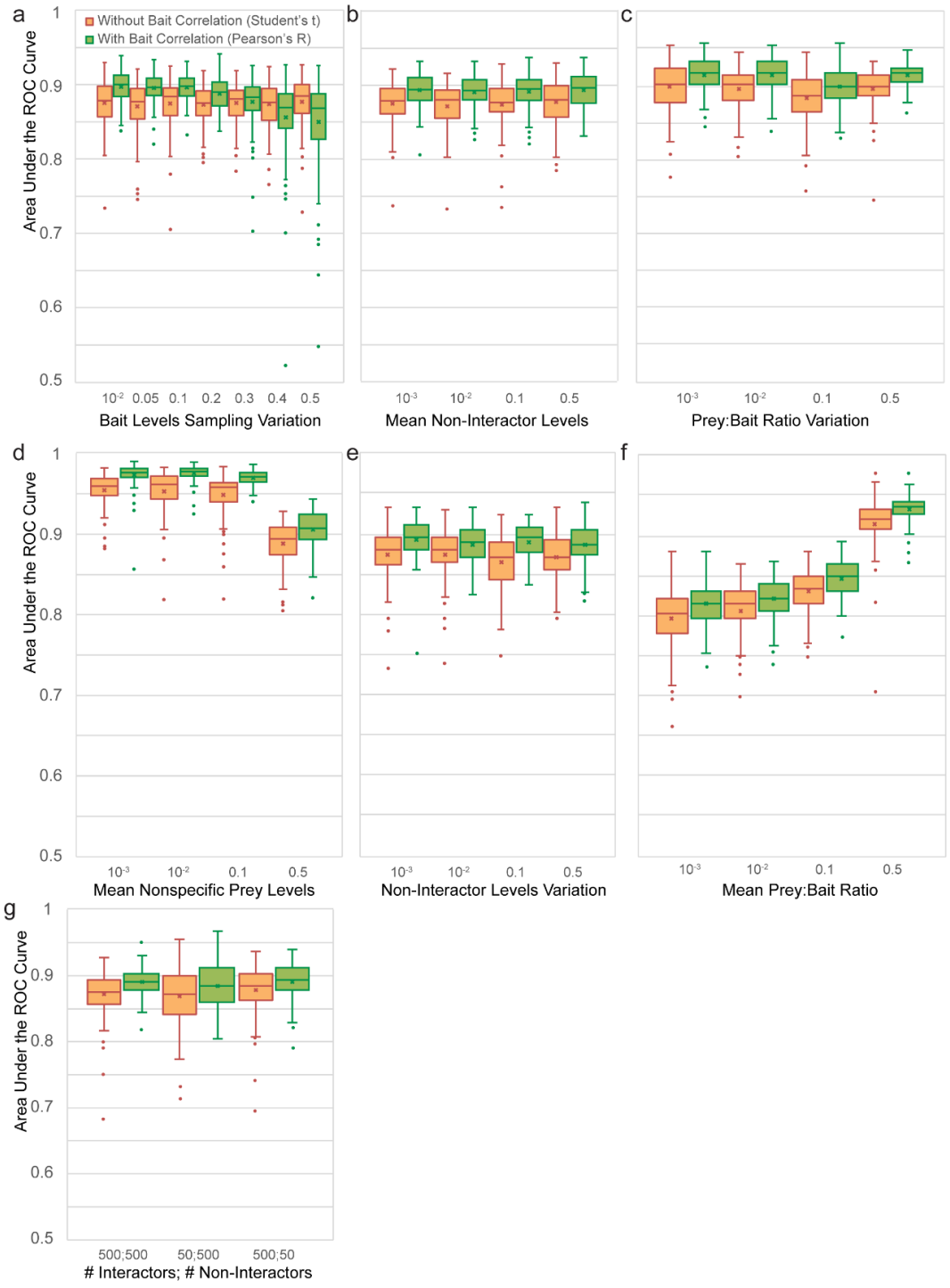

a
b
c
d
e
f
g
Without Bait Correlation (Student's t)
With Bait Correlation (Pearson's R)
Area Under the ROC Curve
1
0.9
0.8
0.7
0.6
0.5
10-2 0.05 0.1 0.2 0.3 0.4 0.5
Bait Levels Sampling Variation
10-3 10-2 0.1 0.5
Mean Non-Interactor Levels
10-3 10-2 0.1 0.5
Prey:Bait Ratio Variation
10-3 10-2 0.1 0.5
Mean Nonspecific Prey Levels
10-3 10-2 0.1 0.5
Non-Interactor Levels Variation
10-3 10-2 0.1 0.5
Mean Prey:Bait Ratio
500;500 50;500 500;50
Interactors; # Non-Interactors

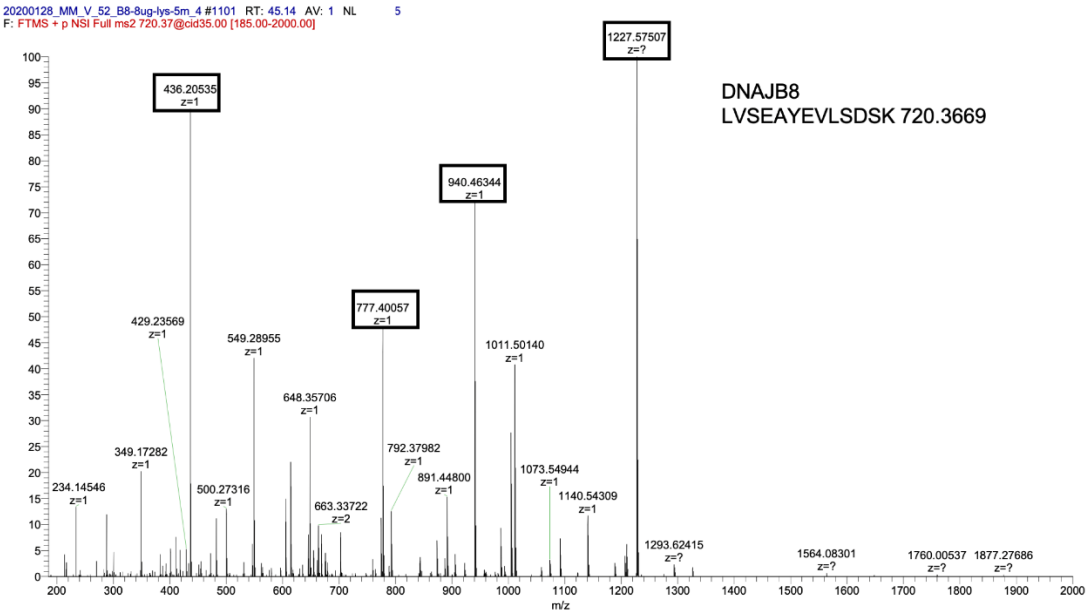
Figure 2.2

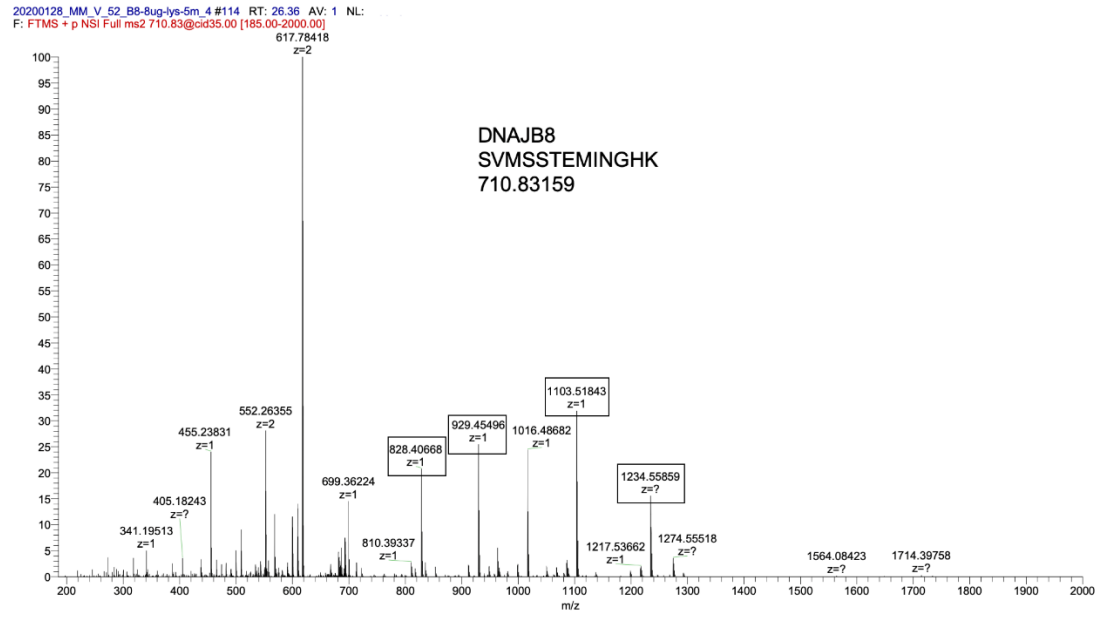
Figure 2.2. Box and whisker plots of the distribution of Areas Under the Curve for Receiver Operating Characteristic Curves generated from simulations (100 replicates) of non-specific (non-interactor) and specific (prey) proteins recovered from bait immunoprecipitation. The following conditions were varied (parameter names in parentheses): Bait Levels Sampling Variation (Bait Levels Channel Standard Deviation), Mean Non-specific Levels of Non-interactors (NonSpecific Noninteract Mean), Prey:Bait ratio Variation (Ratio Standard Deviation), Mean Nonspecific Prey Levels (Non-Specific Prey Mean), Variation in Nonspecific Recovery of Noninteractors (Nonspecific Noninteract Standard Deviation), Mean Prey-to-bait Ratio (Ratio Mean), Number of Non-interactors, and Number of Interactors. Parameter values that were varied are indicated for each plot (abscissa) and calculated with a bait level variance (Bait Levels Standard Deviation) set to 0.3. Other parameters are provided in **Table 2.4**.

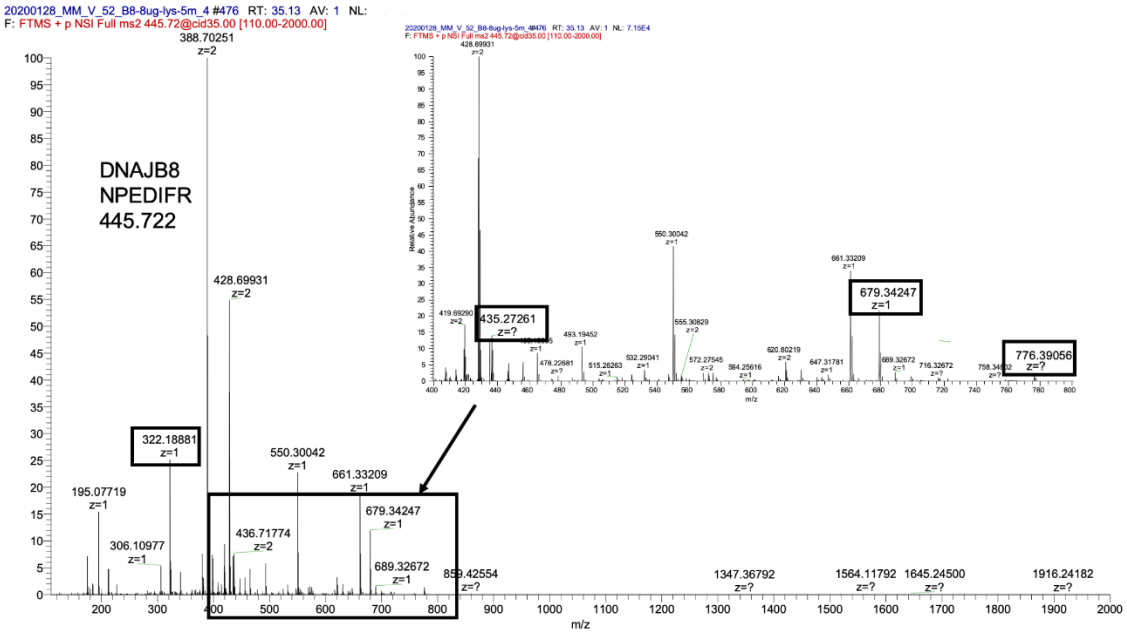
Table 2.4. Parameters for simulations determined from experimental data.			
Parameter	**Symbol**	**Mean (Value in Fig. 2.2)**	**Variance (Value in Fig. 2.2)**
Non-specific Non-interactor Levels	$\mu_i^{ni,ns}$	NonSpecific NoninteractMean (0.02)	NonSpecificNoninteract StandardDeviation (0.02)
Non-specific Non-interactor Levels Sampling Variation	$\epsilon_{ij}^{ni,ns}$	1	NonSpecificNoninteract ChannelStandardDeviation (0.1)
Non-specific Prey Levels	$\mu_k^{prey,ns}$	NonSpecificPreyMean (0.05)	NonSpecificPreyStandard Deviation (0.05)
Non-specific Prey Levels Sampling Variation	$\epsilon_{kj}^{prey,ns}$	1	NonSpecificPreyChannel Standard Deviation (0.1)
Prey:Bait Ratio	R_k	RatioMean (0.2)	RatioStandardDeviation (0.2)
Prey:Bait Ratio Sampling Variation	ϵ_{kj}^{prey}	1	RatioChannelStandard Deviation (0.1)
Non-specific Bait Levels	$\mu^{bait,ns}$	NonSpecificBaitMean (0.01)	NonSpecificBaitStandard Deviation (0.008)
Non-specific Bait Levels Sampling Variation	$\epsilon_j^{bait,ns}$	1	NonSpecificBaitChannel Standard Deviation (0.008)
Bait Levels	μ_j^{bait}	1	BaitLevelsStandard Deviation (indicated)
Bait Levels Sampling Variation	ϵ_j^{bait}	1	BaitLevelsChannel Standard Deviation (0.1)

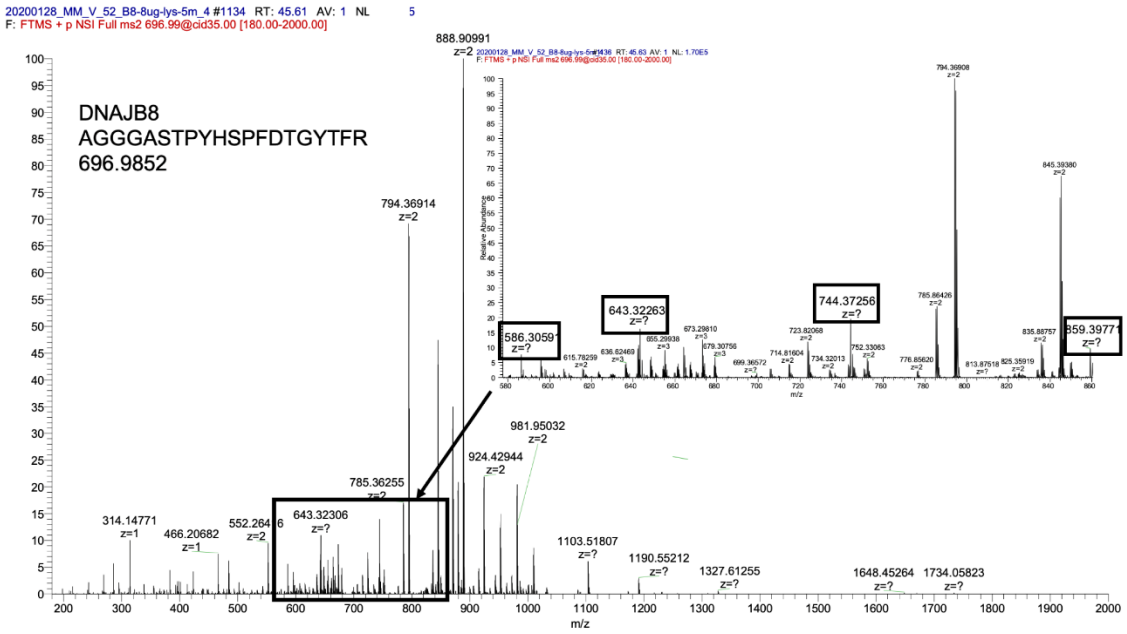
To evaluate this hypothesis, we initially designed a TMT-AP-MS experiment that minimizes non-specific interactors while providing a large number of true interactors. Hsp40 co-chaperones are responsible for recruiting about a third of the proteome to the Hsp70 chaperoning pathway[56]. These protein clients are then handed off from Hsp40 to Hsp70 to promote folding or degradation. Mutation of the Hsp70-binding motif of Hsp40, however, inhibits client release[57]. We constructed ^{Flag}DNAJB8^{H31Q}, where the H31Q mutation blocks association with Hsp70[58]. We further employed the cell permeable crosslinker DSP to immortalize interactions between prey and ^{Flag}DNAJB8^{H31Q} prior to lysis and immunopurification[59]. Stringent washing with the high-detergent buffer RIPA was employed to minimize non-specific interactions with the beads. To evaluate the fidelity of our bait recovery, elution, digestion, and labeling protocol, we transfected three concentrations of ^{Flag}DNAJB8^{H31Q} DNA into HEK293T cells and compared ^{Flag}DNAJB8^{H31Q} TMT bait reporter ion ratios against the amount of ^{Flag}DNAJB8^{H31Q} in the lysates as measured by Parallel Reaction Monitoring (PRM). Both measures provide similar ratios. The two lowest levels of ^{Flag}DNAJB8^{H31Q} indicate similar recovery, while recovery decreases for the highest level of transfected ^{Flag}DNAJB8^{H31Q} (**Figure 2.3 a-c**).

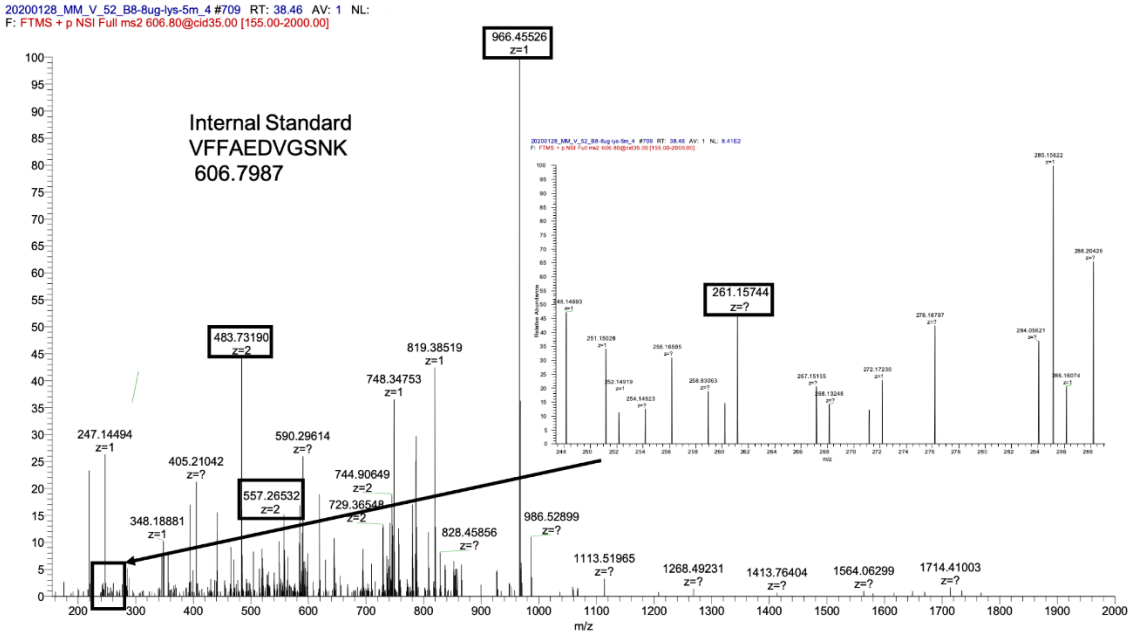
a


20200128_MM_V_52_B8-8ug-lys-5m_4 #1101 RT: 45.14 AV: 1 NL 5
F: FTMS + p NSI Full ms2 720.37@cid35.00 [185.00-2000.00]
DNAJB8
LVSEAYEVLSDSK 720.3669
1227.57507 z=?
436.20535 z=1
940.46344 z=1
777.40057 z=1
429.23569 z=1
549.28955 z=1
1011.50140 z=1
648.35706 z=1
349.17282 z=1
792.37982 z=1
1073.54944 z=1
234.14546 z=1
500.27316 z=1
891.44800 z=1
663.33722 z=2
1140.54309 z=1
1293.62415 z=?
1564.08301 z=?
1760.00537 z=?
1877.27686 z=?
m/z


20200128_MM_V_52_B8-8ug-lys-5m_4 #114 RT: 26.36 AV: 1 NL:
F: FTMS + p NSI Full ms2 710.83@cid35.00 [185.00-2000.00]
DNAJB8
SVMSSTEMINGHK
710.83159
617.78418 z=2
1103.51843 z=1
552.26355 z=2
929.45496 z=1
455.23831 z=1
1016.48682 z=1
828.40668 z=1
1234.55859 z=?
699.36224 z=1
405.18243 z=?
341.19513 z=1
810.39337 z=1
1217.53662 z=1
1274.55518 z=?
1564.08423 z=?
1714.39758 z=?
m/z


20200128_MM_V_52_B8-8ug-lys-5m_4 #476 RT: 35.13 AV: 1 NL:
F: FTMS + p NSI Full ms2 445.72@cid35.00 [110.00-2000.00]
DNAJB8
NPEDIFR
445.722
388.70251 z=2
428.69931 z=2
322.18881 z=1
550.30042 z=1
661.33209 z=1
679.34247 z=1
195.07719 z=1
306.10977 z=1
436.71774 z=2
689.32672 z=1
859.42554 z=?
1347.36792 z=?
1564.11792 z=?
1645.24500 z=?
1916.24182 z=?
435.27261 z=?
776.39056 z=?
Relative Abundance
m/z


20200128_MM_V_52_B8-8ug-lys-5m_4 #1134 RT: 45.61 AV: 1 NL
F: FTMS + p NSI Full ms2 696.99@cid35.00 [180.00-2000.00]
DNAJB8
AGGGASTPYHSPFDTGYTFR
696.9852
888.90991 z=2
794.36914 z=2
981.95032 z=2
924.42944 z=2
785.36255 z=2
643.32306 z=?
314.14771 z=1
466.20682 z=1
552.26416 z=2
1103.51807 z=?
1190.55212 z=?
1327.61255 z=?
1648.45264 z=?
1734.05823 z=?
586.30591 z=?
643.32263 z=?
744.37256 z=?
859.39771 z=?
Relative Abundance
m/z


20200128_MM_V_52_B8-8ug-lys-5m_4 #709 RT: 38.46 AV: 1 NL:
F: FTMS + p NSI Full ms2 606.80@cid35.00 [155.00-2000.00]
Internal Standard
VFFAEDVGSNK
606.7987
966.45526
z=1
483.73190
z=2
819.38519
z=1
748.34753
z=1
247.14494
z=1
590.29614
z=?
405.21042
z=?
557.26532
z=2
744.90649
z=2
729.36548
z=2
348.18881
z=1
828.45856
z=?
986.52899
z=?
1113.51965
z=?
1268.49231
z=?
1413.76404
z=?
1564.06299
z=?
1714.41003
z=?
261.15744
z=?
m/z

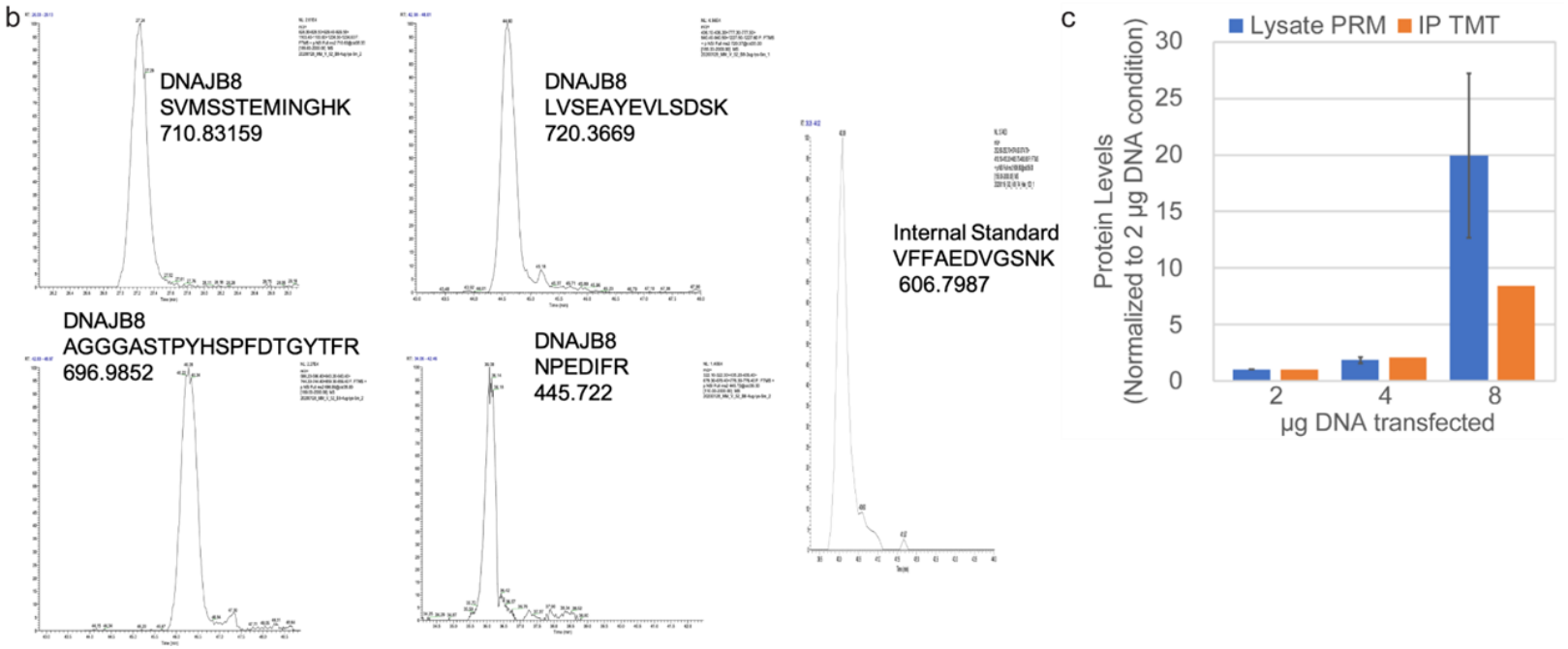
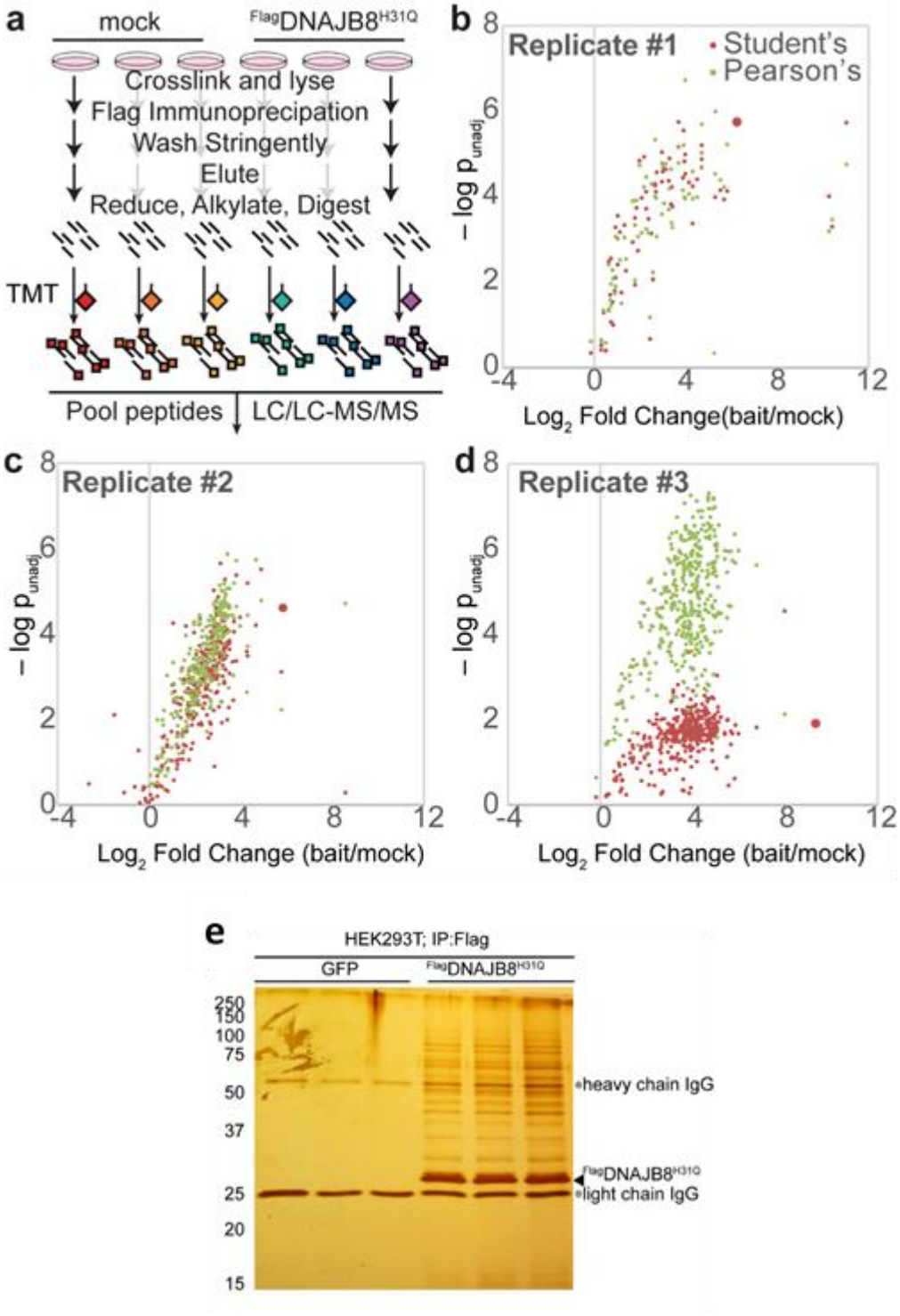


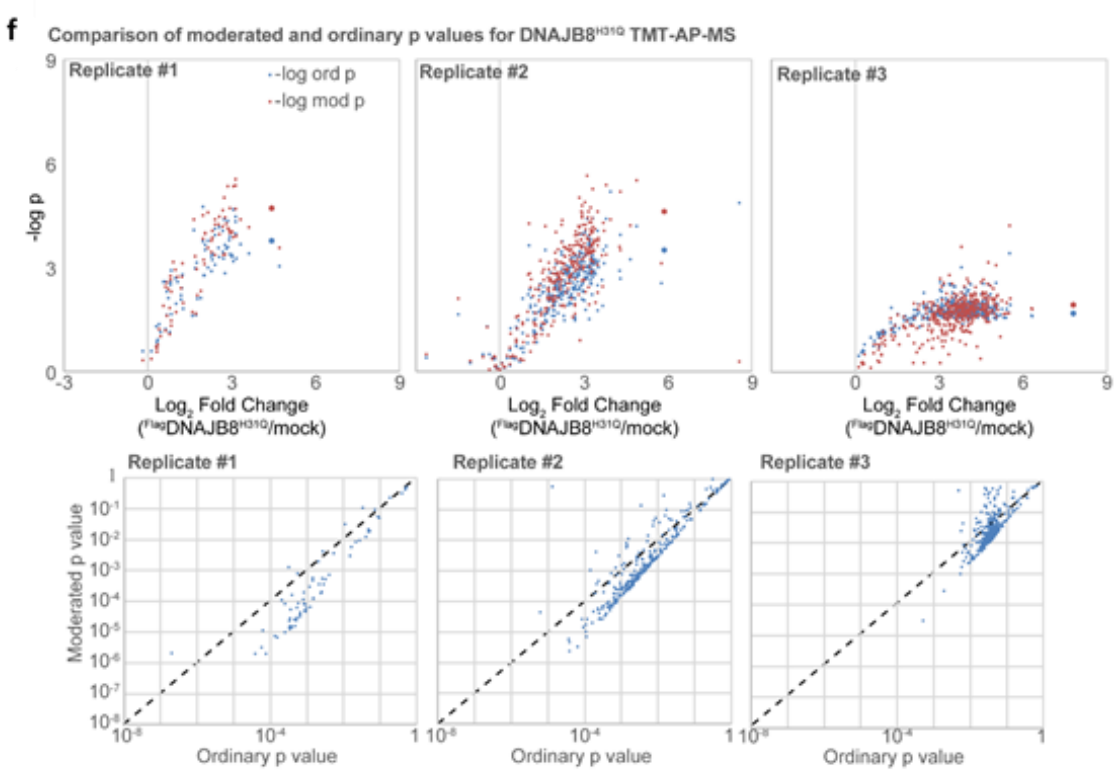
Figure 2.3

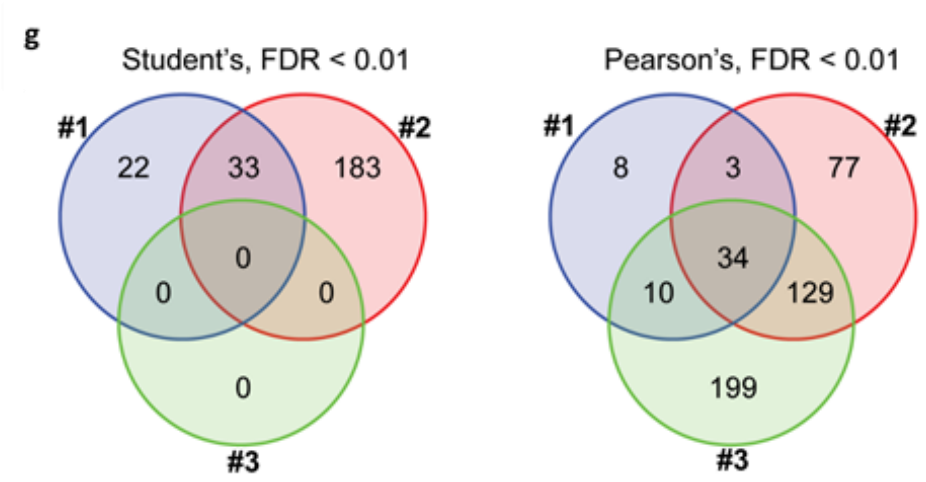
Figure 2.3. a) MS2 spectra for ^{Flag}DNAJB8^{H31Q} peptides used for parallel reaction monitoring (PRM). **b)** Representative PRM chromatograms of ^{Flag}DNAJB8^{H31Q} peptides. **c)** Plot comparing the TMT reporter ion ratios for eluted ^{Flag}DNAJB8^{H31Q} to the ^{Flag}DNAJB8^{H31Q} levels in the lysates as ascertained by PRM. Cells were transfected with 2 µg, 4 µg, or 8 µg DNA encoding ^{Flag}DNAJB8^{H31Q}, crosslinked with DSP, quenched, and lysed. Aliquots of lysate were processed for PRM mass spectrometry. The remaining lysate was normalized for total protein and loaded onto M2 anti-FLAG magnetic Dynabeads for immunoprecipitation. Beads were washed well with RIPA buffer prior to elution. Eluates were processed for mass spectrometry and labeled with TMT tags. Standard deviations for PRM are across peptides.

We generated three replicate six-plex sets as follows. Nine plates each of HEK293T cells were transfected with either bait (^{Flag}DNAJB8^{H31Q}) or mock (GFP) (**Figure 2.4a**). After DSP crosslinking, lysing and immunopurification, the eluates were reduced, alkylated, digested, and TMT-labeled. TMT-labeled peptides were pooled in six-plex to yield three replicates that were then analyzed by shotgun proteomics (**Figure 2.4 b-d**, **e**). Note that here, "replicate" refers to a single run that includes six independent "samples". Student's t-test, followed by Benjamini-Hochberg analysis[52] to determine q-values (q_{BH}) and false discovery rates (FDR), was applied to determine likely DNAJB8^{H31Q} interactors. When the number of samples is low, ordinary t-tests suffer from poor estimation of variance. This estimate is improved by moderating the variance of each individual protein's integrated TMT ion intensity with the global variance[47,60]. Consistent with prior reports[60], we find that moderation slightly decreases the p-value for most proteins, while sharply increasing the p-value for proteins featuring anomalously low variance (**Figure 2.4 f**). The first two replicates yield several dozen significant interactors of DNAJB8^{H31Q} (using a threshold of FDR < 1%), and the second replicate captures 60% (33 out of 55) of the interactors identified in the first replicate (**Figure 2.4 g**). In the third replicate, however, no significant interactors are identified. This is due to higher bait level variance; the coefficient of variation of the bait in this replicate is 60% as opposed to <15% for the first two replicates. Even the DNAJB8^{H31Q} *bait* is not significantly different between the mock and bait transfection conditions (**Figure 2.4 d**). Rather, in each replicate the p-value for the bait itself (large points in **Figure 2.4 b-d**) defines the approximate lower limit for what p-values are calculated for the various prey. The cause of this variance could be due

to any number of factors: variance in efficiencies of transfection, lysis, immunoprecipitation, elution, and digestion. When we use Pearson's correlation to derive t-statistics, we find little change in p-values for Replicates #1 and #2, where bait variability is low (**Figure 2.4 b**, **c**, **h**). ^{Flag}DNAJB8^{H31Q} interactors in Replicate #3, however, have far lower p-values when derived from Pearson's correlation than from a Student's t-test (**Figure 2.4 d**, **h**). The q_{BH} values for the high variability sample are similarly decreased when t-statistics are generated from using Pearson's correlation (**Figure 2.4 i**). Now, 92% (34 out of 37) of prey shared between Replicates #1 and #2 appear in the set of Replicate #3 prey with FDR < 1% (**Figure 2.4 g**). There is little change in the overlap between Replicates #1 and #2. Of the 33 proteins that fall below the 1% threshold for both Replicates #1 and #2 using the Student's t-statistic, 32 fall below the threshold for both replicates when using the Pearson's correlation derived t-statistic. This demonstrates that using Pearson's correlation to determine t-statistics from TMT-AP-MS data can account for bait level variation, as predicted by our simulations (**Figure 2.1**).


a
mock
FlagDNAJB8H31Q
Crosslink and lyse
Flag Immunoprecipation
Wash Stringently
Elute
Reduce, Alkylate, Digest
TMT
Pool peptides
LC/LC-MS/MS
b
Replicate #1
Student's
Pearson's
– log punadj
Log2 Fold Change(bait/mock)
c
Replicate #2
– log punadj
Log2 Fold Change (bait/mock)
d
Replicate #3
– log punadj
Log2 Fold Change (bait/mock)
e
HEK293T; IP:Flag
GFP
FlagDNAJB8H31Q
250
150
100
75
50
37
25
20
15
heavy chain IgG
FlagDNAJB8H31Q
light chain IgG


f
Comparison of moderated and ordinary p values for DNAJB8H31Q TMT-AP-MS
Replicate #1
Replicate #2
Replicate #3
-log ord p
-log mod p
-log p
Log2 Fold Change
(FlagDNAJB8H31Q/mock)
Moderated p value
Ordinary p value


g
Student's, FDR < 0.01
Pearson's, FDR < 0.01
#1
#2
#3
22
33
183
0
0
0
0
8
3
77
34
10
129
199

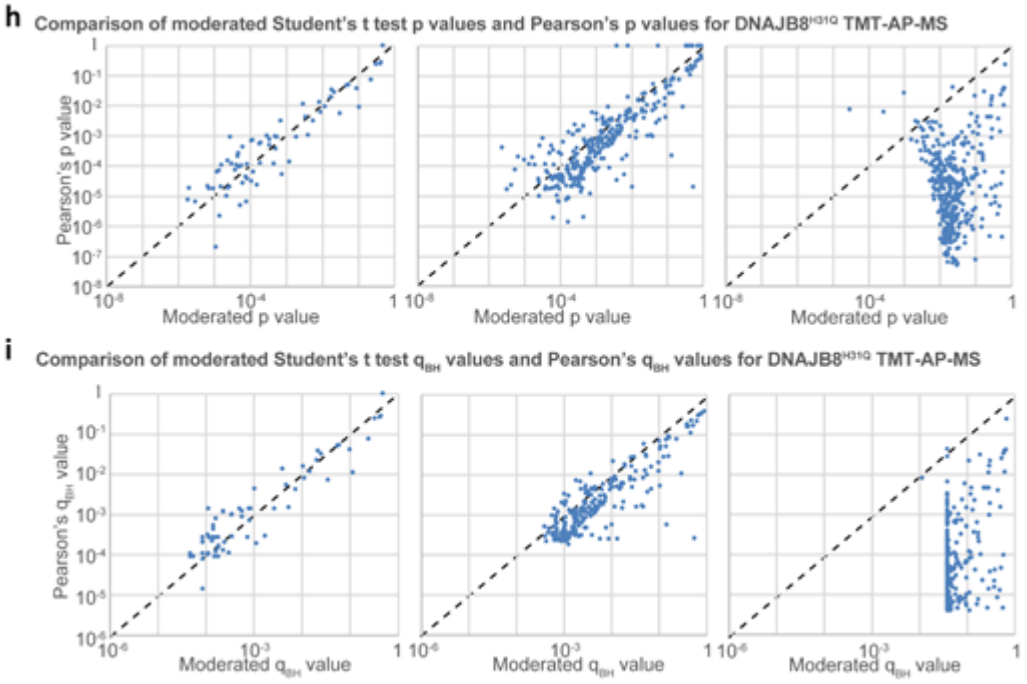


Figure 2.4

Figure 2.4. **a**) Schematic of TMT-AP-MS to characterize DNAJB8^{H31Q} interactors. Cells were crosslinked in 1 mM DSP for 30 min. and quenched with Tris buffer prior to lysis. **b-d**) Volcano plots for DNAJB8^{H31Q} interactome replicates collected as described in panel **a**, with unadjusted p-values determined from unpaired, two-way moderated Student's t (top) and Pearson's correlation (bottom). The large point indicates the bait, using the p-value determined from Student's t analysis **e)** Silver stained SDS-PAGE gel of eluates from three mock (GFP) and three ^{Flag}DNAJB8^{H31Q} transfected samples. Cells were treated with crosslinker prior to lysis, and immunoprecipitates were washed with high stringency RIPA buffer prior to elution. **f**) Volcano plots for ^{Flag}DNAJB8^{H31Q} immunoprecipitations with p-values determined by ordinary (blue) and moderated (red) Student's t-tests. Direct comparison of p-values are plotted below, with the dotted line demonstrating unity. **g**) Comparison of identified prey (FDR < 0.01) between replicates for the two analysis methods, with FDR determined from the method of Benjamini and Hochberg[52]. **h**) Direct comparison of p-values for moderated Student's t test and Pearson's correlation-based t test. The dotted line demonstrates unity. **i**) Direct comparison of q_{BH} values for moderated Student's t test and Pearson's correlation-based t test. The dotted line demonstrates unity.

We hypothesized that deliberately varying the bait level and then determining significance by employing Pearson-derived t-statistics might further improve interactor identification. Although this approach loses the ability to determine "fold change" between mock and bait expression conditions, this fold change reflects non-specific interactions of a potential prey with beads as much as it reflects specific interactions with bait, and hence is not inherently useful for determining meaningful interactions. To test our hypothesis, we transfected DNA encoding ^{Flag}DNAJB8^{H31Q} over a range of concentrations (0 µg, 2 µg, 4 µg, 6 µg, 8 µg, 10 µg DNA per 10 cm dish) into HEK293T cells, and quantified interactors by TMT-AP-MS for three replicates (**Figure 2.5 a**, **b**). For each replicate and at the highest DNA concentration, the measured amount of ^{Flag}DNAJB8^{H31Q} by TMT reporter ion ratios decreases as compared to the amount of ^{Flag}DNAJB8^{H31Q} protein at lower DNA concentrations, suggesting that the cell does not well handle this high amount of ^{Flag}DNAJB8^{H31Q}. Surprisingly, this dosing strategy yields similar but not superior results to the more traditional bait vs. mock approach (**Figure 2.5 a**, **c**). Nevertheless, a larger fraction of prey was identified as significant in each run for the dosing approach as opposed to the bait vs. mock approach (**Figure 2.5 d**), indicating that dosing might be valuable when robustness across data sets is key.

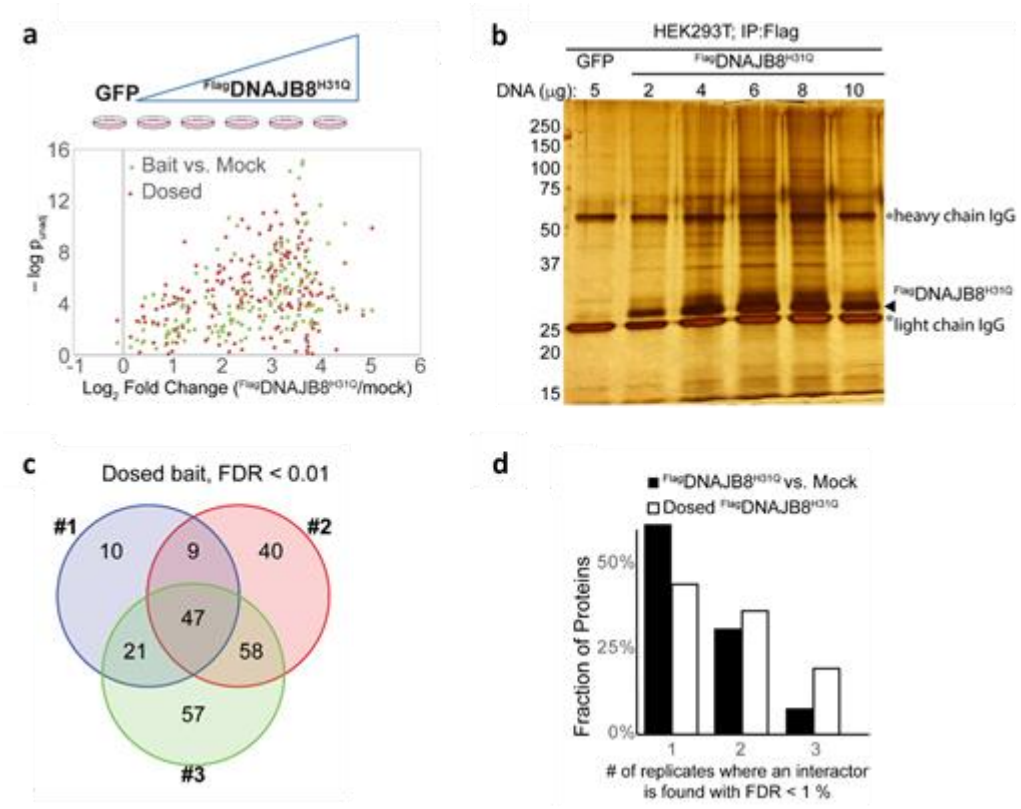


Figure 2.5

Figure 2.5. Comparison of p-values obtained through Pearson's correlation analysis from experiments analyzing three bait vs. three mock conditions, or intentional bait dosing for DNAJB8^{H31Q} AP-MS. **a**) Transfection schematic for the dosed bait approach (upper), and comparison of unadjusted p-values obtained from either the combined bait vs. mock experiments (n = 3) and the dosed bait approach (n = 3). **b**) Silver stained SDS-PAGE gel of eluates from one mock (GFP) sample and five samples dosed with increasing concentration of ^{Flag}DNAJB8^{H31Q}. Cells were treated with crosslinker prior to lysis, and immunoprecipitates were washed with high stringency RIPA buffer prior to elution. **c**) Comparison of identified prey (FDR < 0.01) between replicates. **d**) Histogram of the fraction of identified prey that was found in the given number of replicates.

In our simulated data (**Figure 2.1**), accuracy of interactor assignment using Pearson-derived t-statistics increased with higher bait variation, but we did not observe any such improvement in practice, perhaps due to a loss of the assumed linearity between recovered prey and bait levels. Within each replicate, we normalized the interactor:bait ratio for each recovered interactor to the ratio with the highest levels of DNAJB8^{H31Q} by TMT intensity. If an interactor varies linearly with bait levels over this concentration range, then we would expect this ratio to remain constant. Instead, we find that the mean interactor:bait ratio decreases with increasing bait levels, at least for the lower levels of bait (**Figure 2.6**). This non-linearity is severe enough to suppress bait-interactor correlation and explains why deliberately dosing bait over a wide range of levels does not improve interactor t-statistics. In molecular terms, it is likely that at higher levels of bait expression, the bait begins to saturate endogenous interactors, thus decreasing the interactor:bait ratio.

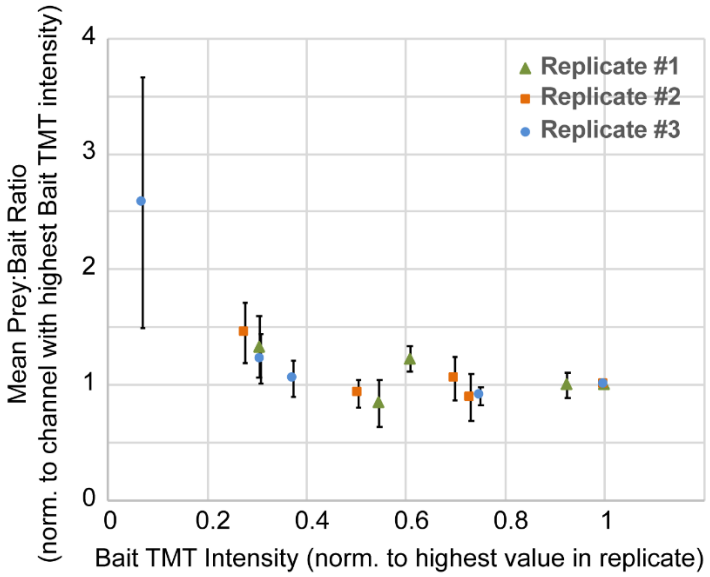


Figure 2.6

Figure 2.6. Average (mean) interactor-to-bait TMT reporter ion intensity ratios for all identified interactors (FDR < 0.01) from bait dosing experiments (as in **Figure 2.5 a**), normalized to the channel with the highest levels of DNAJB8^{H31Q} as determined from TMT intensities. The abscissa provides the Bait TMT Intensity for each channel, normalized within replicates to the TMT channel with the highest DNAJB8^{H31Q} intensity. Averages are performed across all interactors, and then across the three replicates. Error bars represent standard deviation across all prey (#1: n = 87; #2: n = 154; #3: n = 184).

A lack of a validated DNAJB8 interactor data set makes it challenging to judge the accuracy of our discovered interactors. However, as an Hsp40 chaperone, DNAJB8 should associate with a relatively destabilized proteome. To evaluate that possibility, we compared our results to a recently published data set from Walker et al. that directly measured proteome-wide $\Delta G_{folding}$ by Stability of Proteins by Rate of Oxidation (SPROX)[61]. In SPROX, H_2O_2 is added to cell lysate, and the relative oxidation of methionines to the sulfoxide is determined by quantitative shotgun proteomics[62]. Methionine oxidation of a peptide reveals the extent to which that peptide is solvent exposed. If the dependence of methionine oxidation on chaotrope concentration follows a 2-state transition, then $\Delta G_{folding}$ can be inferred. We divided the identified proteins from DNAJB8^{H31Q} immunoprecipitates into likely interactors ($q_{BH} < 0.01$; 476 proteins) and less reliable interactors ($q_{BH} > 0.01$; 159 proteins) on the basis of their Benjamini-Hochberg q-values. Out of these, 163 and 46 proteins respectively had stabilities reported in Walker et al. The likely DNAJB8^{H31Q} interactors are significantly destabilized compared to the less reliable DNAJB8 ^{H31Q} interactors (**Figure 2.7 a**). The converse holds as well; the $\Delta G_{folding}$ for peptides found in our DNAJB8 interactors is significantly less negative than for the peptides not found in our interactors (**Figure 2.7 b**). Hence, the interactors that we observe co-immunoprecipitating with DNAJB8^{H31Q} represent a more destabilized proteome, consistent with DNAJB8^{H31Q}'s functional role as a chaperone for misfolded protein.

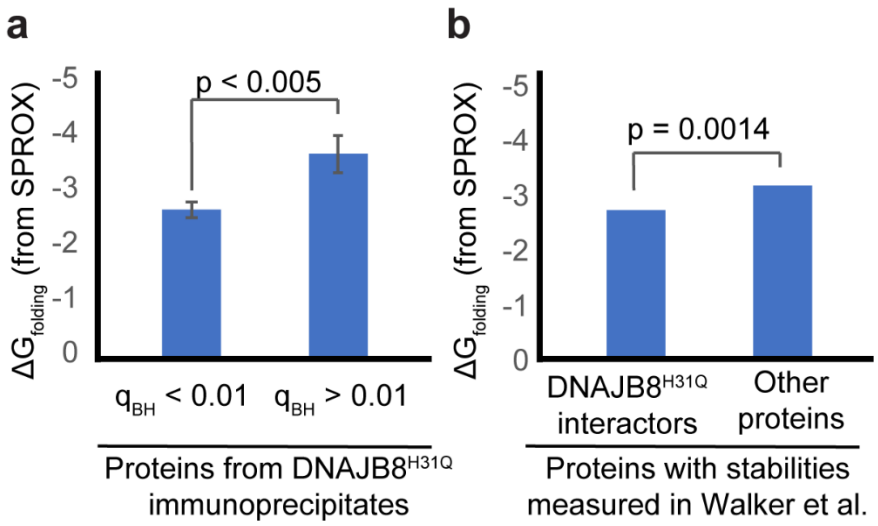


Figure 2.7

Figure 2.7. SPROX-derived $\Delta G_{folding}$ for the DNAJB8^{H31Q}-interacting proteome. **a**) Mean $\Delta G_{folding}$ for likely DNAJB8^{H31Q} -interacting prey compared to the rest of the identified proteome in immunoprecipitates. Likely prey are those with a $q_{BH} < 0.01$. **b)** Mean $\Delta G_{folding}$ for likely ($q_{BH} > 0.01$) DNAJB8^{H31Q} -interacting prey as compared to all other proteins with $\Delta G_{folding}$ identified by SPROX. Data were assessed by two-tailed Student's t-test, and error bars represent standard deviations.

We considered whether less stringent AP-MS conditions, wherein a substantial number of identified proteins will be non-specific interactors, would affect the performance of bait-prey correlation in identifying specific prey. Towards this end, we chose a bait protein that reversibly binds its clients, 14-3-3ζ[64]. To increase the number of potential Type I errors, we did not crosslink and only washed with a gentle, low-detergent wash buffer. Twelve plates each of HEK293T cells were transfected with either ^{Flag}14-3-3ζ or mock (GFP) and immunoprecipitated with anti-Flag beads. Eluates were reduced, alkylated, digested, and peptides TMT-labeled. These peptides were pooled to generate four replicates that each contained 3 samples with bait and 3 mock, and each sample was analyzed by MuDPIT LC-MS/MS. Fewer interactors in general are observed in each individual replicate, as compared to the ^{Flag}DNAJB8^{H31Q} pulldowns (**Figure 2.8 a-e**). There is no global change in interactor p-values between the Pearson and Student approaches, nor is there a change in the number of proteins passing the FDR < 0.01 threshold (**Figure 2.9 a**). Unlike DNAJB8, there are many 14-3-3ζ interaction sets in the literature[65,66]. We generated a true positive interactor list from BioGrid, requiring that proteins be recovered with 14-3-3ζ in at least three different AP-MS studies or at least two AP-IB studies (58 total proteins, or which 19 were found in at least two of our runs)[22,63]. To generate the true negative interactor list, we considered proteins that were reported as significant Dynabead contaminants from total human cell lysate in the Contaminant Repository from Affinity Proteomics (CRAPome)[21], using a filter of appearing in at least 8/24 reported control runs and removing known 14-3-3ζ interactors

(983 total proteins, of which 55 are observed in at least 2 of our runs). Most proteins without filtering appear in three or fewer runs, with a steep drop as the threshold is increased. We chose 8/24 as the threshold based on the inflection point in the cumulative distribution function relating proteins observed vs. number of CRAPome runs. Using these sets, we generated Receiver Operator Characteristic Curves for each replicate AP-MS experiment, using either Student's t-statistics, or incorporating bait correlation by using Pearson's-derived t-statistics (**Figure 2.9 b**). While bait correlation had no effect on the Area Under the Curve for two replicates, it substantially improved differentiation between false and true positives for the other two replicates (**Figure 2.9 c**). Not surprisingly, given our findings with DNAJB8^{H31Q} (**Figure 2.5**), dosing in variable levels of bait did not improve prey identification as opposed to the bait vs. mock experimental setup. We further considered combining all 14-3-3ζ replicates to identify consistent high-quality interactors (Bonferonni-adjusted p-value < 0.001; 34 proteins) and unlikely interactors (unadjusted p-value > 0.7; 42 proteins). Here, we are using Bonferonni adjustment because it is a highly conservative metric. With this set of interactors/non-interactors, Receiver Operating Curves show larger areas under the curve when evaluated using Pearson's correlation-derived t-statistics as opposed to Student's t-statistics (**Figure 2.9 d**, **e**). Finally, as with DNAJB8^{H31Q}, the dosing approach yields more reproducible high-confidence interactors between replicates than the more traditional bait vs. mock approach (**Figure 2.9 f**).

The large number of non-interactors in the 14-3-3ζ TMT-AP-MS experiments allow us to estimate the mean and standard deviation of non-interactor TMT reporter ion ratios.

Nonspecific levels of bait TMT reporter ion ratios in non-transfected mock samples, and both levels and distributions of prey in mock samples without bait transfection, were similarly determined from the DNAJB8^{H31Q} TMT-AP-MS experimental data (see **Methods 2.2.7**). We used this experimental data to test the normality assumption underlying the data in Figure 1, finding that non-interactor mean intensities and prey-bait ratios are both close to a truncated normal distribution (**Figures 2.10 a**, **b**). Interestingly the mean intensities of DNAJB8^{H31Q} interactors in the bait-free conditions deviate sharply below what would be expected from a truncated normal distribution (**Figures 2.10 c**). High levels of protein in one TMT channel can allow quantification of that protein in other channels, even if the levels in those channels would not normally be adequate for data dependent isolation or quantification[67]. In this case, high levels of interactors in the bait pull-downs could be allowing quantification of negligible levels in the untransfected samples. Using these measured parameters, we revisited the simulations of **Figure 2.1** to determine whether under realistic parameters we still observe improvement in prey identification by using bait correlation. The use of Pearson's correlation-derived t-statistics continues to outperform Student's t-statistics over a wide range of bait level variances (**Figure 2.10 d**).

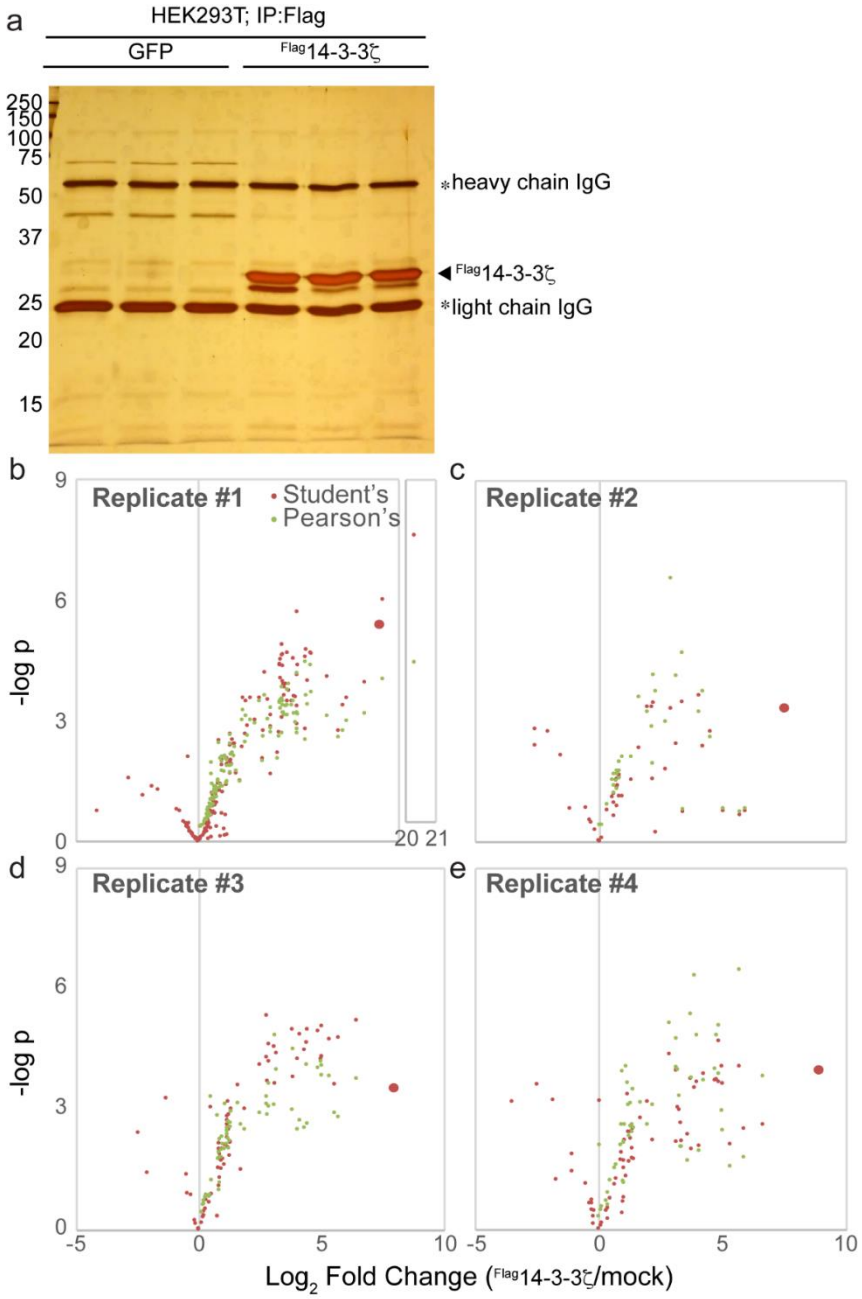


Figure 2.8

Figure 2.8. **a**) Representative silver stain of SDS-PAGE separated Flag immunoprecipitates from HEK293T cells overexpressing mock (GFP) or ^{Flag}14-3-3ζ as indicated. Immunoprecipitates were washed with gentle (low detergent) buffer. **b-d**) Volcano plots for ^{Flag}14-3-3ζ interactome replicates, with unadjusted p-values determined from unpaired, two-way moderated Student's t and Pearson's correlation.

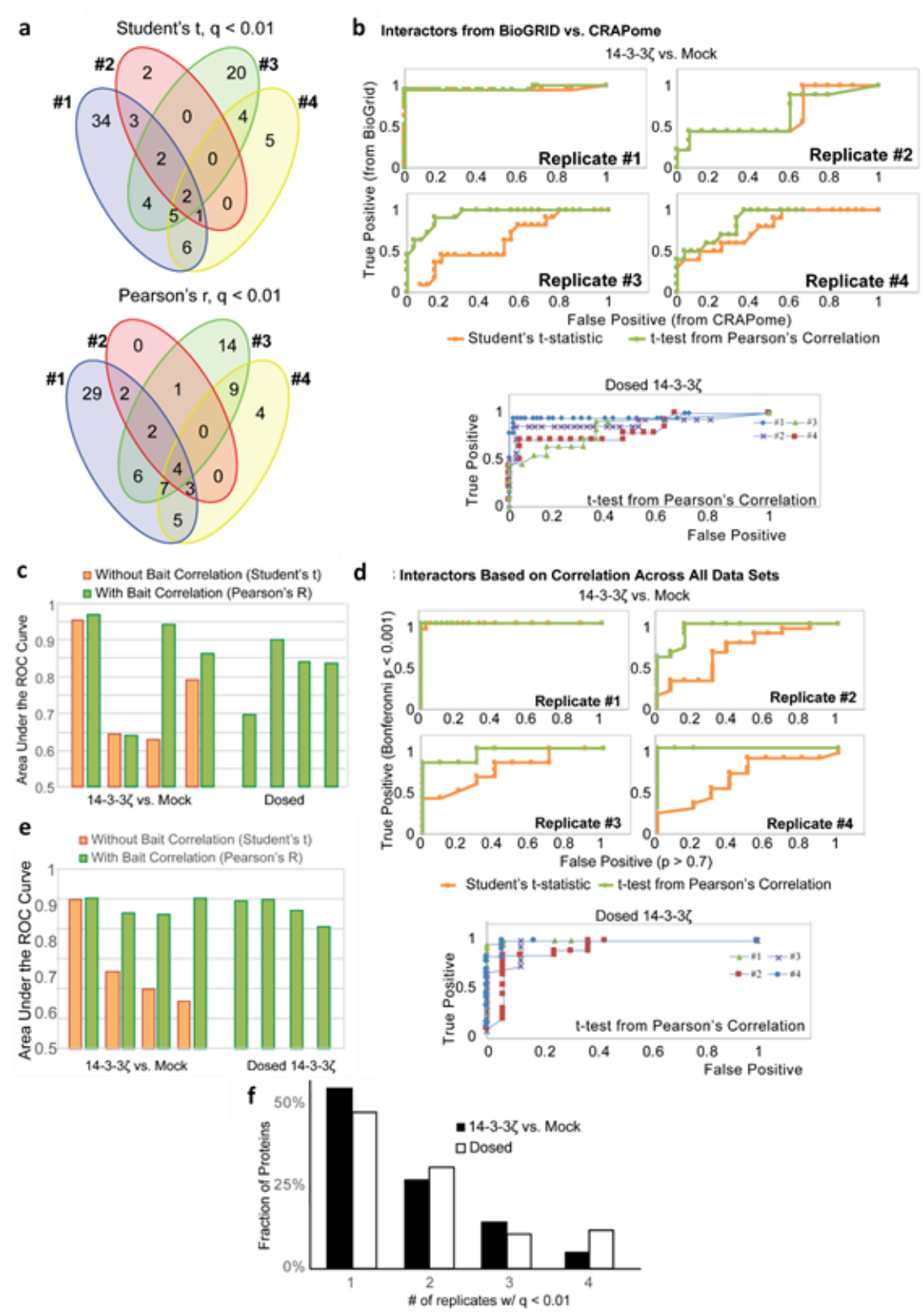

a
Student's t, q < 0.01
#1
#2
#3
#4
34
3
2
20
4
5
0
2
0
4
5
1
0
6
Pearson's r, q < 0.01
#1
#2
#3
#4
29
2
0
14
9
4
1
2
0
6
7
4
3
0
5
b
Interactors from BioGRID vs. CRAPome
14-3-3ζ vs. Mock
True Positive (from BioGrid)
Replicate #1
Replicate #2
Replicate #3
Replicate #4
False Positive (from CRAPome)
Student's t-statistic
t-test from Pearson's Correlation
Dosed 14-3-3ζ
True Positive
t-test from Pearson's Correlation
False Positive
c
Without Bait Correlation (Student's t)
With Bait Correlation (Pearson's R)
Area Under the ROC Curve
14-3-3ζ vs. Mock
Dosed
d
Interactors Based on Correlation Across All Data Sets
14-3-3ζ vs. Mock
True Positive (Bonferonni p < 0.001)
Replicate #1
Replicate #2
Replicate #3
Replicate #4
False Positive (p > 0.7)
Student's t-statistic
t-test from Pearson's Correlation
Dosed 14-3-3ζ
True Positive
t-test from Pearson's Correlation
False Positive
e
Without Bait Correlation (Student's t)
With Bait Correlation (Pearson's R)
Area Under the ROC Curve
14-3-3ζ vs. Mock
Dosed 14-3-3ζ
f
Fraction of Proteins
14-3-3ζ vs. Mock
Dosed
# of replicates w/ q < 0.01

Figure 2.9

Figure 2.9. a) Comparison of identified prey ($q_{BH} < 0.01$) between replicates for the two analysis methods. **b**) Receiver Operating Characteristic (ROC) Curves for ^{Flag}14-3-3ζ immunoprecipitations using true interactors as determined from BioGRID[63]. True positive interactors were proteins that were found in at least three different AP-MS studies or at least two AP-IB studies. True negative interactors were total human cell lysate proteins identified as significant Dynabead contaminants reported in the CRAPome, excluding interactors from BioGRID[21]. Replicate numbers for the two statistical approaches to analyzing the Bait vs. Mock experiments correspond to the same experiment. **c**) Comparison of Student's and Pearson's-derived t-statistics for differentiating true and false positive interactors of 14-3-3ζ from TMT-AP-MS. Areas under the Curve (AUCs) for Receiver Operating Characteristic Curves for 14-3-3ζ interactors identified by Student's (orange) and Pearson's-derived (green) t-statistics. Four replicates are shown for bait vs. mock experiments (as in **Figure 2.4 a**) and bait dosing experiments (as in **Figure 2.5 a**). Higher AUCs indicate higher accuracy at distinguishing true (from BioGrid[63]) and false (from CRAPome[21]) interactors. **d**) ROC Curves for ^{Flag}14-3-3ζ immunoprecipitations. For each curve, the high confidence interactions (Bonferonni-adjusted p-value <0.001 over all replicate runs) were taken as the true interactors, while proteins with an unadjusted p-value above 0.7 were taken as false prey. Replicate numbers for the two statistical approaches to analyzing the Bait vs. Mock experiments correspond to the same experiment. **e**) Bar graph illustrating Area Under the Curve for ROC Curves in **Figure 2.8**. Four replicates of each Bait vs. Mock and Dosed experiments are shown. **f)** Histogram of the fraction of identified prey that was found in the given number of replicates.

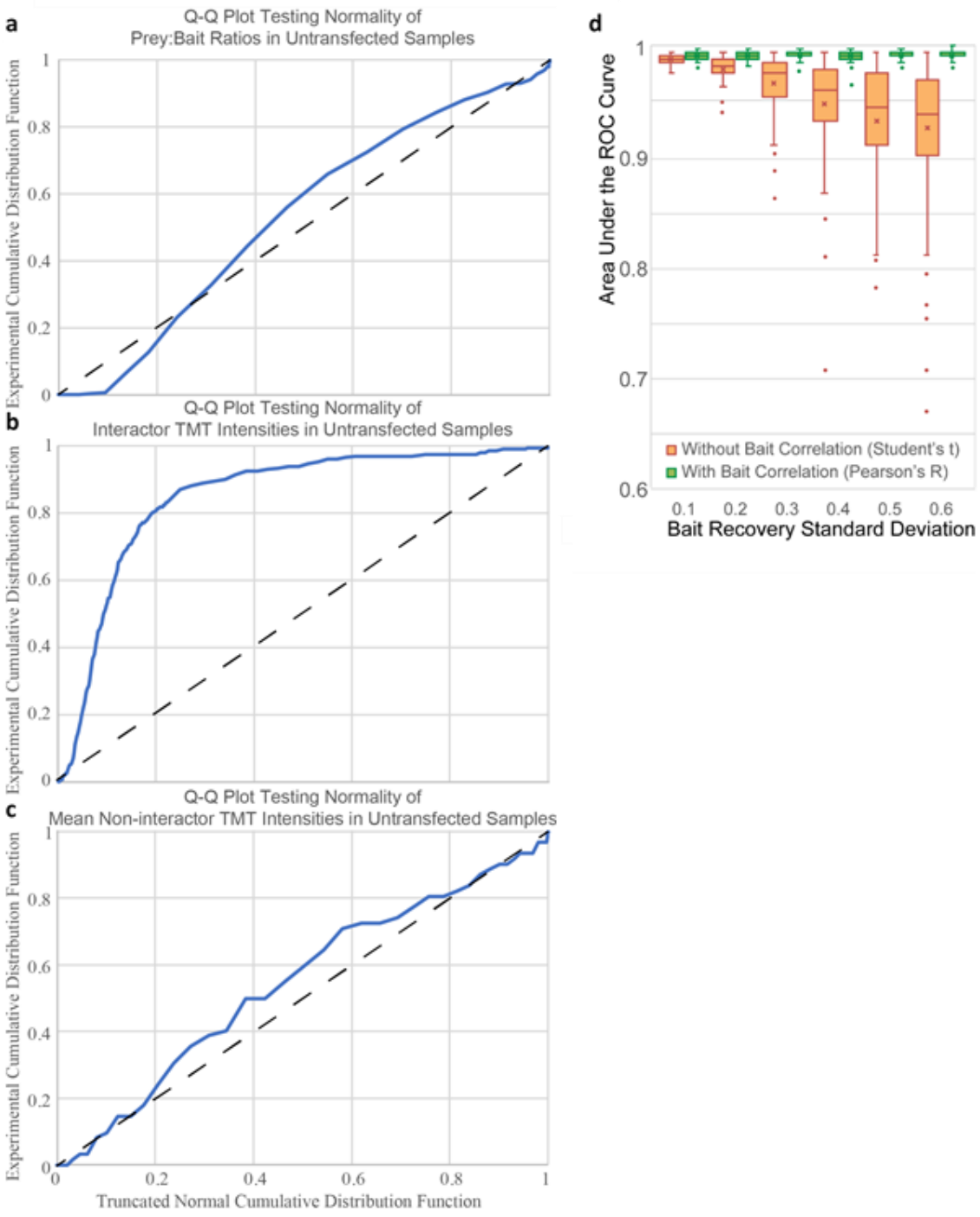

a
Q-Q Plot Testing Normality of
Prey:Bait Ratios in Untransfected Samples
Experimental Cumulative Distribution Function
b
Q-Q Plot Testing Normality of
Interactor TMT Intensities in Untransfected Samples
Experimental Cumulative Distribution Function
c
Q-Q Plot Testing Normality of
Mean Non-interactor TMT Intensities in Untransfected Samples
Experimental Cumulative Distribution Function
0 0.2 0.4 0.6 0.8 1
Truncated Normal Cumulative Distribution Function
d
Area Under the ROC Curve
1 0.9 0.8 0.7 0.6
Without Bait Correlation (Student's t)
With Bait Correlation (Pearson's R)
0.1 0.2 0.3 0.4 0.5 0.6
Bait Recovery Standard Deviation

Figure 2.10

Figure 2.10. **a-c**) Q-Q plots comparing the cumulative distribution functions of experimental data against truncated normal distribution determined from the experimental mean and standard deviations. Plots that deviate sharply from unity, represented as a dashed line, indicate that the data is not well-described by a normal distribution. **d)** Box and whisker plots of interactor identification accuracy, as determined using t-statistics derived from either Student's t test or Pearson's correlation, as indicated. Areas Under the Curve (AUCs) for Receiver Operating Characteristic Curves, which reflect interactor accuracy, were generated from simulations (100 replicates) of 500 non-specific (non-interactor) and 500 specific (prey) proteins recovered from bait immunoprecipitation under conditions of bait level variation. The abscissa represents the standard deviation of the distribution from which individual bait levels are drawn. Parameters are drawn from the new data sets generated in this work, as described in **Methods 2.2.2** and **Table 2.4.**

2.4 Conclusions

To some extent, all proteins that share a compartment interact inside the crowded environment of the cell. After membrane disruption, even proteins from separate compartments that would never encounter each other within the cell can display high affinities in the lysate. Distinguishing meaningful interactions continues to be challenging. The highest-confidence sets include stringent controls, multimodal characterization, and independent validation. Practically, however, an individual researcher seeking high-value interactor targets for a single bait needs methodology that is simple and reliable. The development of the CRAPome suite enabled reasonable confidences to be inferred for individual label-free experiments by spectral counting and comparison to a mock experiment database, but this approach does not mitigate the low sensitivity of spectral counting. Isobaric quantification provides an accessible approach to quantitatively compare several replicates in a single run. We have demonstrated that variation in transfected bait levels poses a challenge to reliable interactor identification during TMT-AP-MS. Incorporating bait correlation by deriving t-statistics from Pearson's correlation improves the sensitivity and specificity of prey identification. This approach allows small-scale TMT-AP-MS experiments to be "rescued" under conditions of bait variation.

2.5 Acknowledgements

pcDNA5/FRT/TO V5 DNAJB8 (Addgene # 19531) was generously shared by H. Kampinga. eGFP.pDEST30 plasmid was generously shared by R.L. Wiseman.

2.6 References

(1) de Lichtenberg, U.; Jensen, L. J.; Brunak, S.; Bork, P. *Science* **2005**, *307* (5710), 724–727.

(2) Fields, S.; Song, O. K. *Nature* **1989**, *340* (6230), 245–246.

(3) Phizicky, E. M.; Fields, S. *Microbiol. Rev.* **1995**, *59* (1), 94–123.

(4) Scott, J. D.; Pawson, T. *Science* **2009**, *326* (5957), 1220–1224.

(5) von Mering, C.; Krause, R.; Snel, B.; Cornell, M.; Oliver, S. G.; Fields, S.; Bork, P. *Nature* **2002**, *417* (6887), 399–403.

(6) Chien, C. T.; Bartel, P. L.; Sternglanz, R.; Fields, S. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88* (21), 9578–9582.

(7) Fields, S.; Sternglanz, R. *Trends Genet.* **1994**, *10* (8), 286–292.

(8) Parrish, J. R.; Gulyas, K. D.; Finley, R. L. *Curr. Opin. Biotechnol.* **2006**, *17* (4), 387–393.

(9) Barrios-Rodiles, M.; Brown, K. R.; Ozdamar, B.; Bose, R.; Liu, Z.; Donovan, R. S.; Shinjo, F.; Liu, Y. M.; Dembowy, J.; Taylor, I. W.; et al. *Science* **2005**, *307* (5715), 1621–1625.

(10) Braun, P.; Tasan, M.; Dreze, M.; Barrios-Rodiles, M.; Lemmens, I.; Yu, H. Y.; Sahalie, J. M.; Murray, R. R.; Roncari, L.; de Smet, A. S.; et al. *Nat. Methods* **2009**, *6* (1), 91–98.

(11) Tang, Y.; Qiu, J.; Machner, M.; LaBaer, J. *Curr. Protoc. Cell Biol.* **2017**, *74*, 15.21.1-15.21.14.

(12) Sinz, A. *Mass Spectrom. Rev.* **2006**, *25* (4), 663–682.

(13) Iacobucci, C.; Piotrowski, C.; Aebersold, R.; Amaral, B. C.; Andrews, P.; Bernfur, K.; Borchers, C.; Brodie, N. I.; Bruce, J. E.; Cao, Y.; et al. *Anal. Chem.* **2019**, *91* (11), 6953–6961.

(14) Kratchmarova, I.; Blagoev, B.; Haack-Sorensen, M.; Kassem, M.; Mann, M. *Science* **2005**, *308* (5727), 1472–1477.

(15) Gingras, A. C.; Gstaiger, M.; Raught, B.; Aebersold, R. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (8), 645–654.

(16) Gavin, A. C.; Maeda, K.; Kuhner, S. *Curr. Opin. Biotechnol.* **2011**, *22* (1), 42–49.

(17) Pankow, S.; Bamberger, C.; Calzolari, D.; Martinez-Bartolome, S.; Lavallee-Adam, M.; Balch, W. E.; Yates, J. R. *Nature* **2015**, *528* (7583), 510-+.

(18) Tarassov, K.; Messier, V.; Landry, C. R.; Radinovic, S.; Molina, M. M. S.; Shames, I.; Malitskaya, Y.; Vogel, J.; Bussey, H.; Michnick, S. W. *Science* **2008**, *320* (5882), 1465–1470.

(19) Budayeva, H. G.; Cristea, I. M. *Adv. Exp. Med. Biol.* **2014**, *806*, 263–282.

(20) Lavallée-Adam, M.; Cloutier, P.; Coulombe, B.; Blanchette, M. *J. Proteome Res.* **2011**, *10* (2), 886–895.

(21) Mellacheruvu, D.; Wright, Z.; Couzens, A. L.; Lambert, J. P.; St-Denis, N. A.; Li, T.; Miteva, Y. V.; Hauri, S.; Sardiu, M. E.; Low, T. Y.; et al. *Nat. Methods* **2013**, *10* (8), 730-+.

(22) Huttlin, E. L.; Ting, L.; Bruckner, R. J.; Gebreab, F.; Gygi, M. P.; Szpyt, J.; Tam, S.; Zarraga, G.; Colby, G.; Baltier, K.; et al. *Cell* **2015**, *162* (2), 425–440.

(23) Choi, H.; Larsen, B.; Lin, Z. Y.; Breitkreutz, A.; Mellacheruvu, D.; Fermin, D.; Qin, Z. S.; Tyers, M.; Gingras, A. C.; Nesvizhskii, A. I. *Nat Methods* **2011**, *8* (1), 70–73.

(24) Sowa, M. E.; Bennett, E. J.; Gygi, S. P.; Harper, J. W. *Cell* **2009**, *138* (2), 389–403.

(25) Huttlin, E. L.; Bruckner, R. J.; Navarrete-Perea, J.; Cannon, J. R.; Baltier, K.; Gebreab, F.; Gygi, M. P.; Thornock, A.; Zarraga, G.; Tam, S.; et al.preprint *Systems Biology*, **2020**.

(26) Langley, S. R.; Mayr, M. *J. Proteomics* **2015**, *129*, 83–92.

(27) Nesvizhskii, A. I. *Proteomics* **2012**, *12* (10), 1639–1655.

(28) Plate, L.; Rius, B.; Nguyen, B.; Genereux, J. C.; Kelly, J. W.; Wiseman, R. L. *Cell Chem. Biol.* **2019**, *26* (7), 913-925.e4.

(29) Papachristou, E. K.; Kishore, K.; Holding, A. N.; Harvey, K.; Roumeliotis, T. I.; Chilamakuri, C. S. R.; Omarjee, S.; Chia, K. M.; Swarbrick, A.; Lim, E.; et al. *Nat. Commun.* **2018**, *9*.

(30) Keilhauer, E. C.; Hein, M. Y.; Mann, M. *Mol. Cell. Proteomics* **2015**, *14* (1), 120–135.

(31) Hosp, F.; Scheltema, R. A.; Eberl, H. C.; Kulak, N. A.; Keilhauer, E. C.; Mayr, K.; Mann, M. *Mol. Cell. Proteomics* **2015**, *14* (7), 2030–2041.

(32) Youn, J.-Y.; Dunham, W. H.; Hong, S. J.; Knight, J. D. *Mol. Cell* **2018**, *69* (3), 517-532.e11. https://doi.org/10.1016/j.molcel.2017.12.020.

(33) Joshi, P.; Greco, T. M.; Guise, A. J.; Luo, Y.; Yu, F.; Nesvizhskii, A. I.; Cristea, I. M. *Mol. Syst. Biol.* **2013**, *9*, 672.

(34) Hageman, J.; Kampinga, H. H. *Cell Stress Chaperones* **2009**, *14* (1), 1–21.

(35) Klock, H. E.; Lesley, S. A. *Methods Mol Biol* **2009**, *498*, 91–103.

(36) Shoulders, M. D.; Ryno, L. M.; Genereux, J. C.; Moresco, J. J.; Tu, P. G.; Wu, C.; Yates, J. R.; Su, A. I.; Kelly, J. W.; Wiseman, R. L. *Cell Rep.* **2013**, *3* (4), 1279–1292.

(37) Washburn, M. P.; Wolters, D.; Yates, J. R. *Nat. Biotechnol.* **2001**, *19* (3), 242–247.

(38) Zecha, J.; Satpathy, S.; Kanashova, T.; Avanessian, S. C.; Kane, M. H.; Clauser, K. R.; Mertins, P.; Carr, S. A.; Kuster, B. *Mol Cell Proteomics* **2019**, *18* (7), 1468–1478.

(39) Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.

(40) Xu, T.; Park, S. K.; Venable, J. D.; Wohlschlegel, J. A.; Diedrich, J. K.; Cociorva, D.; Lu, B.; Liao, L.; Hewel, J.; Han, X.; et al. *J. Proteomics* **2015**, *129*, 16–24.

(41) Tabb, D. L.; McDonald, W. H.; Yates, J. R. *J Proteome Res* **2002**, *1* (1), 21–26.

(42) Park, S. K.; Venable, J. D.; Xu, T.; Yates, J. R. *Nat Methods* **2008**, *5* (4), 319–322.

(43) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. *Nucleic Acids Res.* **2019**, *47* (D1), D442–D450.

(44) Zauber, H.; Kirchner, M.; Selbach, M. *Nat. Methods* **2018**, *15* (3), 156–157.

(45) Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J. *Mol. Cell. Proteomics* **2012**, *11* (11), 1475–1488.

(46) Sherrod, S. D.; Myers, M. V.; Li, M.; Myers, J. S.; Carpenter, K. L.; MacLean, B.; MacCoss, M. J.; Liebler, D. C.; Ham, A.-J. L. *J. Proteome Res.* **2012**, *11* (6), 3467–3479.

(47) Smyth, G. K. *Stat. Appl. Genet. Mol. Biol.* **2004**, *3* (1), 1–25.

(48) Krammer, F.; Pica, N.; Hai, R.; Margine, I.; Palese, P. *J. Virol.* **2013**, *87* (12), 6542–6550.

(49) Oberg, A. L.; Mahoney, D. W.; Eckel-Passow, J. E.; Malone, C. J.; Wolfinger, R. D.; Hill, E. G.; Cooper, L. T.; Onuma, O. K.; Spiro, C.; Therneau, T. M.; et al. *J. Proteome Res.* **2008**, *7* (1), 225–233.

(50) Fisher, R. A. *Statistical Methods for Research Workers*, 14th ed., revised and enlarged.; Oliver and Boyd: Edinburgh, **1970**.

(51) Hogg, R. V.; Tanis, E. A. *Probability and Statistical Inference*, 2nd ed.; Macmillan ; Collier Macmillan: New York : London, **1983**.

(52) Benjamini, Y.; Hochberg, Y. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **1995**, *57* (1), 289–300.

(53) Yekutieli, D.; Benjamini, Y. *J. Stat. Plan. Inference* **1999**, *82* (1–2), 171–196.

(54) Storey, J. D.; Tibshirani, R. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (16), 9440–9445.

(55) Swets, J. *Science* **1988**, *240* (4857), 1285–1293.

(56) Kampinga, H. H.; Craig, E. A. *Nat. Rev. Mol. Cell Biol.* **2010**, *11* (10).

(57) Jin, Y.; Awad, W.; Petrova, K.; Hendershot, L. M. *EMBO J.* **2008**, *27* (21), 2873–2882.

(58) Hageman, J.; Rujano, M. A.; van Waarde, M. A. W. H.; Kakkar, V.; Dirks, R. P.; Govorukhina, N.; Oosterveld-Hut, H. M. J.; Lubsen, N. H.; Kampinga, H. H. *Mol. Cell* **2010**, *37* (3), 355–369.

(59) Liu, H.; Zhang, H.; Niedzwiedzki, D. M.; Prado, M.; He, G.; Gross, M. L.; Blankenship, R. E. *Science* **2013**, *342* (6162), 1104–1107.

(60) Kammers, K.; Cole, R. N.; Tiengwe, C.; Ruczinski, I. *EuPA Open Proteomics* **2015**, *7*, 11–19.

(61) Walker, E. J.; Bettinger, J. Q.; Welle, K. A.; Hryhorenko, J. R.; Ghaemmaghami, S. *Proc. Natl. Acad. Sci.* **2019**, *116* (13), 6081–6090.

(62) West, G. M.; Tucker, C. L.; Xu, T.; Park, S. K.; Han, X.; Yates, J. R.; Fitzgerald, M. C. *Proc. Natl. Acad. Sci.* **2010**, *107* (20), 9078–9082.

(63) Chatr-aryamontri, A.; Oughtred, R.; Boucher, L.; Rust, J.; Chang, C.; Kolas, N. K.; O'Donnell, L.; Oster, S.; Theesfeld, C.; Sellam, A.; et al. *Nucleic Acids Res.* **2017**, *45* (D1), D369–D379.

(64) Lim, G. E.; Johnson, J. D. *Adipocyte* **2016**, *5* (2), 232–237.

(65) Johnson, C.; Tinti, M.; Wood, N. T.; Campbell, D. G.; Toth, R.; Dubois, F.; Geraghty, K. M.; Wong, B. H. C.; Brown, L. J.; Tyler, J.; et al. *Mol. Cell. Proteomics* **2011**, *10* (10), M110.005751.

(66) Mugabo, Y.; Sadeghi, M.; Fang, N. N.; Mayor, T.; Lim, G. E. *J. Biol. Chem.* **2018**, *293* (18), 6736–6750.

(67) Budnik, B.; Levy, E.; Harmange, G.; Slavov, N. *Genome Biol.* **2018**, *19* (1), 161.

Chapter 3: Cellular Protein Client Recovery by Human Hsp40s DNAJB1 and DNAJB8

3.1 Introduction

Each protein has a designated role in cellular maintenance. In general, when a cell cannot sustain adequate functional protein there is an increased risk of disease states[1,2]. Deficiencies in proteostasis resulting from misfolded or aggregated protein could decrease the amount of functional protein. Some of these deficiencies can be categorized as loss-of-function or gain-of-function diseases states. Loss-of-function disease states may result from proteins that cannot fold properly and so the cell lacks (has lost) that protein[2]. Gain-of-function disease states may result from the overproduction of proteins, leading to a build-up of aggregates that are no longer functional[2]. Although misfolded protein is not solely responsible for these types of diseases states, it is one contributor. Fortunately, the cell has a proteostasis network that prevents and ameliorates the effects of misfolded or aggregated protein[1]. The most abundant group of proteins that are upregulated under conditions of stress are heat shock proteins (Hsps).

There are several types of Hsps in eukaryotes, each with their own functions and localization within the cell[3,4]. Hsp70 is the most ubiquitous Hsp in the cell and has a well-known cycle of client (misfolded protein) capture and release[5]. In this cycle, misfolded clients are first recruited by an Hsp40. The Hsp40 J-domain mediates binding to Hsp70 and stimulates Hsp70 ATPase activity. ATP hydrolysis promotes client unfolding and transfer to Hsp70. The cycle is completed when a nucleotide exchange factor promotes release of the unfolded client from Hsp70, which now has an opportunity to fold to its

native state. Several studies have demonstrated the variety of functions of Hsp70 in the cell such as promoting trafficking of proteins to parts of the cell, preventing proteins from aggregation, or maintaining translation under stress[6-8]. Examples of organelle specific Hsp70s that have different functions in the cell are HSPA1A/HSP70, HSPA5/BiP, and HSPA9/Grp75. HSPA1A/HSP70 can be found in the cytosol and nucleus and its production is activated after response to stresses such as heat, heavy metal exposure, or oxidation[9]. HSPA5/BiP is the ER Hsp70 and has the same general function as its cytosol/nucleus counterpart[10]. HSPA9/Grp75 in the mitochondria also promotes folding under stress condition and has demonstrated cell cycle control capability[11]. The functions and mechanisms of Hsp70 have been widely studied for several decades, however, human Hsp40s have been studied less extensively, with little knowledge about how they bind clients or the identity of their clients. Recently, work on Hsp40 client binding has been done on *Thermus thermophilus* class B Hsp40 (ttHsp40)[12]. In this study, authors found that different Hsp40s have different numbers of client-binding sites. Depending on the number of binding sites, Hsp40s may be entirely or partially displaced from Hsp70 during client handoff, as Hsp70 can occupy client-binding sites on the Hsp40. Decreased availability of client binding sites could lead to a lower affinity for clients. Other studies have focused on how individual domains in Hsp40 proteins can modulate client recognition[13].

Because Hsp40 proteins are highly diverse and are not always 40 kDa, recently the term J-domain protein (JDP) has become the accepted signifier for the class for their highly conserved J-domain[14]. The J-domain, found with the characteristic HPD motif between

helices II and III, is responsible for stimulating the ATPase activity of Hsp70 and thus many of the JDPs are dependent on this domain to assist in protein refolding[14,15]. The distinction between classes of Hsp40s (A, B and C), is based on the regions present in each JDP. All classes have the J-domain, class A has a zinc-finger region and dimerization domain following the G/F-rich region after the J-domain, class B has a G/F-rich region after the J-domain, and class C is any JDP that does not fit in class A or B[14-18]. Beyond the two main regions, J-domain and G/F-rich region, there is a diverse array of domains among the JDPs that may affect its binding to Hsp70. Research demonstrating these differences in domain functions supports the hypothesis that different JDPs would have their own client pools. Because JDPs must bind client proteins and are crucial for the Hsp70 cycle machinery, it is necessary to uncover the extent to which client pools among the JDPs differ. Some studies have shown that JDPs can interact with clients or Hsp70 differently[19-21]. For example, some JDPs are sent to capture and bring clients to the Hsp70 chaperone, other JDPs are membrane-bound and only interact with Hsp70 to promote ATP hydrolysis, and others may even bind and release clients without interacting with Hsp70 at all[4].

In this work we compare client protein recognition between two JDPs from the class B family, DNAJB1 and DNAJB8. In **Figure 3.1A** and **3.1B**, we show the J-domains of DNAJB1 and DNAJB8, respectively. From these figures it is demonstrated that the J-domain of DNAJB1 and DNAJB8 are similar in size and structure as expected from this highly conserved region characteristic of all JDPs. DNAJB1 and DNAJB8 have been previously described as having varied anti-aggregation activities and reliance on the J-

domain[22,23]. DNAJB1 can suppress aggregation of short polyglutamine sequences and this activity requires Hsp70 interaction; ablation of the J-domain HPD motif, abolishes anti-aggregation affects[22,24]. DNAJB1 can also heterodimerize with other JDPs to act as a disaggregase[25]. DNAJB8 (as well as the structurally similar DNAJB6) can prevent polyglutamine aggregation of longer sequences, and even maintains this activity without the J-domain as long as the C-terminus is intact for this function[22]. The described cellular functions of DNAJB1 and DNAJB8 diverge as well. DNAJB1, located in the nucleus and cytosol, is the most widely studied and abundant human class B JDP and inhibits aggregation associated with spinocerebellar ataxia type 3, in an Hsp70-dependent manner[17,22,26]. DNAJB8, located in the cytosol, is less studied, but its role has been implicated in Parkinson's disease by reducing aggregation of a Parkin RING1 domain mutant[27]. In contrast to other polyglutamine containing proteins, DNAJB8 requires a functional Hsp70 interaction to prevent mutant Parkin aggregation[22]. Most human class B JDPs have similar sequence and domain architecture to either DNAJB1 or DNAJB8, suggesting that insights into the functional differences between these two proteins might also apply across the class. DNAJB6 and DNAJB8 have both demonstrated anti-aggregation activity and are similar in size and structure[23]. DNAJB2a, DNAJB6, and DNAJB8 have been implicated in Parkinson's disease[27]. DNAJB4 and DNAJB5 both have high homology with DNAJB1 and all three have poor anti-aggregation activity[23]. In our study (**Chapter 2**), we generated a high-confidence list of hundreds of DNAJB8 clients, which we now leverage to more carefully compare DNAJB8-type JDP clients

binding to the DNAJB1-type client binding to illustrate the diversity of this class of co-chaperones.

We believe the roles of DNAJB1 and DNAJB8 will affect how they bind to misfolded clients. We seek to determine the conditions for which these two JDPs can effectively recover client proteins using affinity purification experiments. We hope to differentiate these two JDPs by identifying their respective interactomes using our Tandem Mass Tag-Affinity Purification-Mass Spectrometry (TMT-AP-MS) platform. In this work we discover that DNAJB1 and DNAJB8 have unique interactomes which can be further studied in terms of their ability to probe for the misfolded proteome. DNAJB1 clients are Hsp70-related, while clients of DNAJB8 include the bulk of the proteome. In heat shock studies on DNAJB8, it is demonstrated that this JDP has potential for identifying misfolded proteins.

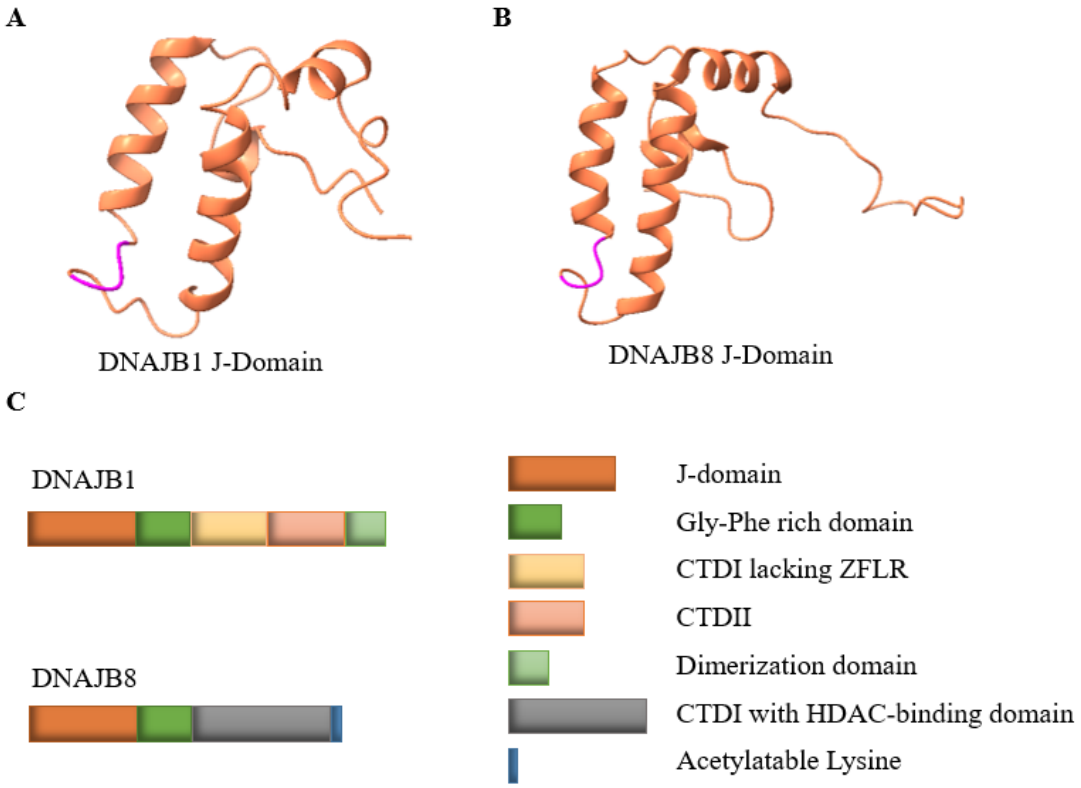


Figure 3.1

Figure 3.1. 3D solution NMR structures of the J domains for DNAJB1 (**A**, PDB: 1HDJ)[28,30] and DNAJB8 (**B**, PDB: 2DMX)[29,30]. The HPD motif responsible for Hsp70 interaction is highlighted in magenta. Panel (**C**) shows the domains present in each JDP[4]. Both proteins contain the highly conserved J-domain and Gly-Phe rich domain[15]. DNAJB1 has a dimerization domain, while DNAJB8 contains an HDAC-binding domain that has been shown in the literature to be critical to its anti-aggregation activity[23].

3.2 Materials and Methods

3.2.1 Materials.

Reagents. Biochemical reagents and buffer components were purchased from Fisher Scientific, VWR, or Millipore Sigma. Millipore water and sterilized consumables were used for all biochemical experiments.

Molecular Cloning. DNAJB1 was amplified from cDNA derived from HEK293T cells (ATCC) using TRIzol (Thermo Fisher Scientific) and inserted into the pFLAG.CMV2 vector by PIPE cloning[31]. The H32Q mutation was introduced into DNAJB1 using site-directed mutagenesis. The creation of DNAJB8 and its mutant (H31Q) is DNAJB8 was amplified from pcDNA5/FRT/TO V5 DNAJB8[17] and inserted into the pFLAG.CMV2 vector by PIPE cloning[31]. eGFP.pDEST30 cloning has been reported[32]. DNAJB1 and DNAJB8 constructs were analytically digested and sequenced (Retrogen) to confirm identity. All cloning enzymes and buffers were purchased from New England Biolabs and primers were purchased from IDT. Primer sequences are shown in **Table 3.1** and **Table 3.2**.

Human Tissue Culture. HEK293T cells (ATCC) were cultured in DMEM (Corning). All DMEM without L-Glutamine was supplemented with 10% fetal bovine serum (FBS; Seradigm), 2 mM L-Glutamine (Corning), and penicillin (100 IU/mL)-streptomycin (100 μg/mL; Corning). DMEM including 5 mM L-Glutamine was supplemented with 10% fetal bovine serum, and penicillin (100 IU/mL)-streptomycin (100 μg/mL).

3.2.2 Immunoprecipitation. Calcium phosphate transfection was used to introduce plasmid DNA into cells. Every experiment involving DNAJB8 used one 10 cm plate per

condition. 4-plex experiments involving DNAJB1 used two -10 cm plates per condition to account for its lower expression. Cells were harvested from confluent 10 cm dishes at 36 to 48 h post-transfection. For 4-plex experiments, dithiobis succinimidyl propionate (DSP) crosslinking was used as indicated. Cells were incubated in 1 mM DSP/1% DMSO in PBS or 1% DMSO in PBS (vehicle) for 30 min with rotation at ambient temperature, and then quenched by addition of Tris pH 8.0 (to 90 mM) and rotation for 15 min. For DSP crosslinking dependence experiments, the same ratio of DMSO/PBS was used, with only the concentration of DSP changing. After crosslinking, or directly after harvest for experiments without crosslinking, cells were lysed for 30 min on ice in lysis buffer supplemented with fresh 1 x protease inhibitor cocktail (Roche). High stringency lysis was performed in RIPA buffer (150 mM NaCl, 50 mM Tris pH 7.5, 1% Triton x100, 0.5% sodium deoxycholate, 0.1 % SDS) for all experiments involving DNAJB8. Low stringency lysis was performed with 0.1% Triton x-100 in TBS (10 mM Tris pH 7.5, 150 mM NaCl) for the three “Low stringency” 4-plex experiments involving DNAJB1. Lysate was separated from cell debris by centrifugation at 21,100 x g for 15 min at 4 °C. Protein was quantified by Bradford assay (Bio-Rad). Lysates were pre-cleared with 15 µL sepharose-4B beads (Millipore Sigma) for 30 min at 4 °C, followed by immunoprecipitation with 15 µL M2 anti-FLAG Magnetic Beads (Millipore Sigma) and overnight rotation at 4 °C. Beads were washed four times with lysis buffer the next day for DNAJB8 or three days later for DNAJB1. Proteins were eluted from the beads by boiling in 30 µL of Laemmli concentrate (120 mM Tris pH 6.8, 60% glycerol, 12% SDS,

brilliant phenol blue to color). About 17% of eluates were reserved for silver stain analysis, while the remainder was prepared for mass spectrometry.

3.2.3 Silver Stain. Eluates were boiled for 5 min at 100 °C with 0.17 M DTT, loaded into 1.0 mm, 12% polyacrylamide gels, and separated by SDS-PAGE. Gels were rinsed in Millipore water for 5 min. Gels were left overnight in fixing solution (10% acetic acid, 30% ethanol), washed 3 x 20 min in 35% ethanol, sensitized (0.02% sodium thiosulfate) for 2 min, washed with Millipore water 3 x 2 min, and stained for 30 min to overnight in Ag staining solution (0.2% $AgNO_3$, 0.076% formalin). Gels were washed 2 x 1 min with Millipore water and developed (6% sodium carbonate, 0.05% formalin, 0.0004% sodium thiosulfate) until bands reached desired intensity and imaged on a white-light transilluminator (UVP).

3.2.4. TMT-MuDPIT. Immunoprecipitates were prepared for TMT-AP-MS according to standard protocols[33,34]. After TMT labeling, each TMT reaction was quenched with 0.4% ammonium bicarbonate. Labeled digests were combined and fractionated by SCX in line with a reversed-phase analytical column to enable two-dimensional separation prior to electrospray ionization. Peptides were analyzed using an LTQ Orbitrap Velos Pro in data-dependent mode. The top ten peaks from each full precursor scan were fragmented by HCD to acquire fragmentation spectra. Peptide-spectra matches were evaluated by ProLuCID[35,36] using a Uniprot proteome database supplemented with common contaminants and a full decoy set and filtered (DTA Select version 2.0.2737)[37] to 1% false discovery rate for peptide identifications. TMT reporter ion ratios were quantified in

Census[38], and only unique peptides were considered. Full TMT MuDPIT conditions and parameters are detailed in **Chapter 2**.

3.2.5 Statistical Methods. TMT intensity ratios were analyzed using Excel. For 4-plex Box-and-Whisker plots, WT and H31Q/H32Q TMT channels were normalized to WT without crosslinking. For heat shock experiments with DNAJB8^{H31Q} and DNAJB8^{WT}, TMT intensities of identified proteins were normalized to bait intensities. Box and whisker plots are presented with lines marking median values, X marking average values, boxes from the first to third quartiles, whiskers extending to minimum and maximum values (excluding outliers), and outliers defined at points greater than 1.5-fold the interquartile range beyond the first and third quartiles.

Because the Pearson's coefficient of non-correlated data (the null hypothesis) is normally distributed[39], a simple t-statistic can be directly calculated from: $t = R\sqrt{\frac{n-2}{1-R^2}}$ (for derivation see ref 40), where $n - 2$ is the degrees of freedom and R is the correlation coefficient, $r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$. This t-statistic is equivalent to the ratio of the measured slope from the linear fit divided by its standard error. p-values are then inferred by comparing the t-statistic to the two-tailed Student's t-distribution with $n - 2$ degrees of freedom.

q-values (q_{BH}) were determined from p-values using the Benjamini-Hochberg methodology, $q_k = p_k \frac{N}{k}$, where k is the rank for each protein, arranged as increasing p-value, and N is the total number of proteins examined (representing the number of hypotheses)[41].

Table 3.1. Primers used for Molecular Cloning (DNAJB8)	
Primer Name	Sequence
DNAJB8 PIPES Vector Fwd	5’-GAC AGC AAG TAG GCG AAT TCA TCG ATA GAT CTG-3’
DNAJB8 PIPES Vector Rev	5’-CTT CGT AGT AGT TAG CCA TAA GCT TGT CGT CAT CGT C-3’
DNAJB8 PIPES Insert Fwd	5’-GAC GAT GAC GAC AAG CTT ATG GCT AAC TAC TAC GAA G-3’
DNAJB8 PIPES Insert Rev	5’-CAG ATC TAT CGA TGA ATT CGC CTA CTT GCT GTC-3’
DNAJB8 H31Q SDM Fwd	5’- CTT CGT TGG CAG CCC GAC AAG AAC CCT GAC AAT AAG-3’
DNAJB8 H31Q SDM Rev	5’- GTT CTT GTC GGG CTG CCA ACG AAG GGC-3’

Table 3.2 Primers used for Molecular Cloning (DNAJB1)	
Primer Name	Sequence
DNAJB1 PIPES Vector Fwd	5’- CTT GAG CAG GTT CTT CCA ATA GCG AAT TCA TCG ATA GAT CTG-3’
DNAJB1 PIPES Vector Rev	5’- GTA GTA GTC TTT ACC CAT GAC CTT GTC GTC ATC GTC TTT G-3’
DNAJB1 PIPES Insert Fwd	5’- CAA AGA CGA TGA CGA CAA GGT CAT GGG TAA AGA CTA CTA C-3’
DNAJB1 PIPES Insert Rev	5’- CAG ATC TAT CGA TGA ATT CGC TAT TGG AAG AAC CTG CTC AAG-3’
DNAJB1 H31Q SDM Fwd	5’- CTA CCA ACC GGA CAA GAA CAA GGA GCC CGG-3’
DNAJB1 H31Q SDM Rev	5’- CCG GTT GGT AGC GCA GCG CC-3’

3.3 Results and Discussion

3.3.1 Crosslinker is not necessary for affinity of clients to DNAJB8. We previously identified > 540 high-confidence interactors (Benjamini-Hochberg false discovery rate < 0.05) of DNAJB8^{H31Q} (H31Q is a mutation of the HPD motif to QPD to inactivate the J-domain) using Tandem Mass Tag-Affinity Purification-Mass Spectrometry (TMT-AP-MS) (**Chapter 2**). These interactors were relatively destabilized as compared to the bulk proteome, indicating that DNAJB8 co-purification is a way to access the bulk destabilized proteome. In this study, we used DSP crosslinking and TMT-AP-MS to identify interactors of DNAJB8^{H31Q}. We expected that ablation of the J-domain would prevent client handoff to Hsp70, increasing both the recovery of clients and decreasing the recovery of Hsp70 and Hsp70-associated chaperones. Crosslinking was applied to maintain transient client interactions. However, the role of the J-domain in maintaining client association by DNAJB8 is not established, and crosslinking decreases general protein recovery during lysis and DNAJB8 immunorecovery (**Figure 3.2**). If JDPs are to serve as recognition elements for misfolded proteins, then we need conditions that maximize interactor yield. Hence, we evaluated the proteome-wide effect of crosslinking and J-domain activity on DNAJB8 client recovery using a TMT-AP-MS experiment (**Figure 3.3A**).The multiplexed capability of TMT tagging allows us to quantitatively compare interactor intensities for both wild-type and mutant (H31Q) DNAJB8, with and without crosslinking, in the same LC-MS run. High stringency washing (RIPA) is used to minimize non-specific interactions with the beads and consequently false positives[42]. In these 4-plex experiments, we compare wild type and mutant DNAJB8. We hypothesized

that mutant DNAJB8 would allow a higher recovery of proteins by immunoprecipitation, compared to wild type, since mutant would keep its clients and not give them to Hsp70. Non-client recovery, conversely, would not be improved or diminished by use of wild type or mutant DNAJB8. We also expected crosslinking to affect the recovery of proteins by increasing recovery for some proteins, not affecting other proteins, and not decreasing recovery relative to DNAJB8 for any proteins. TMT-AP-MS is the best method for testing this because we can uncover hundreds of proteins that interact with our bait in a single sample set.

We performed three replicate sets of 4-plex TMT-AP-MS experiments for DNAJB8, with similar results (**Figure 3.3B-D**). 206 proteins were identified in at least two replicates, with 141 (~68 %) of these known interactors from our previous study that used only a single condition (**Chapter 2**). The interactors that did not match our list of $DNAJB8^{H31Q}$ clients could be those that interact only with $DNAJB8^{WT}$ or are recovered only in the presence of crosslinker. Because this list is dominated by known specific DNAJB8 interactors, we included all proteins in the following analysis. We show Box-and-Whisker plots comparing three conditions: $DNAJB8^{WT}$ with crosslinking (WT +), $DNAJB8^{H31Q}$ without crosslinking (H31Q -), and $DNAJB8^{H31Q}$ with crosslinking (H31Q +), each normalized to $DNAJB8^{WT}$ in the absence of crosslinker (WT -). Surprisingly, interactor recoveries for $DNAJB8^{WT}$ and $DNAJB8^{H31Q}$ are similar. Most proteins do not reproducibly increase affinity to $DNAJB8^{H31Q}$ (**Figure 3.3E, F**). As expected, Hsp70 affinity decreases with J-domain inactivation (orange point in **Figures 3.3E, F**). Furthermore, crosslinking decreases interactor recovery for both DNAJB8 baits

(**Figure 3.3G**). Variation in bait levels can strongly influence interactor recovery and reproducibility[43]. Hence, we also considered normalizing TMT reporter ion intensities to DNAJB8 recovery (**Figure 3.3H**). Bait normalization does modestly increase the apparent relative recovery of interactors with crosslinking, due to decreased bait recovery after crosslinking (**Figure 3.3I, J**). However, bait normalization increases variance in protein recovery with crosslinking (**Figure 3.3H**, WT+ and H31Q+), as indicated by the box width. This demonstrates that DNAJB8 client recovery is unaffected by J-domain inactivation and is impaired by cellular crosslinking. DNAJB8, and its similar subfamily member DNAJB6, have been shown to form oligomers[44], which may contribute to its ability to bind and hold clients[13]. Although RIPA wash buffer was developed to aggressively break most protein-protein interactions, oligomeric Hsp40-client complexes have been previously found to be RIPA-resistant[45]. Since DSP decreases total protein recovery, clients of DNAJB8 could be lost if binding to DNAJB8 oligomers with crosslinker, causing the observed larger variation in interactor recovery.

We considered that perhaps 1 mM DSP, though a typical concentration, is too aggressive and that an optimized concentration might allow greater interactor recovery during TMT-AP-MS. To confirm whether crosslinker is unnecessary for DNAJB8 client recovery, we performed a crosslinking dependence (**Figure 3.4A-C**). Across the proteome, the interactor recoveries smoothly follow bait recovery at varying DSP concentration (**Figure 3.4C**), indicating that even lower levels of crosslinker fail to improve bait recovery.

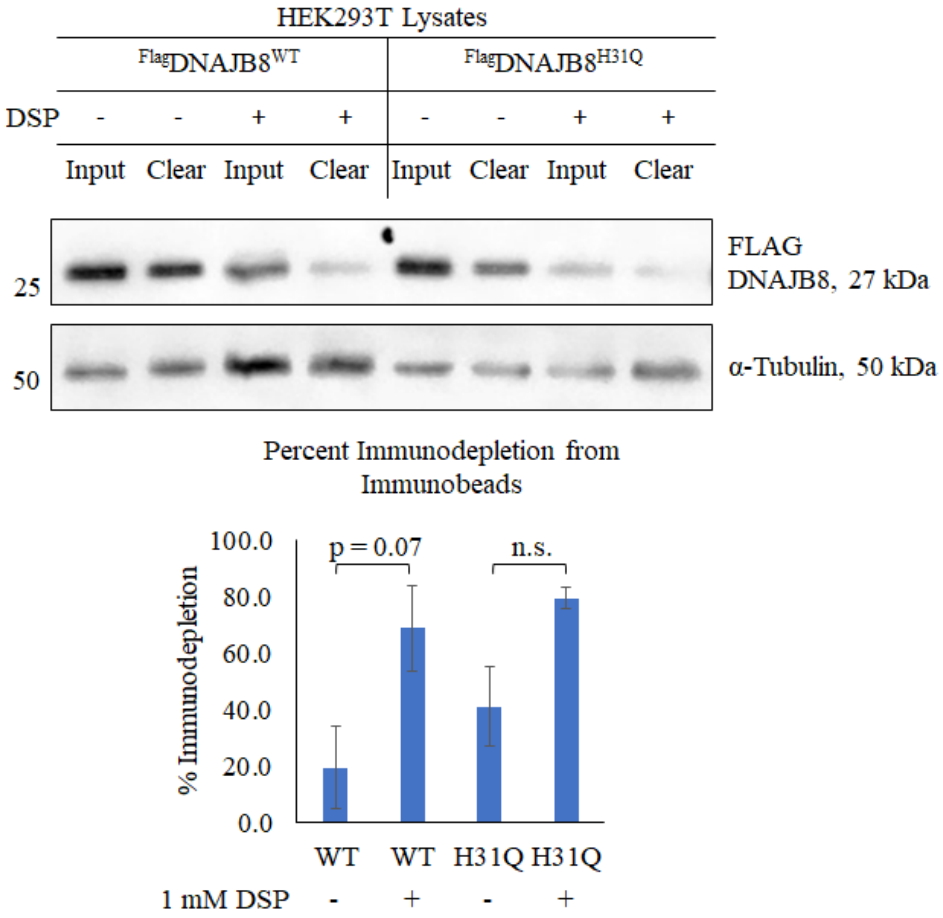
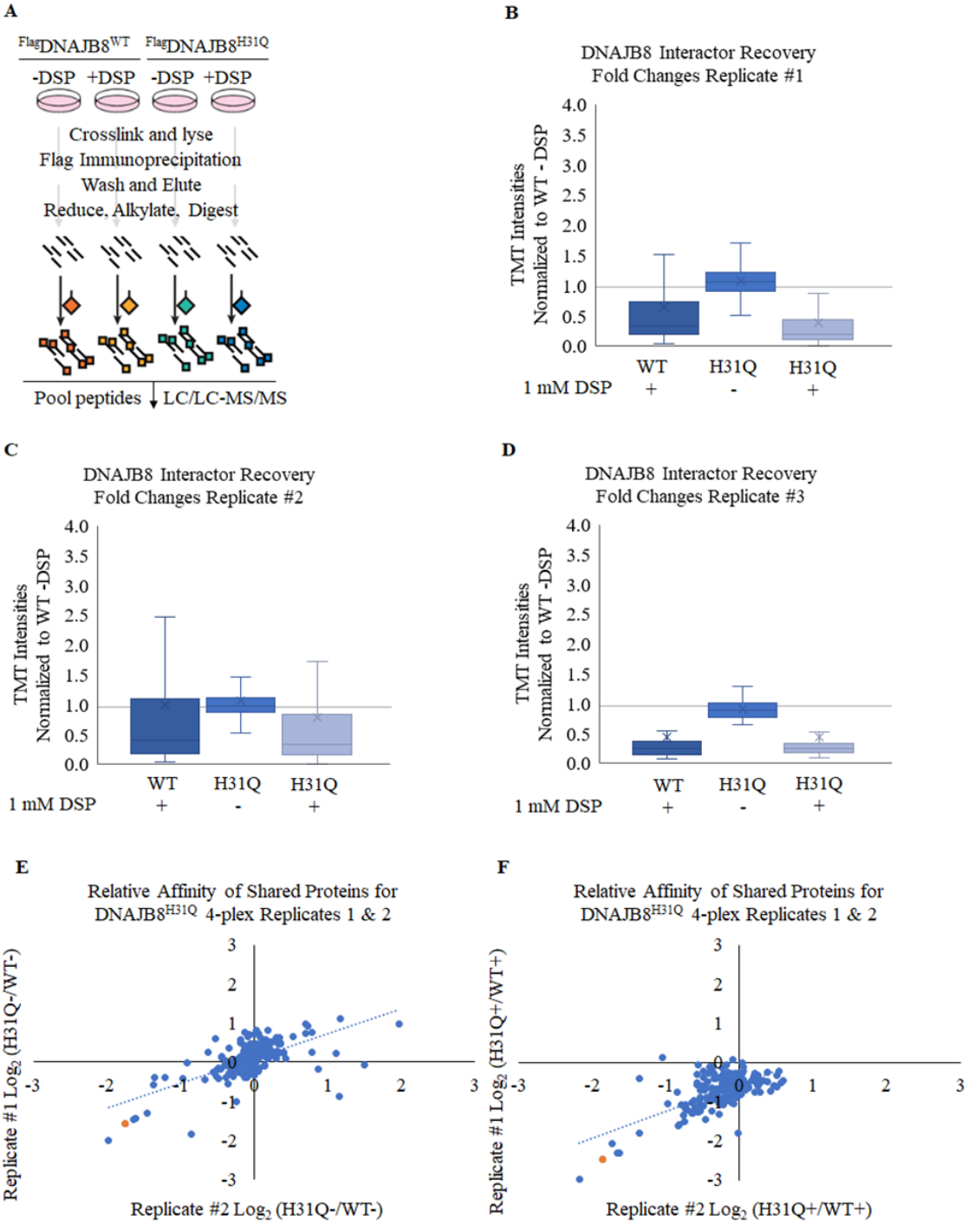


Figure 3.2

Figure 3.2 Western blot of DNAJB8^{WT} and DNAJB8^{H31Q} levels in HEK293T lysate before (Input) and after (Clear) incubation with immunobeads in one replicate. Bar graph representation of percent immunodepletion for three replicates of this experiment depicted below. Error bars represent standard deviation between the three replicates. Brackets indicate a two-tailed t-statistic ($\alpha = 0.05$) between immunodepletion of crosslinking (1 mM DSP) versus no crosslinking conditions.


A
FlagDNAJB8WT
FlagDNAJB8H31Q
-DSP +DSP -DSP +DSP
Crosslink and lyse
Flag Immunoprecipitation
Wash and Elute
Reduce, Alkylate, Digest
Pool peptides
LC/LC-MS/MS
B
DNAJB8 Interactor Recovery
Fold Changes Replicate #1
TMT Intensities
Normalized to WT - DSP
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
WT
H31Q
H31Q
1 mM DSP
+
-
+
C
DNAJB8 Interactor Recovery
Fold Changes Replicate #2
TMT Intensities
Normalized to WT -DSP
WT
H31Q
H31Q
1 mM DSP
+
-
+
D
DNAJB8 Interactor Recovery
Fold Changes Replicate #3
TMT Intensities
Normalized to WT -DSP
WT
H31Q
H31Q
1 mM DSP
+
-
+
E
Relative Affinity of Shared Proteins for
DNAJB8H31Q 4-plex Replicates 1 & 2
Replicate #1 Log2 (H31Q-/WT-)
Replicate #2 Log2 (H31Q-/WT-)
3
2
1
0
-1
-2
-3
F
Relative Affinity of Shared Proteins for
DNAJB8H31Q 4-plex Replicates 1 & 2
Replicate #1 Log2 (H31Q+/WT+)
Replicate #2 Log2 (H31Q+/WT+)

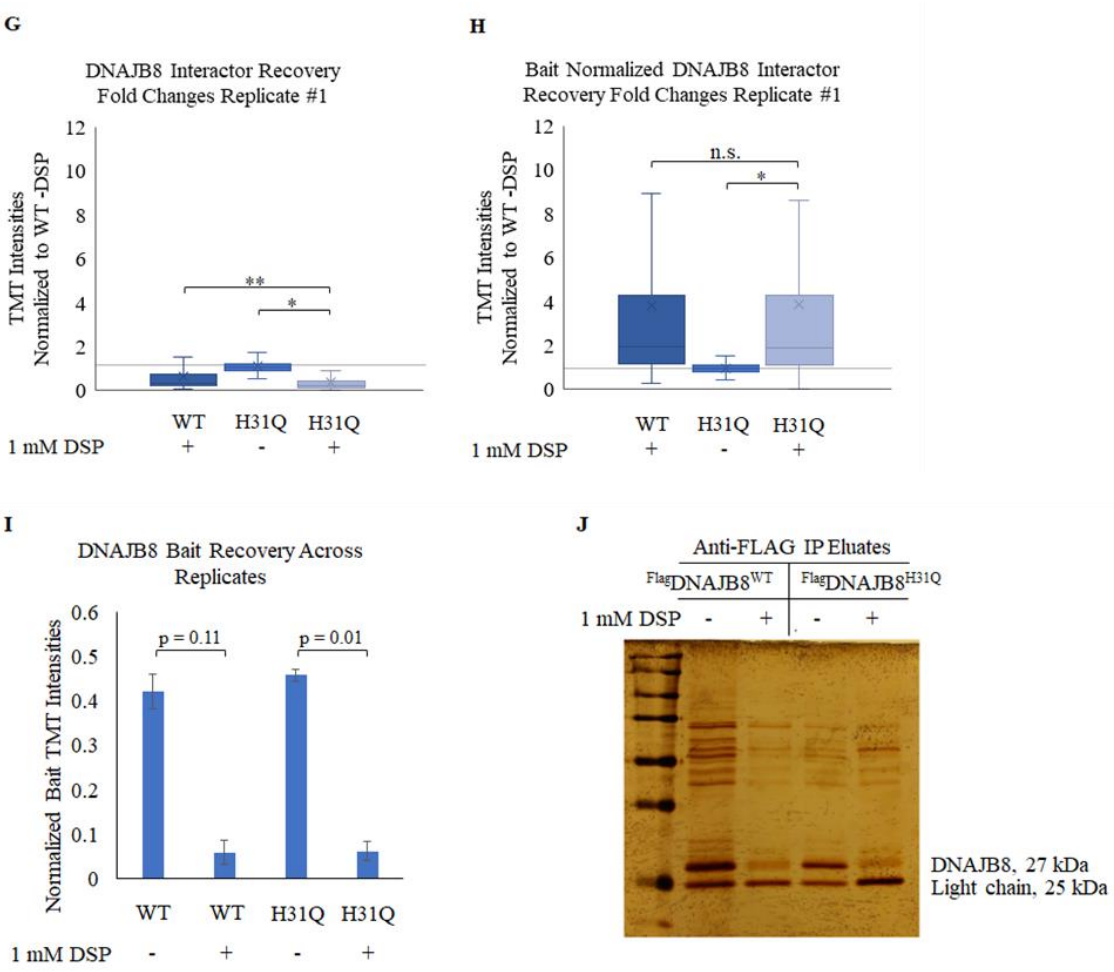

G
DNAJB8 Interactor Recovery
Fold Changes Replicate #1
TMT Intensities
Normalized to WT -DSP
12
10
8
6
4
2
0
**
*
WT
H31Q
H31Q
1 mM DSP
+
-
+
H
Bait Normalized DNAJB8 Interactor
Recovery Fold Changes Replicate #1
TMT Intensities
Normalized to WT -DSP
n.s.
*
WT
H31Q
H31Q
1 mM DSP
+
-
+
I
DNAJB8 Bait Recovery Across
Replicates
Normalized Bait TMT Intensities
0.6
0.5
0.4
0.3
0.2
0.1
0
p = 0.11
p = 0.01
WT
WT
H31Q
H31Q
1 mM DSP
-
+
-
+
J
Anti-FLAG IP Eluates
FlagDNAJB8WT
FlagDNAJB8H31Q
1 mM DSP
-
+
-
+
DNAJB8, 27 kDa
Light chain, 25 kDa

Figure 3.3

Figure 3.3. A) Experimental setup for TMT-AP-MS 4-plex assay for DNAJB8. **B-D**) Box-and-Whisker plots (outliers not shown) from Replicate #1 (**B**), 2 (**C**), and 3 (**D**) of the DNAJB8 4-plex experiments showing TMT intensities for three conditions: $DNAJB8^{WT}$ with crosslinker, $DNAJB8^{H31Q}$ without crosslinker, and $DNAJB8^{H31Q}$ with crosslinker, all normalized to $DNAJB8^{WT}$ without crosslinker. Outliers not shown for visual clarity. x marks the median value. Unity is marked with a gray line across the plot. **E**) Scatter plot of fold change (H31Q/WT) TMT Intensities for recovered proteins without crosslinking. HSPA1B is colored orange. **F**) Scatter plot of fold change (H31Q/WT) TMT Intensities for recovered proteins with crosslinking. HSPA1B is colored orange. **G**) Box-and-Whisker plots (outliers not shown) from Replicate #1. Brackets indicate significance from two-tailed t-statistics with $\alpha = 0.05$ (*3 $x10^{-60}$, **6 $x10^{-17}$). **H**) Box-and-Whisker plot from high stringency buffer experiments showing bait normalized TMT intensities for three conditions: $DNAJB8^{WT}$ with crosslinker, $DNAJB8^{H31Q}$ without crosslinker, and $DNAJB8^{H31Q}$ with crosslinker, all normalized to $DNAJB8^{WT}$ without crosslinker. Brackets indicate significance from two-tailed t-statistics (*1 $x10^{-17}$). **I**) Bar graph representation of DNAJB8 bait recovery from TMT intensities for 4-plex high stringency experiments. Bait TMT Intensities were normalized to the sum of all four bait TMT intensities in a single run, for each run. **J**) Representative silver stain of SDS-PAGE separated Flag immunoprecipitates from HEK293T cells overexpressing the indicated plasmids for DNAJB8 4-plex experiments.

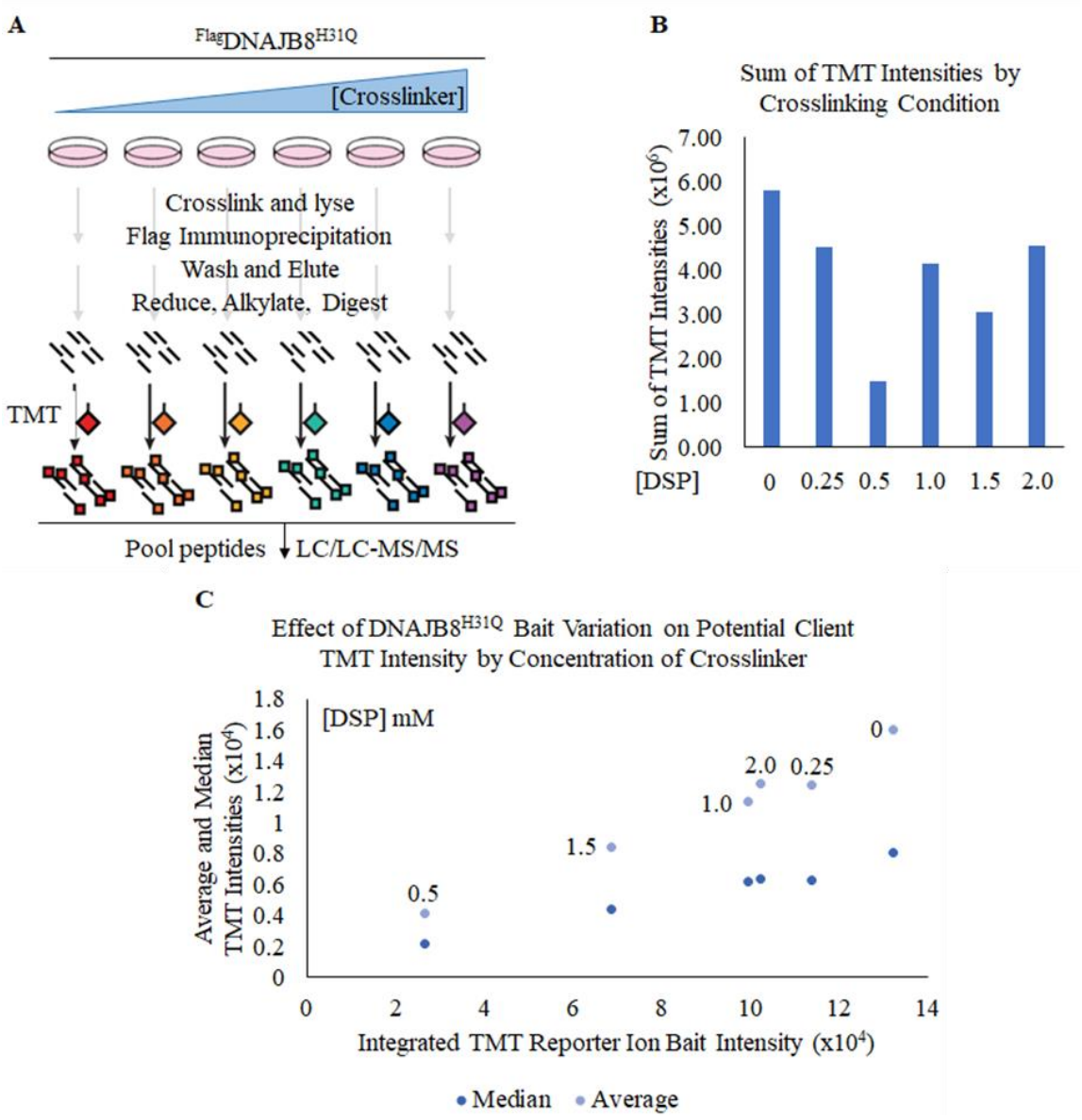


Figure 3.4

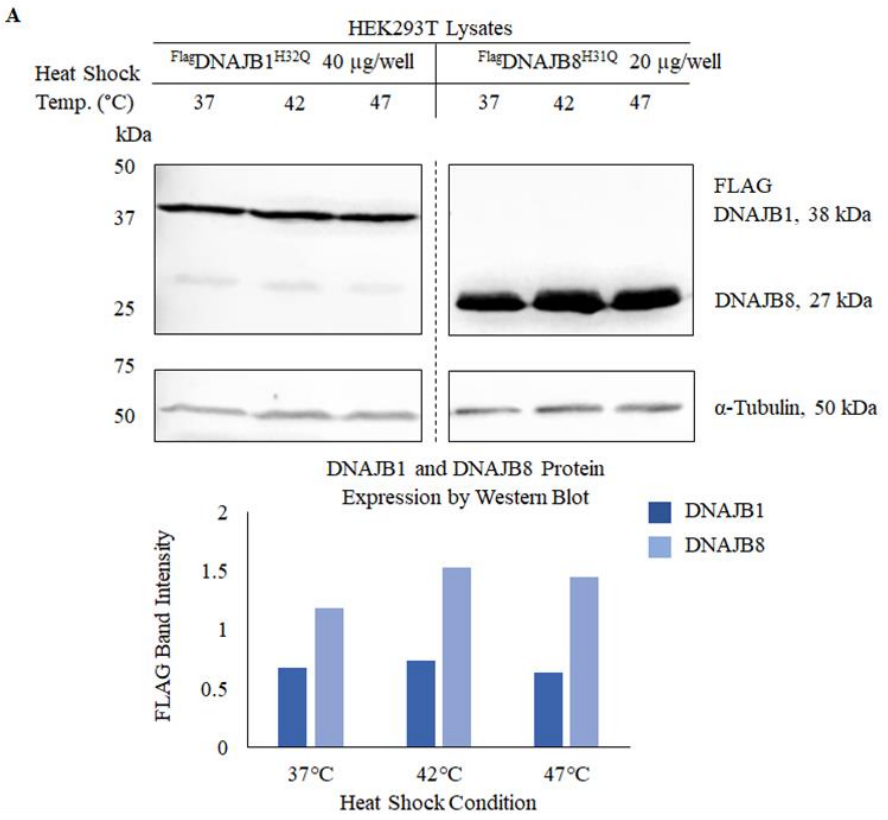
Figure 3.4 A) Experimental setup for TMT-AP-MS crosslinking dependence for DNAJB8^{H31Q}. **B**) Bar graph representation showing the sum of TMT intensities of recovered proteins by crosslinking condition. **C**) Scatter plot depicting the average and median protein intensities versus the bait TMT intensities at the different crosslinking conditions: 0, 0.25, 0.5, 1.0, 1.5, and 2.0 mM DSP.

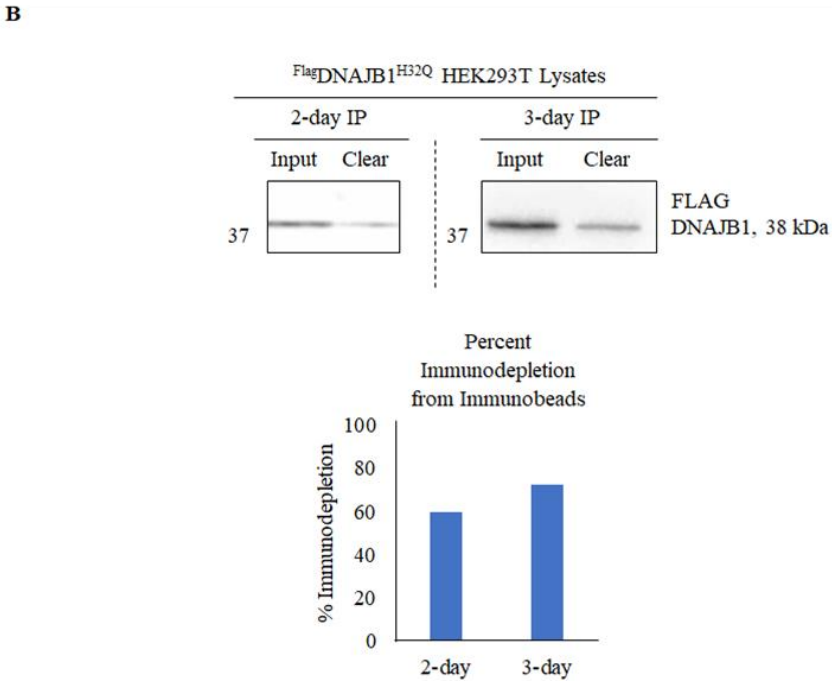
3.3.2 Crosslinking decreases DNAJB1 immunodepletion. To compare the capability of DNAJB1 as a probe for destabilized protein to that of DNAJB8, we prepared DNAJB1 in the same vector with an N-terminal Flag Tag. DNAJB1 is the most abundant class B JDP in HEK293 cells, while DNAJB8 is not expressed at all[17]. Surprisingly, we find that DNAJB1^{H32Q} expression levels consistently lag well below DNAJB8^{H31Q} expression levels (**Figure 3.5 A**) based on anti-Flag immunoblotting of whole cell lysate following transient transfection of HEK293T cells. This difference is consistent across multiple plasmid preparations of each construct, and could reflect differences in translation, degradation, or stability of the two JDPs. The presence of misfolded proteins might influence the relative recovery of each JDP by promoting binding to insoluble aggregates[44]. To test this hypothesis, we evaluated steady state levels of each protein in lysate following a brief 30 min. heat shock; no difference in recovery is observed for either DNAJB1^{H32Q} nor DNAJB8^{H31Q}, suggesting that the presence of misfolded protein does not impact recovery of overexpressed JDPs from lysate.

As with any AP experiment, the efficiency of the pull-down must be optimized to ensure immunodepletion of the intended bait protein[46]. We expected the two JDPs to have similar immunoprecipitation because they are from the same subfamily. Each protein was overexpressed in HEK293T cells and immunoprecipitation evaluated by the clearance of the bait protein lysates.

Unlike for DNAB8 that overexpresses well for a single day immunopreciptation, DNAJB1 gave improved immunodepletion after a 3-day immunoprecipitation as compared to 2 days (**Figure 3.5 B**). Chaperone-client interactions are frequently

transient, as weak interactions are adequate to allow clients to be delivered through chaperoning cycles[47]. Treatment of cells with a covalent crosslinker, such as DSP, prevents dissociation of baits and interactors during cell lysis immunoprecipitation, and bead washing steps[48]. An added benefit of crosslinking is to allow high stringency washing of the beads, lowering non-specific protein binding and incorrect assignment of non-interacting proteins as bait interactors[46]. However, aggressive crosslinker treatment could inhibit immunodepletion due to presence of crosslinker-reactive lysines in the FLAG epitope. We determined the effect of DSP crosslinking and J-domain activity on DNAJB1 immunodepletion (**Figure 3.5 C**). While inactivation of the J-domain does not affect immunodepletion in the absence of crosslinker, we see a moderate decrease in DNAJB1 immunodepletion efficiency in the presence of crosslinker. Immunodepletion of DNAJB1 from immunobeads was greater than immunodepletion of DNAJB8, likely due to the difference in expression (**Figure 3.2**). Lower bait recovery in the presence of crosslinker can be expected to lower interactor recovery. To understand in more detail how crosslinking and J-domain inactivation affect DNAJB1 client recovery, we applied TMT-AP-MS.


A
HEK293T Lysates
FlagDNAJB1H32Q 40 µg/well
FlagDNAJB8H31Q 20 µg/well
Heat Shock Temp. (°C)
37
42
47
37
42
47
kDa
50
37
25
75
50
FLAG
DNAJB1, 38 kDa
DNAJB8, 27 kDa
α-Tubulin, 50 kDa
DNAJB1 and DNAJB8 Protein Expression by Western Blot
DNAJB1
DNAJB8
FLAG Band Intensity
2
1.5
1
0.5
0
37°C
42°C
47°C
Heat Shock Condition


B
FlagDNAJB1H32Q HEK293T Lysates
2-day IP
3-day IP
Input
Clear
Input
Clear
37
37
FLAG
DNAJB1, 38 kDa
Percent Immunodepletion from Immunobeads
% Immunodepletion
100
80
60
40
20
0
2-day
3-day

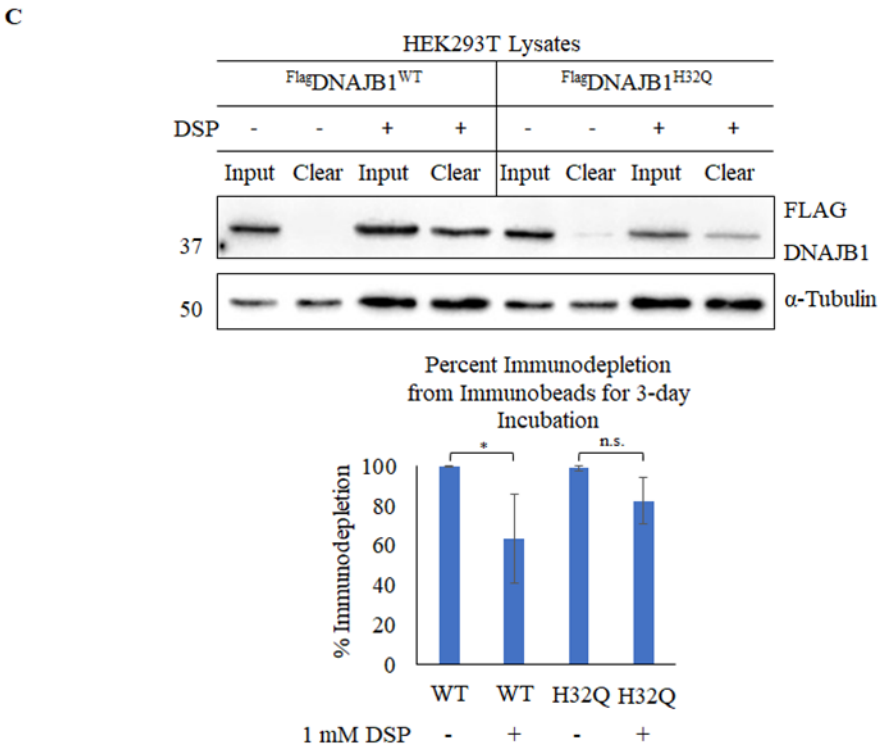


Figure 3.5

Figure 3.5 A) Western blot images of DNAJB1 (38 kDa) and DNAJB8 expression (27 kDa) for six heat shock conditions (37°C, 42°C, 47°C, 52°C, 57°C, 62°C). HEK293T cells were transfected with DNAJB1^{H32Q} or DNAJB8^{H31Q}, heated at the indicated temperatures for 30 min, then harvested and lysed. 40 µg of HEK293T lysates were loaded per well for DNAJB1^{H32Q}. 20 µg of HEK293T lysates were loaded per well for DNAJB8^{H31Q}. A lower amount of DNAJB8^{H31Q} was loaded to allow comparison with DNAJB1^{H32Q} on the same gel. Bar graph representation of the Western blot signal intensity for this single blot of DNAJB1^{H32Q} and DNAJB8^{H31Q} at the shared heat shock conditions (37°C, 42°C, 47°C) is depicted below the blot. **B**) DNAJB1^{H32Q} immunodepletion from immunobeads after two-day and three-day immunobead incubation. Bar graph representation of percent immunodepletion for these single separate experiments is depicted below the blots. **C**) Western blot of DNAJB1^{WT} and DNAJB1^{H32Q} levels in HEK293T lysate before (Input) and after (Clear) incubation with immunobeads in one replicate. Bar graph representation of percent immunodepletion for three replicates of this experiment depicted below. Error bars represent standard deviation between the three replicates. Brackets indicate a two-tailed t-statistic ($\alpha = 0.05$, *p = 0.025) between immunodepletion of crosslinking (1 mM DSP) versus no crosslinking conditions.

3.3.3 Chemical crosslinking improves recovery for a handful of potential clients of DNAJB1.

As with DNAJB8, we characterized the effect of J-domain activity and crosslinking on DNAJB1 interactor recovery using TMT-AP-MS, using a similar experimental design (**Figure 3.3A**). We identify 203 proteins present in each of the three replicates (**Figure 3.6A, B-D**). Because DNAJB1 is unable to bind substrate and Hsp70 at the same time[12], and generally has been reported to rely on Hsp70 for anti-aggregation activity, we hypothesized that the J-domain would be more essential for client binding for DNAJB1 than it is for DNAJB8. In that case, most interactors would be recovered in higher yield with DNAJB1^{H32Q} than with the wild type. This is not what was observed. As seen with DNAJB8, DNAJB1^{WT} and DNAJB1^{H32Q} recover the same levels of their interactors in the absence of crosslinking (**Figure 3.6E**) and bait recovery is decreased with crosslinking (**Figure 3.6H, I**). We also again expected crosslinking to affect the recovery of proteins by increasing recovery for some proteins (potential interactors of DNAJB1) and not affecting other proteins (non-interactors). Surprisingly, although DNAJB1^{H32Q} (with crosslinking) has a higher recovery of proteins than DNAJB1^{WT} for some proteins, the recoveries are not increasing for most of the proteome (≥ 50 % proteins have a fold change < 2 for at least two replicates), and crosslinking appears to create a larger variation across those conditions, as depicted by box height with fold changes varying from 0-7 (**Figure 3.6F**, H32Q +DSP). A list of proteins with improved recovery with crosslinking is shown in **Table 3.3**[49-51]. DNAJB1^{H32Q} (without crosslinking) is more consistent across all proteins. We considered optimizing crosslinker concentration to

maximize interactor recovery using TMT-AP-MS (similar setup to **Figure 3.4A**). Although crosslinking does not have a large effect on recovery of proteins with DNAJB1 (**Figure 3.7A**), the bait levels are greatly affected by crosslinking (**Figure 3.7B**). Although we found a slight increase in interactor recovery for the mutant version of DNAJB1, confirming our hypothesis that crosslinking improves recovery for some proteins, it is not true for the bulk of the proteome that binds DNAJB1. Furthermore, varying crosslinker concentration does not lead to meaningful changes in interactor recovery (**Figure 3.7A**).

It is possible that the similarity between the recovered DNAJB1^{WT} and DNAJB1^{H32Q} protein interactors is due to weak affinity of the chaperone: that interactors are readily released in the absence of crosslinking, and that any clients that are blocked from hand-off to Hsp70 are lost during washing steps. Furthermore, crosslinking decreases immunodepletion and bait recovery. Hence, we performed a similar series of experiments using a low-stringency wash condition. From our earlier study on the effect of crosslinking on immunodepletion of DNAJB1 from immunobeads, we wanted to investigate the effect of buffer stringency on recovery of interactors. If clearance from immunobeads was better without crosslinking, then that would be a condition that would give the greatest recovery of interactors. Therefore, by using a less-stringent buffer, DNAJB1 may be able to hold onto its interactors without the assistance of crosslinker. To consider the effect of stringency of the lysis and wash buffer on recovery of prey for DNAJB1^{WT} and DNAJB1^{H32Q}, we repeated the 4-plex experiment under low stringency

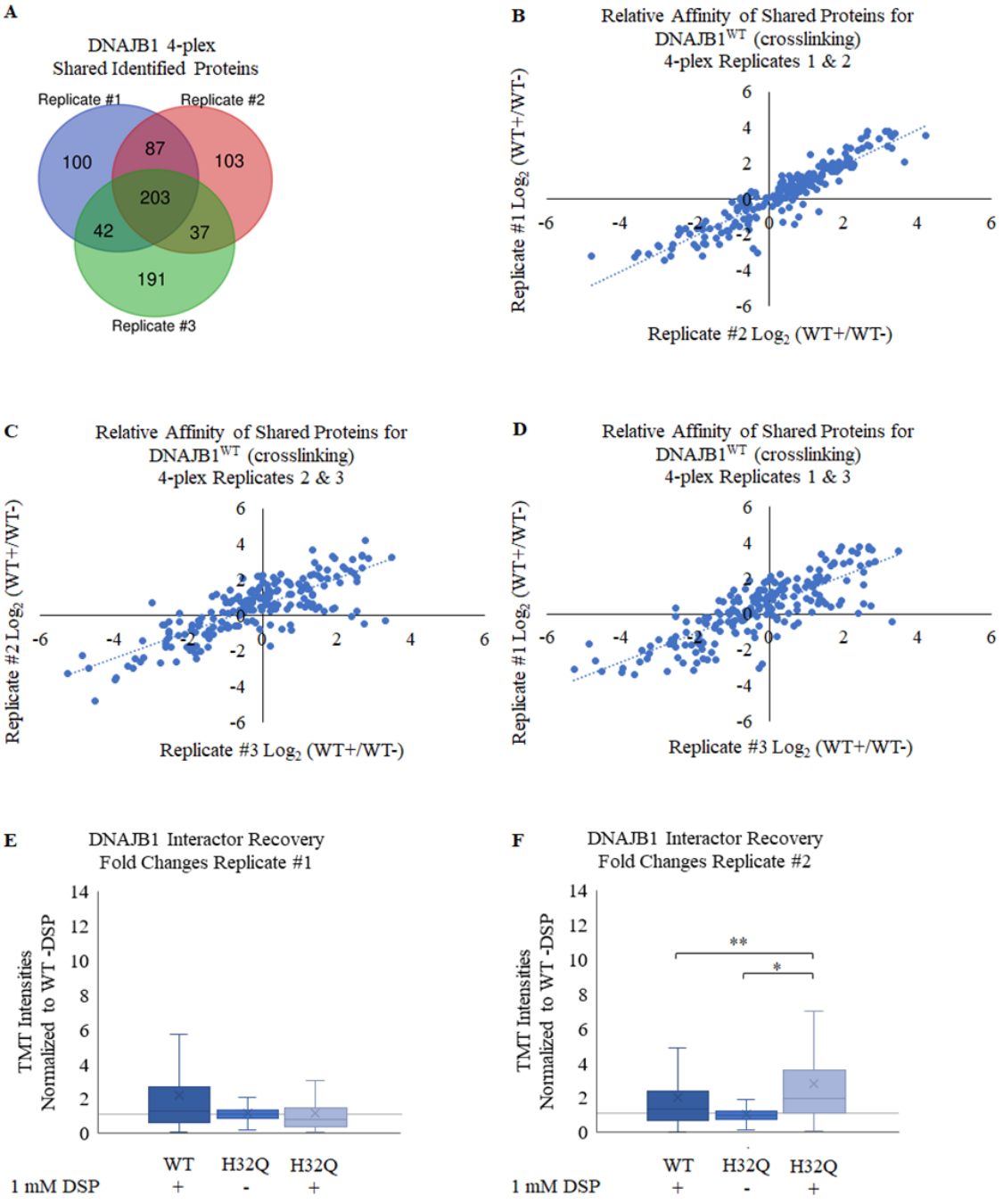
conditions. A low stringency buffer would increase non-specific binding but may allow for more proteins to be recovered without crosslinker.

There is little observed dependence of DNAJB1 interactors on J-domain activity and we little effect on recovery of proteins with crosslinking (**Figure 3.8A-C**). We see a slight affect for the wild type with crosslinking (**Figure 3.8A**, WT +) compared to mutant with and without crosslinking for higher recovery of proteins, but the large deviation in the whisker implies that this increase in recovery with crosslinking is not representative of most recovered interactors. In **Figure 3.8D-F**, the lower stringency has similar reproducibility between replicates. Stringency of the buffer overall did not improve recovery.

For DNAJB1 to be a useful probe in identifying clients, we need both reliable bait-normalized results and a condition with the highest bait-normalized recovery, which can be evaluated by monitoring the bait TMT intensity under the chosen conditions. Thus far, the conditions we have studied have diminished bait recovery results for DNAJB1. According to data from our 4-plex experiments, crosslinker appears to help with protein recovery for some proteins, but most are unaffected. We do observe that crosslinker decreases bait recovery, so for our first exploration of the DNAJB1 interactome, it may be that DNAJB1 would be best investigated by TMT-AP-MS without crosslinking. Additionally, the 4-plex data suggests that DNAJB1^{WT} and DNAJB1^{H32Q} do not have significant differences in recovery of prey, therefore either construct could be used for uncovering DNAJB1 interactors.

Table 3.3. Protein Class of DNAJB1 Client Recovery Improved with Crosslinking	
GO Protein Class*	Gene Name (Fold Change WT+DSP/WT-DSP)
Chaperone	HSPE1 (4.3, 12.6, 2.6), HSP90AA1 (10.5, 5.6, 5.2), HSP90AB1 (6.5, 5.2, 5.5)
Cytoskeletal protein	ARPC4 (10.3, 7.9, 2.6)
Metabolite interconversion enzyme	CKB (11.9, 18.8, 6.9), TPI1 (7.2, 9.9, 6.5), LDHA (8.2, 6.3, 2.9), LDHB (7.8, 6.6, 4.5), MAT2A (5.9, 4.1, 2.1), GAPDH (3.7, 4.8, 2.8), PAICS (3.7, 3.6, 4.3), PRDX6 (4.2, 2.8, 2.4)
Scaffold-adaptor protein	YWHAB (12.6, 10.5, 6.5), YWHAE (11.7, 6.1, 4.7), YWHAG (11.4, 9.9, 3.7), YWHAH (13.9, 9.3, 3.8), YWHAQ (13.7, 8.8, 5.4), YWHAZ (13.7, 6.3, 6.5)

*Protein class match from Gene Ontology Consortium database[49-51]


A
DNAJB1 4-plex
Shared Identified Proteins
Replicate #1
Replicate #2
Replicate #3
100
87
103
203
42
37
191
B
Relative Affinity of Shared Proteins for
DNAJB1WT (crosslinking)
4-plex Replicates 1 & 2
Replicate #1 Log2 (WT+/WT-)
Replicate #2 Log2 (WT+/WT-)
C
Relative Affinity of Shared Proteins for
DNAJB1WT (crosslinking)
4-plex Replicates 2 & 3
Replicate #2 Log2 (WT+/WT-)
Replicate #3 Log2 (WT+/WT-)
D
Relative Affinity of Shared Proteins for
DNAJB1WT (crosslinking)
4-plex Replicates 1 & 3
Replicate #1 Log2 (WT+/WT-)
Replicate #3 Log2 (WT+/WT-)
E
DNAJB1 Interactor Recovery
Fold Changes Replicate #1
TMT Intensities
Normalized to WT -DSP
WT
H32Q
H32Q
1 mM DSP
+
-
+
F
DNAJB1 Interactor Recovery
Fold Changes Replicate #2
TMT Intensities
Normalized to WT -DSP
**
*
WT
H32Q
H32Q
1 mM DSP
+
-
+

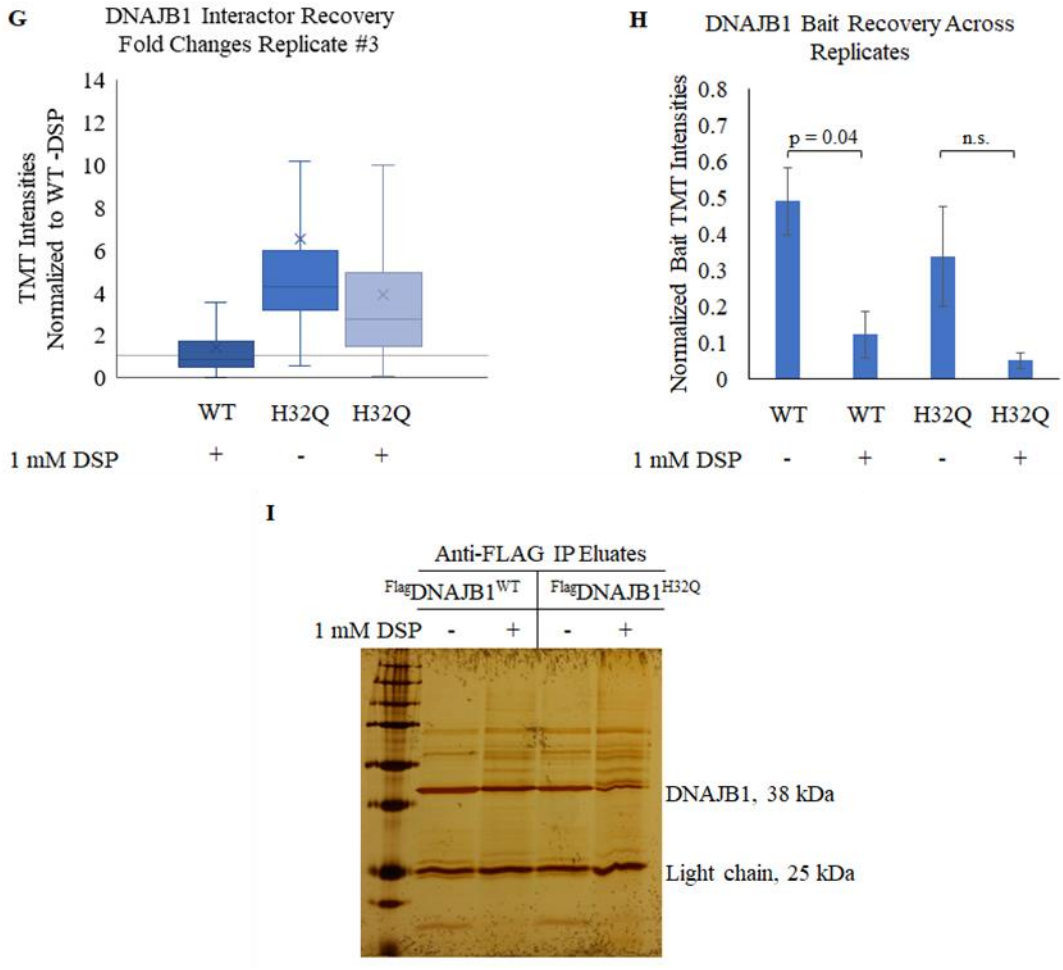

G
DNAJB1 Interactor Recovery
Fold Changes Replicate #3
TMT Intensities
Normalized to WT -DSP
14
12
10
8
6
4
2
0
WT
H32Q
H32Q
1 mM DSP
+
-
+
H
DNAJB1 Bait Recovery Across
Replicates
Normalized Bait TMT Intensities
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
p = 0.04
n.s.
WT
WT
H32Q
H32Q
1 mM DSP
-
+
-
+
I
Anti-FLAG IP Eluates
FlagDNAJB1WT
FlagDNAJB1H32Q
1 mM DSP
-
+
-
+
DNAJB1, 38 kDa
Light chain, 25 kDa

Figure 3.6

Figure 3.6 A) Venn diagram showing shared identified proteins across the three DNAJB1 4-plex replicates. **B**-**D**) Scatter plot showing relative affinity shared proteins recovered for DNAJB1 high stringency 4-plex replicates. Values plotted are fold changes of the WT +DSP normalized to WT -DSP for proteins that the two replicates share. **E-G**) Box-and-Whisker plot from high stringency buffer experiments (Replicates #1, 2 and 3) showing TMT intensities for three conditions: $DNAJB1^{WT}$ with crosslinker, $DNAJB1^{H32Q}$ without crosslinker, and $DNAJB1^{H32Q}$ with crosslinker, all normalized to $DNAJB1^{WT}$ without crosslinker. x marks the median value. Unity is marked with a gray line across the plot. Brackets indicate significance from two-tailed t-statistics with $\alpha = 0.05$ (*4 $x10^{-35}$, **1 $x10^{-15}$). **H**) Bar graph representation of DNAJB1 bait recovery from TMT intensities for 4-plex high stringency experiment. Bait TMT Intensities were normalized to the sum of all four bait TMT intensities in a single run, for each run. **I**) Representative silver stain of SDS-PAGE separated Flag immunoprecipitates from HEK293T cells overexpressing the indicated plasmids for high stringency DNAJB1 4-plex experiments.

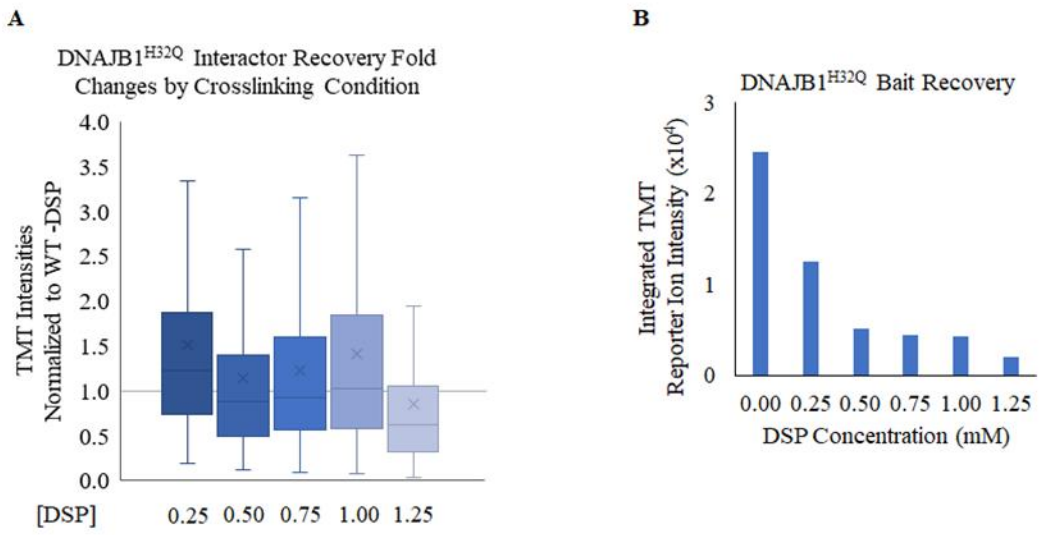


Figure 3.7

Figure 3.7. A) Box-and-Whisker plot from crosslinking dependence showing TMT intensities for five conditions: DNAJB1^{H32Q} with 0.25 mM crosslinker, DNAJB1^{H32Q} with 0.50 mM crosslinker, DNAJB1^{H32Q} with 0.75 mM crosslinker, DNAJB1^{H32Q} with 1.00 mM crosslinker, and DNAJB1^{H32Q} with 1.25 mM crosslinker, all normalized to DNAJB1^{H32Q} without crosslinker. **B**) Bar graph representation of DNAJB1 bait recovery from TMT intensities at varying crosslinker concentrations.

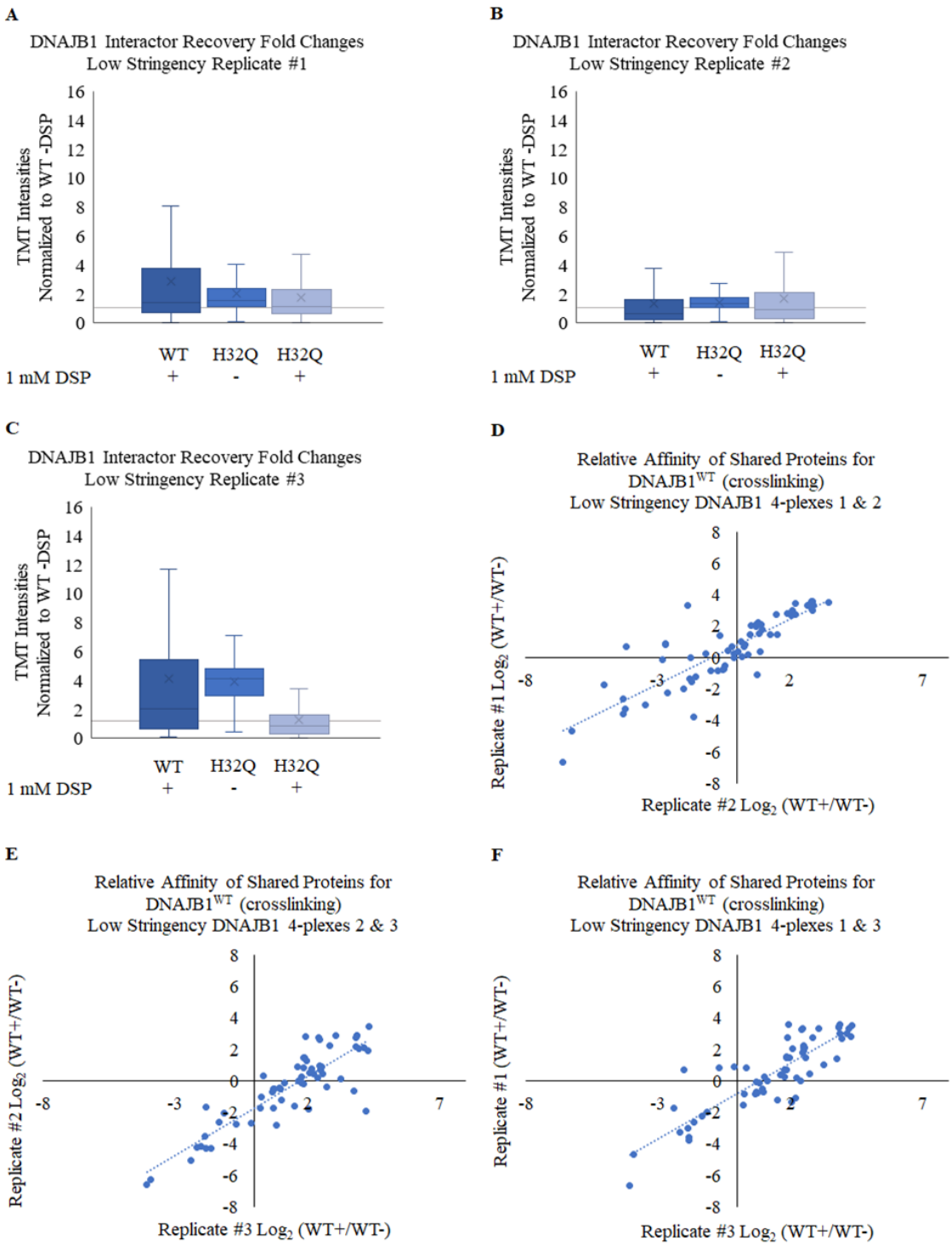

A
DNAJB1 Interactor Recovery Fold Changes
Low Stringency Replicate #1
TMT Intensities
Normalized to WT -DSP
WT
H32Q
H32Q
1 mM DSP
+
-
+
B
DNAJB1 Interactor Recovery Fold Changes
Low Stringency Replicate #2
C
DNAJB1 Interactor Recovery Fold Changes
Low Stringency Replicate #3
D
Relative Affinity of Shared Proteins for
DNAJB1WT (crosslinking)
Low Stringency DNAJB1 4-plexes 1 & 2
Replicate #1 Log2 (WT+/WT-)
Replicate #2 Log2 (WT+/WT-)
E
Relative Affinity of Shared Proteins for
DNAJB1WT (crosslinking)
Low Stringency DNAJB1 4-plexes 2 & 3
Replicate #2 Log2 (WT+/WT-)
Replicate #3 Log2 (WT+/WT-)
F
Relative Affinity of Shared Proteins for
DNAJB1WT (crosslinking)
Low Stringency DNAJB1 4-plexes 1 & 3
Replicate #1 (WT+/WT-)
Replicate #3 Log2 (WT+/WT-)

Figure 3.8

Figure 3.8. A-C) Box-and-Whisker plot (outliers not shown) from low stringency Replicate #1 (**A**), 2 (**B**), and 3 (**C**) of the DNAJB1 4-plex experiments showing TMT intensities for three conditions: $DNAJB1^{WT}$ with crosslinker, $DNAJB1^{H32Q}$ without crosslinker, and $DNAJB1^{H32Q}$ with crosslinker, all normalized to $DNAJB1^{WT}$ without crosslinker. **D**) Scatter plot showing relative affinity of recovered interactors for $DNAJB1^{WT}$ for low stringency 4-plex Replicates #1 and 2. Values plotted are fold changes of conditions normalized to WT -DSP. **E**) Scatter plot showing relative affinity of recovered interactors for $DNAJB1^{WT}$ for low stringency 4-plex Replicates #2 and 3. Values plotted are fold changes of WT +DSP normalized to WT -DSP **F**) Scatter plot showing relative affinity of recovered interactors for $DNAJB1^{WT}$ for low stringency 4-plex Replicates #1 and 3. Values plotted are fold changes of WT +DSP normalized to WT -DSP.

3.3.4 DNAJB1^{WT} interactors are predominantly chaperones.

We performed three replicate sets of 6-plex Bait (DNAJB1^{WT}) vs. Mock (GFP) experiments (**Figure 3.9A, B-D**) to uncover the potential clients of DNAJB1. Using our previously reported method of bait correlation with Pearson's R, followed by Benjamini-Hochberg analysis (Mei 2020), we compiled a list of interactors (19 proteins, FDR < 0.05) for DNAJB1^{WT} from three TMT experiments (**Figure 3.9E**). We found that DNAJB1^{WT} engages with chaperone proteins (**Table 3.4**). It is not surprising that DNAJB1 interacts with mostly chaperone proteins because it is the most abundant Hsp40 that is well-known for its role in the Hsp70 cycle. Additionally, DNAJB1 function is typically Hsp70-dependent and so it is expected that it will be bound to Hsp70 chaperones, which could make DNAJB1 substrate binding sites less available for clients. Therefore, uncovering true clients would be more difficult. If clients must compete with Hsp70 for DNAJB1 binding, then Hsp70 is not only accelerating client dissociation, but may also be suppressing client association.

Once we obtained a list of interactors for DNAJB1, we reanalyzed the 4-plex experiments to determine whether crosslinking improves specifically interactor recovery for the most reproducible interactors. We found that only 2/20 interactors, HSPH1 and NUDC are present and enriched with crosslinking in the three high stringency replicates (**Table 3.5**). One interactor, DNAJC7 was only found in two replicates, but appears to be enriched with crosslinking. It is likely that crosslinking did not enrich the identified interactors because they are mostly chaperones bound to DNAJB1, and not the weaker, more transient interactions between DNAJB1 and true clients.

Table 3.4. List of DNAJB1^{WT} Interactors	
Gene Name (Fold Change Bait/Mock)	
BAG2 (217, 103, 58.9)	HSPA2 (18.5, 14.3, 17.8)
*BCKDK (9.4, 9.8, 9.9)	HSPA4 (7.9, 5.6, 6.3)
*CTSA (11.5, 25.4, 17.3)	HSPA5 (9.8, 8.5, 6.6)
DNAJB4 (22.0, 13.1, 50.7)	HSPA8 (23.3, 14.2, 16.7)
DNAJB5 (112, 195, 39.0)	HSPA9 (5.8, 6.0, 4.7)
DNAJC7 (4.0, 3.0, 2.4)	HSPH1 (11.5, 8.1, 6.6)
GLB1 (16.4, 18.3, 28.3)	MLF2 (67.2, 10.4, 18.1)
*HSPA1B (20.1, 12.9, 13.6)	NUDC (24.4, 20.6, 18.4)
*HSPA1L (17.6, 13.8, 15.3)	STUB1 (21.6, 8.7, 13.8)
TTLL12 (13.8, 12.1, 7.9)	

*Interactors not found in BioGRID Database for DNAJB1 interactors[52]

Table 3.5. DNAJB1 Interactors' Recoveries Improved with Crosslinking									
	Replicate #1			Replicate #2			Replicate #3		
Condition*	WT+	H32Q-	H32Q+	WT+	H32Q-	H32Q+	WT+	H32Q-	H32Q+
Gene Name	Fold Changes								
DNAJC2	N/A	N/A	N/A	2.6	0.8	3.3	1.1	2.4	5.3
HSPH1	3	0.9	2.6	2.9	1.1	3.2	2.1	2.6	9.0
NUDC	7.2	1.1	1.5	5.6	1.0	1.7	3.3	3.7	6.1

*Values normalized to WT -DSP, "-" = absence of DSP, "+" = presence DSP

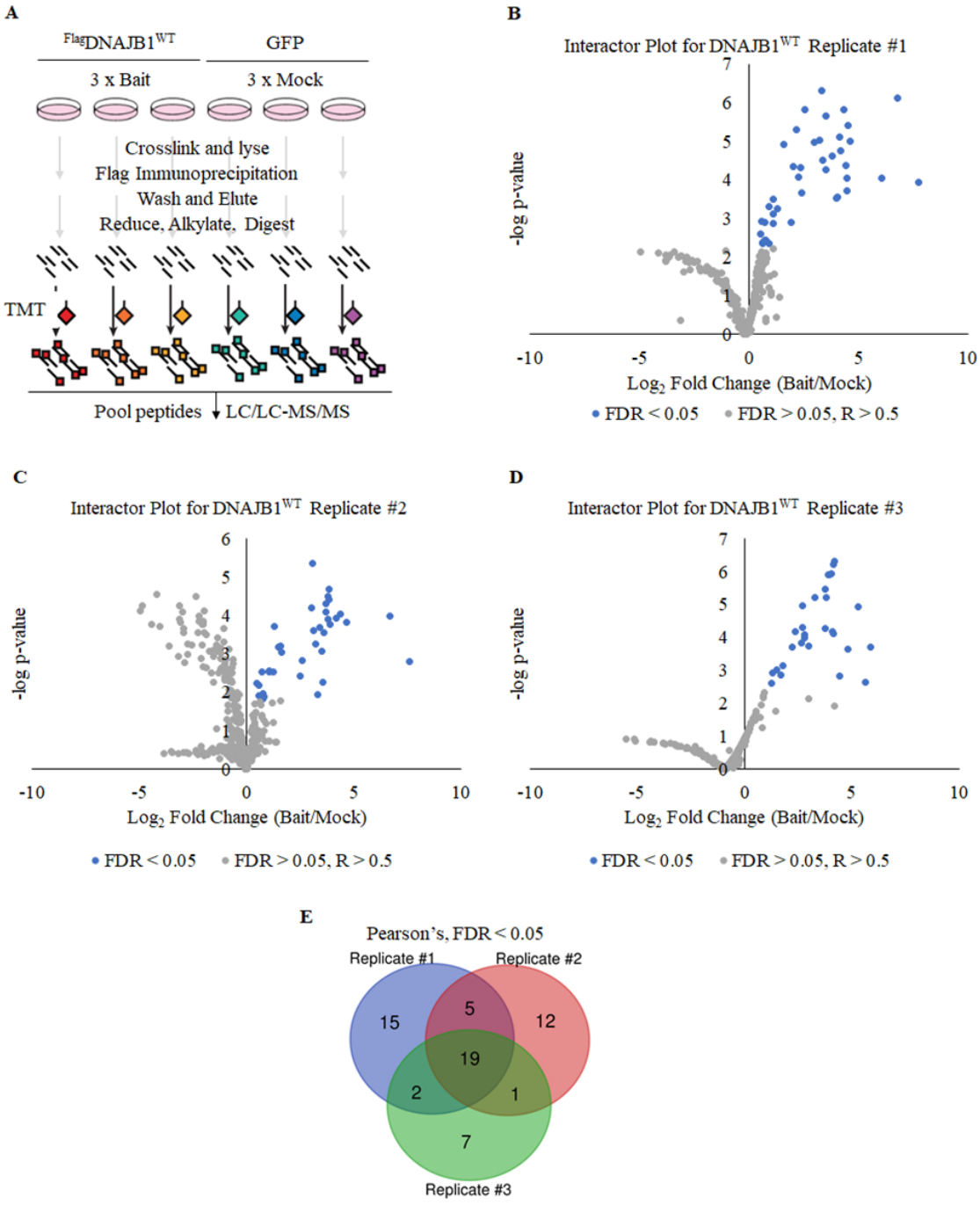

A
FlagDNAJB1WT
GFP
3 x Bait
3 x Mock
Crosslink and lyse
Flag Immunoprecipitation
Wash and Elute
Reduce, Alkylate, Digest
TMT
Pool peptides
LC/LC-MS/MS
B
Interactor Plot for DNAJB1WT Replicate #1
-log p-value
Log2 Fold Change (Bait/Mock)
FDR < 0.05
FDR > 0.05, R > 0.5
C
Interactor Plot for DNAJB1WT Replicate #2
-log p-value
Log2 Fold Change (Bait/Mock)
FDR < 0.05
FDR > 0.05, R > 0.5
D
Interactor Plot for DNAJB1WT Replicate #3
-log p-value
Log2 Fold Change (Bait/Mock)
FDR < 0.05
FDR > 0.05, R > 0.5
E
Pearson's, FDR < 0.05
Replicate #1
Replicate #2
Replicate #3
15
5
12
19
2
1
7

Figure 3.9

Figure 3.9. A) Experimental setup for Bait vs. mock TMT-AP-MS experiments for DNAJB1^{H32Q} without crosslinking. **B**) Volcano plot of Replicate #1 for DNAJB1^{WT}. Interactors with FDR < 0.05 for Replicate #1 are shown in blue. Less confident interactors with FDR > 0.05 or are with Pearson's R < 0.5 are shown in gray. **C**) Volcano plot of Replicate #2 for DNAJB1^{WT}. Interactors with FDR < 0.05 for Replicate #2 are shown in blue. Less confident interactors with FDR > 0.05 or are with Pearson's R < 0.5 are shown in gray. **D**) Volcano plot of Replicate #3 for DNAJB1^{WT}. Interactors with FDR < 0.05 for Replicate #3 are shown in blue. Less confident interactors with FDR > 0.05 or are with Pearson's R < 0.5 are shown in gray. **E**) Venn diagram showing shared proteins recovered across all three replicates. The three replicates share 19 proteins that have FDR < 0.05. q-values were determined by Benjamini-Hochberg analysis using p-values derived from Pearson's R.

3.3.5 DNAJB8^{WT} in the absence of crosslinker yields additional clients compared to DNAJB8^{H31Q}.

Based on the 4-plex crosslinking wild type versus mutant comparison experiments, DNAJB8^{WT} and DNAJB8^{H31Q} did not appear to have much difference in the amount or type of protein pulled down. We wanted to compare the clients of each to determine if there was a difference in the proteome acquired from wild type versus mutant immunoprecipitations. We performed bait (DNAJB8^{WT}) versus mock experiments as we did in the previous section with DNAJB1^{WT} and compared the clients we identified by TMT-AP-MS from three replicates (**Figure 3.10A-C, D**) with a set of known DNAJB8^{H31Q} interactors from our previous study (**Figure 3.10E**). About half (115/249) of DNAJB8^{WT} interactors were identified from the DNAJB8^{H31Q} study, and half of the interactors are unique (**Figure 3.10E**). There are a couple possible reasons for these two forms of DNAJB8 to be able to access different proteomes. First, in the wild type form of DNAJB8, the J-domain can function properly and bind Hsp70s. If DNAJB8^{WT} can bind Hsp70, we may be observing clients of Hsp70 that come into proximity with our DNAJB8^{WT} bait and allow DNAJB8^{WT} to interact with proteins it may otherwise not bind on its own. Yet, the fold changes for Hsp70-related proteins are similar for DNAJB8^{WT} as they are for DNAJB8^{H31Q} (with fold changes ranging from 5-40 for these proteins), so it is unclear if interaction with Hsp70 would increase the association with many client proteins. Second, these DNAJB8^{WT} bait vs. mock experiments are performed in the absence of crosslinker, therefore, the bait recovery is improved. As discussed for the 4-plex experiments, crosslinker has an adverse effect on bait levels. The interactor list

compiled from DNAJB8^{H31Q} experiments, were performed in the presence of 1 mM DSP crosslinker, which would have resulted in lower recovery of bait. With less bait, less prey may be captured, and we would access less of the proteome. If DNAJB8^{WT} can identify more clients than DNAJB8^{H31Q}, then its ability to bind misfolded proteins needs to be explored. Other than our previous study (**Chapter 2**) the reported interactors for DNAJB8 have been sparse, only 15 reported (see BioGrid[52]). To discover the full extent of the Hsp40-Hsp70 versatility to assist and restore misfolded proteins, the interactome for the different JDPs must be known. Therefore, we want the best condition that recovers the greatest number of clients for DNAJB8 in our investigation.

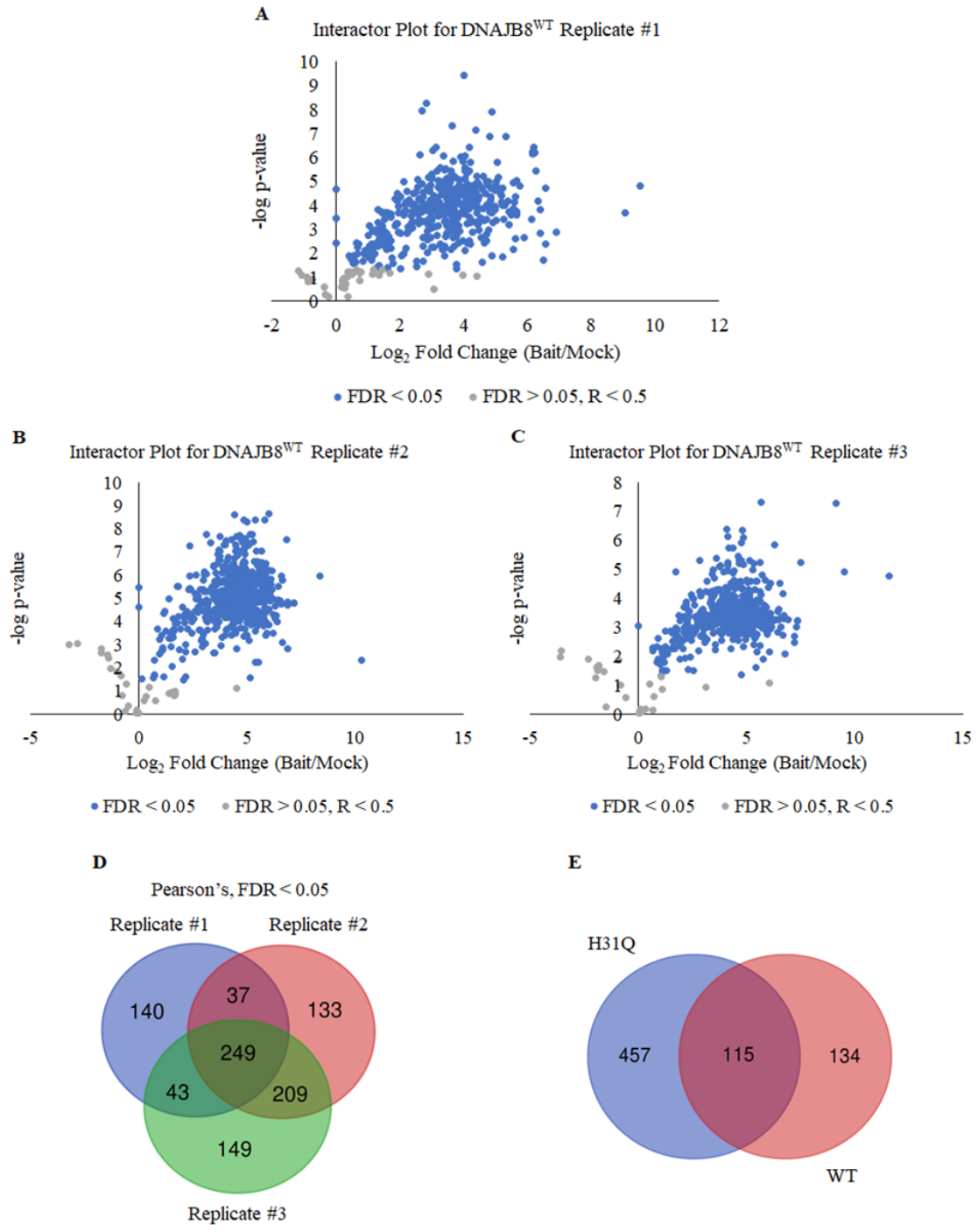

A
Interactor Plot for DNAJB8WT Replicate #1
-log p-value
Log2 Fold Change (Bait/Mock)
FDR < 0.05
FDR > 0.05, R < 0.5
B
Interactor Plot for DNAJB8WT Replicate #2
-log p-value
Log2 Fold Change (Bait/Mock)
FDR < 0.05
FDR > 0.05, R < 0.5
C
Interactor Plot for DNAJB8WT Replicate #3
-log p-value
Log2 Fold Change (Bait/Mock)
FDR < 0.05
FDR > 0.05, R < 0.5
D
Pearson's, FDR < 0.05
Replicate #1
Replicate #2
Replicate #3
140
37
133
249
43
209
149
E
H31Q
WT
457
115
134

Figure 3.10

Figure 3.10. A) Volcano plot of Replicate #1 for DNAJB8^{WT}. Interactors with FDR < 0.05 for Replicate #1 are shown in blue. Less confident interactors with FDR > 0.05 or are with Pearson's R < 0.5 are shown in gray. **B**) Volcano plot of Replicate #2 for DNAJB8^{WT}. Interactors with FDR < 0.05 for Replicate #2 are shown in blue. Less confident interactors with FDR > 0.05 or are with Pearson's R < 0.5 are shown in gray. **C**) Volcano plot of Replicate #3 for DNAJB8^{WT}. Interactors with FDR < 0.05 for Replicate #3 are shown in blue. Less confident interactors with FDR > 0.05 or are with Pearson's R < 0.5 are shown in gray. **D**) Venn diagram showing overlap between DNAJB8^{WT} Bait vs. mock experiment replicates. **E**) Venn diagram showing the overlap between DNAJB8^{H31Q} and DNAJB8^{WT} clients. DNAJB8^{H31Q} interactors are generated from a list of clients with FDR < 0.05 from the previous study discussed in **Chapter 2**. q-values were determined by Benjamini-Hochberg analysis using p-values derived from Pearson's R.

3.3.6 Mutant DNAJB8 is a better probe for targeting misfolded protein than wild type.

It is possible that the two baits might be best suited for probing stability of different subpopulations of DNAJB8 clients. To evaluate this hypothesis, we determined how DNAJB8 binding is affected by a stress that induces extensive and promiscuous protein misfolding. A set of heat shock experiments was performed by overexpressing either DNAJB8^{WT} or DNAJB8^{H31Q} in HEK293T cells and heating the cells varying temperatures. The heat shock would act as a stress to the cells and trigger the action of the Hsp40s to bind affected client proteins. We chose a 30 min heat shock incubation because longer times could decrease cell viability. Additionally, cells were harvested immediately after heat treatment to avoid recovery time that would cause pleiotropic effects, such as induction of the heat shock response transcriptional program.

Three sets of 6-plex experiments were performed for both DNAJB8^{WT} and DNAJB8^{H31Q}, with each plate of HEK293T cells incubated at a different temperature for 30 minutes (**Figure 3.11A**). If either wild type or mutant DNAJB8 is effective at probing for misfolded proteins, we expect to see higher recovery of those clients at higher temperatures; greater stress means more misfolding and therefore more binding to the Hsp40. For DNAJB8^{WT}, Box-and-Whisker plots of TMT intensities for each temperature show no change in proteins identified upon heat shock at any temperature (**Figure 3.11B-D**). It is possible that interaction with Hsp70 allows clients to be handed off to Hsp70, which may be why we do not see misfolded clients interacting with DNAJB8^{WT}. For these same experiments for DNAJB8^{H31Q}, we see an increase in the recovery of most

clients at 47°C (**Figure 3.12A, B**). When we compare the shared interactors from DNAJB8^{H31Q} heat shock experiments to the known DNAJB8^{H31Q} clients, we see that 209 (~60 %) of those identified match the known interactors (**Figure 3.12C**). Of these 209 proteins, 116 (~55 %) and 131 (~63%) have increased recovery (> 1.5-fold change 47 °C/37 °C) at 47 °C for Replicates #1 and #3, respectively. Replicate #2 has minimal change in recovery across all temperatures (**Figure 3.12D**), but also had particularly low TMT signals. This suggests that DNAJB8^{H31Q} is capable of targeting protein substrates affected by heat stress.

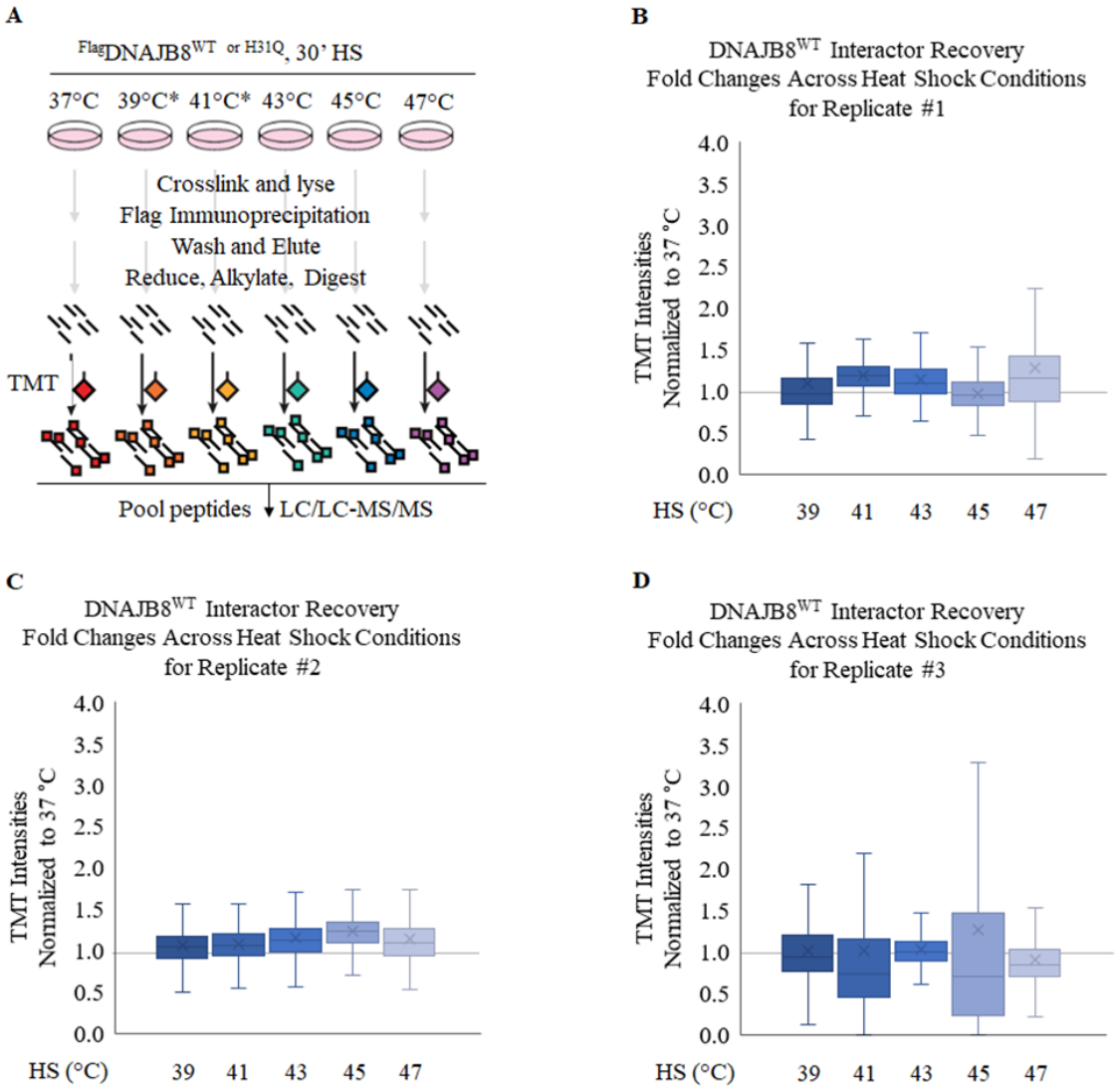


Figure 3.11

Figure 3.11. A) Experimental setup for DNAJB8 heat shock temperature dependence. Each 10 cm plate was incubated for 30 minutes at the indicated temperature. The indicated heat shock temperatures were used for both DNAJB8^{WT} and DNAJB8^{H31Q}, except for 39 °C and 41 °C: only 40°C was used for DNAJB8^{H31Q} experiments, not 39 °C or 41 °C. **B-D**) Box-and-Whisker plots for Replicates #1 (**B**), 2 (**C**), and 3 (**D**) from DNAJB8^{WT} heat shock experiments showing bait normalized TMT intensities for five conditions: 39 °C, 41 °C, 43 °C, 45 °C, and 47 °C, all normalized to 37 °C. Outliers not shown for visual clarity.

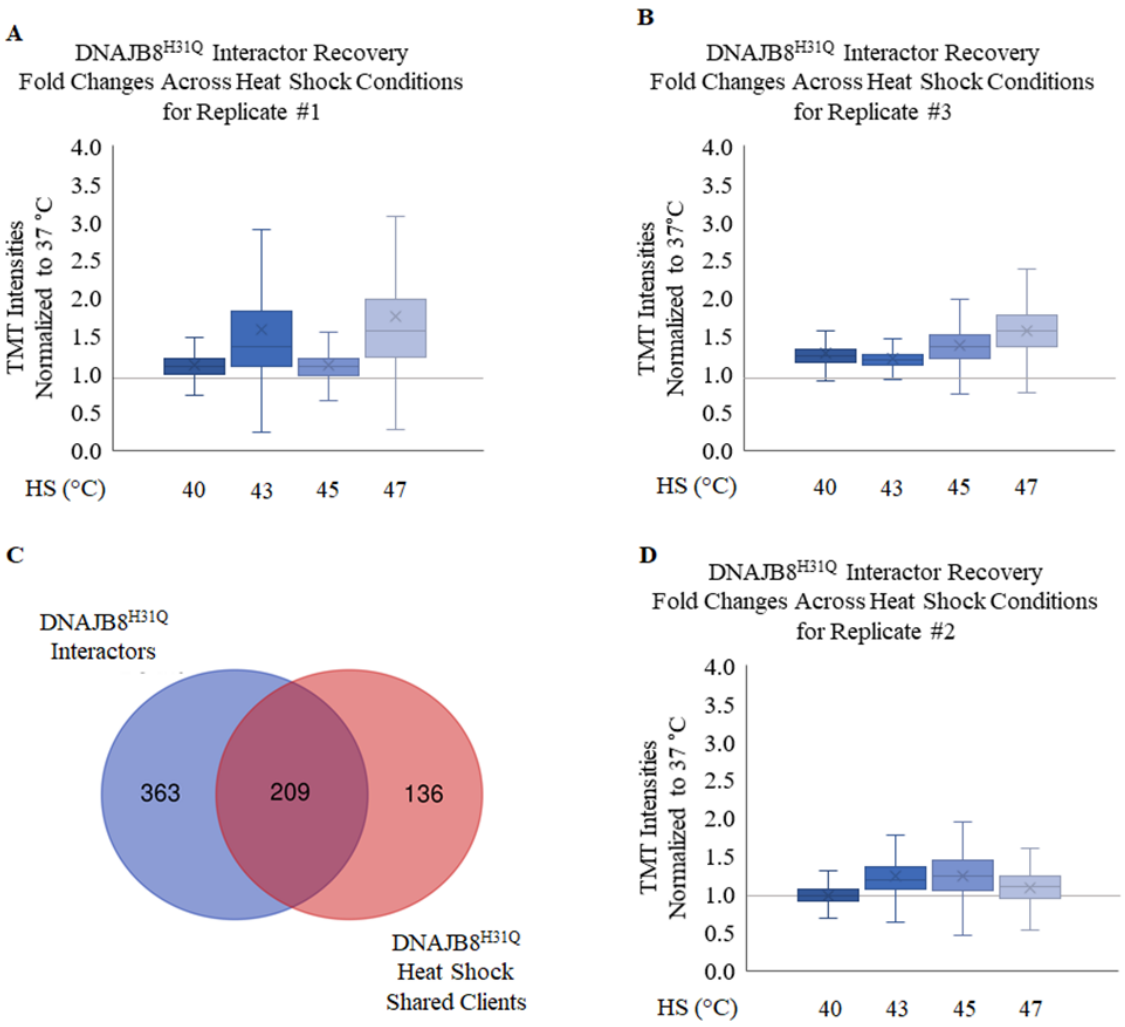


Figure 3.12

Figure 3.12. A) Box-and-Whisker plots for Replicates #1 (**B**) and 2 (**C**) from DNAJB8^{H31Q} heat shock experiments showing bait normalized TMT intensities for four conditions: 40 °C, 43 °C, 45 °C, and 47 °C, all normalized to 37 °C. For the final 47 °C condition, proteins that had a higher fold change than the GFP 47 °C condition were filtered out. Outliers not shown for visual clarity. **C**) Venn diagram comparing the shared interactors in the DNAJB8^{H31Q} heat shock experiments to the known interactors of DNAJB8^{H31Q}. **D**) Box-and-Whisker plots for Replicates #3 from DNAJB8^{H31Q} heat shock experiments showing bait normalized TMT intensities for four conditions: 40 °C, 43 °C, 45 °C, and 47 °C, all normalized to 37 °C. For the final 47 °C condition, proteins that had a higher fold change than the GFP 47 °C condition were filtered out.

3.4 Conclusions

We have differentiated client recovery under crosslinking conditions for the Hsp40s, DNAJB1 and DNAJB8. Crosslinker improves recovery of some proteins at the cost of bait recovery. For TMT-AP-MS experiments, bait recovery is essential for identifying clients. DNAJB1 bait recovery suffers in the presence of crosslinker and is unable to recover many potential interactors, with or without crosslinking, for the wild type or mutant forms. Considerable improvements and optimization would need to be performed to make DNAJB1 an appropriate bait for client recovery. After uncovering DNAJB1 interactors by our TMT-AP-MS experiments, we find that crosslinker (at least at 1 mM DSP) does not help improve interactor recovery. These experiments also produced few interactors (only 19) when compared to the number of DNAJB8 interactors (>500) we can identify. Since DNAJB1 predominantly binds chaperone proteins, it may not be an appropriate bait protein if we want to explore proteins that misfold under stress conditions. DNAJB8 has already been optimized for this assay and has more reliable bait recovery under many conditions. We determined that DNAJB8^{WT} co-precipitates a partially different proteome from the J-domain inactive mutant DNAJB8^{H31Q}. Provided the interactors of DNAJB8^{H31Q} are not affected by crosslinking, this difference in the client pool between wild type and mutant could be attributed to Hsp70 interaction with DNAJB8^{WT}. Although DNAJB8^{WT} can pull-down a different set of interactors than DNAJB8^{H31Q}, heat shock studies show that DNAJB8^{WT} is not able to retain misfolded clients with increasing temperatures. DNAJB8^{H31Q} demonstrates some level of client recovery with heat stress, but further studies with more replicates would need to be

performed to confirm. Hsp40s are the most diverse group of HSPs, with over 40 members that capture and carry unfolded clients to Hsp70. Identifying what clients these Hsp40s bind to will bring us closer to understanding why the Hsp70 cycle is efficient in aiding the misfolded proteome.

3.5 Acknowledgements

pcDNA5/FRT/TO V5 DNAJB8 (Addgene # 19531) was generously shared by H. Kampinga. eGFP.pDEST30 plasmid was generously shared by R.L. Wiseman.

3.6 References

(1) Powers E. T., Morimoto R. I., Dillin A.; Kelly J. W.; Balch W. E. *Annu Rev Biochem*. **2009**,78, 959-991.

(2) Narayan, P.; Ehsani, S.; Lindquist, S. *Nat. Chem. Biol.* **2014**, *10* (11), 911–920.

(3) Morimoto, R. I. *Science* **1993**, *259*, 1409–1410.

(4) Kampinga, H. H.; Craig, E. A. *Nat. Rev. Mol. Cell Biol.* **2010**, *11* (8), 579–592.

(5) Szabo, A.; Langer, H.; Schroder, H.; Flanagan, J.; Bukau, B.; Hartl, F. U. *PNAS*. **1994**. *91* (22): 10345-10349.

(6) Meimaridou, E.; Gooljar, S. B.; Chapple, J. P. *J Mol Endocrinol*. **2009**, *42* (1), 1-9.

(7) Fernández-Fernández, M. R.; Gragera, M.; Ochoa-Ibarrola, L.; Quintana-Gallardo, L.; Valpuesta, J. M. *FEBS Lett*. **2017**, *591* (17), 2648-2660.

(8) Shalgi, R.; Hurt, J. A.; Krykbaeva, I.; Taipale, M.; Lindquist, S.; Burge, C. B. *Mol Cell* **2013**, *49* (3), 439-452.

(9) Morimoto, R. I. *Genes Dev*. **1998**, *12* (24), 3788-3796.

(10) Ellgaard, L.; Helenius, A. *Nat Rev Mol Cell Biol*. **2003**, *4* (3), 181-191.

(11) Gao, Z.; Niu, X.; Zhang, Q.; et al. *Oncotarget* **2017**, *8* (35), 58536-58552.

(12) Jiang, Y.; Rossi, P.; Kalodimos, C. G. *Science (80).* **2019**, *365* (6459), 1313–1319.

(13) Ryder, B. D.; Matlahov, I.; Bali, S.; Vaquer-Alicea, J.; van der Wel, P. C. A. *BioRxiv*. **2020**.

(14) Kampinga, H. H.; Andreasson, C.; Barducci, A.; et al. *Cell Stress Chaperones* **2019**, *24* (1), 7-15.

(15) Hennessy, F.; Cheetham, M. E.; Dirr, H. W.; Blatch, G. L. *Cell Stress Chaperones* **2000**, *5* (4), 347–358.

(16) Cheetham, M. E.; Caplan, A. J. *Cell Stress Chaperones* **1998**, *3* (1), 28-36.

(17) Hageman, J.; Kampinga, H. H. *Cell Stress Chaperones* **2009**, *14* (1), 1–21.

(18) Ohtsuka, K.; Hata, M. *Cell Stress Chaperones* **2000**, *5* (2), 98-112.

(19) Ellis, J. *Nature* **1987**, 328, 378–9.

(20) Hartl, F. U.; Hayer-Hartl, M. *Nat Struct Mol Biol*. **2009**, 16, 574–81.

(21) Bukau, B.; Weissman, J.; Horwich, A. *Cell* **2006**, 125, 443–51.

(22) Gillis, J.; Schipper-Krom, S.; Juenemann, K.; et al. *J. Biol. Chem.* **2013**, *288* (24), 17225–17237.

(23) Hageman, J.; Rujano, M. A.; van Waarde, M. A., et al. *Mol. Cell* **2010**, *37* (3), 355–369.

(24) Davis, A. K.; Pratt, W. B.; Lieberman, A. P.; Osawa, Y. *Cell Mol Life Sci*. **2020**, *77* (6), 977-996.

(25) Nillegoda, N. B.; Stank, A.; Malinverni, D.; et al. *Elife* **2017**, 6, e24560.

(26) Seidel, K.; Siswanto, S.; Brunt, E. R.; den Dunnen, W.; Korf, H. W.; Rüb, U. *Acta Neuropathol*. **2012**, *124* (1), 1-21.

(27) Kakkar, V.; Kuiper, E. F. E.; Pandey, A.; Braakman, I.; Kampinga, H. H. *Sci. Rep.* **2016**, *6* (September), 1–12.

(28) Qian, Y.Q.; Patel, D.; Hartl, F.U.; McColl, D.J. *J Mol Biol*. **1996**, 260, 224-235.

(29) Ohnishi, S.; Tochio, N.; Koshiba, S.; Inoue, M.; Kigawa, T.; Yokoyama, S. To be published.

(30) Goddard, T. D.; Huang, C. C.; Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Morris, J. H.; Ferrin, T. E. *Protein Sci*. **2018**, *27* (1), 14-25.

(31) Klock, H. E.; Lesley, S. A. *Methods Mol Biol.* **2009**, 498, 91–103.

(32) Shoulders, M. D.; Ryno, L. M.; Genereux, J. C.; et al. *Cell Rep*. **2013**, *3* (4), 1279–1292.

(33) Plate, L.; Rius, B.; Nguyen, B.; Genereux, J. C.; Kelly, J. W.; Wiseman, R. L. *Cell Chem. Biol*. **2019**, *26* (7), 913-925.e4.

(34) Washburn, M. P.; Wolters, D.; Yates, J. R. *Nat. Biotechnol*. **2001**, *19* (3), 242–247.

(35) Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom*. **1994**, *5* (11), 976–989.

(36) Xu, T.; Park, S. K.; Venable, J. D.; et al. *J. Proteomics* **2015**, 129, 16–24.

(37) Tabb, D. L.; McDonald, W. H.; Yates, J. R. *J Proteome Res* **2002**, *1* (1), 21–26.

(38) Park, S. K.; Venable, J. D.; Xu, T.; Yates, J. R. *Nat Methods* **2008**, *5* (4), 319–322.

(39) Fisher, R. A. Statistical Methods for Research Workers, 14th ed., revised and enlarged.; *Oliver and Boyd*: Edinburgh, **1970**.

(40) Hogg, R. V.; Tanis, E. A. Probability and Statistical Inference, 2nd ed.; Macmillan ; *Collier Macmillan*: New York : London, **1983**.

(41) Benjamini, Y.; Hochberg, Y. *J. R. Stat. Soc. Ser. B-Stat. Methodol*. **1995**, *57* (1), 289–300.

(42) Makowski, M. M.; Willems, E.; Jansen, P. W.; Vermeulen, M. *Mol Cell Proteomics* **2016**, *15* (3), 854-865.

(43) Papachristou, E. K.; Kishore, K.; Holding, A. N.; Harvey, K.; et al. *Nat. Commun*. **2018**, *9* (1), 2311.

(44) Söderberg, C. A. G.; Månsson, C.; Bernfur, K.; et al. *Sci Rep*. **2018**, *8* (1), 5199.

(45) Genereux, J. C.; Qu, S.; Zhou, M.; et al. *EMBO J.* **2015**, *34* (1), 4-19.

(46) Gingras, A. C.; Gstaiger, M.; Raught, B.; Aebersold, R. *Nat. Rev. Mol. Cell Biol*. **2007**, *8* (8), 645–654.

(47) Schmidt, C.; Beilsten-Edmands, V.; Robinson, C. V. *Oncotarget* **2015**, *6* (21), 18276-18281.

(48) Wang, H.; Li, S.; Wang, J.; Chen, S.; Sun, X. L.; Wu, Q. *Elife* **2018**, 7, e35672.

(49) Ashburner, M.; Ball, C. A.; Blake, J. A.; et al. *Nat Genet*. **2000**, *25* (1), 25-9.

(50) The Gene Ontology Consortium. *Nucleic Acids Res*. **2019**, *47* (D1), D330-D338.

(51) Mi, H.; Huang, X.; Muruganujan, A.; Tang, H.; Mills, C.; Kang, D.; Thomas, P. D. *Nucleic Acids Res*. **2019**, *47* (D1), D419-D426.

(52) Chatr-aryamontri, A.; Oughtred, R.; Boucher, L.; et al. *Nucleic Acids Res*. **2017**, *45* (D1), D369–D379.

Chapter 4: Perspectives and Concluding Remarks

4.1 Perspectives

It has been suggested that JPDs, not Hsp70, may determine the folding fates of their clients, targeting their substrates for refolding or degradation[1-3]. JDPs can also function differently depending on the clients they bind. As discussed in **Chapter 1**, DNAJB6 can suppress aggregation of poly-Q containing proteins without Hsp70 interaction yet requires Hsp70 interaction to suppress parkin C289G aggregation. Additionally, DNAJB1 can target some clients for degradation, inhibit aggregation, or recruit clients to Hsp70 for refolding[4-6]. JDPs can and do determine what happens to misfolded substrates and it is essential that we understand the extent of their substrate-related functions. In this section, we will explore a few of the other class B JDPs and relate them to what is known from our previous studies in **Chapter 2** and **3**.

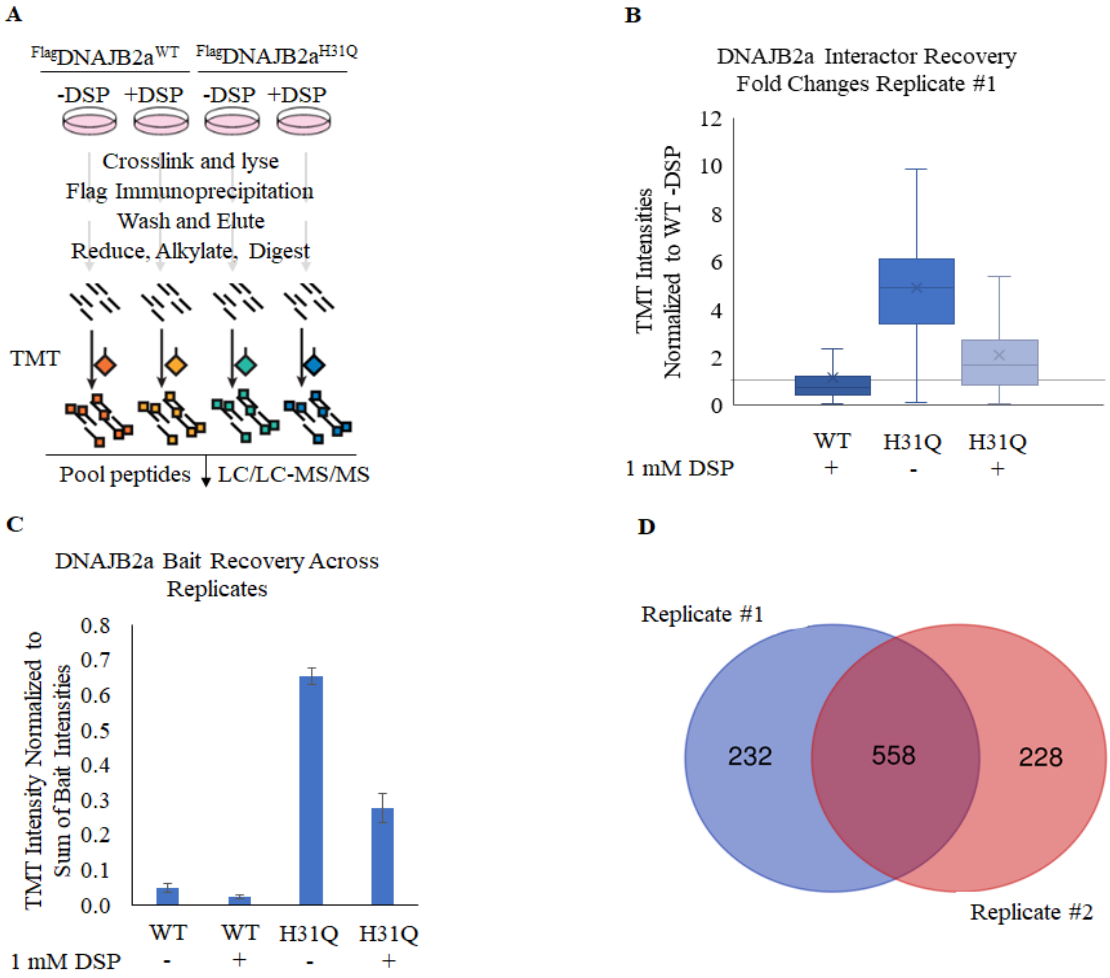
4.1.1 Uncovering the DNAJB2a Interactome: Is it DNAJB8-like, DNAJB1-like, or both?

DNAJB2 is another JDP member that has demonstrated ability to suppress aggregation of the parkin RING1 C289G mutant[7]. It contains the conserved regions of all JDPs, the J-domain and the G/F-rich region but has C-terminal domains that differ from the previously discussed DNAJB1 and DNAJB8[1,8]. DNAJB2 has the typical ß-barrel in its first C-terminal domain and is followed by a ubiquitin-interacting motif[8]. This ubiquitin-interacting feature allows DNAJB2 to direct its clients toward degradation, such as with poly Q proteins[9]. DNAJB2 can even suppress refolding of misfolded clients in an Hsp70-dependent manner by directing them toward degradation pathways[10].

Since DNAJB2 can suppress aggregation like DNAJB6 and DNAJB8, it may be more DNAJB8-like, rather than DNAJB1-like. To evaluate how similar (or different) DNAJB2 is to DNAJB1 and DNAJB8, we can perform the same crosslinking experiment as described in **Chapter 3** in addition to interactome studies introduced in **Chapter 2**. Using the same experimental setup (**Figure 4.1A**) with DNAJB2a^{WT} and DNAJB2a^{H31Q}, we would expect to see no difference in client preference for wild type or mutant, as with DNAJB8, and that crosslinker decreases bait and client recovery. Interestingly, preliminary testing with two replicates of this 4-plex experiment shows that proteins are recovered more effectively by DNAJB2a^{H31Q} (**Figure 4.1B**) and the bait recovery profile is quite different than either DNAJB8 or DNAJB1 with far more bait recovered with H31Q (**Figure 4.1C**). It may be that bait recovery is driving client recovery. The identities of recovered proteins are reproducible (**Figure 4.1D, E-H**) and it appears as though DNAJB2a interactors show a preference toward DNAJB2a^{H31Q} with crosslinking (**Figure 4.1B**). This indicates that J-domain inactivation improves client recovery, contrasting with DNAJB8 client affinity. More work should be done to validate this finding.

Another set of experiments that would allow us to uncover DNAJB2a substrates would be bait correlation by TMT-AP-MS, as discussed in **Chapter 2**. Preliminary findings, from a dosed experiment (**Figure 4.2A**) with DNAJB2a^{H31Q} overexpressed at varying amounts in HEK293T cells, followed by crosslinking post-harvest, show that DNAJB2a^{H31Q} clients share ~78 % similarity with DNAJB8^{H31Q} interactors (**Figure 4.2B**). Additionally, when we compare to DNAJB1^{WT} interactors (only 19 found in

Chapter 3), DNAJB1^{WT} shares only ~16 % similarity with DNAJB2a^{H31Q} (**Figure 4.2 B**). Although these initial studies imply that DNAJB2a is more like DNAJB8 than DNAJB1, we cannot draw clear conclusions from these findings or verify clients of DNAJB2a without further interactome replicates.



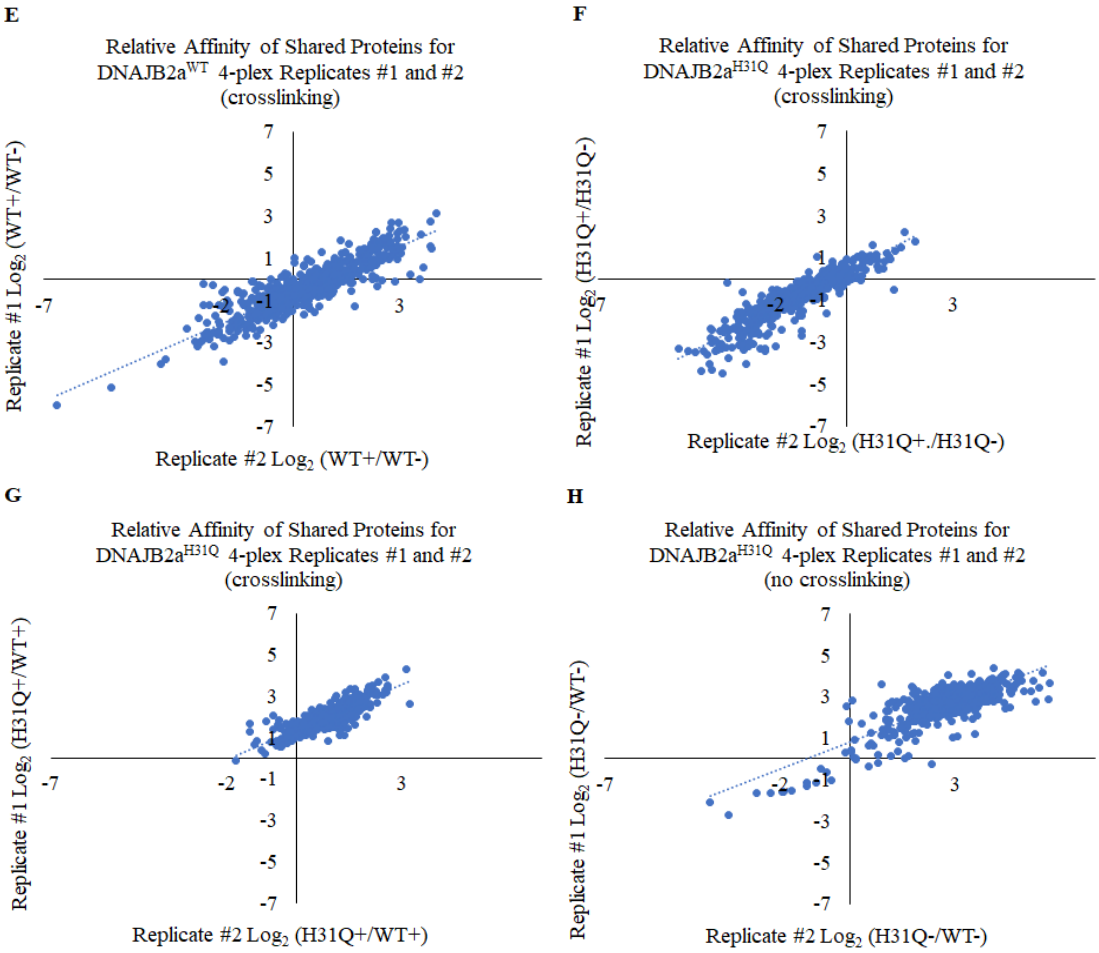

E
Relative Affinity of Shared Proteins for DNAJB2a^WT 4-plex Replicates #1 and #2 (crosslinking)
Replicate #1 Log2 (WT+/WT-)
Replicate #2 Log2 (WT+/WT-)
F
Relative Affinity of Shared Proteins for DNAJB2a^H31Q 4-plex Replicates #1 and #2 (crosslinking)
Replicate #1 Log2 (H31Q+/H31Q-)
Replicate #2 Log2 (H31Q+./H31Q-)
G
Relative Affinity of Shared Proteins for DNAJB2a^H31Q 4-plex Replicates #1 and #2 (crosslinking)
Replicate #1 Log2 (H31Q+/WT+)
Replicate #2 Log2 (H31Q+/WT+)
H
Relative Affinity of Shared Proteins for DNAJB2a^H31Q 4-plex Replicates #1 and #2 (no crosslinking)
Replicate #1 Log2 (H31Q-/WT-)
Replicate #2 Log2 (H31Q-/WT-)

Figure 4.1

Figure 4.1. A) Experimental setup for TMT-AP-MS 4-plex assay for DNAJB2a. **B**) Box-and-Whisker plot from high stringency buffer experiment showing TMT intensities for three conditions: DNAJB2a^{WT} with crosslinker, DNAJB2a^{H31Q} without crosslinker, and DNAJB2a^{H31Q} with crosslinker, all normalized to DNAJB2a^{WT} without crosslinker. x marks the median value. Unity is marked with a gray line across the plot. Outliers not shown for visual clarity. **C**) Bar graph representation of DNAJB2a bait recovery from integrated bait TMT reporter ion intensities. Bait TMT reporter ion intensities were normalized to the sum of all bait TMT report ion intensities for all bait channels for a single run to allow comparison between runs. **D**) Venn diagram comparing proteins recovered in both DNAJB2a 4-plex replicates. **E**) Scatter plot showing relative affinity of shared proteins for DNAJB2a^{WT} in Replicates #1 and 2. Values plotted are fold changes of conditions WT +DSP normalized to WT -DSP for proteins that the two replicates share. **F-H**) Scatter plots showing relative affinity of shared protein across the two replicates for DNAJB2a^{H31Q}. The data plotted are fold changes of conditions H31Q +DSP normalized to H31Q -DSP (**F**), H31Q +DSP normalized to WT +DSP (**G**), and H32Q -DSP normalized to WT -DSP (**H**).

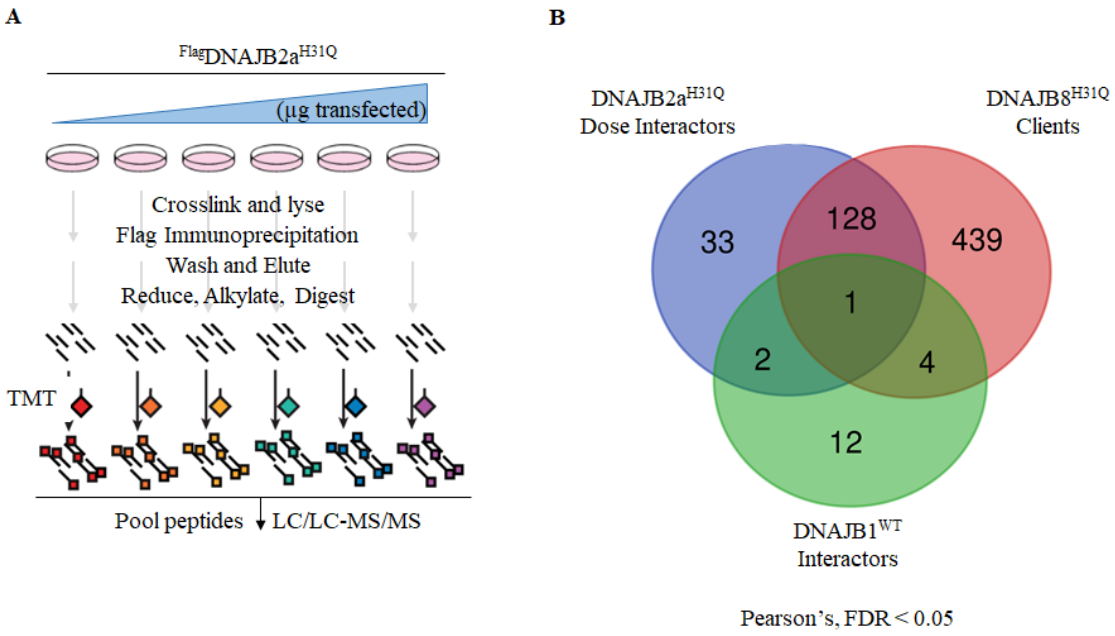


Figure 4.2

Figure 4.2. A) Experimental setup for TMT-AP-MS dosing assay for DNAJB2a^{H31Q}. HEK293T cells were transfected with 0 µg (GFP), 2 µg, 4 µg, 6 µg, or 8 µg of DNAJB2a^{H31Q}. Cells were crosslinked with 1 mM DSP post-harvest. **B**) Venn diagram showing shared proteins between interactors in the single DNAJB2a^{H31Q} dose run and the known clients of DNAJB8^{H31Q} (n=12), and interactors of DNAJB1^{WT} (n=3) found in Chapter 3.

4.1.2 Uncovering the DNAJB4 Interactome: Comparison to DNAJB1

DNAJB4 has similar sequence homology to DNAJB1, with nearly identical domain structure[1]. In the literature, DNAJB4 has been implicated in tumor suppression of lung, breast, and colon cancers[11-14]. This function makes DNAJB4 a potential drug target for cancer[15]. Although DNAJB4 is a good candidate for cancer studies, it has shown poor anti-aggregation activity in poly Q containing proteins[16]. With both its inability to suppress aggregation and its similarity in structure to DNAJB1, it is hypothesized that DNAJB4 may have similar client preference to DNAJB1.

A preliminary experiment using the 4-plex setup (see DNAJB2a **Figure 4.1A**) with $DNAJB4^{WT}$ and $DNAJB4^{H32Q}$ shows that bait recovery is decreased with crosslinker (**Figure 4.3A**), just like DNAJB1 and DNAJB8 in **Chapter 3**, but the extent of decrease cannot be commented on without further replicates. The client recovery in the presence and absence of crosslinker, too, cannot be understood without more experimentation. When comparing the list of $DNAJB1^{WT}$ interactors from **Chapter 3**, to the potential clients recovered from the DNAJB4 experiment, $DNAJB1^{WT}$ shares ~63 % similarity to the proteins recovered with DNAJB4 (**Figure 4.3B**). This similarity of potential interactors of DNAJB4 with DNAJB1 is encouraging, but bait correlation interactome studies (**Chapter 2**) need to be performed to identify true interactors of DNAJB4 to allow for a more confident comparison between DNAJB4 and DNAJB1.

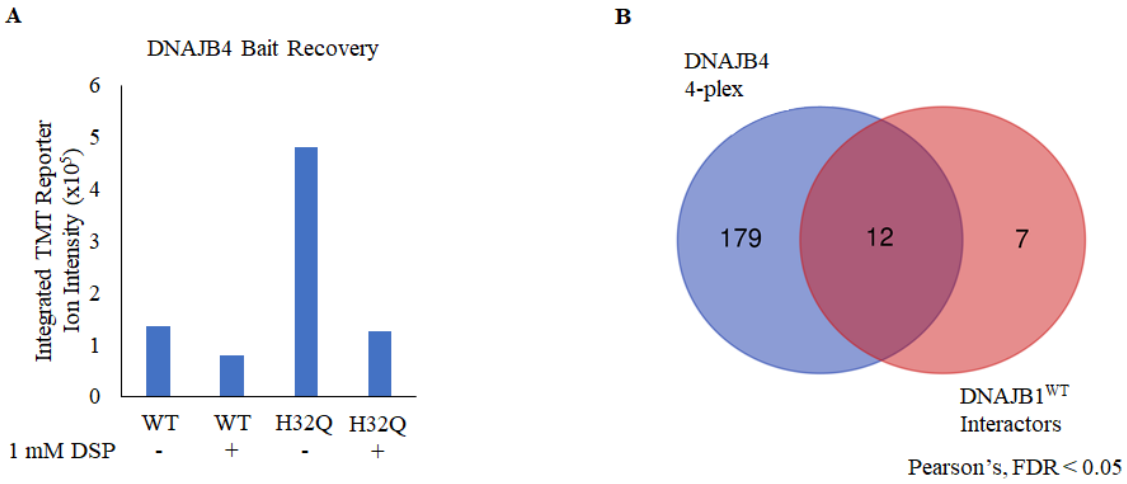


Figure 4.3

Figure 4.3. A) Bar graph representation of DNAJB4 bait recovery from integrated bait TMT reporter ion intensities. **B**) Venn diagram comparing proteins recovered from a single DNAJB4 4-plex run with the interactors (n=3) of DNAJB1^{WT} uncovered from Chapter 3.

4.1.3 Uncovering the DNAJB6b Interactome: Comparison to DNAJB8

As discussed in the Introduction (**Chapter 1**), DNAJB6 has several functions related to misfolded substrates, the most prominent being its aggregation suppression capabilities[4,7,16-21]. In many of these studies, DNAJB8 behaves like DNAJB6, particularly when it comes to proteins containing poly Q and the parkin C289G mutant[4,7,16]. Therefore, we wonder whether DNAJB6 and DNAJB8 would share most of the same clients and similar client binding properties. We hypothesize that DNAJB6 client recovery would be affected by crosslinking and that J-domain inactivation would have no effect on client recovery, just as we observed with DNAJB8 in **Chapter 3**. Preliminary experiments, overexpressing DNAJB6b^{WT} and DNAJB6b^{H31Q} in cells, do show a similar decrease in bait recovery with crosslinking like DNAJB8 (**Figure 4.4A**, **Figure 3.3I**). Clients do not appear to be affected by J-domain inactivation (**Figure 4.4B** H31Q - and H31Q +), but this cannot be certain with only one replicate. Depending on interactor recovery with crosslinking, it may be that we can uncover DNAJB6b interactors without crosslinking, as with DNAJB8.

However, we do have a set of preliminary experiments using bait correlation to determine interactors of DNAJB6^{H31Q} in the presence of 1 mM DSP crosslinker (**Figure 4.5A-C**). Using the same experimental setup from Bait vs. Mock experiments in **Chapters 2** and **3**, cells overexpressing DNAJB6b^{H31Q} or mock (GFP) were crosslinked post-harvest, followed by TMT-AP-MS. These initial interactome replicates show that 200 of the 231 (shared among two replicates, **Figure 4.5C**) DNAJB6b^{H31Q} interactors coincide with DNAJB8^{H31Q} clients (**Figure 4.5D**). Conversely, when compared to DNAJB1^{WT}

interactors, only 2 out of 19 are shared with DNAJB6b^{H31Q} (**Figure 4.5E**). Initial 4-plex (**Figure 4.4**) and interactome (**Figure 4.5**) experiments suggest that DNAJB6b^{H31Q} is more DNAJB8-like, as hypothesized, and should be further explored for confirmation.

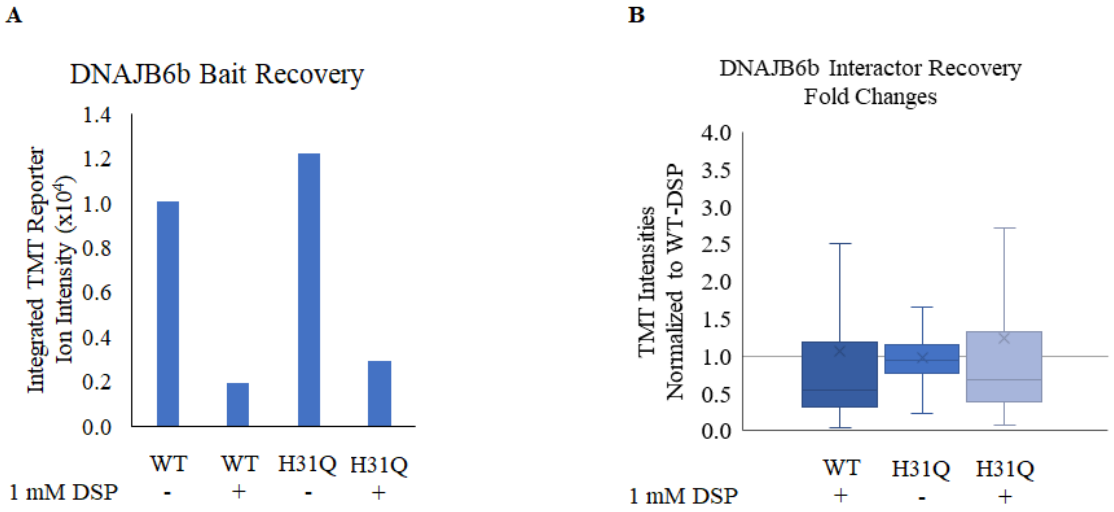


Figure 4.4

Figure 4.4. A) Bar graph representation of DNAJB6b bait recovery from integrated bait TMT reporter ion intensities. **B**) Box-and-Whisker plot from a 4-plex high stringency buffer experiment showing TMT intensities for three conditions: DNAJB6b^{WT} with crosslinker, DNAJB6b^{H31Q} without crosslinker, and DNAJB6b^{H31Q} with crosslinker, all normalized to DNAJB6b^{WT} without crosslinker. x marks the median value. Unity is marked with a gray line across the plot.

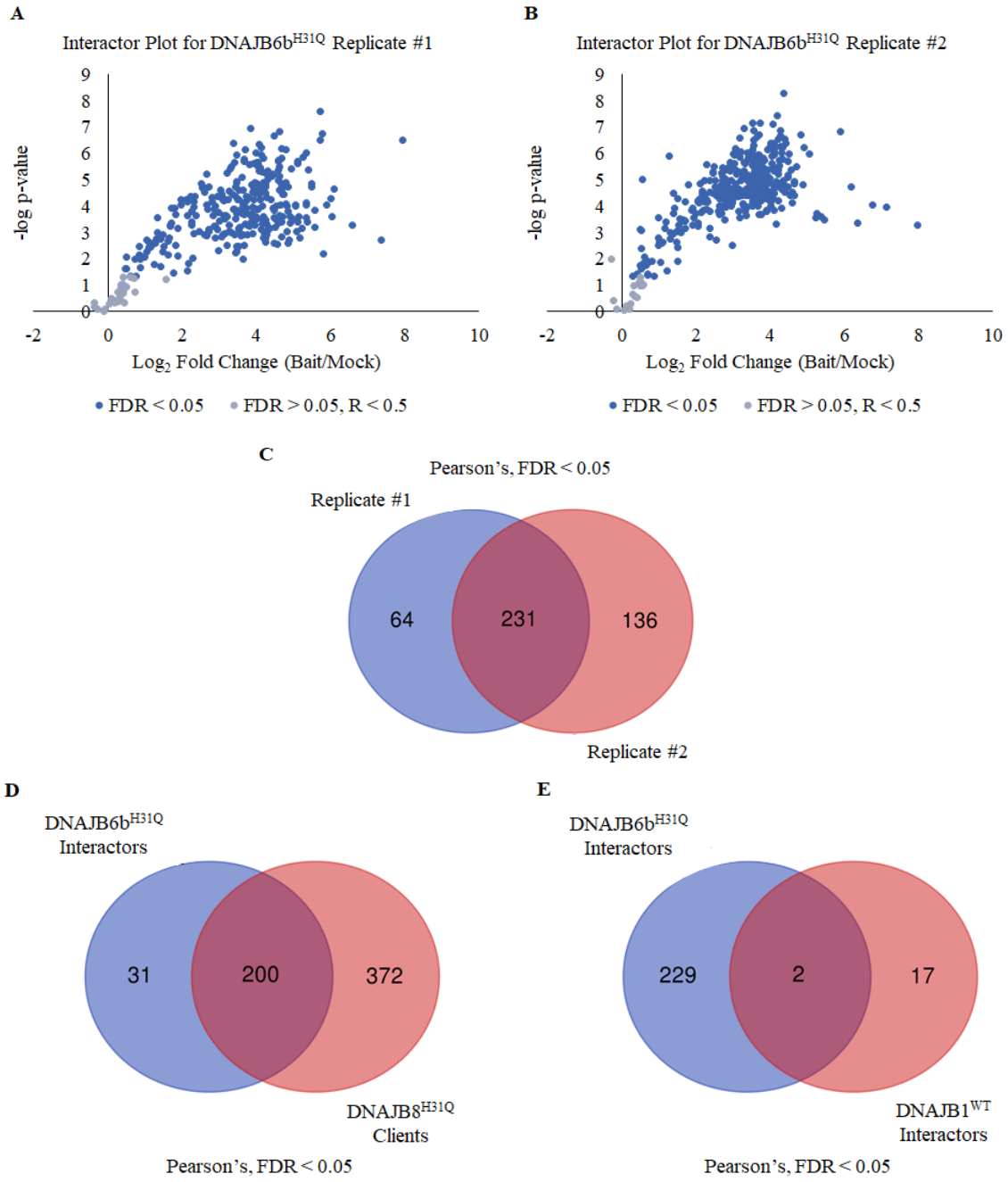

A
Interactor Plot for DNAJB6b^{H31Q} Replicate #1
-log p-value
Log$_2$ Fold Change (Bait/Mock)
FDR < 0.05
FDR > 0.05, R < 0.5
B
Interactor Plot for DNAJB6b^{H31Q} Replicate #2
-log p-value
Log$_2$ Fold Change (Bait/Mock)
FDR < 0.05
FDR > 0.05, R < 0.5
C
Pearson's, FDR < 0.05
Replicate #1
64
231
136
Replicate #2
D
DNAJB6b^{H31Q} Interactors
31
200
372
DNAJB8^{H31Q} Clients
Pearson's, FDR < 0.05
E
DNAJB6b^{H31Q} Interactors
229
2
17
DNAJB1^{WT} Interactors
Pearson's, FDR < 0.05

Figure 4.5

Figure 4.5. A-B) Representative volcano plots of Replicates #1 (**A**) and #2 (**B**) from bait vs. mock TMT-AP-MS experiments in cells overexpressing DNAJB6b^{H31Q} or GFP. Interactors (FDR < 0.05) are shown in blue. Non-interactors have FDR > 0.05 or have Pearson's R value < 0.5. **C**) Venn diagram showing that the two bait vs. mock replicates share 231 proteins with FDR < 0.05. **D**) Venn diagram comparing interactors (n=2) of DNAJB6b^{H31Q} and DNAJB8^{H31Q} (n=12). **E**) Venn diagram comparing interactors (n=2) of DNAJB6b^{H31Q} and DNAJB1^{WT} (n=3).

4.1.4 Future Directions of Project

Validation of ongoing interactome and crosslinking studies on DNAJB2a, DNAJB4, and DNAJB6b is necessary to elucidate client preferences of each JDP. It is promising that domain structure in the JDPs seem to indicate client specificity. For example, DNAJB6 and DNAJB8 have analogous structures and appear to bind the same clients. As studied by Jiang et al., client binding is dependent on the number of binding sites available to clients in the structure of the JDP, and is also dependent on interaction with Hsp70, as binding with Hsp70 prevents client binding[22]. It may be that DNAJB6 and DNAJB8 have related functions and client preference because of their client binding site availability and the formation of oligomeric structures that regulate client and Hsp70 binding[23]. Additionally, interactome studies could be performed with known clients, such as poly Q containing proteins for the DNAJB8-like JDPs, that way we can observe how client pools change with introduction of a destabilized substrate. In a study by Ryu et al., they identified clients of Hsp70 and Hsc70 using a ubiquitin-mediated proximity tagging strategy, and found that introducing a misfolded substrate (SOD1) to cells changed the binding partners of Hsp70/Hsc70.[24] It could be interesting to introduce poly Q containing proteins to cells expressing the JDPs and see how the landscape of interactors change. Furthermore, this project could elaborate on heat shock studies that were touched upon in **Chapter 3**. The clients affected by heat shock could be identified and would lead to identification of a misfolded proteome for each JDP.

This project should seek to uncover the extent of client selectivity among the different class B JDPs. If we can group the class B JDPs by structure, client specificity, and

function then researchers have a specific set of JDP members for disease or cancer targeted studies. For example, if we know that DNAJB1 can act as a disaggregase, or that DNAJB4 is implicated in tumor suppression, then this group (DNAJB1, DNAJB4, DNAJB5) can be used to monitor such activity. Conversely, if we know DNAJB6 can suppress aggregation of different neurodegenerative disorders, then that group (DNAJB2, DNAJB6, DNAJB8), can be directed toward study of those disorders. Determination of client and functional selectivity of class B JDPs could lead to similar success with other classes.

4.2 Concluding Remarks

In this work, we have demonstrated the capability of bait correlation (Pearson's R) TMT-AP-MS to identify high-confidence interactors of DNAJB8^{H31Q}. Using Pearson's R to generate p-values, we can "rescue" TMT-AP-MS experiments that may have varied bait levels. We use TMT-AP-MS to characterize the affect the J-domain has on the recovery of client proteins for DNAJB1 and DNAJB8, and the affect that chemical crosslinking has both bait and client recovery. We discovered that DNAJB1^{WT} and DNAJB8 (wild type or mutant) have different client pools, with DNAJB1^{WT} interacting with mostly chaperones and DNAJB8 interacting with the bulk proteome. We also found that DNAJB8^{H31Q} can reveal heat destabilized substrates not identified by its wild type version. We have established a platform for uncovering interactors of JDPs. This can contribute to what is known about JDPs and what types of clients they bind.

Furthermore, this study shows progress toward determining the extent to which JDPs control the fate of substrates and functional diversity of Hsp70 by targeting misfolded substrates to degradation, disaggregation, or other JDP-specific function.

4.3 References

(1) Kampinga, H. H.; Craig, E. A. *Nat. Rev. Mol. Cell Biol.* **2010**, *11* (8), 579–592.

(2) Ajit Tamadaddi, C.; Sahi, C. *Cell Stress Chaperones* **2016**, *21* (4), 563-570.

(3) Kampinga HH, Andreasson C, Barducci A, et al. *Cell Stress Chaperones* **2019**, *24* (1), 1-15.

(4) Gillis, J.; Schipper-Krom, S.; Juenemann, K.; et al. *J. Biol. Chem.* **2013**, *288* (24), 17225–17237.

(5) Seidel, K.; Siswanto, S.; Brunt, E. R.; den Dunnen, W.; Korf, H. W.; Rüb U. *Acta Neuropathol.* **2012**, *124* (1), 1-21.

(6) Davis, A. K.; Pratt, W. B.; Lieberman, A. P.; Osawa, Y. *Cell Mol Life Sci*. **2020**, *77* (6), 977-996.

(7) Kakkar, V.; Kuiper, E. F. E.; Pandey, A.; Braakman, I.; Kampinga, H. H. *Sci. Rep.* **2016**, *6* (September), 1–12.

(8) Chapple, J. P.; van der Spuy, J.; Poopalasundaram, S.; Cheetham M. E. *Biochem Soc Trans*. **2004**, *32* (Pt 4), 640-642.

(9) Westhoff, B.; Chapple, J. P.; van der Spuy, J.; Hohfeld, J.; Cheetham, M. E. *Curr Biol*. **2005**, *15*, 1058–64.

(10) Michels AA, et al. *J Biol Chem*. **1997**, 272, 33283–9.

(11) Chen, C. H.; Chang, W. H.; Su, K. Y.; et al. *Oncogene* **2016**, *35* (43), 5674-5685.

(12) Chang, T.P.; Yu, S.L.; Lin, S. Y.; et al. *Cancer Res.* **2010**, *70* (4), 1656-67.

(13) Acun, T.; Doberstein, N.; Habermann, J. K.; et al. *OMICS*. **2017**, *21* (5), 257-265.

(14) Liu, Y.; Zhou, J.; Zhang, C.; et al. *Int J Clin Exp Pathol*. **2014**, *7* (3), 969-77.

(15) Sterrenberg, J. N.; Blatch, G. L.; Edkins, A. L. *Cancer Lett*. **2011**, *312* (2), 129-42.

(16) Hageman, J.; Rujano, M. A.; van Waarde, M. A.; et al. *Mol. Cell* **2010**, *37* (3), 355–369.

(17) Aprile, F.A.; Källstig, E.; Limorenko, G.; Vendruscolo, M.; Ron, D.; Hansen, C. *Sci Rep*. **2017**, *7* (1), 9039.

(18) Söderberg, C. A. G.; Månsson, C.; Bernfur, K. et al. *Sci Rep*. **2018**, *8* (1), 5199.

(19) Deshayes, N.; Arkan, S.; Hansen, C. *Int J Mol Sci*. **2019**, *20* (18), 4495.

(20) Österlund, N.; Lundqvist, M.; Ilag, L. L.; Gräslund, A.; Emanuelsson, C. *J Biol Chem*. **2020**, *295* (24), 8135-8144.

(21) Rodríguez-González, C; Lin, S.; Arkan, S.; Hansen, C. *Sci Rep*. **2020**, *10* (1), 8130.

(22) Jiang, Y.; Rossi, P.; Kalodimos, C. G. *Science (80).* **2019**, *365* (6459), 1313–1319.

(23) Ryder, B. D.; Matlahov, I.; Bali, S.; Vaquer-Alicea, J.; van der Wel, P. C. A. *BioRxiv* **2020**.

(24) Ryu, S. W.; Stewart, R.; Pectol, D. C.; et al. *PLoS Biol*. **2020**, *18* (7), e3000606.

Appendix

Mathematica Code:

```
Replicate:=1;
AUCPearson:={};
AUCStudent:={};
NonspecificNoninteractMean:=0.3;
NonspecificNoninteractStandardDeviation:=0.6;
NonspecificNoninteractChannelStandardDeviation:=0.2;
BaitLevelsStandardDeviation:=0.3;
BaitLevelsChannelStandardDeviation:=0.2;
NonspecificBaitMean:=0.1;
NonspecificBaitStandardDeviation:=0.01;
NonspecificBaitChannelStandardDeviation:=0.2;
NonspecificPreyMean:=0.1;
NonspecificPreyStandardDeviation:=0.6;
NonspecificPreyChannelStandardDeviation:=0.2;
RatioMean:=0.3;
RatioStandardDeviation:=0.3;
RatioChannelStandardDeviation:=0.2;
nNI:=500;
nPrey:=50;
While[Replicate<101,
NonspecificNoninteractList:=Evaluate[RandomVariate[TruncatedDistribution[{0,\[Infini
ty]},NormalDistribution[NonspecificNoninteractMean,NonspecificNoninteractStandardD
eviation]],nNI]];
NonspecificNoninteractChannelStandardDeviationList:=Evaluate[RandomVariate[Trunc
atedDistribution[{0,\[Infinity]},NormalDistribution[1,NonspecificNoninteractChannelSta
ndardDeviation]],6*nNI]];
NonspecificPreyList:=Evaluate[RandomVariate[TruncatedDistribution[{0,\[Infinity]},No
rmalDistribution[NonspecificPreyMean,NonspecificPreyStandardDeviation]],nPrey]];
NonspecificPreyChannelStandardDeviationList:=Evaluate[RandomVariate[TruncatedDis
tribution[{0,\[Infinity]},NormalDistribution[1,NonspecificPreyChannelStandardDeviatio
n]],6*nPrey]];
PreyRatioList:=Evaluate[RandomVariate[TruncatedDistribution[{0,\[Infinity]},NormalD
istribution[RatioMean,RatioStandardDeviation]],nPrey]];
PreyRatioChannelStandardDeviationList:=Evaluate[RandomVariate[TruncatedDistributi
on[{0,\[Infinity]},NormalDistribution[1,RatioChannelStandardDeviation]],6*nPrey]];No
nspecificBaitList:=Evaluate[RandomVariate[TruncatedDistribution[{0,\[Infinity]},Norm
alDistribution[NonspecificBaitMean,NonspecificBaitStandardDeviation]],6]];
NonspecificBaitChannelStandardDeviationList:=Evaluate[RandomVariate[TruncatedDist
ribution[{0,\[Infinity]},NormalDistribution[1,NonspecificBaitChannelStandardDeviation
```

```
]],6]];BaitLevelsList:=Evaluate[RandomVariate[TruncatedDistribution[{0,\[Infinity]},No
rmalDistribution[1,BaitLevelsStandardDeviation]],3]];
BaitLevelsChannelStandardDeviationList:=Evaluate[RandomVariate[TruncatedDistributi
on[{0,\[Infinity]},NormalDistribution[1,BaitLevelsChannelStandardDeviation]],3]];NIT
MT:=Transpose[{Array[NonspecificNoninteractList[[#1]]*NonspecificNoninteractChan
nelStandardDeviationList[[#1]]&,nNI],Array[NonspecificNoninteractList[[#1]]*Nonspec
ificNoninteractChannelStandardDeviationList[[#1+nNI]]&,nNI],Array[NonspecificNoni
nteractList[[#1]]*NonspecificNoninteractChannelStandardDeviationList[[#1+2.*nNI]]&,
nNI],Array[NonspecificNoninteractList[[#1]]*NonspecificNoninteractChannelStandardD
eviationList[[#1+3.*nNI]]&,nNI],Array[NonspecificNoninteractList[[#1]]*NonspecificN
oninteractChannelStandardDeviationList[[#1+4.*nNI]]&,nNI],Array[NonspecificNoninte
ractList[[#1]]*NonspecificNoninteractChannelStandardDeviationList[[#1+5.*nNI]]&,nN
I]}];
BaitTMT:=Join[Array[NonspecificBaitList[[#1]]*NonspecificBaitChannelStandardDevia
tionList[[#]]+BaitLevelsList[[#]]*BaitLevelsChannelStandardDeviationList[[#]]&,3],Arr
ay[NonspecificBaitList[[#1+3.]]*NonspecificBaitChannelStandardDeviationList[[#+3.]]
&,3]];
PreyTMT:=Transpose[{Array[NonspecificPreyList[[#1]]*NonspecificPreyChannelStand
ardDeviationList[[#1]]+PreyRatioList[[#1]]*PreyRatioChannelStandardDeviationList[[#
1]]*(BaitLevelsList[[1]]+NonspecificBaitList[[1]])&,nPrey],Array[NonspecificPreyList[[
#1]]*NonspecificPreyChannelStandardDeviationList[[#1+nPrey]]+PreyRatioList[[#1]]*P
reyRatioChannelStandardDeviationList[[#1+nPrey]]*(BaitLevelsList[[2]]+NonspecificB
aitList[[2]])&,nPrey],Array[NonspecificPreyList[[#1]]*NonspecificPreyChannelStandard
DeviationList[[#1+2.*nPrey]]+PreyRatioList[[#1]]*PreyRatioChannelStandardDeviation
List[[#1+2.*nPrey]]*(BaitLevelsList[[3]]+NonspecificBaitList[[3]])&,nPrey],Array[Non
specificPreyList[[#1]]*NonspecificPreyChannelStandardDeviationList[[#1+3.*nPrey]]+P
reyRatioList[[#1]]*PreyRatioChannelStandardDeviationList[[#1+3.*nPrey]]*(Nonspecifi
cBaitList[[4]])&,nPrey],Array[NonspecificPreyList[[#1]]*NonspecificPreyChannelStand
ardDeviationList[[#1+4.*nPrey]]+PreyRatioList[[#1]]*PreyRatioChannelStandardDeviat
ionList[[#1+4.*nPrey]]*(NonspecificBaitList[[5]])&,nPrey],Array[NonspecificPreyList[[
#1]]*NonspecificPreyChannelStandardDeviationList[[#1+5.*nPrey]]+PreyRatioList[[#1]
]*PreyRatioChannelStandardDeviationList[[#1+5.*nPrey]]*(NonspecificBaitList[[6]])&,
nPrey]}];
PearsonPrey=Array[PearsonCorrelationTest[PreyTMT[[#]],BaitTMT,"PValue"]&,nPrey]
;
TTestArrayPrey=Array[TTest[{PreyTMT[[#,1;;3]],PreyTMT[[#,4;;6]]}]&,nPrey];
PearsonNI=Array[PearsonCorrelationTest[NITMT[[#]],BaitTMT,"PValue"]&,nNI];
TTestArrayNI=Array[TTest[{NITMT[[#,1;;3]],NITMT[[#,4;;6]]}]&,nNI];
PearsonFP[logalpha_]:=Count[PearsonNI,u_/;u<10.^logalpha];
PearsonTP[logalpha_]:=Count[PearsonPrey,u_/;u<10.^logalpha];
StudentFP[logalpha_]:=Count[TTestArrayNI,u_/;u<10.^logalpha];
StudentTP[logalpha_]:=Count[TTestArrayPrey,u_/;u<10.^logalpha];
beta:=Array[.1*#-10.&,100];
```

```
AppendTo[AUCStudent,Total[Array[(StudentTP[beta[[#]]]+StudentTP[beta[[#+1.]]])*.5
*(StudentFP[beta[[#+1.]]]-StudentFP[beta[[#]]])&,99]]/(nPrey*nNI)];
AppendTo[AUCPearson,Total[Array[(PearsonTP[beta[[#]]]+PearsonTP[beta[[#+1.]]])*.
5*(PearsonFP[beta[[#+1.]]]-PearsonFP[beta[[#]]])&,99]]/(nPrey*nNI)];
Replicate++]
```