The Effect of Quaternary Structure on Small Heat Shock Proteins: Chaperone Activity of GST-Induced Dimers of Human HspB1 and HspB5

By

Caley Butler

Thesis Advisor: Kathryn McMenimen

A Paper Presented to the

Faculty of Mount Holyoke College in

Partial Fulfillment of the Requirements for

The Degree of Bachelor of Arts with

Honor

Department of Biochemistry

South Hadley, MA 01075

May 2018

ACKNOWLEDGEMENTS

To Katie McMenimen. Thank you for being such a constant source of support throughout my time at Mount Holyoke, both inside the lab and out. I wouldn't have been able to accomplish any of this if I didn't know that you were always there to help, whether to offer advice, answer questions, or just remind me to breathe.

To Hannah Arbach and Elizabeth De Leon, for being my mentors and teaching me everything I know about working in a lab. Your passion for science reminded me why we keep going in the face of failure, and your knowledge and willingness to help was invaluable in overcoming the challenges of research.

To Alina Smithe, for being one of the kindest, smartest, hardest-working people I know. You have inspired me to go out of my comfort zone and accomplish more than I would have thought possible, and I can't imagine what my time at Mount Holyoke would have been like without you. I am so incredible lucky to have you as a friend.

To the rest of the McMenimen Group, for making this lab such a joy to be a part of. Your support is what made this project possible. To everyone that came before me, thank you for inspiring my research, and to everyone coming after, I can't wait to see what you accomplish.

To Alan van Giessen, for being there for me from the very first day I stepped foot on this campus. I came in wanting to do everything, and you were the person that told me I never had to choose – if I wanted to do biology and chemistry, if I wanted to study abroad and be pre-med, you believed I could do it all and helped plan a way to make it work. Thank you for being on my thesis committee.

To Chris Rivers, for teaching me a language and so much more than that. Your passion for teaching has exposed me to so many topics and ideas that I never would have encountered otherwise. Taking French has pushed me out of my comfort zone so many times, from trying to speak up in a class full of people more fluent than me to living in France for a semester. However, those experiences have also been the most rewarding, and I wouldn't have been able to do any of that without your support. Thank you for being on my thesis committee.

To all of my Mount Holyoke professors and everyone who supported me and inspired me in my time here, thank you. There are far too many people here to

thank individually, but you are what makes Mount Holyoke so special, and I cannot thank you enough for that.

Most importantly, to my family. Macrina, I wouldn't have gotten to where I am today without having you to turn to for advice. Thank you for being such an incredible role model, sister, and friend. Francis, thank you for always believing in me, and for inspiring me to try new things. Your adventures inspire me to try new things, even if it's just running a 5k and not jumping out of a plane. Dad, thank you for your emails whenever you find something interesting to share. It's nice to have a little reminder of the real world in my inbox inspiring me to always learn new things. Mom, thank you for sending me a card every single week of college. No matter how crazy things got, I always had something to look forward to every week. Thank you for your endless support, your hours-long phone calls, your pep talks at all hours of the night, and your daily good morning texts. I can't express how much all of that means to me. None of this would have been possible without all of you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Small heat shock proteins (sHsps) are a class of ATP-independent molecular chaperones that maintain protein homeostasis by preventing denatured proteins from aggregating and causing cell damage.[1] When cells are stressed (whether by heat, oxidation, infection, or a variety of other stressors), proteins denature and expose hydrophobic surfaces. In the aqueous cellular environment, this denaturation promotes protein aggregation linked to many diseases such as Alzheimer's or cataracts.[2] sHsps bind to and form complexes with these unfolded client proteins in a way that prevents their aggregation and can promote their refolding or degradation with the help of other cellular chaperones.[3]

sHsps exist *in vivo* as polydisperse oligomers composed of numerous subunits that can mix and exchange with other subunits, and this dynamic nature has made the determination of a specific mechanism of sHsp action difficult.[3] A given amount of sHsps can bind up to an equal amount of denatured proteins in terms of molecular weight; in order to have enough surface area available to bind an equal amount of denatured proteins, it has been suggested that dissociated sHsp subunits are responsible for observed chaperone activity, while large oligomers merely act as molecular storage for the smaller active chaperone subunits.[1, 3] X-ray crystallography and mass spectrometry studies have identified the dimer as the most-likely dissociated subunit of large oligomers.[3]

HspB1 and HspB5 are both widely produced in human tissues as a stress response to prevent client protein aggregation. In order to pin down the role of the dimer in chaperone activity, glutathione s-transferase (GST) was genetically linked as a fusion protein to the N-terminus regions of both proteins in order to constrain HspB1 and HspB5 forms to a dimer. Using purified GST fusion dimers and wild type HspB1 and HspB5, the *in vitro* chaperone activity of dimers as compared to polydisperse oligomers was measured through the use of a UV-Vis aggregation assay. The results suggest the fusion proteins function as active molecular chaperones, and furthermore, the two different fusion proteins demonstrate different chaperone activity in relation to multiple different substrate proteins.

PART I: INTRODUCTION

With 20,000 to 25,000 proteins in the human proteome, the maintenance of proteostasis (proteome homeostasis) is critical for normal function.[4] Regulation is required throughout protein synthesis, folding, localization, and degradation to preserve the protein balance of healthy cells, and mistakes in this system range from the unnoticeable to the catastrophic.[5]

I-1 Protein Synthesis

Of the 3 billion nucleotide base pairs that make up the human genome, approximately one percent code for proteins.[6,7] With the aid of another 10 percent of promoter and enhancer sequences, transcription factors, and the mediator complex, expressed protein-coding genes are transcribed into messenger RNA (mRNA) by the RNA polymerase II enzyme within the nucleus.[7,8] While the mRNA is being transcribed, other cellular machinery is processing the mRNA in preparation for export from the nucleus: noncoding introns are removed by a spliceosome and a 5'-cap and a 3' poly(A) tail are added to stabilize and protect mRNA from degradation (Figure 1).[9]

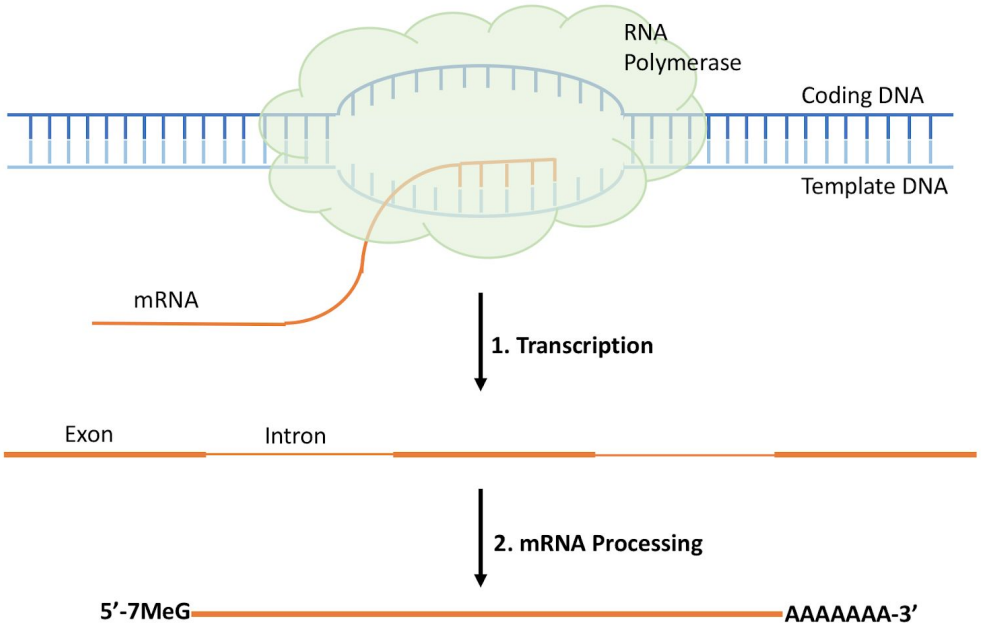


Figure 1: Overview of DNA transcription and RNA processing. With the help of elongation factors, RNA polymerase unwinds double-stranded DNA and transcribes mRNA from the template strand. The primary transcript is processed for nuclear export through splicing (the removal of introns, shown as thin RNA in the figure) and the addition of a 5' 7-methyl guanosine cap and a 3' poly(A) tail.

At its most basic, translation of this mRNA template into a protein consists of three steps: initiation, elongation, and termination. A preinitiation complex forms from a ribosomal subunit and initiation factors (at least 10 are required throughout the initiation process) which then binds a ternary complex of Met-tRNA, GTP, and an initiation factor to form an initiation complex.[10] Guided by protein factors, the ribosomal subunit of the initiation complex binds the mRNA close to the 5' terminus.[11] Ribosomal scanning subunits move along the bound mRNA until they find the AUG start sequence.[11] Each tRNA has an anticodon complementary to the mRNA codon corresponding to the attached amino acid, which allows the

translation of an mRNA's codon sequence into a peptide's amino acid sequence.[10,11] Met-tRNA binds to the start codon and the other ribosomal subunit is attached, completing the ribosomal complex.[10,11]

A completed ribosome contains three sites: the aminoacyl-tRNA site (A site), the peptidyl-tRNA site (P site) and the exit site (E site).[10] The Met-tRNA and mRNA start codon begin in the P site, and the aminoacyl-tRNA matching the next mRNA codon binds to the A site with the help of elongation factors.[10] The ribosome catalyzes a peptidyl transferase reaction, moving the peptide chain in the P site to be attached via peptide bond to the aminoacyl-tRNA in the A site.[10 11] The ribosome then moves along the mRNA, so the deacylated tRNA is shifted to the E site and released, and the newly formed peptidyl-tRNA is moved to the P site, leaving the A site open for the next aminoacyl-tRNA (Figure 2).[10,11] This process is repeated until a stop codon enters the A site, at which point a release factor causes the hydrolysis of the aminoacyl linkage in the P site, leading to the release of the completed polypeptide chain and the dissociation of the ribosomal complex.[10,11]

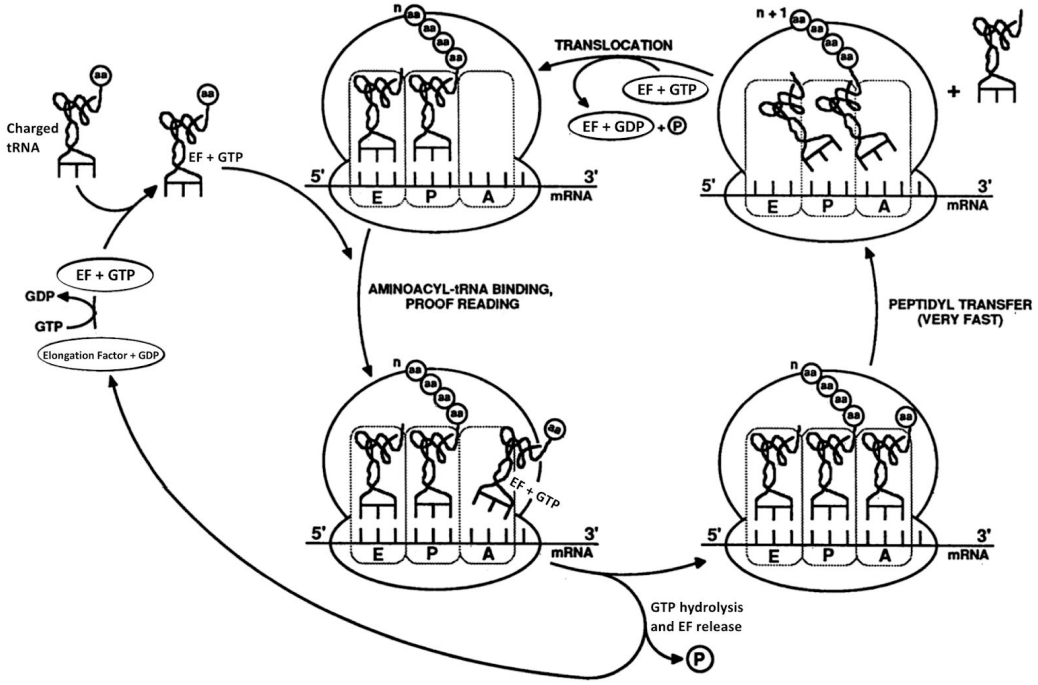


Figure 2: Overview of RNA translation elongation. With the help of elongation factors (EFs), transfer RNA (tRNA) carrying an amino acid comes into the empty A site of the ribosome; the specific tRNA bound is determined by codon-anticodon pairing to ensure the addition of the correct amino acid. Elongation factors that helped with tRNA binding dissociate following GTP hydrolysis. A peptide bond forms between the growing polypeptide chain in the P site and the new amino acid in the A site. The ribosome shifts along the template so the empty tRNA in the E site is release, the newly empty tRNA from the P site is moved to the E site, and the tRNA carrying the growing peptide is moved to the P site, leaving the A site open for the cycle to repeat. This translocation is made more energetically favorable through GTP hydrolysis by more elongation factors. Adapted from Merrick, 1992.[11]

I-2 Protein Folding

The amino acid sequence of the polypeptide chain determines the three-dimensional structure of a completed native-state protein.[4] This structure is what defines and enables the function of a protein, but the structural conformation needed for protein function must be balanced with the thermodynamic stability of

a folded protein.[4] The intersection of these two needs drives the folding process of a protein and determines its completed form.[4] The best representation of protein folding in current literature stems from energy landscapes, visual representations of the energy profiles of folding (and misfolding) intermediates (Figure 3).[12] The native state of a protein needs to be conformationally stable relative to similar but inactive forms, and so the path of a protein along an energy landscape must "funnel" the protein into the low energy valley of its functional conformation.[12]

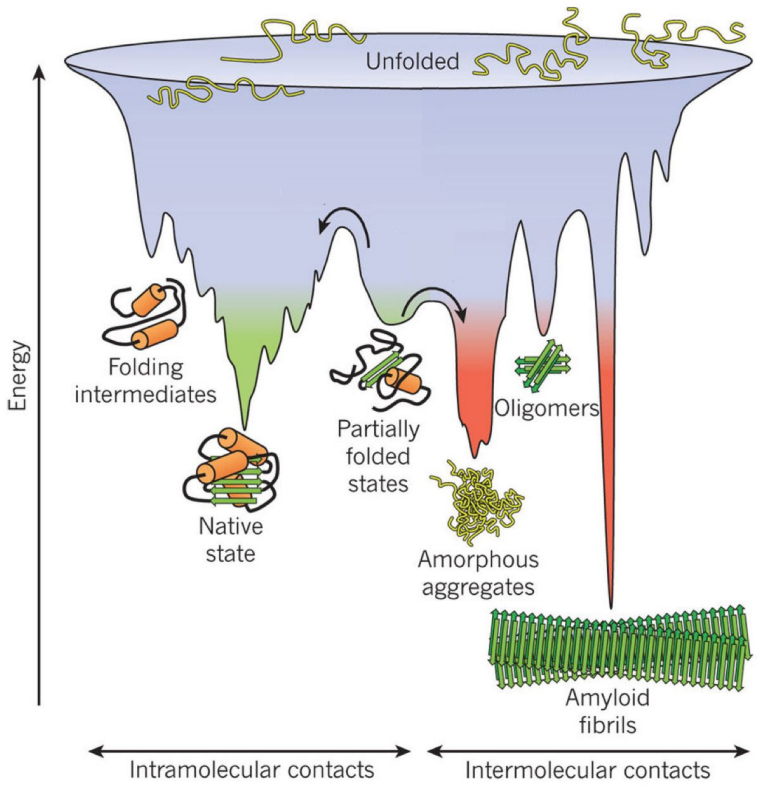


Figure 3: Energy landscape of protein folding. Hydrophobic interactions cause initial collapse of unfolded polypeptide into a molten globule folding nucleus. This partially folded intermediate explores different energetic states. Kinetically favorable interactions may lead to correct native state folding (green) or to misfolding and aggregation (red). Energy barriers to rearrangement may prevent misfolded states from recovering native conformation, regardless of which final state is energetically more favorable. Adapted from Hartl *et al.*, 2011.[13]

Hydrophobic interactions drive the initial chain collapse of the polypeptide into a globular structure, with non-polar amino acids buried in the interior.[13] From there, many weak, non-covalent bonds form and break as the protein adjusts structure.[13] The term "molten globule" was coined to describe this intermediate, encompassing all possible transiently formed bonds within the globular structure.[14] In general, the molten globule is a stable intermediate with secondary structure similar to the final native state, and adaptable tertiary structure.[14] Non-native interactions may stabilize the molten globule enough to form a folding intermediate "well" in the energy landscape of folding. However, native interactions are usually more stable, so increasing native contacts within the molten globule limit its potential conformations and drive the protein towards the completed native structure.[13] Certain 'key residues' interact to support the formation of native interactions, but it should be noted that once a compact globule has formed with non-native interactions, it can be difficult to overcome the energy barriers to structural changes and the protein may be stuck in a misfolded state.[12]

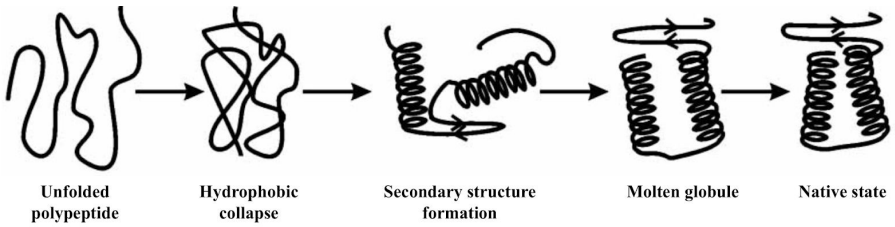


Figure 4: Correct folding pathway. The unfolded polypeptide undergoes hydrophobic collapse leading to the development of secondary structures. These rearrange with different tertiary interactions in the molten globule form until the native state is attained. Adapted from Hatahet and Ruddock, 2009.[15]

In addition to energetics, protein folding *in vivo* must take into consideration the timing and location of folding steps. One distinction is made between co-translational and post-translational folding (Figure 5).[13] Since the formation of peptide bonds is catalyzed within the ribosome, any structural elements that form as the polypeptide is being synthesised must be small enough to fit through exit channel of the ribosome.[13] This limits co-translational structure to small secondary and tertiary elements and prevents the long-range interactions involved in cooperative domain folding.[13] Since a small protein cannot reach its native state through co-translational folding, co-translational folding in practice is understood to be the parallel folding of large multidomain proteins.[12, 13] In this process, also referred to as sequential folding, single domain structural elements fold co-translationally as they exit the ribosome, and the independently folded elements are only brought together at the end of folding to establish the full intra- and inter-domain contacts of the native multidomain protein structure.[12, 13] The independent folding of structural elements and the avoidance of a giant molten globule state of the entire protein helps prevent non-native contacts that would easily lead to misfolding of such large proteins.[13] Interestingly, the slower elongation rate of eukaryotic ribosomes as compared to bacterial ribosomes and the presence of pause sites during translation gives the protein more time to fold co-translationally, and the support of this sequential folding mechanism is a likely influence in the evolution of many large, multidomain proteins in eukaryotes.[13]

Small, single domain proteins are unable to benefit from such a sequential folding process, so most folding occurs post-translationally (meaning after the ribosomal release of the completed polypeptide) for these proteins.[13]

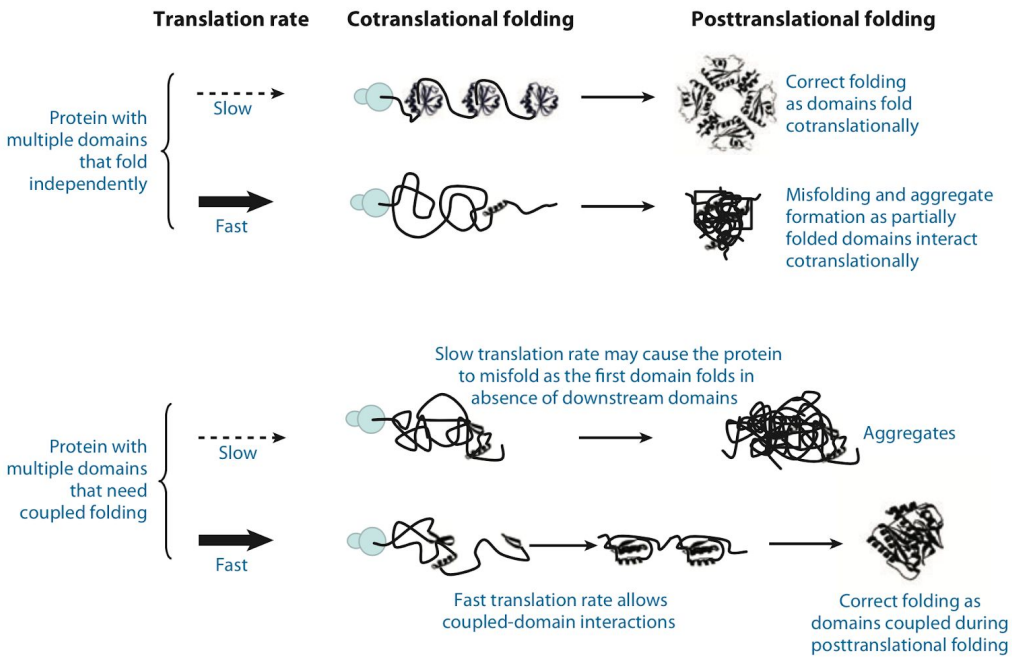


Figure 5: Translation rate influence on co-translational versus post-translational folding. Slow translation rates favor cotranslational folding as different domains have time to reach native structure before forming final tertiary or quaternary interactions. Fast translation rates favor post-translational folding as domains are too far apart to interact during translation for proteins that require coupled domain folding (interactions between partially-folded domains that are necessary for correct folding, even if those domains don't associate with each other in the final structure). Figure from Braakman and Bulleid, 2011.[16]

Translation *in vivo* is unique in that proteins may need to be transported to cellular locations other than the cytoplasm for protein folding, and folding in those compartments may differ from cytoplasmic folding.[4] Many proteins are targeted to the endoplasmic reticulum (ER) for translation, which has an environment

uniquely well-suited to assist in protein folding and post-translational folding and localization (Figure 6).[17] For instance, the ER has an oxidative environment that favors the formation of disulfide bonds, a high Ca^{2+} concentration that helps stabilize calcium-binding proteins and prepare proteins for secretion, and enzymes necessary for post-translational modifications such as glycosylation.[17] Proteins targeted for ER folding contain a signal sequence that pauses translation until the protein can be translocated to the ER, assisted by the ER translocon protein complex, at which point folding resumes.[18]

Even proteins designated for ER folding may continue to change structure in other environments, whether during Golgi complex packaging or even later.[16] While the specifics of many different folding environments are the topic of ongoing research, the point of interest is rather that the variability of protein folding pathways make the arrival of a protein at its correct, completed, native structure an incredibly complicated process.

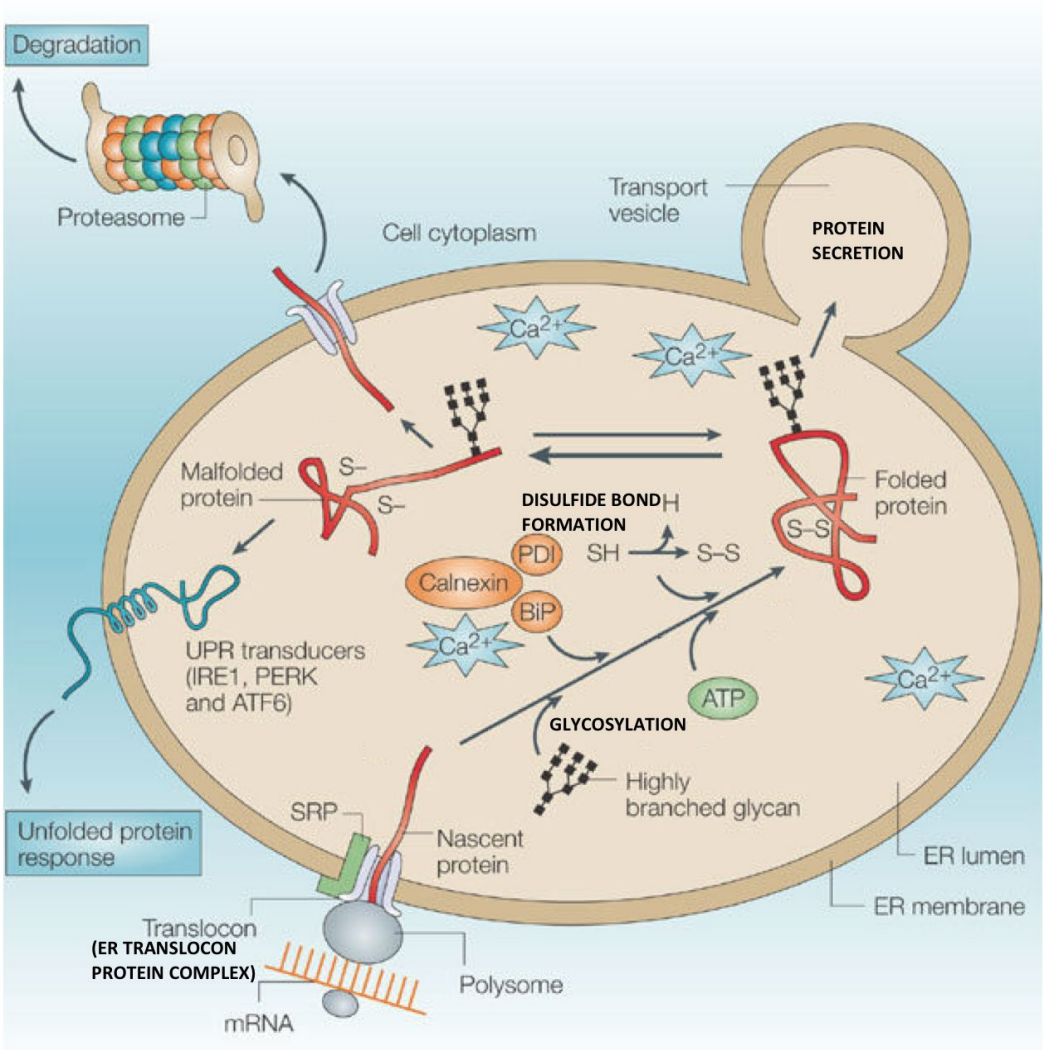


Figure 6: Endoplasmic reticulum involvement in protein folding. ER involvement begins with recognition of the nascent protein and transport into the ER by the *ER TRANSLOCON PROTEIN COMPLEX*. Enzymes assist with post-translational modifications such as *GLYCOSYLATION*, the oxidative environment favors *DISULFIDE BOND FORMATION* and Ca^{2+} assists with protein stability prior to *PROTEIN SECRETION*. The ER also has response systems for the degradation of misfolded proteins, which will be addressed later. Adapted from Ma and Hendershot, 2004.[19]

I-3 Mistakes in Protein Folding

What has been described above is the "ideal" procession of a protein from genetic code to final three-dimensional conformation. In a perfect world, there would be no errors in this process, and so it would remain that simple. However, there are a seemingly endless number of variables that disfavor this process. The same slow elongation that favors co-translational folding of multidomain proteins makes unfolded single domain proteins more susceptible to non-native interactions, since native interactions aren't completely available until after the protein is fully synthesized.[13] The molten globule form that speeds up protein folding may lock proteins into non-native states with if energy barriers to rearrangement are too high.[12] The crowded cellular environment may make correct folding more difficult.[16] Mutations could affect the critical residues necessary to form the folding nucleus that supports native folding.[12] Regardless of the mechanism by which protein misfolding occurs, it has serious implications in disturbing protein homeostasis and leading to disease.

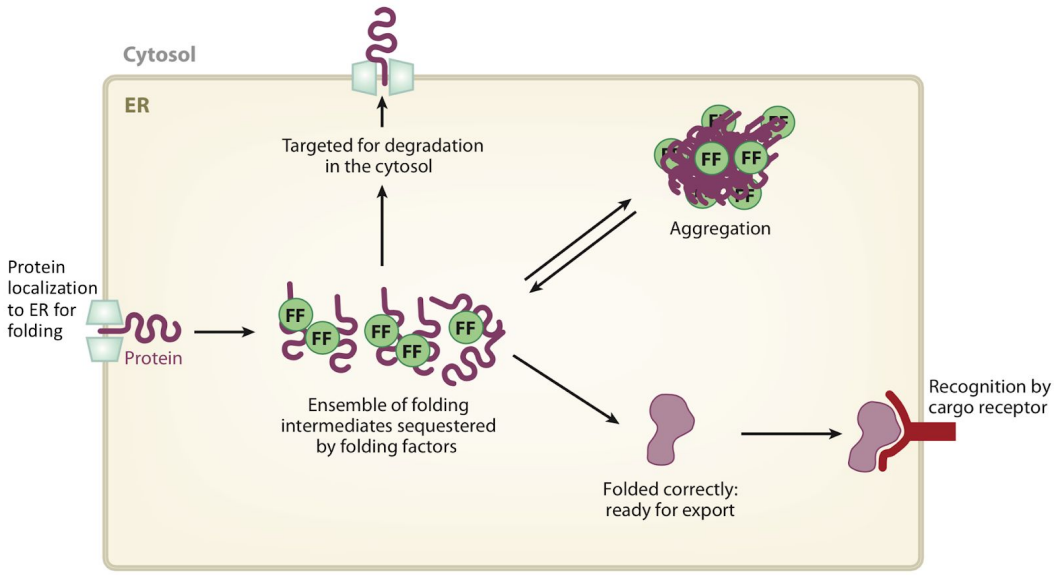


Figure 7: Outcomes of protein folding. Correctly folded proteins are delivered from the ER to their final location of function. Misfolded proteins clump together in aggregates and/or are degraded by cellular systems. Similar end results are available for proteins that fold in the cytoplasm as opposed to in the ER. FF represents folding factors that assist with correct folding in the ER. Cargo receptors help deliver proteins to their final destination. Adapted from Braakman and Bulleid, 2011.[16]

I-4 Protein Misfolding and Aggregation Diseases

When proteins misfold or unfold, they expose hydrophobic "sticky" surfaces.[20] A protein will self-associate if multiple unfolded or misfolded proteins come in contact and stick together, and any such association of non-native proteins is referred to as aggregation.[20] While some highly-structured aggregates may have a protective function by storing and sequestering misfolded proteins, most are toxic, and continued aggregation and tissue-deposition in the place of healthy proteins leads to cell death (Figure 8).[5, 20, 21]

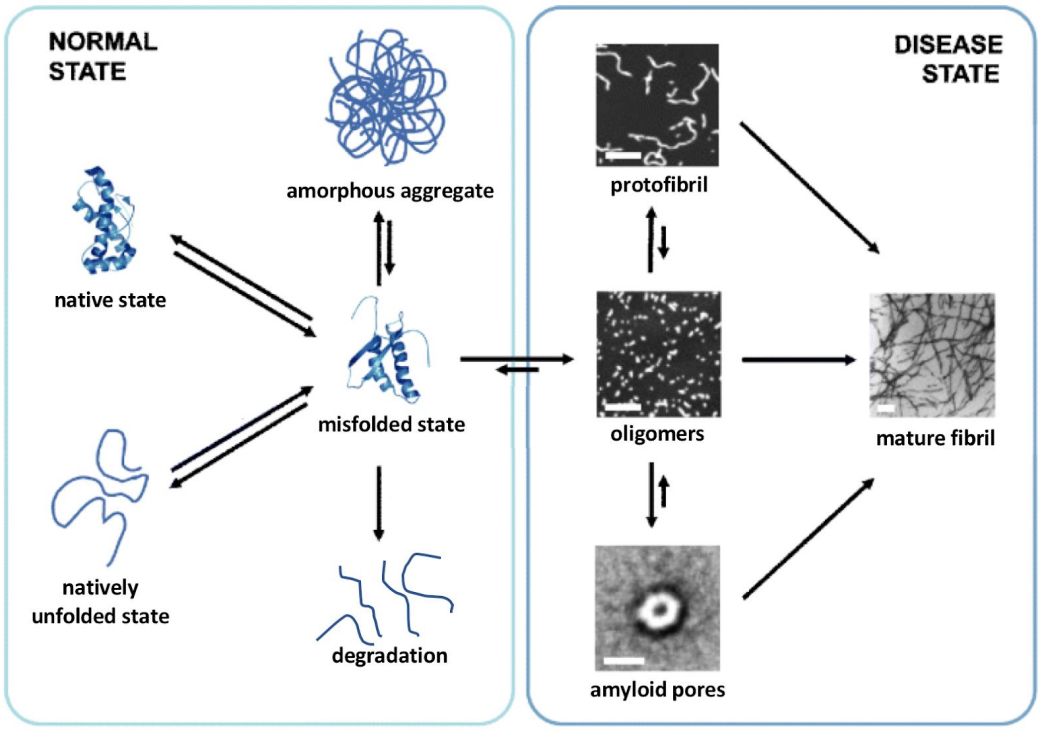


Figure 8: The role of protein misfolding in disease. The cell has defense mechanisms against misfolded proteins, but an excess of such misfolding can lead to the formation of toxic aggregates such as protofibrils or amyloid pores. Scale bars represent 100 nm, except for the amyloid pores, where it represents 10 nm. Adapted from Jahn and Radford, 2005.[12]

Diseases that result from non-native folding of proteins and their subsequent aggregates are referred to as conformational diseases, and include everything from a wide variety of neurodegenerative diseases (such as Alzheimer's, Parkinson's, Huntington's, and ALS) to prion diseases, diabetes, and cataracts (Table 1).[5, 21, 22, 23] Such diseases were first identified by the presence of aggregates, but it wasn't until later that analysis showed the aggregation-prone proteins were, in fact, causing the diseases.[5] The connection between such disparate diseases is not a result of any similarity in protein sequence, structure, or

function; in fact, most diseases are associated with a unique protein, such as α-synuclein with Parkinson's, ataxin-2 with amyotrophic lateral sclerosis (ALS), or beta-amyloid and Tau with Alzheimer's.[5, 20, 24] Instead, the connection between these diseases arises from the fact that all the involved proteins are susceptible to misfolding or unfolding that lead to cytotoxic aggregates.[24]

Table 1: Selection of protein aggregates and their associated diseases.
Adapted from Carrell and Lomas, 1997.[21]

Fibrils and Aggregates		**Amyloidoses**	
Protein	**Diseases**	**Protein**	**Diseases**
Prion proteins	Creutzfeldt-Jakob disease	Transthyretin	Familial amyloid neuropathy
Glutamine repeats	Huntington's disease	Immunoglobulin light chain	Systemic/Nodular AL amyloidosis
Hemoglobin	Sickle cell anemia	$\beta_2$microglobulin	Prostatic amyloid
β-amyloid protein	Alzheimer's disease	Cystatin C	Icelandic cerebral angiopathy

Another interesting connection between many conformational diseases is their age-related onset.[25] The proteostasis network, as refers to all the cellular components involved in maintaining protein homeostasis, experiences a decline in efficiency with aging (Figure 9).[20, 25] As this occurs, there is less cellular support for proteins to fold or stay folded in their native conformations.[20] Proteins that misfold and unfold further upset the proteostasis of the cell, and the overburdening of the proteostasis network components causes even more proteins

to adopt non-native conformations, so the effect is amplified in a toxic cycle.[24] Neurons are particularly susceptible to such proteostasis imbalances for several reasons. As long-lived cells, they both accumulate more damage from environmental stress than other cells, and they are unable to "dilute" any toxic proteins through cell division.[5] Their extended form means they are easily affected by any protein localization or trafficking problems as relates to proteostasis, and any issues with protein-membrane interactions are amplified by their small volume to surface area ratio.[5] In conjunction with the age-related decline of the proteostasis network, these neuronal traits explain why neurodegenerative diseases are particularly characterized by a late onset.

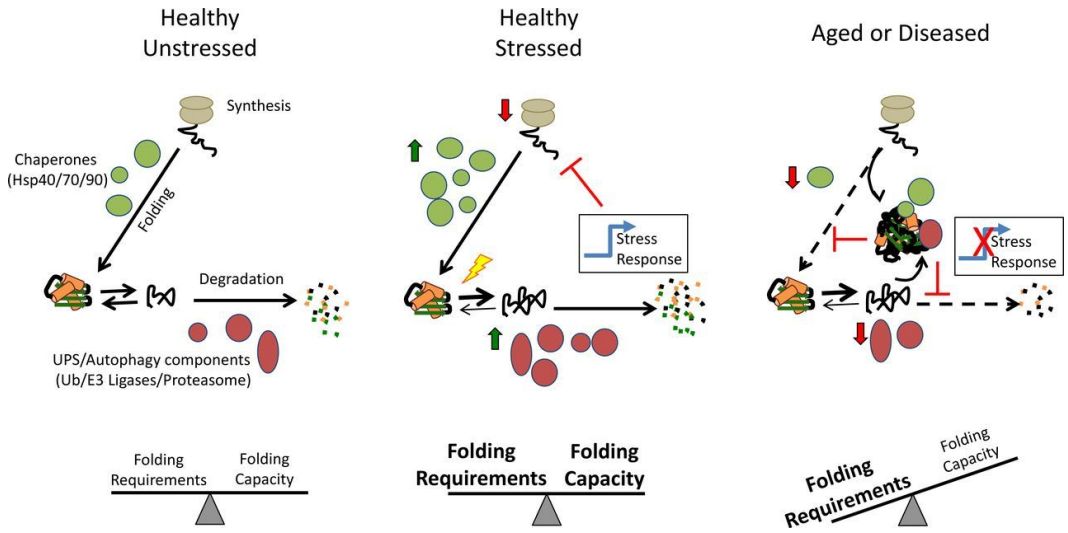


Figure 9: Proteostasis response decline with aging. Healthy systems upregulate cellular components that assist with correct protein folding and misfolded protein degradation when stressed. This response declines with aging or disease, leading to increased protein aggregation as cellular systems are unable to keep up with protein production and misfolding. Figure from Klaips *et al.*, 2017.[26]

While there are no current accounts as to the number of people affected by conformational diseases worldwide, a study of neurological disorders found that in 2005, there were 30 million people suffering from Alzheimer's (and other dementias) or Parkinson's.[27] By 2030, this number was projected to rise to over 50 million.[27] In the United States alone, more than 25 million people are currently living with cataracts, with this amount expected to double by 2050.[28] Other conformational diseases may be less prevalent (ALS and Huntington's each affect an estimated 30,000 Americans), but given the immense number and variety of conformational diseases, even these contribute significantly to the incidence of conformational disease in a global context.[29, 30]

I-5 Proteostasis

The concept of protein homeostasis, commonly called proteostasis, has been referenced several times now, particularly in relation to protein misfolding and disease. However, proteostasis encompasses the entire span of a protein's "life," all the way from RNA synthesis to protein degradation, regardless of whether or not the protein achieves a healthy native state (Figure 10).[24] The proteostasis network refers to the collection of cellular factors necessary to maintain homeostasis, and comprises an estimated 2000 components.[26] The cellular lifetime of a protein, as described above, can be split into three major events: protein synthesis, protein folding, and protein degradation.[26]

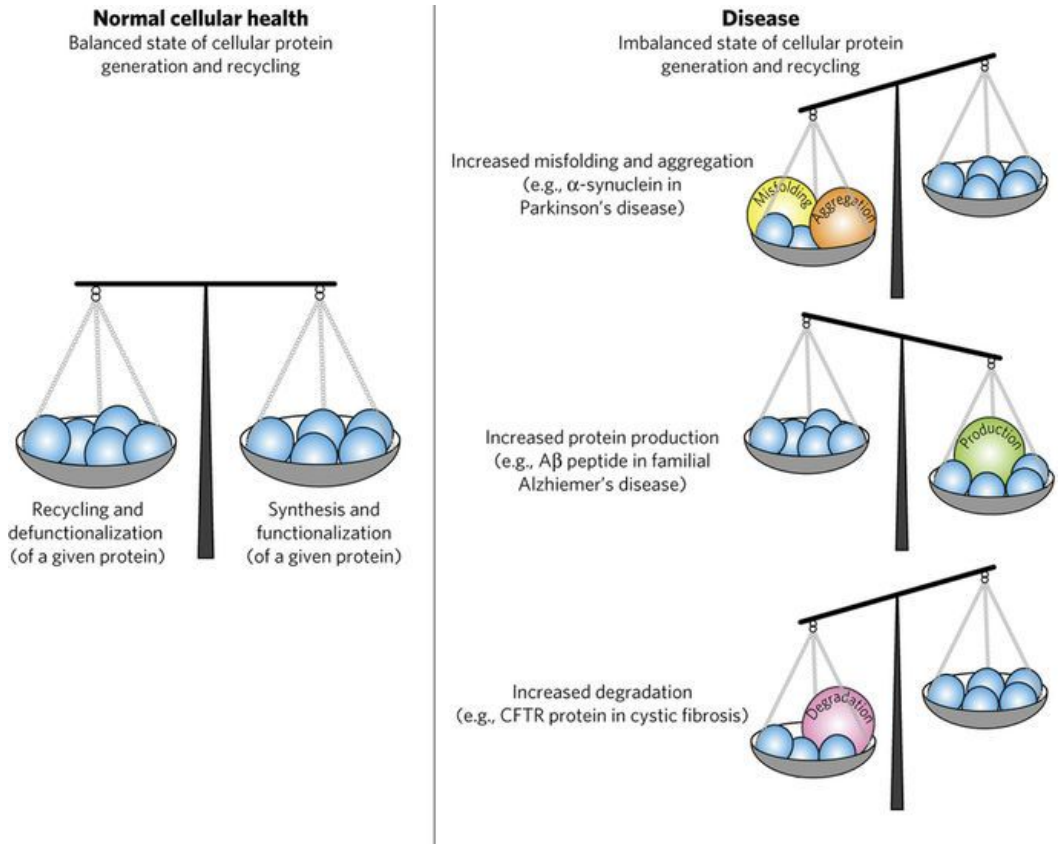


Figure 10: Proteostasis events and their role in disease. The proteostasis system controls the events in the cellular lifespan of a protein: protein synthesis, folding, and degradation. An upset in the homeostasis of any one of these events that is not corrected by the proteostasis network leads to disease. Figure from Narayan *et al.*, 2014.[5]

Regulation of protein synthesis is mostly accomplished prior to synthesis itself, either by transcription or translation factors or through mechanisms such the methylation of DNA, the phosphorylation of initiation factors, or the degradation of mRNA through RNA interference.[11, 31, 32] Protein folding is generally assisted by two factors: protein folding catalysts and molecular chaperones.[15] In accordance with the energy landscapes of protein folding, protein folding

catalysts increase the rate of productive folding while molecular chaperones decrease the rate of non-productive folding.[15] Many proteins exist only as metastable structures, even in their native form, so chaperones also assist with the maintenance of protein structure throughout their functional lifetime.[26] At the end of their cellular lifetime, proteins are primarily degraded by the ubiquitin-proteasome system or the autophagy system.[13] There is a delicate balance between all of the factors involved in proteostasis, and a change in proteome status (such as an overabundance of non-native protein aggregates) that is not corrected by the proteostasis network can lead to diseases such as those described above.[26]

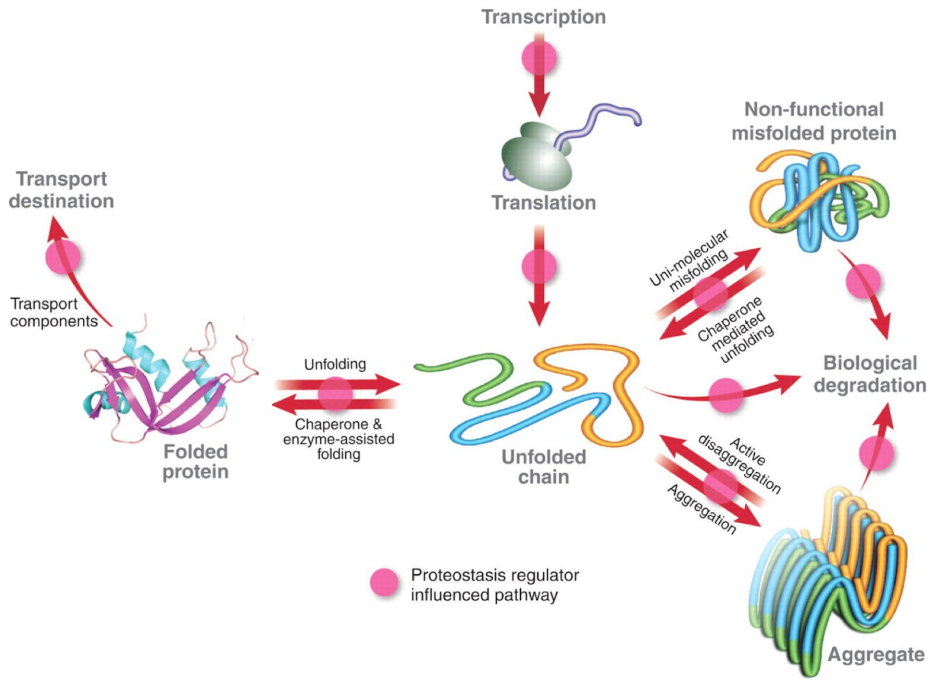


Figure 11: Responsibilities of the proteostasis network. Proteostasis factors are active at every stage of the protein "life" in the cell, from synthesis (transcription and translation) to folding and localization to degradation of proteins that are misfolded or at the end of their cellular life. Adapted from Balch *et al.*, 2008.[25]

I-6 Protein Degradation

There are many different pathways by which protein degradation can occur, but most in the proteostasis network fall in the categories of either the ubiquitin-proteasome system (UPS) or autophagy.[4] About 80-90% of proteins are degraded by the UPS, most of which are short-lived, non-native or damaged proteins.[33] Ubiquitin is a small, 76-amino-acid protein that is used to target proteins for degradation.[34] Ubiquitin is covalently linked to proteins once they have been singled out for degradation, usually by protein modifications such as phosphorylation.[34] Three classes of enzymes are involved in this process to activate ubiquitin, transfer it to a carrier protein, then link it to a lysine residue of the protein to which it is being ligated.[34] This process is repeated several times for a given protein, except after the initial ubiquitination, new ubiquitins are ligated to lysine residues on the preceding ubiquitin to form a polyubiquitin chain.[34] This polyubiquitin chain is recognized by receptors on the small subunit of the proteasome complex, the protease responsible for degradation.[35] Interestingly, the proteasome is ATP-dependent, and is the only such ATP-dependent protease in eukaryotes outside the mitochondria or chloroplasts.[35] The 26S proteasome consists of a small regulatory subunit and a larger proteolytic core particle, which is where the protein is transported after recognition.[35] There, the protein is hydrolyzed into its small peptide and ubiquitin fragments, which are further broken down into free ubiquitin (by ubiquitin-C-terminal hydrolases or

isopeptidases) and amino acids (by cytosolic peptidases) that can be reused by the cell.[34, 35]

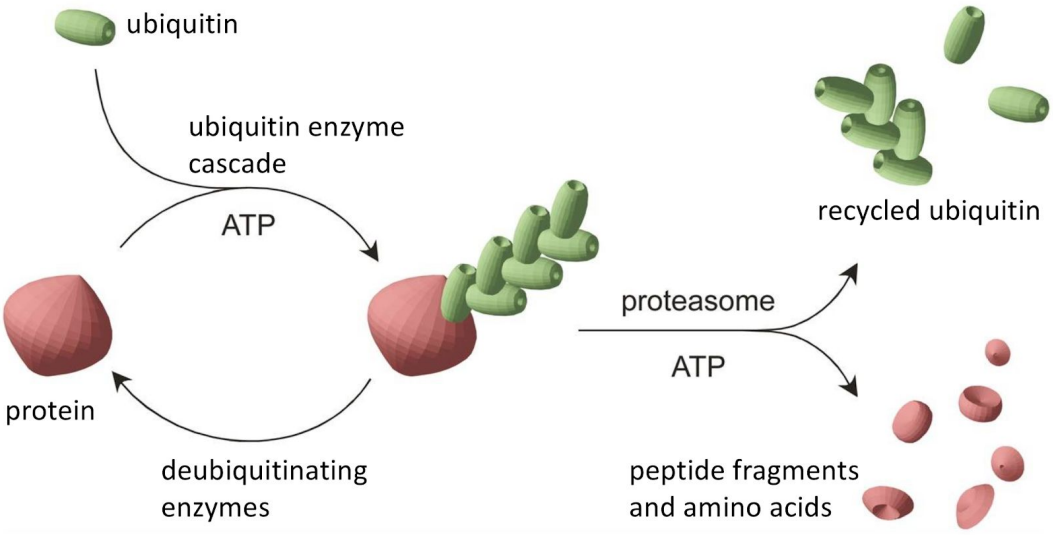


Figure 12: Functional mechanism of the ubiquitin-proteasome system. Proteins singled out for degradation are ubiquitinated by the three enzymes in the ubiquitination cascade. Proteins can be rescued from degradation through deubiquitination. The ATP-dependent proteasome degrades ubiquitinated substrates into their components, which are then recycled for reuse in the cell. Adapted from Finley, 2009.[35]

The UPS can act by itself, or as part of a broader process of degradation, such as with endoplasmic reticulum-associated protein degradation (ERAD).[36] As mentioned previously, many proteins are targeted to the ER for translation and folding.[17] If proteins become misfolded following their localization to the ER, ER quality control proteins escort the misfolded protein to transport channels out of the ER. The ER has its own transmembrane ubiquitin ligase enzymes that ubiquitinate the protein as it is exported back into the cytosol, where it can then be

degraded by the UPS.[16, 36] ERAD is part of a larger ER mechanism referred to as the unfolded protein response (UPR), which works exclusively to maintain ER homeostasis in response to stress, much as the proteostasis network maintains protein homeostasis throughout the entire cell (Figure 13).[37]

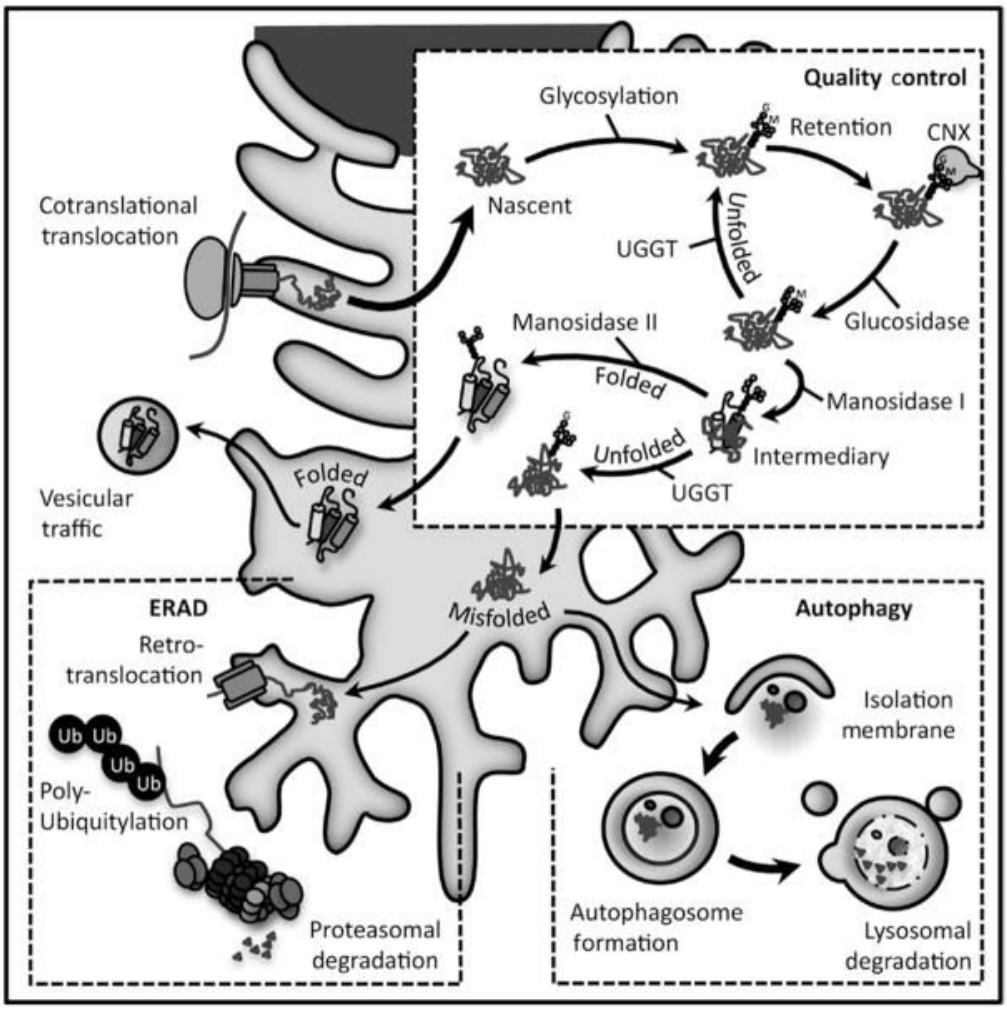


Figure 13: ERAD and UPR in the context of ER involvement in the proteostasis network. Following protein localization to the endoplasmic reticulum, ER enzymes and environmental factors that make up its quality control system assist in protein folding. Natively-folded proteins are secreted from the ER for delivery to their final location. Misfolded or unfolded proteins are marked for degradation and are retro-translocated back into the cytosol to be degraded by the UPS or autophagy. Figure from Bravo *et al.*, 2013.[37]

In contrast with the individual protein degradation of the UPR, autophagy is a bulk degradation pathway, in which multiple proteins (usually long-lived or aggregates) and other cellular components are all collected and degraded at once.[33, 38] There are three different types of autophagy, but they only differ in the mechanism by which substrates are transported to the lysosome, which is where they undergo proteolytic degradation by lysosomal acid proteases.[39] The products of degradation are transported back to the cytoplasm by lysosomal permeases and transporters to be recycled by the cell.[39] In macro-autophagy, a double membrane-vesicle known as the autophagosome forms around all of the cellular components to be degraded, and then fuses with the lysosomal membrane, releasing the components into the lysosome (Figure 14).[39] Micro-autophagy is similar in mechanism, but it is the lysosomal membrane itself that collects cytosolic components through invagination.[39] Such large-scale processes are proposed to assist with degradation of accumulated non-native proteins when the proteasomal degradation pathway is overwhelmed.[38] The last type is chaperone-mediated autophagy, which allows for the lysosomal degradation of specific targeted proteins.[39] In this process, the targeted protein complexes with a chaperone protein which is recognized by a lysosomal membrane receptor, and then the substrate protein is translocated into the lysosome for degradation.[39]

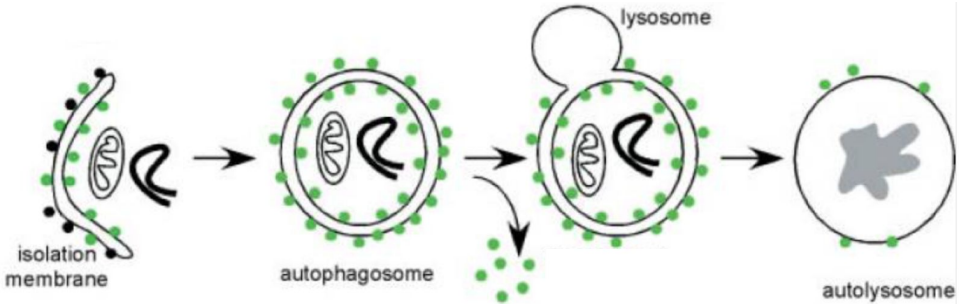


Figure 14: Steps of degradation by autophagy. Proteins to be degraded are collected in the double-membrane autophagosome. The autophagosome fuses with the lysosome, where proteins and other cellular elements are digested into their component parts. This process is assisted by proteins (green and black dots) that help guide the degradative pathway. Adapted from Glick *et al.*, 2010.[39]

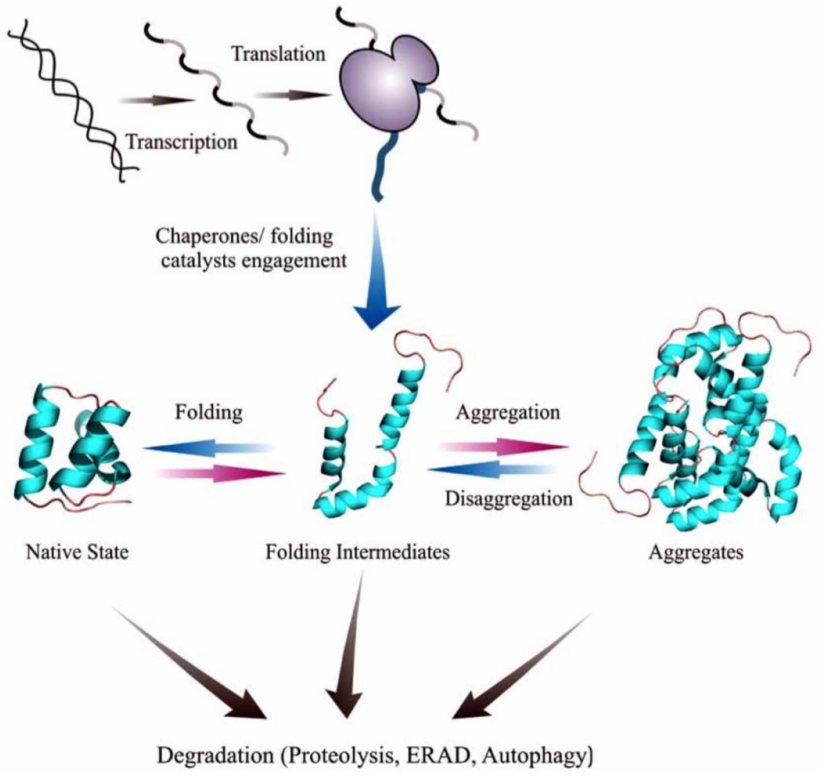


Figure 15: Summary of the cellular life cycle of a protein. After translation into a complete polypeptide, proteins attempt to fold into their native state. If proteins misfold (whether due to cellular error or environmental stress), proteins can form ordered or disordered aggregates. Such aggregates are toxic, so cellular systems attempt to prevent or dispose of aggregates through degradation by proteolysis, ERAD and the UPS, or autophagy. These are also the means by which healthy proteins are degraded at the end of their cellular life. Adapted from Hatahet and Ruddock, 2009.[15]

I-7 Molecular Chaperones

Protein structure, as described above, is determined by the amino acid sequence of the protein and the energetics of folding, both of which mostly allow proteins to achieve folding into their native state without assistance.[13] However, *in vivo* folding usually presents a much greater challenge than *in vitro* modelling of folding would suggest, as the crowded cellular environment and the long folding times of larger proteins favor misfolding and aggregation.[13] A class of proteins referred to as molecular chaperones have evolved to assist with protein folding processes within the cell to overcome these problems.[13] As a most basic definition, the roles of molecular chaperones are: to prevent aggregation, refold non-native folded states, and dissociate protein aggregates (Figure 16).[20] During the initial folding process, these responsibilities can be generally summed up by defining two mechanisms of chaperone function: the direct inhibition of non-native folding and the rescue of non-native intermediates.[15]

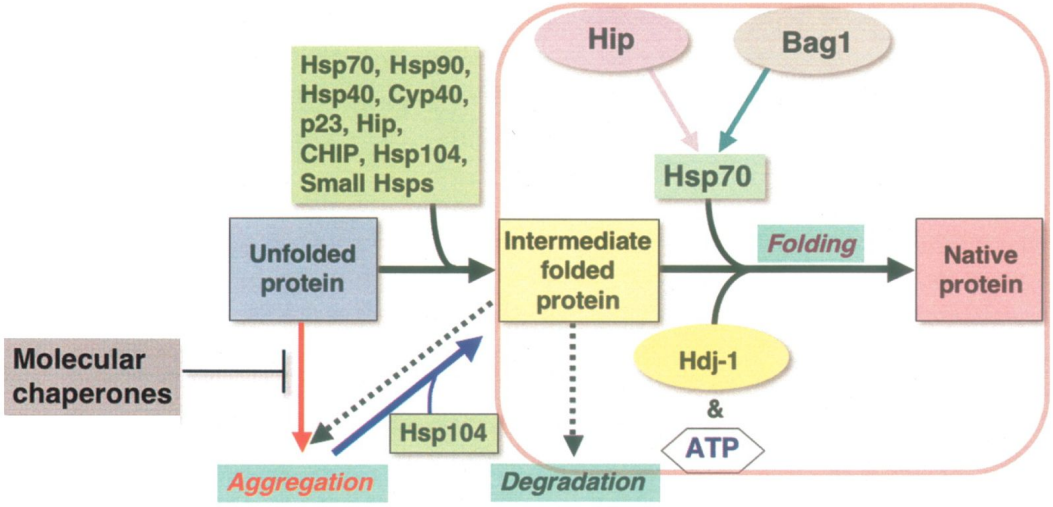


Figure 16: The role of molecular chaperones in protein folding. The Hsp chaperones (listed in green and tan) help guide partially folded proteins to their native conformation and prevent misfolding and toxic aggregation. Chaperone proteins work in conjunction with co-chaperones such as Hip and Bag1. Adapted from Morimoto, 2008.[24]

Chaperone proteins were first discovered as a cellular response to elevated temperatures (termed the heat shock response), and as such, are named heat shock proteins (Hsps) classified by molecular weight: Hsp100s, Hsp90s, Hsp70s, Hsp60s, Hsp40s, and small heat shock proteins (sHsps).[24, 40] Many small proteins can fold immediately following synthesis, with just the aid of structural elements such as the translating ribosome and the Sec61 translocon complex (RTC), a protein-conducting channel that favors folding as proteins exit the ribosome.[18] For proteins that need additional assistance, the process begins with ribosome-binding chaperones such as the nascent-chain-associated complex, trigger factor (TF), and specialized Hsp70s that hold the nascent protein in a folding-competent state.[4, 41]

Downstream of the ribosome, new polypeptides are assisted with folding by the Hsp70 chaperone system, which consists of an Hsp70, an Hsp40 cochaperone, and various nucleotide exchange factors.[4, 41] The Hsp70 system functions to promote folding primarily through a mechanism known as kinetic partitioning, in which the chaperone binds to the hydrophobic, "sticky" regions of unfolded or partially folded proteins to prevent aggregation, followed by ATP-triggered release (Figure 17).[13] The substrate protein folds a little when it is released but is bound again before it can aggregate, and this cycle of binding and release is repeated until the substrate achieves a native conformation.[13] Hsp70 is the chaperone binding the substrate in these cycles, and Hsp40 regulates the process by accelerating ATP hydrolysis and recruiting Hsp70 to substrates.[13] Hsp70 also assists with protein disaggregation, protein regulation, and autophagy.[20] It is interesting to note that while many chaperones primarily function as "holdases," merely holding proteins in intermediate states that favor native folding or holding misfolded proteins to prevent them from joining aggregates, Hsp70 is the primary "foldase," a chaperone involved in active protein folding.[24]

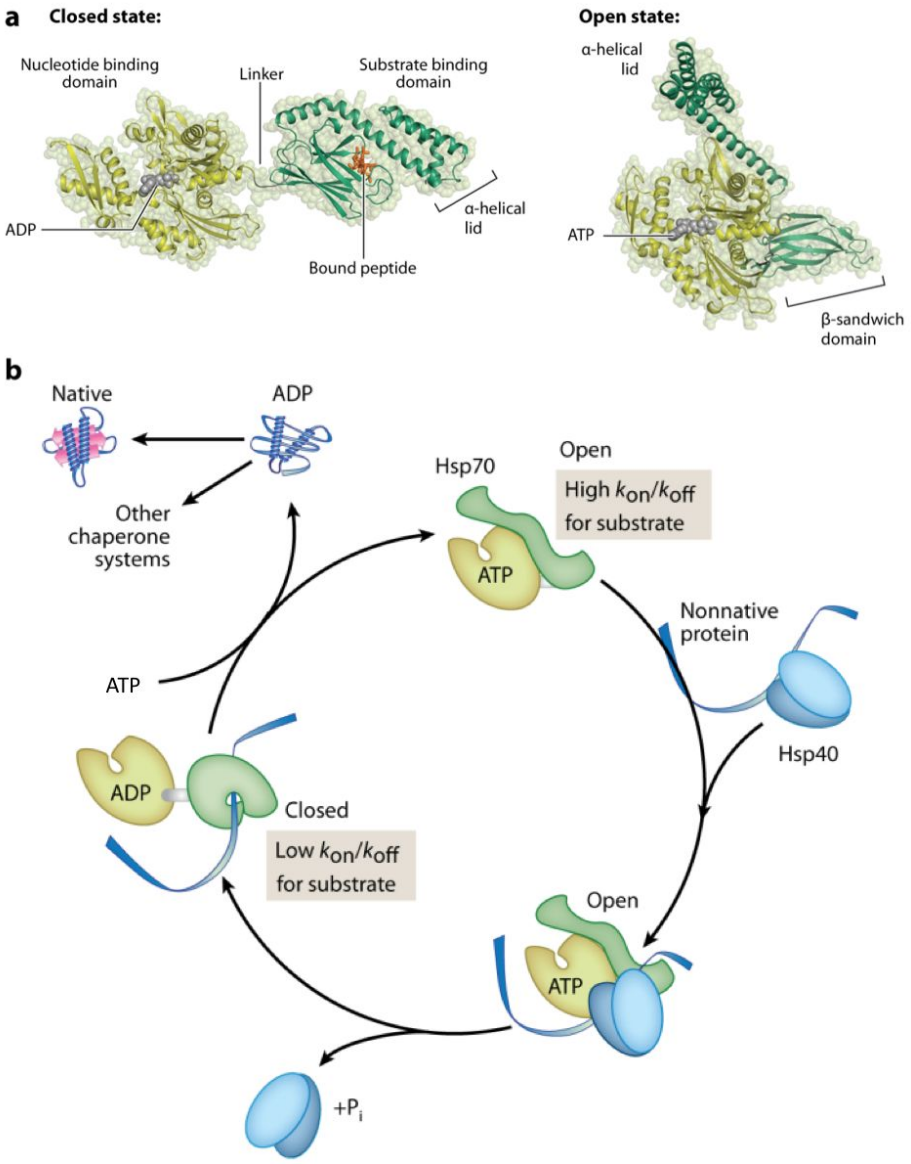


Figure 17: The Hsp70 chaperone system. a. Structure of Hsp70 conformations. The peptide being folded binds to the β-sandwich domain in the open conformation. The nucleotide binding domain binds ATP, which is hydrolyzed to change the conformation of Hsp70. The closed conformation protects the peptide from misfolding and aggregation. **b.** Hsp70 chaperone cycle. Following peptide binding to the open conformation, ATP hydrolysis changes Hsp70 to close the α-helical lid over the substrate, protecting it from aggregation. This process is assisted by Hsp40 (not shown). ATP binding to the nucleotide binding domain, replacing ADP, stabilizes the open conformation and allows release of substrate and binding of new peptide. Adapted from Kim *et al.* 2013.[4]

The Hsp70 system results in the proper folding of about 10-20% of proteins, but more difficult cases continue to the chaperonin system, which folds about 10% of proteins.[41] Chaperonins are large, double-ringed protein complexes that form a barrel-like structure.[41] The interior of the barrel forms an Anfinsen cage, an isolated compartment that protects the folding protein from cytosolic components that promote aggregation, and also may promote folding through steric confinement (Figure 18).[4, 13, 41, 42] There are two groups of chaperonins, but they have the same effective function; Hsp60 (the folding cage) in conjunction with an Hsp10 cochaperone (the lid) falls within Group I, and Group II includes other protein chaperonins.[4]

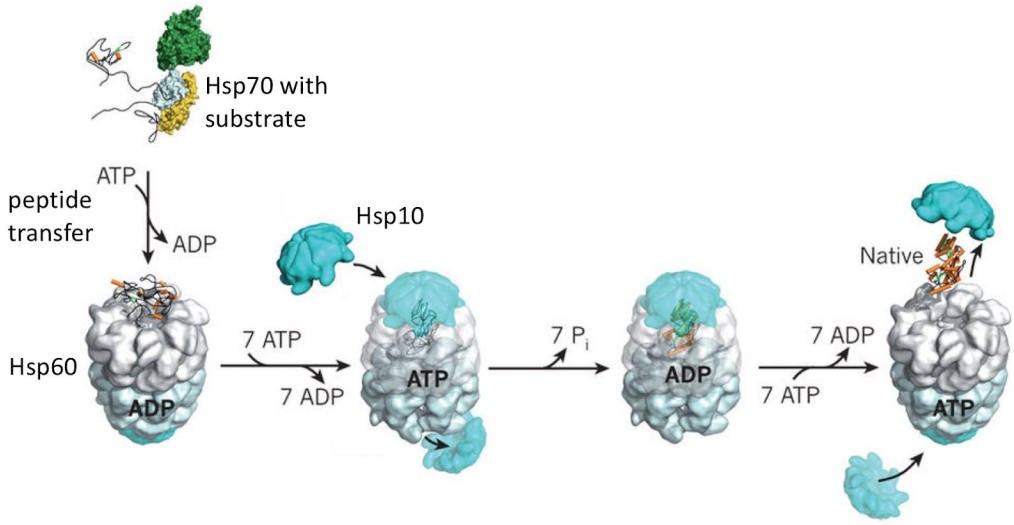


Figure 18: Sample chaperonin folding mechanism. Hsp70 assists with unfolded protein transfer to the folding cage (formed by Hsp60). ATP hydrolysis accompanies chaperone reorganization and lid (Hsp10) binding. The completed chaperonin complex forms an Anfinsen cage favoring native folding. Further ATP binding and hydrolysis allows for the release of the cap and the natively folded protein that was inside the cage. In this example, the chaperonin exists as a dimer, with one subunit in the open conformation and one subunit closed with the lid. Adapted from Hartl *et al.*, 2011.[13]

Hsp90 is the last chaperone directly involved in protein folding and conformational regulation, though it serves a larger function as a proteostasis regulator by controlling signalling pathways involved in such functions as cell-cycle progression, vesicle-mediated transport, or targeted protein degradation.[13] This regulatory function is due to the involvement of Hsp90 in the folding of regulatory proteins, which make up only 2-5% of protein products.[43] The mechanism of Hsp90 function is not yet well-understood, but it functions as a dimer in an ATP-dependent system involving various other regulatory cofactors (Figure 19).[13]

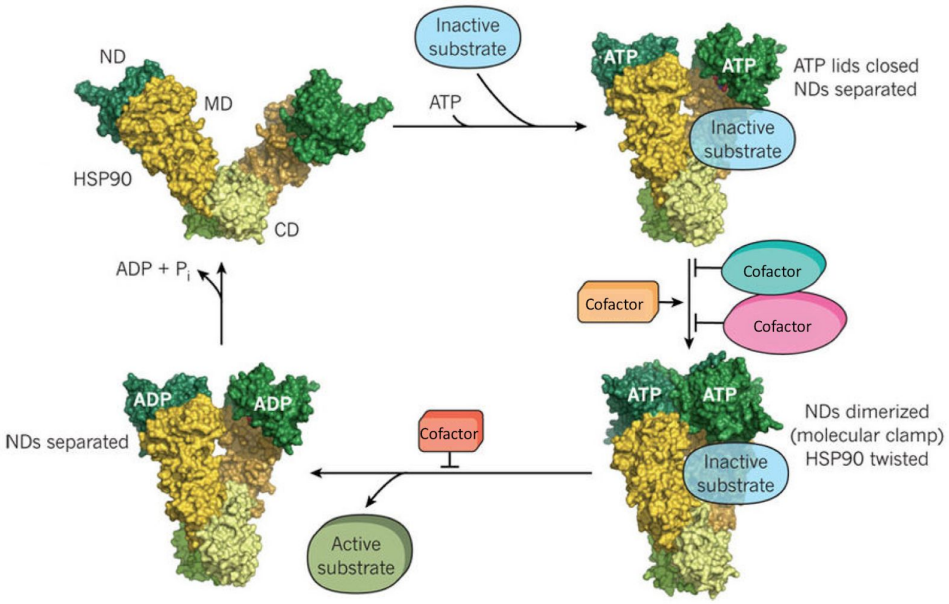


Figure 19: Hsp90-assisted protein folding. The substrate binds the open conformation of Hsp90. ATP binding changes the shape of Hsp90 to the closed conformation, and further conformational changes accompany substrate folding into active form. ATP hydrolysis causes dissociation of Hsp90 dimer and native substrate release. ADP is released and Hsp90 is available for new substrate binding. Various cofactors help regulate this cycle. ND = N-terminal ATPase domain, MD = middle domain, CD = C-terminal domain. Adapted from Hartl *et al.*, 2011.[13]

In addition to assisting with their role in co- or post-translational folding, chaperones are also expressed as part of the heat shock response, a cellular response to unfolded proteins resulting from stressors such as increased temperatures, oxidative stress, or environmental toxins (Figures 20 and 21).[44] Heat shock transcription factors (HSFs) up-regulate proteins involved in the heat shock response in response to such stressors.[15] In stress conditions, Hsp70 can refold aggregated proteins, just as it would assist with *de novo* folding.[44] Hsp100 works in conjunction with Hsp70 and other proteins in a disaggregation system that separates non-native proteins from aggregates and refolds them to their native conformation.[44] Hsp90 works in conjunction with Hsp70 to promote degradation of non-native or aggregated proteins.[20]

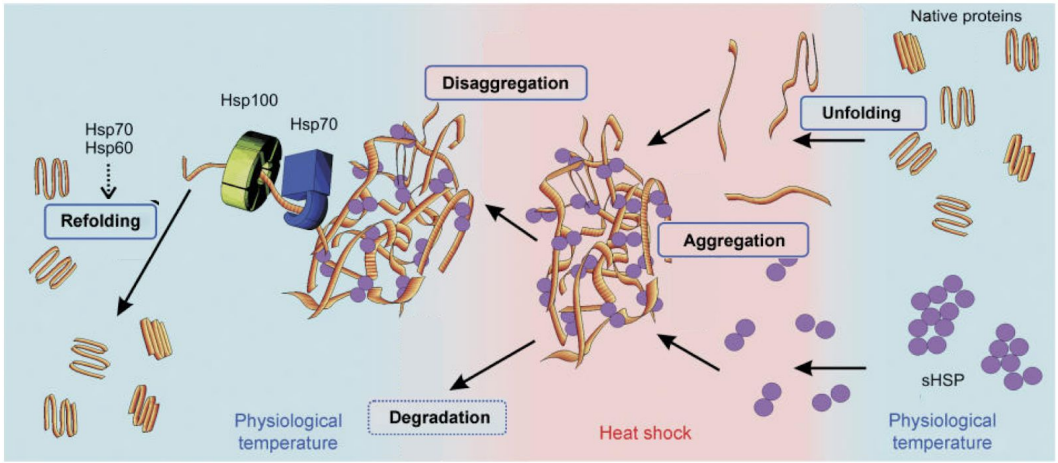


Figure 20: Heat shock response. Heat-denatured proteins aggregate at high temperatures. Heat shock proteins are expressed to prevent aggregation, degrade misfolded or aggregated proteins, or rescue and refold proteins into their native state once cellular conditions have stabilized. Adapted from Liberek *et al.*, 2008.[41]

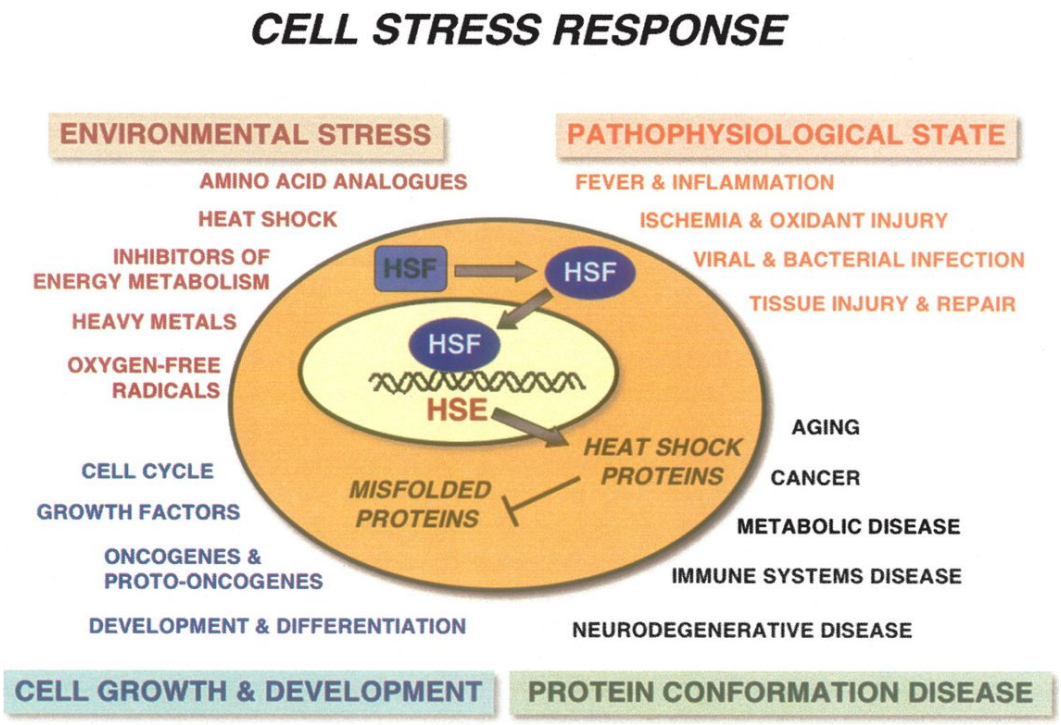


Figure 21: Cellular response to stress. Heat shock genes are expressed in response to several different kinds of stress: environmental stress, disease, and even normal growth development. Heat shock factors (HSF) are transcription factors upregulated in response to stress that activate the heat shock response. This response protects cells against toxic aggregates formed by misfolded proteins. Figure from Morimoto, 2008.[24]

I-8 Small Heat Shock Proteins

Small heat shock proteins (sHsps) form the last class of heat shock chaperones.[44] An extremely widespread and poorly conserved family of proteins, sHsps function as holdases to prevent non-native proteins from aggregating until they can be refolded or degraded by other chaperones in the heat shock system.[44] Unlike many other chaperones, sHsps are ATP-independent in their function as holdases.[44] In general, sHsps function at stoichiometric ratios with regard to their

substrates since all non-native proteins must be isolated to prevent aggregation, a process highly dependent on the concentration of non-native proteins available for aggregation.[44] In accordance with this concept, sHsps have been shown to bind up to an equal molecular weight of unfolding protein.[1] While sHsps prevent substrate proteins from aggregation, they can perform this function by either remaining soluble in cytoplasm or by sequestering themselves in aggregates.[44] The ability of sHsps to join aggregates without allowing their bound substrate to aggregate is particularly interesting, and suggests that sHsps can change the structure of aggregates themselves.[44] Interestingly, sHsps can only rescue partially unfolded substrates; despite their insertion in aggregates, sHsps are unable to protect completely unfolded or aggregated proteins.[45]

Overproduction of sHsps has been shown to increase thermotolerance of a variety of organisms, as is to be expected by their role in heat shock, but deletion of sHsp genes generally doesn't result in a heat-sensitive phenotype, suggesting that other chaperones and the heat shock response can recover enough function to prevent cell death.[41] One proposed mechanism that would explain this finding is that sHsps change the properties of polypeptides to allow Hsp100 and Hsp70 to disaggregate and refold the proteins more efficiently; sHsp deletion mutants would have the same heat shock protection response, it would merely be less efficient.[41]

Despite the lack of conserved tertiary structure, sHsps share the unusual characteristic that they assemble primarily as oligomers, but have no set size.[45] Oligomers generally range from 12 to 32 subunits, but larger and smaller oligomers have both been observed.[45] sHsp monomers are generally 12-42 kDa in size, and oligomers usually range from 200-800 kDa.[46] sHsps are described as polydisperse, dynamic oligomers, because they are constantly exchanging subunits and changing oligomeric size, even at a stable equilibrium (Figure 22).[45] sHsps are activated by an increase in temperature, at which point the oligomer is destabilized to form dimers that associate with substrate proteins in large, stable complexes.[41, 47] Interestingly, when neighboring subunits are crosslinked to prevent oligomer dissociation, chaperone activity is unaffected, indicating that large oligomeric forms of sHsps are also capable of chaperone activity.[47]

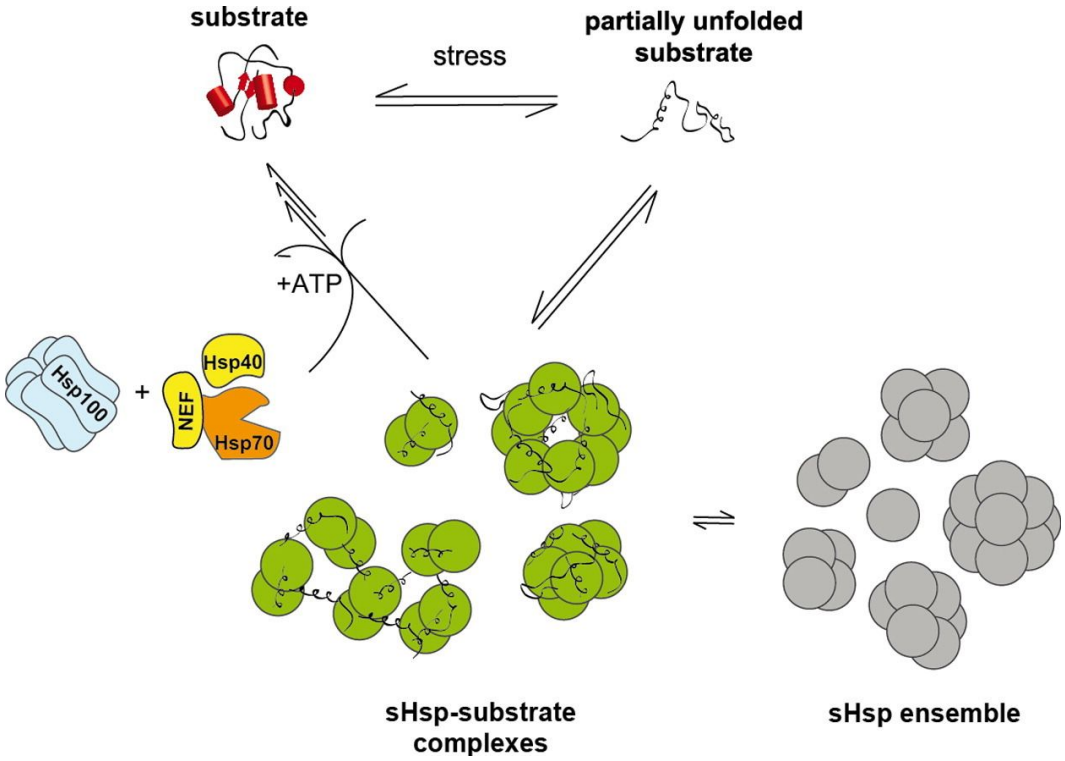


Figure 22: Chaperone activity of sHsp polydisperse oligomers. sHsp-substrate complexes prevent unfolded or misfolded proteins from aggregating. sHsps normally exist in a wide variety of oligomeric forms (grey) but when activated by stress (green), generally form dimers. However, other oligomers may also have chaperone activity and the ability to form sHsp-substrate complexes. Hsp100 and the Hsp70 system with cochaperone Hsp40 and cofactor NEF help refold substrate into native form. Figure from Haslbeck and Vierling, 2015.[45]

Highly variable as they are, sHsps are found in all three domains of life.[45] Prokaryotes usually have only one or two sHsps, but eukaryotes may have more than 20 different sHsps.[45] The primary structure of sHsps consists of a highly variable N-terminal domain (NTD), a relatively conserved α-crystallin domain (ACD), and a highly variable, short C-terminal domain (CTD).[48] The ACD is usually around 100 amino acids in length, and forms a β-sandwich tertiary

structure.[45] The ACD is responsible for sHsp dimerization, as the β6 and/or β7 strands of two β-sandwiches can swap places to connect the sHsps (Figure 23).[46]

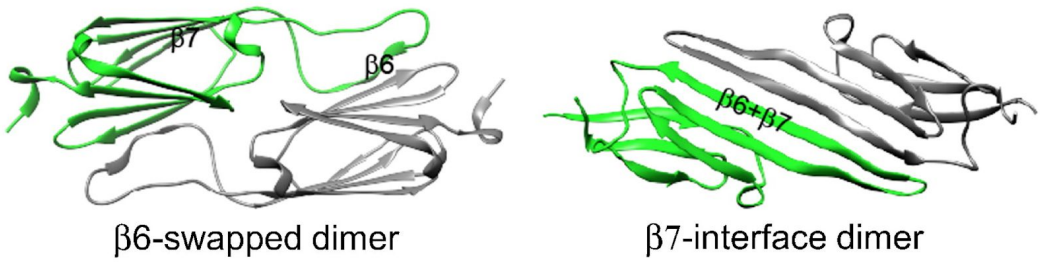


Figure 23: sHsp dimerization. Dimers generally form when β strands in the ACD swap places, as seen in the schematic of the ACD dimer of *M. jannaschii* Hsp 16.5 on the left. Human HspB5, on the right, is unusual in that monomers have elongated β strands that interact as antiparallel β-sheets at the dimeric interface. Each ACD monomer is presented in a different color here. Adapted from Haslbeck and Vierling, 2015.[45]

Though the ACD contains hydrophobic regions that stabilize large oligomers, it is not capable of forming higher order oligomers on its own, so this association is dependent on the NTD and CTD.[45, 46] The CTD, despite its overall variability, does contain a conserved IXI/V motif that is involved in larger oligomer formation.[46] This IXI motif can interact with a groove formed by two β-strands in the ACD to form tetramers or hexamers, and interactions with the NTD result in even larger oligomers (Figure 24).[45]

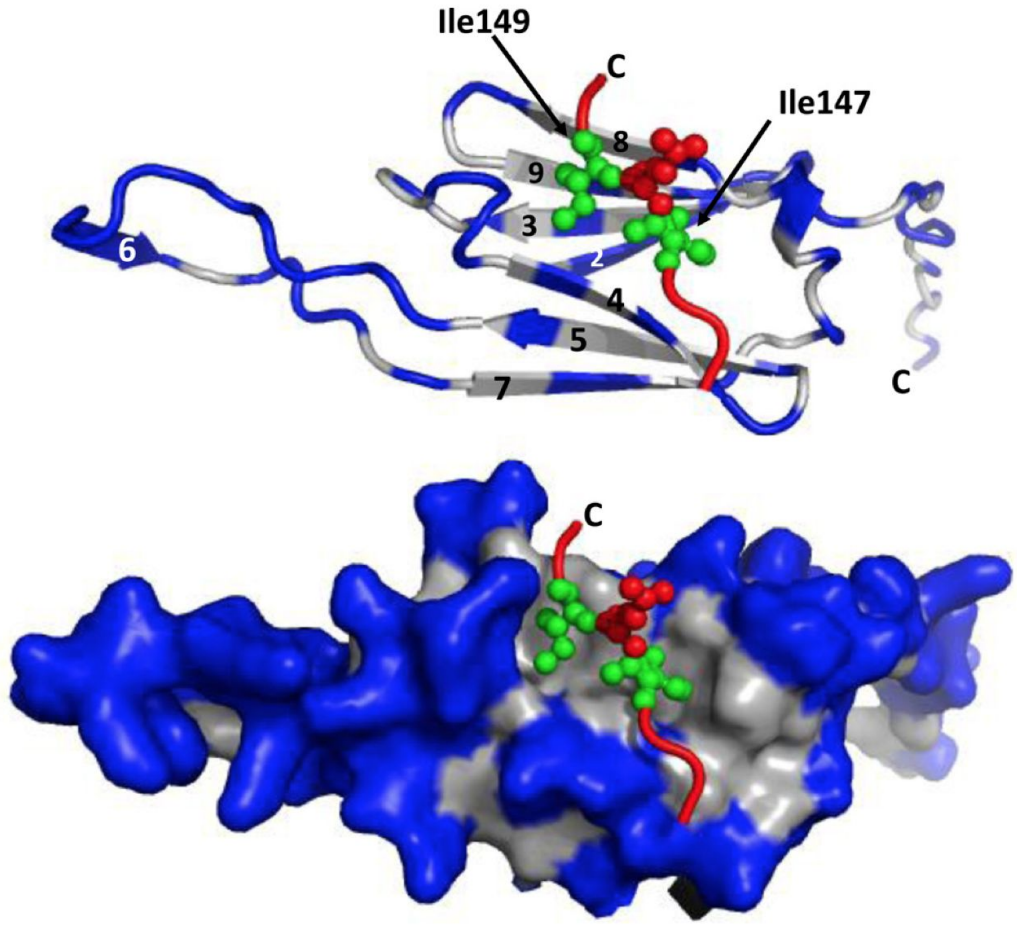


Figure 24: Cartoon (top) and surface (bottom) representation of the CTD IXI motif interaction with an ACD β-strand groove. The IXI motif slots into the hydrophobic groove formed between the β4 and β8 strands of the ACD. The ACD is shown in grey and blue; grey regions indicate the position of hydrophobic residues. β-strands are numbered and the C-terminal end of the ACD is labeled. The red strand is the part of the CTD containing the IXI motif; isoleucine residues are in bright green and the middle amino acid is red. Adapted from Basha *et al.*, 2012.[1]

sHsps are promiscuous in their binding to substrates, and while current research suggests multiple binding sites are present throughout sHsps, the recognition motifs for binding are still relatively unknown.[45] The presence of a wide variety of

sHsps in a given organism, as well as across all domains, suggests that different sHsps may preferentially bind different substrates.[45]

The active form of sHsps is still a topic of debate (Figure 25). The NTD appears to be involved in many substrate binding interactions.[45] Large oligomer formation would prevent the NTD from participating in such interactions, so it has been proposed that the oligomer functions primarily to store sHsps until they are needed.[45] As described above, chaperone activity has been observed with a large oligomeric form, but there are several sHsps that exist only as dimers in the native state, offering support to the idea that the dimer is the primary source of chaperone activity. [45, 47]

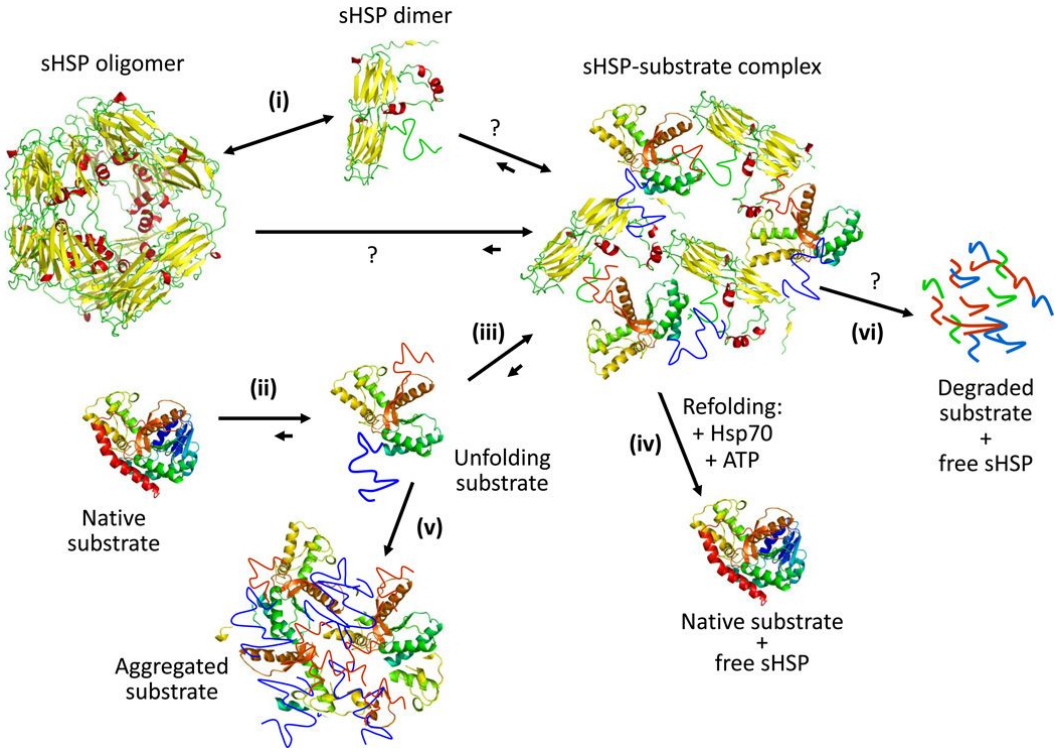


Figure 25: The role of sHsps in the chaperone system. sHsps are polydisperse oligomers (i) that complex with proteins that have misfolded or unfolded (ii, iii) to prevent aggregation (v). The active chaperone sHsps may be dimers or larger oligomers. Following sequestration by sHsps, substrates can be refolded by the Hsp70 system (iv) or degraded (vi). Mechanisms of sHsp chaperone activity, particularly the role of oligomerization in substrate association, are not entirely known. Figure from Basha *et al.*, 2012.[1]

As mentioned above, sHsps are activated by temperature increases, but they can also be regulated by the presence of substrates, phosphorylation, or hetero-oligomer formation.[45] Following Le Chatelier's principle, substrate interaction with small sHsp subunits would drive oligomeric equilibrium to dissociate into more active sHsp subunits.[45] sHsp phosphorylation also tends to drive equilibrium towards smaller subunits, though this occurs due to disruption of the large oligomeric forms themselves.[45] In organisms with multiple different

types of sHsps, the active exchange of oligomeric subunits can result in the formation of hetero-oligomers.[45] The regulatory activity of such hetero-oligomer formation is not well-understood, but some studies have shown increased chaperone activity of sHsps in hetero-oligomers as opposed to homo-oligomers.[45]

I-9 Human sHsps HspB1 and HspB5

A human genome search revealed the existence of 10 different human sHsps (Table 2).[49] The most well known of these human sHsps are HspB1 (also known as Hsp27) and HspB5 (also known as αB-crystallin).[49] HspB1 and HspB5 are both Class I sHsps, characterized by their wide distribution in a variety of tissues (in contrast with the tissue-specific expression of Class II sHsps).[48] Both HspB1 and HspB5 have been shown to be present in increased levels in plaques of conformational neurological diseases such as Alzheimer's, Creutzfeldt-Jakob, or ALS.[45, 50]

Table 2: Human sHsps. Adapted from Bakthisaran *et al.*, 2014.[48]

Name	Subunit Mol. Mass (kDa)	Tissue distribution	Class	Functions
Hsp27 (HspB1)	22.8	Ubiquitous, high levels in heart, striated and smooth muscles	I	Chaperone activity, stabilization of cytoskeleton, anti-apoptotic and anti-oxidant function
HspB2 (MKBP)	20.2	Heart, skeletal and smooth muscle	II	Chaperones DMPK and enhances its kinase activity. Target protein-dependent chaperone activity, myofibrillar integrity, anti-apoptotic function, mitochondrial energetic, anti-apoptotic
HspB3	17.0	Heart, brain, skeletal and smooth muscle	II	Target protein-dependent chaperone activity, Maintaining myofibrillar integrity
αA-crystallin (HspB4)	19.9	Abundant in eye lens. skeletal muscle, liver, spleen, adipose tissue (low level)	II	Chaperone activity, genomic stability, eye lens refractive index
αB-crystallin (HspB5)	20.2	Ubiquitous, abundant in eye lens. High levels in heart and muscle	I	Chaperone activity, stabilization of cytoskeletal and nucleoskeletal matrix, cell cycle, cardioprotection, eye lens refractive index, regulation of muscle differentiation, anti-apoptotic function
Hsp20 (HspB6)	16.8	Ubiquitous, abundant in muscle	I	Smooth muscle relaxation, cardioprotection, chaperone activity, anti-apoptotic
HspB7(cardiovascular heat shock protein)	18.6	Heart and skeletal muscle. Adipose tissue (low level)	II	Chaperone activity, maintaining myofibrillar integrity,
Hsp22 (HspB8, E21G1, α-crystallin C)	21.6	Ubiquitous	I	Chaperone activity, induction of autophagy
HspB9 (Hest shock protein beta-9, cancer/testis antigen 51 (CT51))	17.5	Testis	II	Role in cancer/testis antigen
HspB10 (Outer dense fibre of sperm tails, ODF27, ODFPG, RT7, ODFP, CT133)	28.3	Testis	II	Elastic cytoskeletal structure

The structural organization of these sHsps is not completely understood, but both HspB1 and HspB5 exhibit activity-modifying phosphorylation in response to various stimuli.[51] In HspB1, such phosphorylation decreased oligomeric size, supporting the idea that a smaller subunit is responsible for chaperone activity.[51] In contrast, one study found HspB1 self-associated to form larger oligomers in response to temperature increase, but this increase in oligomeric size still correlated with an increase in chaperone activity.[48, 52] However, temperature increases were also associated with increased subunit exchange for both HspB1 and HspB5, which could support the hypothesis that smaller subunits are made

available for chaperone activity.[48] In addition to temperature-dependence, sHsps also respond to some small molecules. In particular, urea disturbs HspB5 structure with a corresponding increase in chaperone activity.[48]

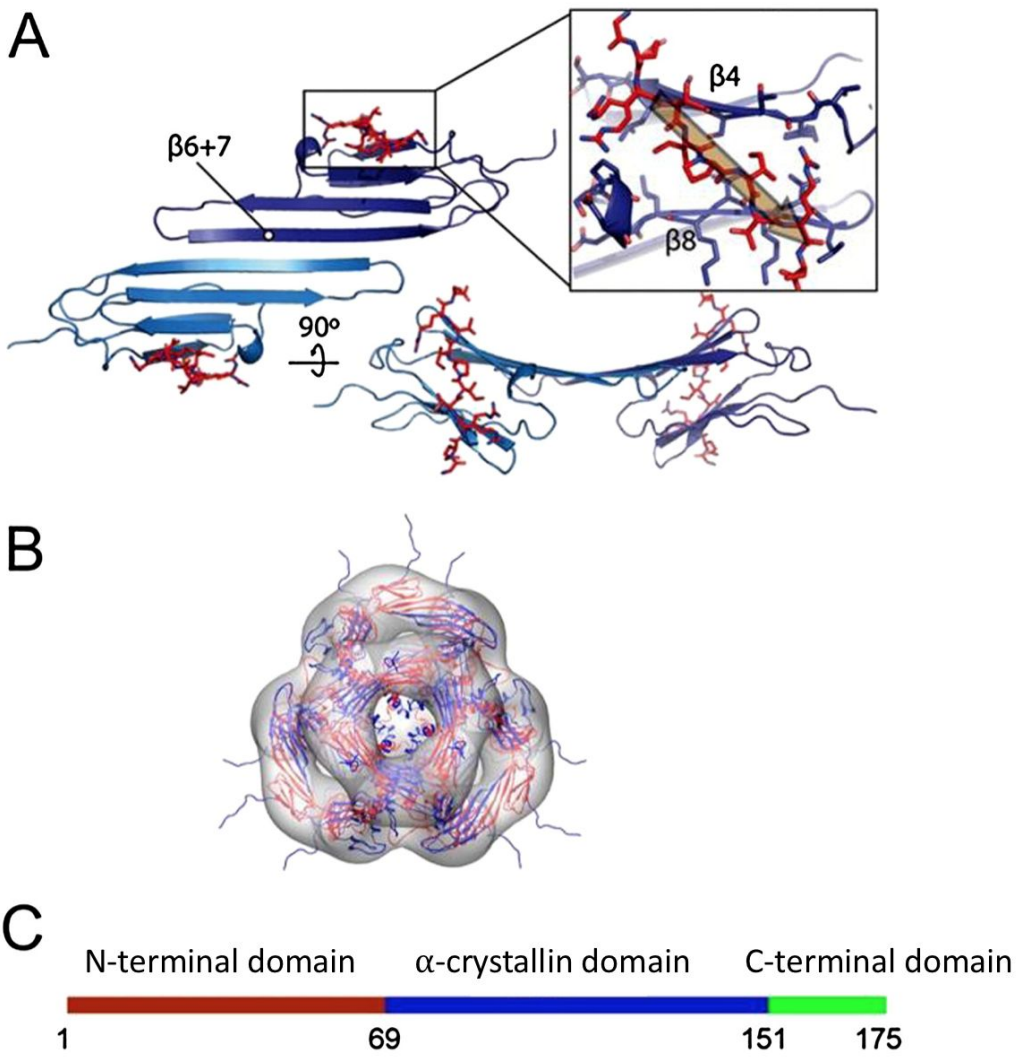


Figure 26: HspB5 structure and oligomerization. A. Dimeric ACD with a close-up on the β4 and β8 strands that make up the β-groove that binds the IXI motif to connect different monomers. **B.** Model of a 24-mer oligomer composed of four hexamers. One is shown on top to illustrate the organization of subunits (alternating pink and purple). **C.** Arrangement of HspB5 domains (NTD in red, ACD in blue, and CTD in green) in the primary sequence. HspB1 has functionally similar structure and domain arrangement. Adapted from Bakthisaran *et al.*, 2014.[48]

Additional structural complexity is present in HspB5, which lacks the β6 strand usually responsible for sHsp dimerization.[1] Instead, HspB5 has an elongated β7 strand that creates a continuous, antiparallel β-sheet between the β4, β5, and β7 strands of each monomer involved in dimerization.[1]

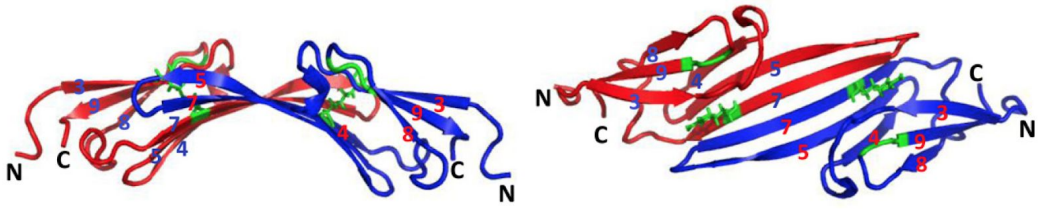


Figure 27: HspB5 ACD dimerization from two perspectives. Monomers are in different colors (red and blue) and β-strands are numbered. Elongated β7 strands interact at the dimer interface as part of a continuous, antiparallel β-sheet. This dimer structure is slightly variable. Conserved residues highlighted in green show different positions in comparison of different dimer structures. Adapted from Basha *et al.*, 2012.[1]

HspB1 and HspB5 are overexpressed in response to cellular stress, but they are also constitutively expressed in some tissues (the eye lens for HspB5, and tissues with high levels of oxidative metabolism such as the heart or skeletal muscles for HspB1).[51] The ubiquitous distribution of HspB1 and HspB5 in conjunction with their upregulation in response to stress has led to their identification as potential therapeutic targets.[51]

I-10 Experimental Aims

As described above, sHsps play a critical role in disease. Therapeutic applications of sHsps could prevent protein aggregation in patients with conformational

diseases, but such applications require a better understanding of the mechanism of sHsp chaperone activity. In particular, the exact relationship between the quaternary organization of sHsps and their chaperone function is still unknown. Based on proposed hypotheses that the dimer is the active subunit, we designed a project to study the role of the dimer in chaperone activity and substrate binding and specificity. Glutathione S-transferase (GST, described below) was genetically linked to two ubiquitous human sHsps, HspB1 and HspB5. GST was connected to the N-terminus regions to disrupt large oligomeric assembly and, as GST naturally forms a dimer, to hopefully constrain the sHsp fusion proteins to dimers, monomers, or small oligomers. Following purification of the two fusion proteins and their wild type counterparts, substrate specificity, interactions and chaperone activity of each sHsp with several model substrates were characterized with chaperone activity assays and size exclusion chromatography analysis.

I-11 Glutathione-S-transferase Fusion Proteins

Glutathione S-transferase (GST) is an enzyme found in many eukaryotes that catalyzes reactions involved in the detoxification of toxic alkylating agents.[53] The primary reaction catalyzed by GST is the addition of reduced glutathione (GSH) to substrates with electrophilic groups, resulting in the formation of a thioether bond between the sulfur and the substrate.[54]

GST can bind a wide variety of ligands, and this versatility is important for being able to detoxify a wide variety of compounds.[55] Part of this versatility comes from the fact that GST is an enzyme "superfamily" that can be divided into many different classes based on sequence and structural properties.[55] However, one thing that remains consistent throughout all classes of GST is that native GST enzymes only exist in the dimeric form.[53]

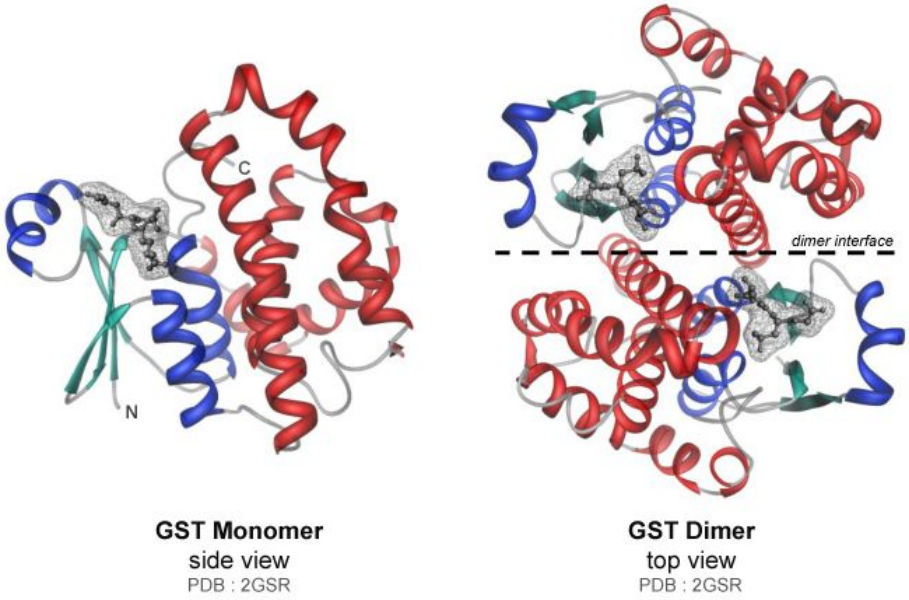


Figure 28: Glutathione S-transferase structure. Monomer (left) and dimer (right). The N-terminal domain is in blue and turquoise and the C-terminal domain is in red. GSH is shown as a dark grey ball and stick figure with a light grey surface structure to indicate the active site position. Figure from Armstrong.[56]

As interesting as GST is in and of itself, it has increasingly been utilized as a biochemical tool to aid in the research of other proteins. One such use is for the purification of proteins fused to GST.[57] Recombinant protein is expressed with

GST located at the N-terminus, where it doubles as both a tag for purification and a chaperone, as it has some function that facilitates correct folding.[57] The fusion protein can then be purified under mild conditions through affinity chromatography, as GST will bind to glutathione coupled to a Sepharose matrix.[57] Following elution with more glutathione, GST can be cleaved from the protein at an engineered protease site, and glutathione and GST can be removed with more chromatography.[57] Throughout the process, using GST for purification supports the purification of stabilized, soluble protein.[57] Some other uses for a GST tag include pull-down assays or immobilization in protein microarrays.[57]

Most of the methods using a GST tag involve the cleavage of GST after it has accomplished its role, so researchers can study the isolated protein without the interference of GST. However, the relative stability and small level of interference of GST in conjunction with its consistent dimer formation has led to new experiments that rely on the dimerization property of GST to restrict the protein of interest to a dimer form as well.[58] GST is particularly favorable to use as a dimeric enzyme since it supports the correct folding of the fusion protein and allows for high yield of the associated protein due to the ability to purify GST fusion proteins.[58] The idea of using GST to force the dimerization of fusion proteins was first introduced in 1995, and tested with moderate success in several different fusions.[59] Further proof of concept followed with a more in-depth study of a cystatin protease inhibitor-GST fusion dimer in 1997.[58] Since then, the use of

GST fusion proteins for induced protein dimerization has become a well-established experimental methodology, studying proteins such as insulin-related kinases or polyubiquitination catalysts.[60, 61]

PART II: MATERIALS AND METHODS

II-1 Materials

Malate dehydrogenase (MDH), insulin, HALT protease inhibitor cocktail, and monoclonal Anti-HspB1 (G3.1) antibody were purchased from Thermo Fisher Scientific (Waltham, MA). Citrate synthase (CS) and dithiothreitol (DTT) were purchased from Sigma-Aldrich (St. Louis, MO). Monoclonal Anti-HspB5 antibody, a polyclonal antibody for glutathione S-transferase, and HRP secondary antibodies were purchased from Abcam (Cambridge, UK). Recombinant human HspB1 and HspB5 with N-terminal His tags were expressed in *Escherichia coli* BL21(DE3) cells using plasmids gifted by Dr. Jason Gestwicki (UCSF). Recombinant GST-HspB1 and GST-HspB5 fusion proteins were expressed using PGex-6P-1 plasmids from GE Healthcare Life Sciences (Marlborough, MA), in which the recombinant proteins were synthesized and cloned into the *Bam*HI and *Eco*RI restriction sites by GenScript (Piscataway, NJ).

II-2 Recombinant Protein Expression and Purification

Competent *E. coli* BL21 (DE3) cells were transformed with plasmids containing recombinant human HspB5 or HspB1 with N-terminal His tags, gifted by Dr. Jason Gestwicki (UCSF), or plasmids containing recombinant GST-HspB1 or GST-HspB5 fusion proteins. Cells were cultured in LB media containing 100

mg/mL ampicillin at 30°C overnight, induced at mid-log phase with isopropyl β-d-thiogalactopyranoside (IPTG) added to a final concentration of 0.5 mM, and protein was expressed for 3 hours at 30°C (HspB1 and HspB5) or 5 hours at 25°C (GST-HspB1 and GST-HspB5).

HspB1 and HspB5 purification protocols were similar to those described by Makley *et al*.[62] Cells were centrifuged at 11,000 rpm for 15 minutes at 4°C, supernatant was removed, and cells were resuspended in lysis buffer (20 mM Tris pH 8.0, 100 mM NaCl, 6 M urea, 5 mM β-mercaptoethanol, 15 mM imidazole) with HALT cocktail protease inhibitor (Thermo Fisher). Cells were sonicated with 6 30-second on/30-second off intervals. An alternative lysis method was B-PER Bacterial Protein Extraction Reagent (Thermo Fisher) using the manufacturer protocol. Cells were centrifuged at 11,000 rpm for 45 minutes at 4°C to remove insoluble material. Proteins were purified from the supernatant using a Ni^{2+} affinity column with Ni-NTA resin and washing with 10 column volumes of wash buffer (lysis buffer with 30 mM imidazole). Proteins were eluted with elution buffer (lysis buffer with 150 mM imidazole) and 5 mM EDTA was added. The solution was centrifuged to concentrate protein, and 1-mL samples were injected onto a Superdex 200 HR 10/30 column equilibrated with Buffer A (20 mM sodium phosphate, pH 7.2, 100 mM NaCl) to refold proteins at 4°C.

Proteins eluted as oligomers with a diameter (165 Å) consistent with literature values.

GST-fusion proteins were also lysed using sonication or B-PER as described above. Protein was dialyzed with 10,000 MW SnakeSkin tubing in 1X PBS (pH 7.4), loaded onto glutathione beads (GE Life Sciences), and eluted with elution buffer (50 mM Tris-HCl, 10 mM reduced glutathione, pH 8.0). Refolding protocol using the Superdex 200 HR 10/30 column was the same as described above. SDS-polyacrylamide gel electrophoresis (PAGE) confirmed protein purification. Protein concentrations were determined for every sample using the Thermo Scientific BSA assay kit prior to further analysis or use in activity assays.

II-3 Gel Electrophoresis and Western Blotting

15 μL protein samples dyed with 5 μL of 1x loading buffer were used for gel electrophoresis. Samples were loaded onto Bio-Rad discontinuous polyacrylamide gels using a 4% (v/v) stacking and 10% (v/v) 4-20% (v/v) resolving gel or 8-16% (v/v) resolving gel for SDS-PAGE and Bio-Rad discontinuous polyacrylamide gels using a 4-20% (v/v) resolving gel for native-PAGE. Gels were run using manufacturer instructions and Coomassie or silver staining (Pierce Kit) was used for visualisation. For western blotting, 5 μM protein samples were run using SDS-PAGE or native-PAGE, and then immunoblotted onto PVDF membranes. Membranes were probed with mouse anti-GST (Thermo Fisher Scientific), mouse

anti-HspB5 (Abcam), or anti-HspB1 (Abcam). A Fast ECL kit (Thermo Scientific) was used to develop the blots and a Bio-imager was used for chemiluminescent signal detection.

II-4 Analytical Size-Exclusion Chromatography

Changes in oligomeric sizes of HspB1, HspB5, and GST fusion proteins with the addition of substrate were determined via size-exclusion chromatography. Samples included sHsps in 2 μM or 12 μM concentrations with 1:1 M ratios of MDH or CS substrates (if used) in 1X PBS (pH 7.4). Samples were heated at 45°C for 30 minutes prior to SEC analysis. 5 μM samples were loaded onto a Superdex 200 10/300 size-exclusion column (GE Healthcare) at a flow rate of 0.3 mL/min at 4°C, following equilibration with 50 mM phosphate and 100 mM NaCl (pH 7.4). Gel filtration standards (Bio-Rad) were used for calibration: bovine thyroglobulin (670 kDa), 10.5 mL, bovine γ-globulin (158 kDa), 14 mL, chicken ovalbumin (44 kDa), 16.5 mL, horse myoglobin (17 kDa), 18.25 mL, and vitamin B12 (1.35 kDa), 21 mL.

II-5 *In Vitro* Chaperone Activity Assay

Chaperone activity of HspB1, HspB5, and GST fusion dimers was assessed by measuring substrate aggregation in the absence and presence of the chaperones. Baseline CS and MDH aggregation in response to heat denaturation at 45°C was measured by solution absorption using methods as described by Ghosh *et al.*[63]

Samples were composed of 1.2 μM chaperone and 1.2μM substrate in 1X PBS (pH 7.3), mixed and placed in a quartz cuvette with a final volume of 1mL. Samples were heated to and maintained at 45°C throughout the assay. When using insulin as a substrate, samples were composed of 75 μM insulin, 25 mM DTT, and 25 μM chaperone in 1X PBS (pH 7.0), and samples were maintained at 25°C throughout the assay. Absorbance at 340 nm as an indication of aggregation was measured every 3 minutes for 1 hour in a Cary 100 UV-Vis spectrophotometer (Agilent) equipped with a multiwell holder, automatic temperature controller, and continuous stirring capabilities.

II-6 Statistical Analysis

Graphing and data analysis was conducted with KaleidaGraph 4.5.2 software. Data represent the mean standard error of the mean for at least 3 independent replicates. Data were normalized to the maximal aggregation of each substrate protein for the chaperone activity data.[64]

PART III: RESULTS

III-1 Purification of Wild Type and GST-Fusion HspB1 and HspB5

Following overexpression in *E. coli* BL21 cells and lysis with sonication or BPER, proteins were purified through dialysis, affinity chromatography, and size-exclusion chromatography as needed. Oligomeric protein size was analyzed with native gel electrophoresis. Figures 29 and 30 show the GST-fusion proteins formed monomers and dimers, while the wild type proteins existed in more diverse oligomeric forms.

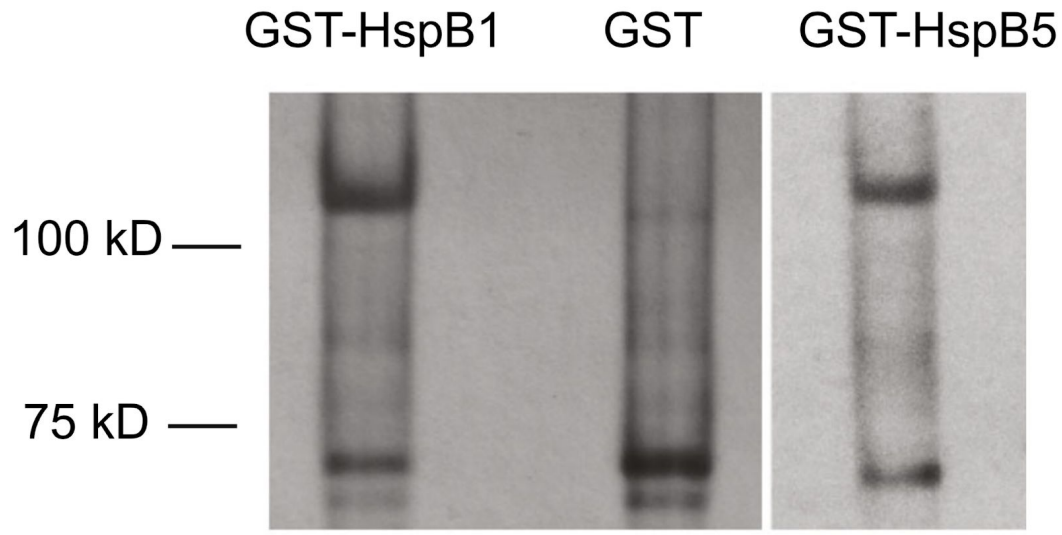


Figure 29: Identification of fusion protein dimers. Coomassie blue staining of 4-20% gradient native PAGE gels shows GST-HspB1 and GST-HspB5 exist in monomer and dimer form in native conditions. The band in lane 2 indicates the dimer form of GST. Figure from Arbach *et al.*, 2017.[64]

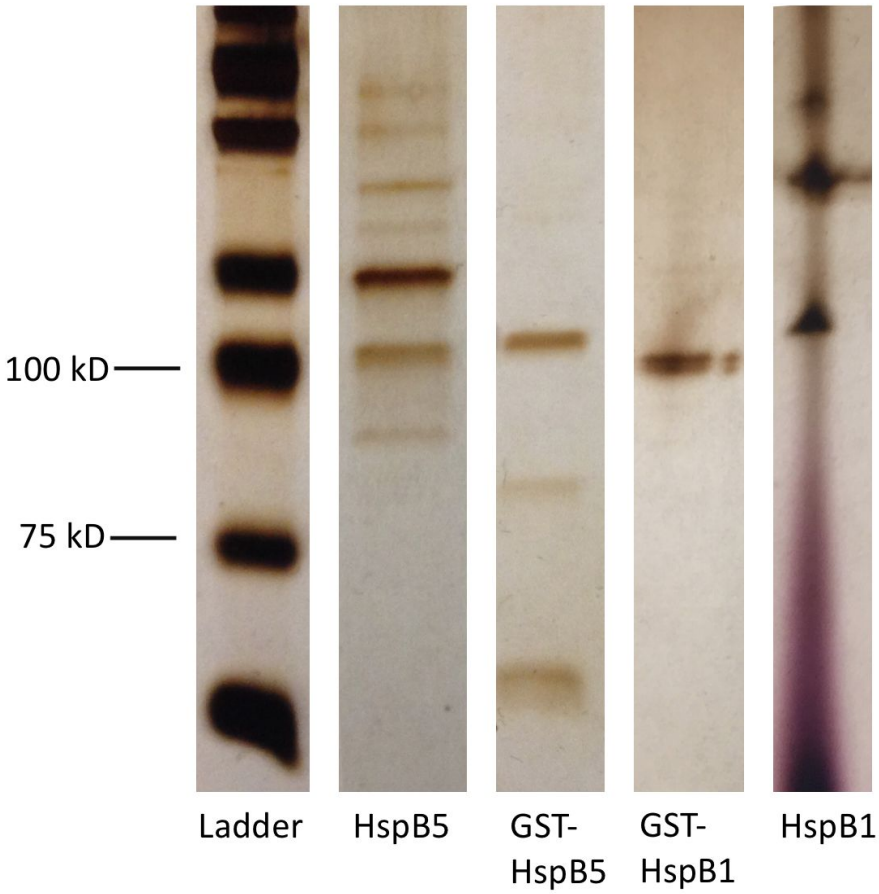


Figure 30: Comparison of wild-type and GST-fusion proteins. A 15% native PAGE gel was silver stained for visualisation. GST-fusion proteins are predominantly in dimer form, approximately 100 kDa. Wild type proteins show a much more variable distribution of oligomeric sizes.

Since native page gel electrophoresis used prior to western blotting showed bands at the expected sizes for wild type and GST-fusion proteins, western blotting was used to confirm protein identity (Figure 31). Anti-HspB1 and anti-HspB5 monoclonal antibodies confirmed the presence of HspB1 and HspB5 in their respective samples and anti-GST antibodies confirmed the presence of GST in specifically the GST-fusion proteins.

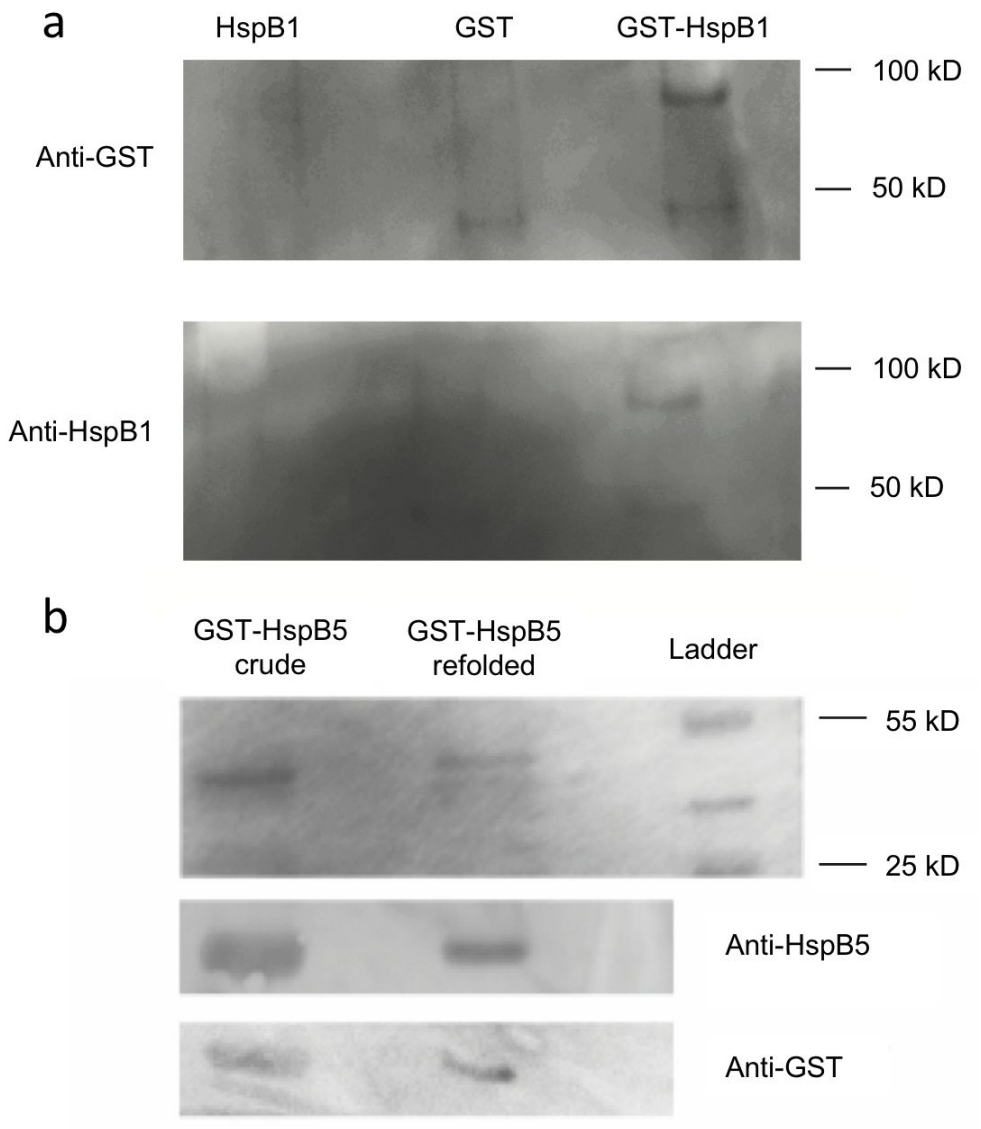


Figure 31: Native PAGE and western blot of HspB1 wild type and GST-fusion proteins (a) and GST-HspB5 fusion protein (b). In both cases, native PAGE preceded western blotting. **a.** Monoclonal anti-GST and Hsp-B1 antibodies were used for western blotting. GST only appeared in the lanes corresponding to GST and GST-HspB1 fusion protein, but not in the lane with wild type HspB1 protein. Similarly, HspB1 did not show up in the GST-only lane. **b.** A coomassie-stained 4-20% gradient native PAGE gel was used as reference for GST-HspB5 fusion band placement. Monoclonal anti-HspB5 and anti-GST antibodies were used for western blotting. Both GST and HspB5 were identified in the bands corresponding to the GST-HspB5 fusion protein. Adapted from Arbach *et al.*, 2017.[64]

Size exclusion chromatography was used with reference to standards to determine GST-fusion protein oligomerization. Both proteins eluted in two primary peaks corresponding to a dimer (~17 mL for GST-HspB1 and ~16 mL for GST-HspB5) and a monomer (~18.5 mL for GST-HspB1 and ~18 mL for GST-HspB5) (Figure 32). It is interesting to note that GST-HspB5 appears to form more monomers than GST-HspB1, according to this chromatogram (Figure 32). Larger oligomers were observed eluting in the range of 8 to 10 mL, and smaller peaks at later volumes were likely due to residual impurities (Figure 32).

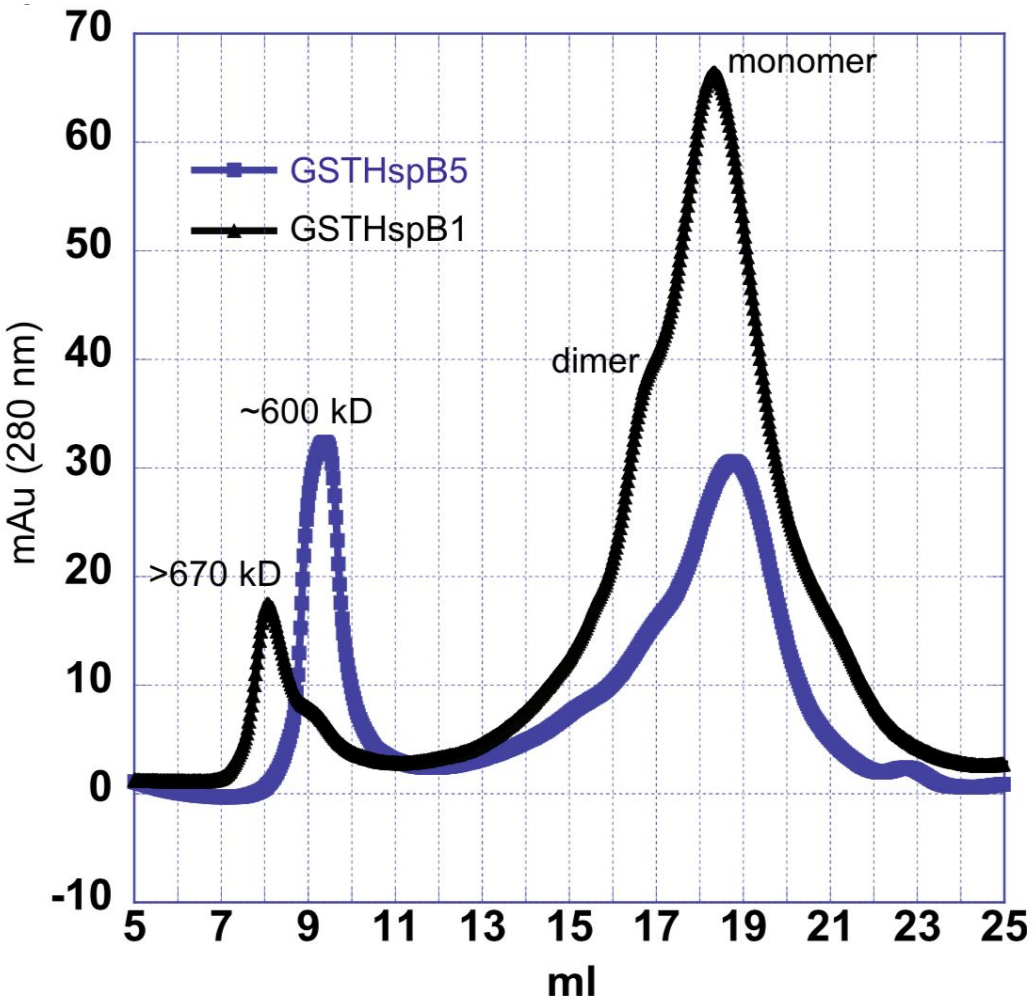


Figure 32: Size exclusion chromatography analysis of GST fusion proteins. SEC analysis was conducted with 2 μM fusion proteins in pH 7.4 phosphate buffer using a Sephadex 200 10/300 GL column at 4°C. GST-HspB1 eluted primarily in peaks at approximately 17 and 18.5 mL, corresponding to dimer and monomer forms. GST-HspB5 similarly eluted in peaks at approximately 16 and 18 mL, corresponding to dimer and monomer forms. The peaks at 8-10 mL are hypothesized to be large oligomers or misfolded aggregates. Figure from Arbach *et al.*, 2017.[64]

Initially, chaperone proteins were evaluated by size exclusion chromatography at a concentration of 12 μM. However, SEC peaks for 12 μM chaperone were shifted left (shorter retention time) relative to peaks for 2 μM chaperone,

indicating higher concentrations induced larger oligomer or aggregate formation and disfavored the dimeric and monomeric forms of interest (Figure 33). Following this trial, all future SEC analysis was conducted with 2 μM chaperone protein to ensure analysis of was primarily of dimeric chaperone activity.

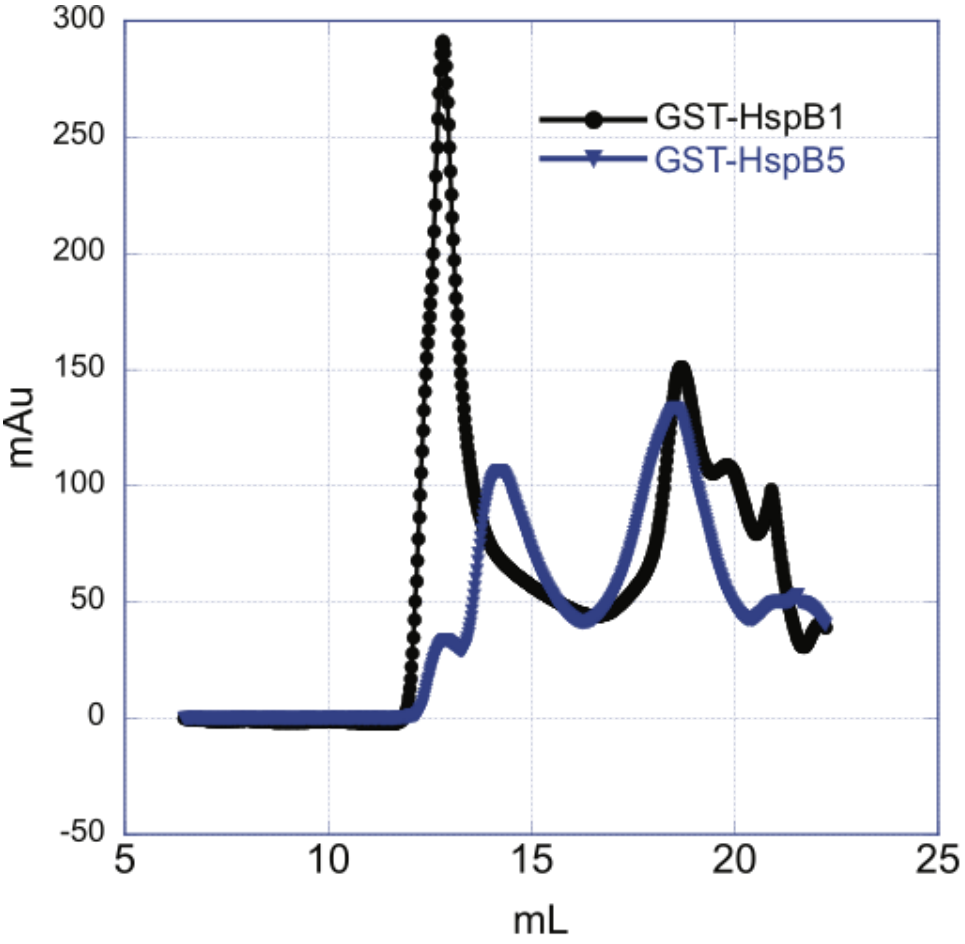


Figure 33: Optimization of size exclusion chromatography analysis of GST fusion proteins. SEC analysis was conducted with 12 μM fusion proteins in pH 7.4 phosphate buffer using a Sephadex 200 10/300 GL column at 4°C. GST-HspB1 eluted predominantly in a peak at 12.5 mL (corresponding to large oligomers or misfolded aggregates), with some later peaks corresponding to dimer and monomer forms. GST-HspB5 eluted in peaks at approximately 12.5, 14, and 18 mL, corresponding to large oligomers or misfolded aggregates with some dimeric or monomeric forms. Figure from Arbach *et al.*, 2017.[64]

III-2 Chaperone Activity Assay with Heat-Denatured Substrate Proteins

For chaperone activity assays, sHsps were combined with an MDH or CS substrate in a 1:1 ratio and heated at 45°C to induce heat-denatured aggregation of the substrate. Light scattering with constant stirring of the solution was used as a measurement of aggregation, with higher light scattering corresponding to increased aggregation, often in the form of visible aggregates. Prior to attempting to measure chaperone activity with substrates, controls confirmed GST alone had no ability to reduce aggregation and neither GST nor the GST-sHsp fusion proteins contributed to the aggregation measurement (Figure 34).

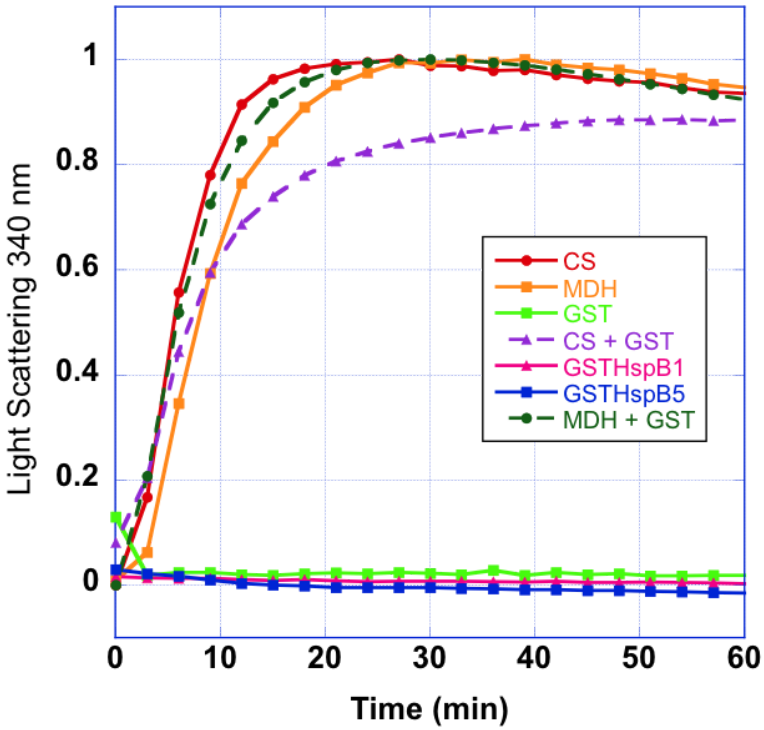


Figure 34: Controls for the heated chaperone activity assay. GST and GST-sHsp fusion proteins showed no aggregation by themselves. GST alone did not reduce aggregation of either substrate. Substrates CS and MDH are included for reference. Figure from Arbach *et al.*, 2017.[64]

In the MDH aggregation assay, the maximum light scattering with wild type HspB1 and HspB5 chaperones was 0.44 and 0.21 respectively, while the maximums with the fusion dimers GST-HspB1 and GST-HspB5 were 0.24 and 0.69 (Figure 35). These were relative to the maximum light scattering of MDH alone, which was normalized to 1.0 (Figure 35). When the same assay was repeated with CS, wild type HspB1, GST-HspB1, and GST-HspB5 all resulted in a maximum light scattering of approximately 0.45, while wild type HspB5 had a maximum of 0.31, again with the maximum scattering for CS alone normalized to 1.0 (Figure 35). Chaperones also exhibited different kinetics; there was a more rapid initial aggregation with HspB1 and GST-HspB5 with MDH and for GST-HspB1 and GST-HspB5 with CS (Figure 35).

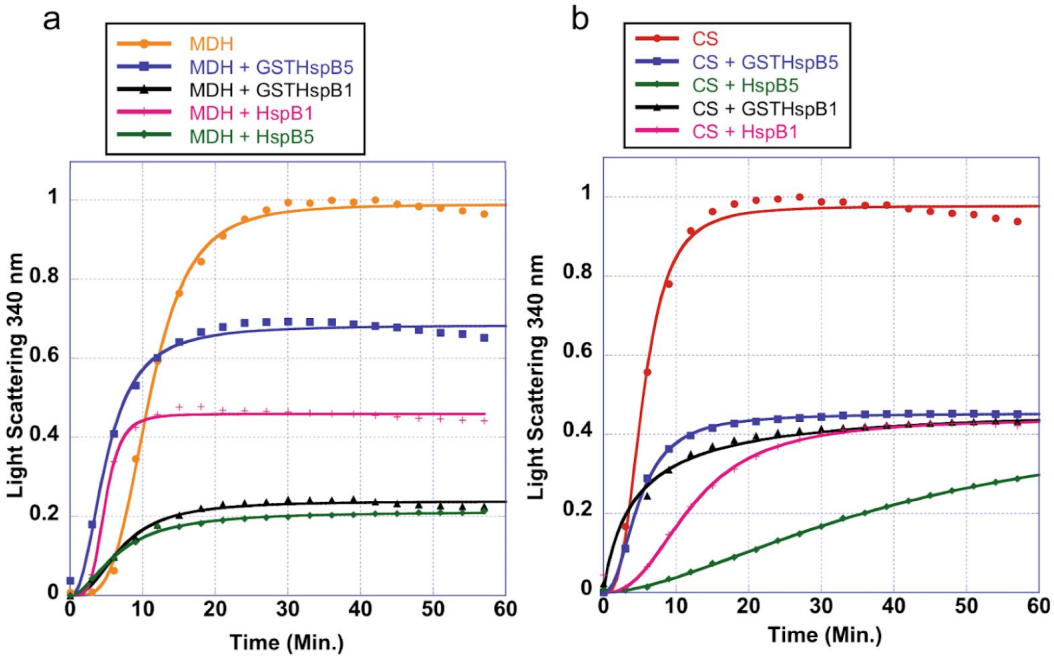


Figure 35: Chaperone activity assay of wild type and GST-fusion sHsps with MDH (a) and CS (b) substrates. All chaperone forms (HspB1, HspB5, GST-HspB1, and GST-HspB5) protect both heat-denatured substrates from aggregation. Faster aggregation kinetics in the first 10 minutes were observed for sHsps less active with a substrate (HspB1 and GST-HspB5 for MDH; GST-HspB1 and GST-HspB5 for CS). Increase aggregation corresponded to increased light scattering at 340 nm. Proteins samples were 1.2 μM with 1:1 rations of chaperone:substrate. Figure from Arbach *et al.*, 2017.[64]

A direct comparison of aggregation for different chaperone-substrate combinations at 30 minutes into the assay allows for a more quantitative summary of chaperone capabilities. The most effective chaperone for the CS substrate was HspB5, which exhibited an ~80% decrease in aggregation (Figure 36). HspB1 and the two GST-fusion chaperones were all similarly effective with an ~60% decrease in aggregation (Figure 36). Chaperones showed more variability with the MDH substrate. HspB5 and GST-HspB1 were most effective, exhibiting an ~80%

reduction in aggregation; HspB1 showed an intermediate ~50% decrease in aggregation, and GST-HspB5 was least effective, only reducing aggregation by ~25% (Figure 36).

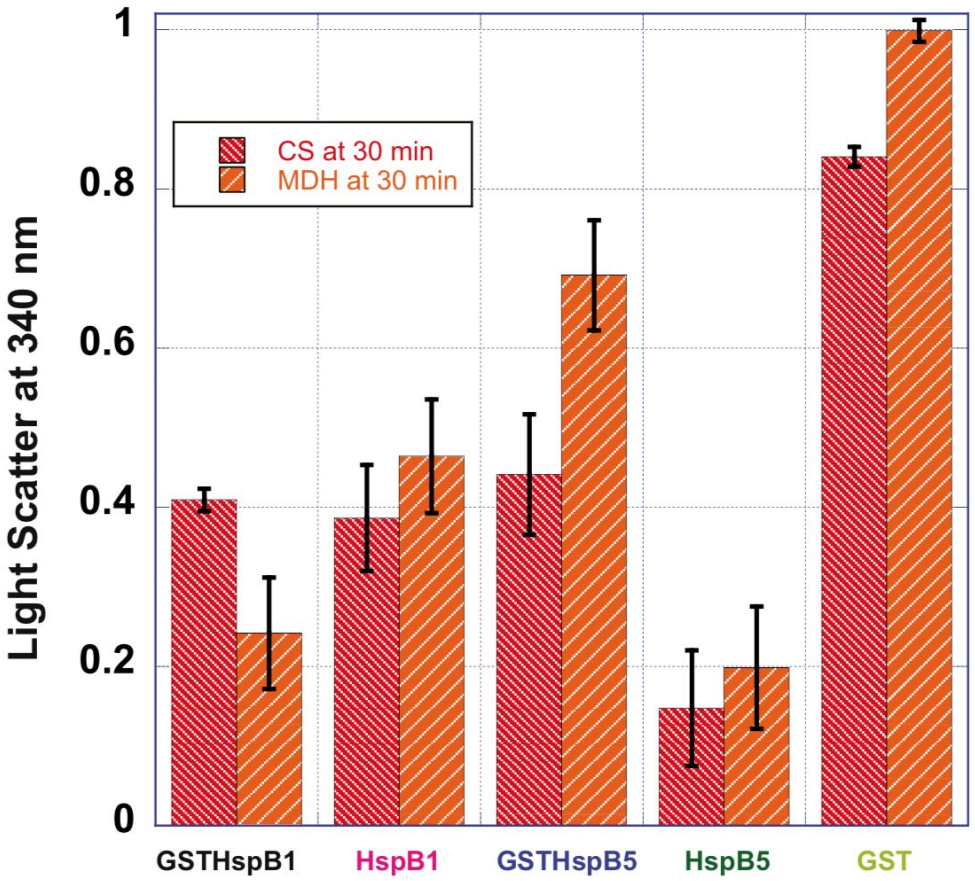


Figure 36: Comparison of chaperone activity for each sHsp with substrates CS or MDH. Light scatter at 340 nm was used as a measure of aggregation after 30 minutes of heating at 45°C. Data were normalized to substrate aggregation without chaperone or a control, with error bars representing the ±SEM for the mean of at least 3 replicates. Figure from Arbach *et al.*, 2017.[64]

III-3 Measurements of Chaperone-Substrate Complexes

Size exclusion chromatography was used to analyze oligomerization and complex formation of chaperone proteins heated alone or with substrates in a 1:1 ratio at 45°C for 30 minutes. Following elution, fractions were collected and further visualized with SDS-PAGE using 8-16% gradient gels in nonreducing conditions. SEC analysis of wild type HspB1 showed primarily elution of very large complexes (>670 kDa). Substrates CS and MDH also formed large complexes (~150 kDa) when heated alone (Figure 37c). When HspB1 was analyzed with either substrate, shifts in elution were observed corresponding to smaller complexes (<100 kDa), and the fact that both HspB1 and the substrates eluted together with a shifted SEC peak indicates complex formation between the chaperone and substrate (Figure 37c). SEC analysis of GST-HspB1 alone showed approximately three peaks (Figure 37d). Unlike HspB1, these peaks predominantly corresponded to smaller complexes (<100 kDa), suggesting dimer or monomer forms instead of larger oligomers. These peaks were, however, similarly shifted upon addition of CS or MDH, indicating the formation of several different-sized complexes further identified by SDS-PAGE (Figure 37b, d).

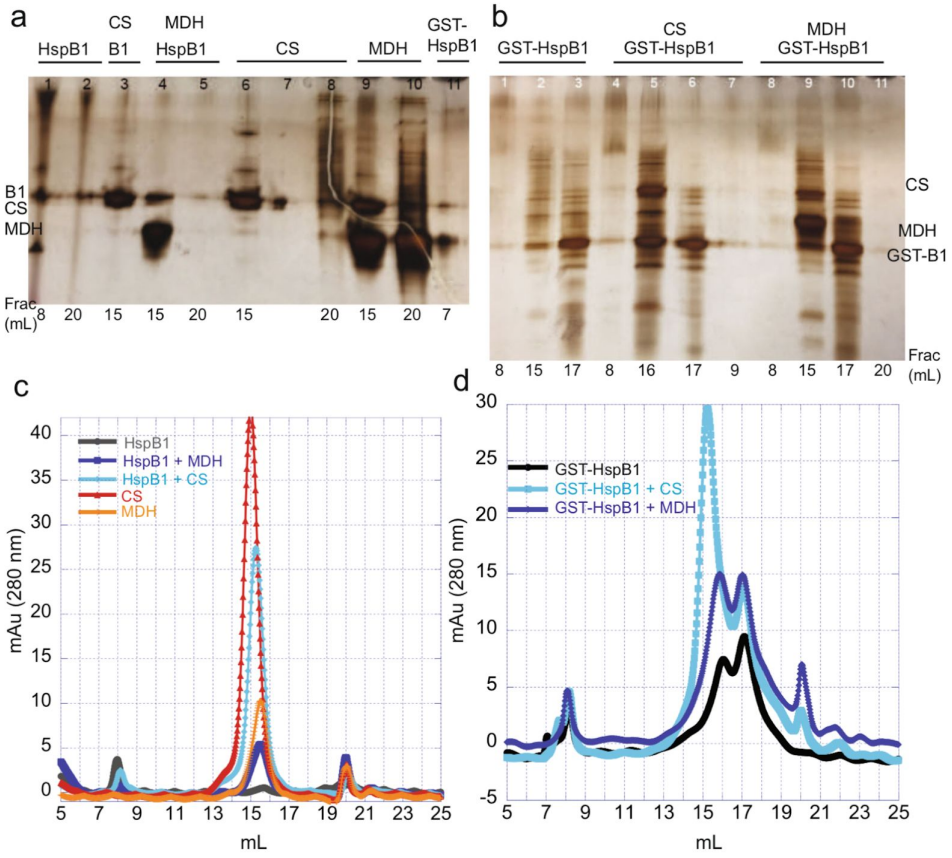


Figure 37: Analysis of HspB1 (left) and GST-HspB1 (right) complex formation with substrates using SEC and SDS-PAGE of fractions. a. SDS-PAGE gel of SEC fractions of HspB1 alone, with CS, or with MDH, CS and MDH alone, and GST-HspB1 alone. **b.** SDS-PAGE gel of SEC fractions of GST-HspB1 alone, with CS, or with MDH. **c.** Corresponding SEC chromatograph of HspB1 alone, with CS, or with MDH, and CS and MDH alone. **d.** Corresponding SEC chromatograph of GST-HspB1 alone, with CS, or with MDH. All chaperone-substrate mixes were 1:1 M ratios and all samples were heated 30 minutes at 45°C prior to SEC evaluation. Figure from Arbach *et al.*, 2017.[64]

The SEC analysis of HspB5 alone is similar to other chaperones, with a large oligomeric complex (>670 kDa) and smaller complexes (~100 kDa and ~25 kDa) (Figure 38c). Of note is the fact that HspB5 eluted in the same fractions as substrates when they were added, indicating complex formation (Figure 38a, c).

Peaks were slightly shifted with the addition of substrate (Figure 38c). GST-HspB5 had a very similar SEC profile as HspB5, but formed fewer large oligomeric complexes (Figure 38d). Small shifts were observed with the addition of substrates, which formed chaperone-substrate complexes as indicated by elution of GST-HspB5 and substrate in the same fractions (Figure 38b, d).

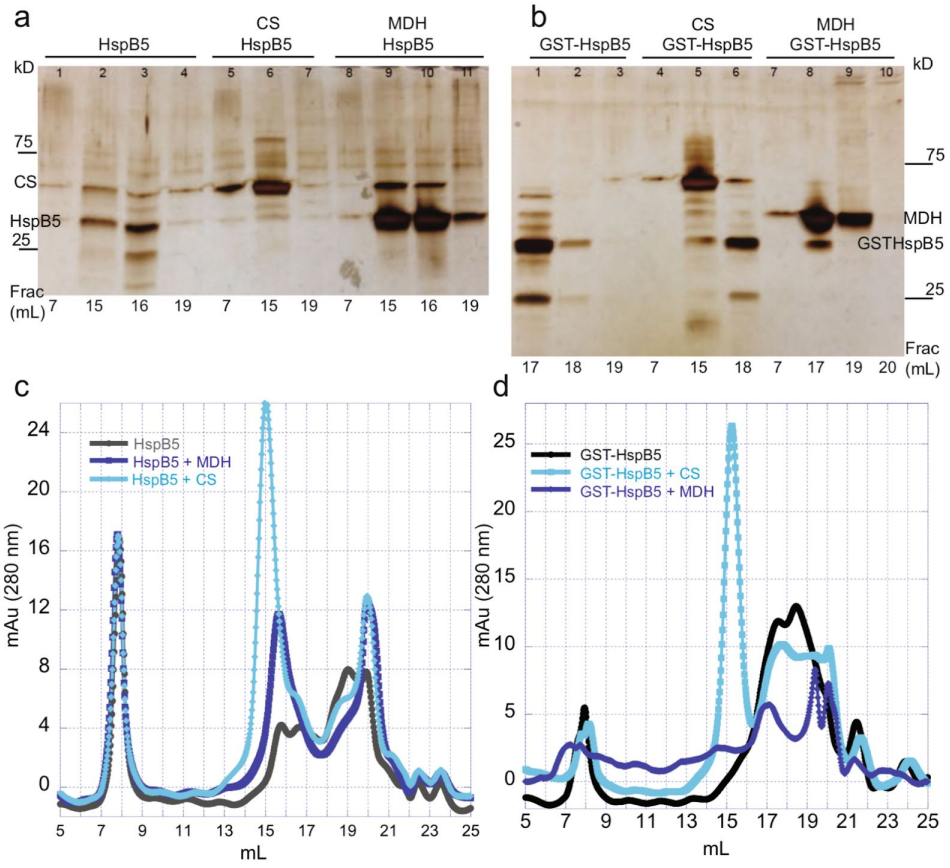


Figure 38: Analysis of HspB5 (left) and GST-HspB5 (right) complex formation with substrates using SEC and SDS-PAGE of fractions. a. SDS-PAGE gel of SEC fractions of HspB5 alone, with CS, or with MDH. **b.** SDS-PAGE gel of SEC fractions of GST-HspB5 alone, with CS, or with MDH. **c.** Corresponding SEC chromatograph of HspB5 alone, with CS, or with MDH. **d.** Corresponding SEC chromatograph of GST-HspB5 alone, with CS, or with MDH. All chaperone-substrate mixes were 1:1 M ratios and all samples were heated 30 minutes at 45°C prior to SEC evaluation. Figure from Arbach *et al.*, 2017.[64]

III-4 Chaperone Activity Assay with Chemically-Denatured Substrate

Following activity assays using heat-denatured substrates, an activity assay was optimized for the use of insulin, a chemically-denatured substrate. UV-Vis light scattering assays were used to measure the aggregation of insulin under a variety of conditions, and the successful aggregation conditions were used in the final chaperone activity assay (Table 3). In lieu of heat-denaturation at 45°C, assays were conducted at 25°C with DTT as a chemical denaturant. Light scattering was still used to measure extent of aggregation.

Table 3: Optimization of insulin aggregation, successful conditions in bold

Insulin Concentration (μM)	10	10	10	10	40	50	50	50	**75**
DTT Concentration (mM)	0	20	20	200	20	20	50	50	**25**
Temperature (°C)	45	25	45	25	25	25	25	37	**25**

Preliminary results include aggregation measurements for insulin alone and insulin with HspB1 in a 3:1 ratio. Insulin aggregation alone had a maximum light scattering of 1.83, and the maximum light scattering with wild type HspB1 was 0.89, exhibiting an ~50% reduction in aggregation (Figure 39). The addition of chaperone did not appear to change aggregation kinetics, as initial rates of aggregation were similar (Figure 39).

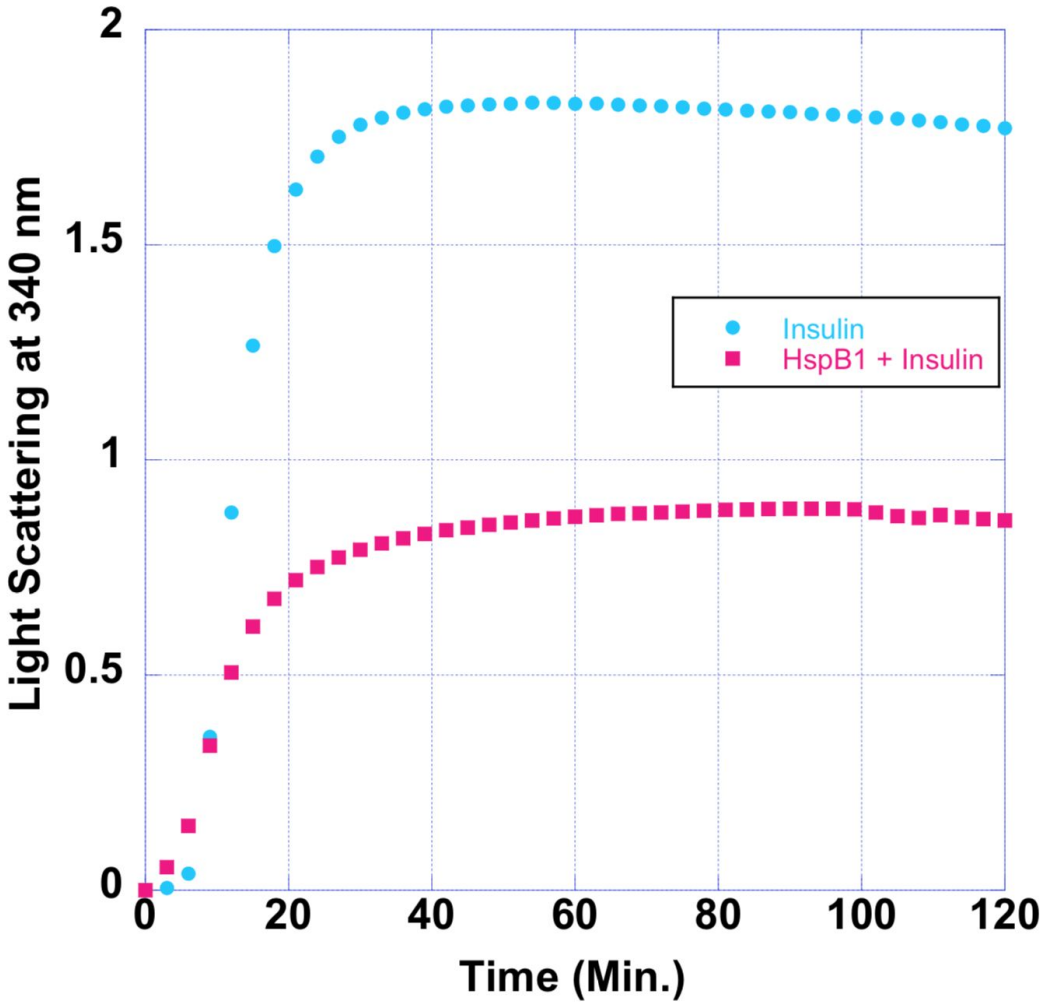


Figure 39: Preliminary chaperone activity assay using insulin as substrate. DTT-treated insulin is denatured and aggregates as measured by increased light scattering at 340 nm. Addition of HspB1 protects insulin from chemically induced aggregation. Samples were composed of 75 μM insulin and 25 mM DTT in 1X PBS (pH 7.0), with or without 25 μM wild type HspB1 chaperone.

PART IV: DISCUSSION

IV-1 Design and Purification of Wild Type and GST-Fusion HspB1 and HspB5: Analysis of Oligomer Formation

The dynamic oligomerization and polydispersity of sHsps makes it difficult to characterize the influence of oligomeric states on chaperone activity. In order to study particular interactions between a specific sHsp oligomer and substrate proteins, and inspired by evidence suggesting the dimer is the smallest oligomer with chaperone activity, we constrained two sHsps, HspB1 and HspB5, to dimer form. GST-fusion proteins were designed to study the chaperone function of sHsps in small oligomeric or monomeric forms as compared to the polydisperse oligomeric forms taken by wild type sHsps. GST was specifically attached to the N-terminus of the sHsp facilitate dimer formation as the N-terminal domain has been implicated in larger oligomer formation, and blocking some access to the N-terminal domain could prevent such quaternary interactions.[45]

Gel electrophoresis and western blotting confirmed the size and identity of GST-sHsp dimer formation as compared to more varied wild type sHsp oligomers, which generally formed large (>250 kDa) complexes. Size exclusion chromatography confirmed the correct formation of dimeric fusion proteins, with the caveat that GST-HspB5 formed more monomers than GST-HspB1. While not

confirmed, a proposed explanation for this observation is that HspB1 contains a cysteine residue in the ACD that could form a disulfide bond with the ACD cysteine of another subunit, and the lack of such crosslinking in HspB5 would lead to fewer HspB5 dimers than HspB1 dimers. Oligomerization did appear to be concentration dependent, as a higher concentration of fusion proteins (12 μM) resulted in a shift towards more larger oligomers and fewer dimers than at a lower concentration (2 μM). Chaperone activity assays were run using the lower concentration to allow the study of dimer formation on chaperone function.

IV-2 Interpreting UV-Vis Light Scattering as a Measure of Aggregation and Chaperone Activity

A chaperone activity assay was designed to measure substrate aggregation in the presence of wild type and GST-fusion chaperones to determine whether the fusion proteins were still active chaperones and, if so, how their chaperone activity compared to wild type sHsps. Substrates and chaperones were combined in a 1:1 ratio at a low concentration to prevent the formation of oligomeric complexes larger than dimers for the GST-fusion proteins. Prior to conducting the chaperone activity assay, controls were established by measuring the heat-induced denaturation and aggregation of substrates CS and MDH. Heating substrates at 45°C resulted in the formation of visible aggregates. When light scattering at 340 nm was measured using UV-Vis spectrophotometry, increased light scattering

was observed over the span of heating until a maximum point, considered to be when all the substrate had formed into aggregates. Therefore, measuring light scattering over the span of heating of chaperones combined with substrates in solution was considered a sufficient measurement of aggregation, and a decrease in maximum light scattering relative to the substrates alone was taken to be a measure of the protective activity of chaperones. In the case of insulin assays, denaturation was a result of DTT addition, not heating, but measurements of light scattering following DTT addition resulted in the same characteristic aggregation curves as with heated CS or MDH.

IV-3 UV-Vis Heated Chaperone Assay Measuring Protective Capacity of Each Chaperone with Malate Dehydrogenase as Substrate

In the heated chaperone activity assay measuring light scattering over time, all four chaperones (wild type HspB1, HspB5, and the GST-fusions) successfully decreased aggregation of both CS and MDH model substrates. With MDH as the substrate, the chaperone activity of the GST fusion proteins was very different from the activity of their wild type counterparts, but they were also very different from each other. This suggests that substrate binding and subsequent chaperone activity is dependent both sequence and oligomerization states of the chaperone involved.

Analysis of aggregation kinetics revealed more insights into chaperone activity. The activity assay with the two least effective chaperones for MDH, HspB1 and GST-HspB5, showed a rapid initial increase in light scattering representative of aggregation before it leveled off and prevented further aggregation like MDH alone would undergo. However, the more effective chaperones did not demonstrate such a rapid initial increase in MDH aggregation when used in the activity assay. In consideration of this difference, it should be noted the chaperones are being subject to the same heat stress as the substrates, which has been shown to change oligomerization states of chaperones.[41, 47] The change in chaperone organization in conjunction with the denaturation of the substrate in response to heating could lead to interactions between HspB1 or GST-HspB5 and MDH that prevent optimal chaperone protective activity. In the case of GST-HspB5, it is also possible that the dimer is not the most efficient state for protecting MDH; a larger HspB5 oligomer may allow interactions that make a larger oligomer the more favorable chaperone state for preventing MDH aggregation.

IV-4 UV-Vis Heated Chaperone Assay Measuring Protective Capacity of Each Chaperone with Citrate Synthase as Substrate

When the model substrate was CS, chaperones exhibited a more similar activity profile than with MDH; with the exception of wild type HspB5 (which was more

active as a chaperone than the others) and following initial kinetic differences, the two GST-fusion proteins and wild type HspB1 all appears to have similar maximal chaperone capacities.

As with MDH, there was a distinction between chaperones that exhibited rapid initial aggregation kinetics and those that did not. However, this was not as clearly linked to chaperone efficacy as was the case with MDH. HspB1, GST-HspB1, and GST-HspB5 all exhibited similar chaperone activity with regards to the final amount of aggregation prevented; HspB5 was more effective than all of the other chaperones at preventing CS aggregation. However, it was only in the activity assays with the fusion proteins, GST-HspB1 and GST-HspB5, that the more rapid initial increase in light scattering was observed. The assay with HspB1 did not exhibit faster aggregation kinetics. This implies that the variable oligomerization dynamics found in wild type sHsps but not in the fusion-constrained dimers is helpful in establishing contacts with CS. It is also possible that the NTR, made less accessible by GST attachment to that domain in the fusion proteins, is necessary for establishing favorable interactions upon initial contact with CS. Regardless, the difference in chaperone activity must be as a result of initial complex formation and determination, as it is only the initial aggregation kinetics that changes between the GST-fusion proteins and the wild type sHsps, not the final overall chaperone capacity. This remains the case except for the difference observed between wild type HspB5 and GST-HspB5 chaperone

efficacy, as wild type HspB5 reduced aggregation by 20% more than GST-HspB5. This difference in chaperone capacity may be explained with similar justifications as for the differences in activity when MDH was the substrate: GST-HspB5 may interact with the substrate less favorably than HspB5 or the dimer may simply not be the optimal form for HspB5 chaperone activity.

IV-5 Chaperone Activity Comparison Using Heat-Denatured Model Substrate Aggregation After a Given Time

To allow for a more quantitative comparison of chaperone activity, the amount of aggregation of each substrate in the presence and absence of each chaperone was analyzed at a specific point in time (after 30 minutes of heating). This revealed more subtle patterns of chaperone activity. Wild type HspB1 and HspB5 had similar protective activity for each substrate (meaning that chaperone activity was similar regardless of the model substrate used, but HspB5 was a more effective chaperone overall than HspB1). HspB5 prevented approximately 80-85% of aggregation for both substrates, while HspB1 only prevented approximately 55-60% of aggregation.

In contrast, GST-HspB1 and GST-HspB5 protective activity varied greatly depending on which model substrate was used. GST-HspB1 was more effective at preventing aggregation of MDH than CS, reducing aggregation by ~80% and ~60% respectively. GST-HspB5 was a more effective chaperone for CS than

MDH, reducing CS aggregation by ~60% and MDH aggregation by ~25%. When comparing fusion proteins with the same substrate as opposed to comparing different substrates with the same chaperone, GST-HspB1 and GST-HspB5 also differed in chaperone activity. GST-HspB1 was more effective at protecting MDH than GST-HspB5 (preventing ~80% of aggregation as compared to ~45%), but both chaperones were similarly effective when CS was the substrate (reducing aggregation by ~60%).

Wild type HspB5 was the most effective chaperone at preventing aggregation of both CS and MDH. Interestingly, the wild type chaperones were not always more effective than the fusion proteins, despite the wider range of oligomeric states available to them. HspB1 and GST-HspB1 displayed similar protective capacities with regard to preventing CS aggregation, but GST-HspB1 decreased aggregation of MDH more than HspB1.

IV-6 Chaperone-Substrate Complex Formation Measured After a Given Time of Heat-Denaturation

In order to analyze complex formation, size exclusion chromatography and SDS-PAGE of eluted fractions were used to study the chaperones in the presence or absence of substrate, after heating at 45°C for 30 minutes. For all four chaperones, protein and substrate were eluted in the same fractions, indicating complex formation between the chaperone and substrate. These complexes were

of several different sizes, as can be seen in both the different SEC peaks and the SDS-PAGE gel lanes showing both chaperone and substrate present in given fractions. Wild type chaperones generally formed more large complexes (>670 kDa), while the GST-fusion proteins primarily eluted as smaller complexes (<100 kDa). It should be noted that in addition to component analysis of fractions through SDS-PAGE, peaks were shifted slightly to the left upon addition of substrates, indicating that chaperones formed larger complexes with substrates and smaller complexes without, which makes sense as the inclusion of substrate in a complex would increase its size relative to the independent chaperone oligomers.

IV-7 Design and Optimization of a UV-Vis Chaperone Activity Assay with a Chemically-Denatured Substrate

All previous experiments had been conducted in nonreducing conditions to avoid disruption of quaternary structure, as both GST and HspB1 form disulfide bonds naturally.[65] However, the variable activity of chaperones with MDH and CS suggested a substrate specificity for chaperone activity that could be better characterized with more model substrates. Insulin was therefore introduced as a novel substrate, but the denaturation of insulin was chemically-induced as opposed to heat-induced. DTT, the reagent used in insulin denaturation, is a powerful reducing agent, which not only affects the tertiary interactions giving

insulin its structure, but also the quaternary interactions leading to GST and HspB1 dimerization, as both rely on disulfide bond crosslinking. As such, analysis of chaperone activity with insulin is not just a method to further characterize the substrate specificity of chaperones, but could also be used to make inference about the importance of the organization of quaternary structure in chaperone activity.

Using existing protocols for insulin aggregation assays as a baseline for design, UV-Vis light scattering measurements were used to measure aggregation and optimize conditions for a chaperone activity assay.[66, 67, 68, 69] The activity assay used samples containing 75 μM insulin, 25 mM DTT, and 25 μM chaperone and was conducted at 25°C instead of 45°C as for the heat-denatured assays of CS and MDH. Due to the extensive optimization process, preliminary results only tested wild type HspB1 as a chaperone. HspB1 successfully demonstrated chaperone activity with insulin, reducing aggregation by ~50% compared to the substrate alone. Despite using the wild type instead of a GST-fusion chaperone, there was still a chance that chemical disruption by DTT would eliminate chaperone activity given the reliance of HspB1 on disulfide bond formation in oligomerization. Further tests with the other chaperones are necessary to truly characterize the effect of DTT on chaperone activity. The HspB5 chaperones would serve as a control for measuring these effects, as HspB5 does not rely on disulfide bond formation. However, there is still an element of substrate specificity to be taken

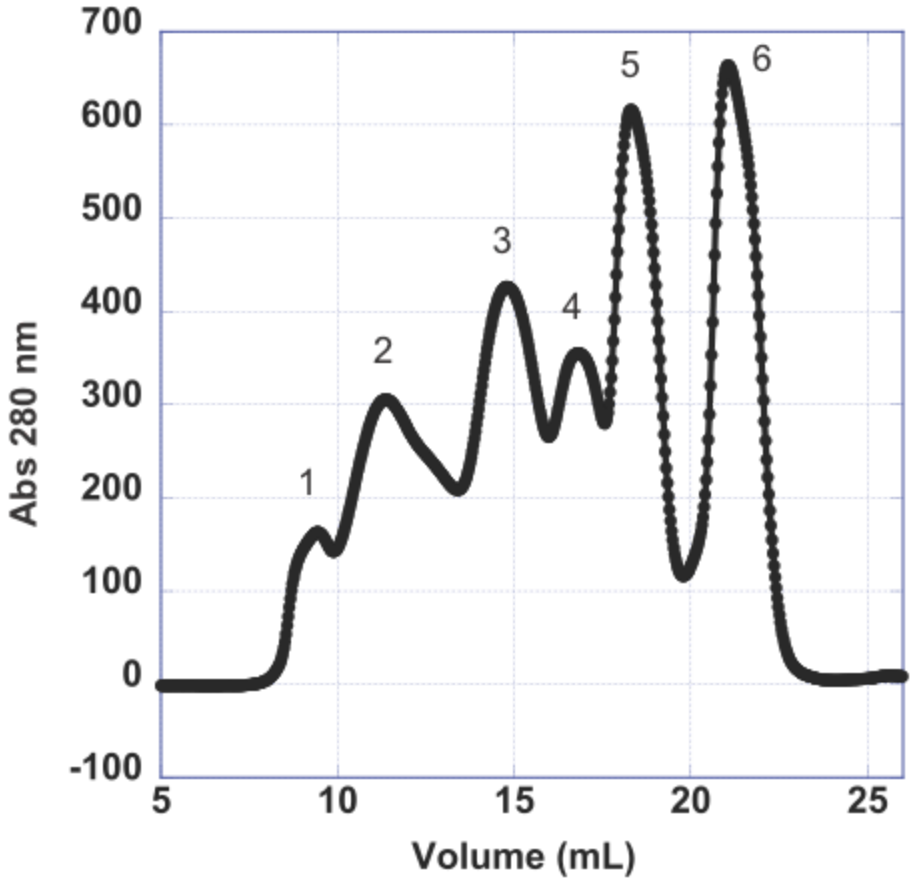
into consideration that could be further influencing HspB5 activities as compared to HspB1 activities. Any results will need to be considered as an intersection of the effects of primary sequence, oligomerization, and crosslinking disruption on substrate specificity and chaperone activity.

IV-8 Overall Conclusions and Future Directions for Research

Through the use of gel electrophoresis, western blotting, and size exclusion chromatography, an N-terminal GST linkage was established as a successful way to constrain naturally polydisperse sHsp oligomers to dimeric and monomeric forms. Chaperone activity assays with heat-denatured model substrates show the GST-fusion sHsps demonstrate comparable chaperone activity to wild type sHsps, confirming suggestions that the dimer is a functionally active oligomer of sHsps and that dynamic interchange between oligomeric states is not necessary for chaperone activity. Some substrate specificity was observed for chaperone activity, but it was not consistent across all sHsps and model substrates tested. Preliminary results suggest a chaperone activity assay with a chemically-denatured model substrate may be helpful in further characterizing the substrate specificity of chaperones. Given the variability of aggregation reduction for different chaperone-substrate combinations, substrate specificity is likely influenced by both oligomerization and primary structure, as measured by differences in the chaperone activity of wild type versus GST-fusion proteins and

HspB1 versus HspB5 chaperones. The McMenimen Lab is continuing to research the mechanism of sHsp chaperone activity with ongoing work with sHsp dimers as well as more intensive characterization of the N-terminal domain of sHsps.

APPENDIX



Size exclusion chromatography protein standards eluted 0.3 mL/min in PBS pH 7.4 for calibration of a Superdex 200 10/300 GL column. Peaks are as follows: (1) aggregate peaks, >670 kDa, (2) bovine thyroglobulin, 670 kDa, (3) bovine γ-globulin, 158 kDa, (4) chicken ovalbumin, 44 kDa, (5) horse myoglobin, 17 kDa, and (6) vitamin B12, 1.35 kDa.

REFERENCES

(1) Basha, E, O'Neill, H, and Vierling, E. 2012. Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions. *Trends in Biochemical Sciences*. **37**(3): 106-117. doi:10.1016/j.tibs.2011.11.005

(2) Sun, Y, and MacRae, TH. 2005. The small heat shock proteins and their role in human disease. *FEBS Journal*. **272**(11): 2613-2627. doi:10.1111/j.1742-4658.2005.04708.x

(3) Treweek, TM, Meehan, S, Ecroyd, H, and Carver, JA. 2015. Small heat-shock proteins: important players in regulating cellular proteostasis. *Cellular and Molecular Life Sciences.* **72**(3): 429-451. doi:10.1007/s00018-014-1754-5

(4) Kim, YE, Hipp, MS, Bracher, A, Hayer-Hartl, M, and Hartl, FU. 2013. Molecular chaperone functions in protein folding and proteostasis. *Annual Review of Biochemistry*. **82**: 323-355. doi:10.1146/annurev-biochem-060208-092442

(5) Narayan, P, Ehsani, S, and Lindquist, S. 2014. Combating neurodegenerative disease with chemical probes and model systems. *Nature Chemical Biology*. **10**: 911-920. doi:10.1038/nchembio.1663

(6) The Human Genome Project Completion: Frequently Asked Questions. https://www.genome.gov/11006943/ (accessed Aug 25, 2017)

(7) Rands, CM, Meader, S, Ponting, CP, and Lunter, G. 2014. 8.2% of the human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. *PLOS Genetics*. **10**(7): e1004525. doi:10.1371/journal.pgen.1004525

(8) Allen, BL, and Taatjes, DJ. 2015. The Mediator complex: a central integrator of transcription. *Nature Reviews Molecular Cell Biology*. **16**(3): 155-166. doi:10.1038/nrm3951

(9) Proudfoot, NJ, Furger, A, and Dye, MJ. 2002. Integrating mRNA processing with transcription. *Cell*. **4**(22): 501-512. doi:10.1016/S0092-8674(02)00617-7

(10) Moldave, K. 1985. Eukaryotic protein synthesis. *Annual Review of Biochemistry*. **54**: 1109-1149. Retrieved from: https://www.ncbi.nlm.nih.gov/pubmed/3896117

(11) Merrick, WC. 1992. Mechanism and regulation of eukaryotic protein synthesis. *Microbiological Reviews*. **56**(2): 291-315. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC372869/

(12) Jahn, TR, and Radford, SE. 2005. The yin and yang of protein folding. *FEBS Journal*. **272**: 5962-5970. doi:10.1111/j.1742-4658.2005.05021.x

(13) Hartl, FU, Bracher, A, and Hayer-Hartl, M. 2011. Molecular chaperones in protein folding and proteostasis. *Nature*. **475**: 324-332. doi:10.1038/nature10317

(14) Kuwajima, K. 1989. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins: Structure, Function, and Genetics*. **6**: 87-103. doi:10.1002/prot.340060202

(15) Hatahet, F, and Ruddock, LW. 2009. Modulating proteostasis: Peptidomimetic inhibitors and activators of protein folding. *Current Pharmaceutical Design*. **15**: 2488-2507. Retrieved from: https://www.ncbi.nlm.nih.gov/pubmed/19601846

(16) Braakman, I, and Bulleid, NJ. 2011. Protein folding and modification in the mammalian endoplasmic reticulum. *Annual Review of Biochemistry*. **80**: 71-99. doi:10.1145/annurev-biochem-062209-093836

(17) Araki, K, and Nagata, K. 2011. Protein folding and quality control in the ER. *Cold Spring Harbor Perspectives in Biology*. **3**(11): 1-25. doi:10.1101/cshperspect.a007526

(18) Skach, WR. 2009. Cellular Mechanisms of Membrane protein folding. *Nature Structural & Molecular Biology*. **16**(6): 606-612. doi:10.1038/nsmb.1600

(19) Ma, Y, and Hendershot, LM. 2004. The role of the unfolded protein response in tumour development: friend or foe? *Nature Reviews Cancer*. **4**: 966-977. doi:10.1038/nrc1505

(20) Hipp, MS, Park, SH, and Hartl, FU. 2014. Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends in Cell Biology*. **24**(9): 506-514. doi:10.1016/j.tcb.2014.05.003

(21) Carrell, RW, and Lomas, DA. 1997. Conformational disease. *The Lancet*. **350**: 134-138. Doi: 10.1016/S0140-6736(97)02073-4

(22) Paulsson, JF, and Westermark, GT. 2005. Aberrant processing of human proislet amyloid polypeptide results in increased amyloid formation. *Diabetes*. **54**(7): 2117-2125. Retrieved from: https://www.ncbi.nlm.nih.gov/pubmed/15983213

(23) Maulucci, G, Papi, M, Arcovito, G, and De Spirito, M. 2011. The thermal structural transition of α-crystallin inhibits the heat induced self-aggregation. *PLoS One*. **6**(5): e18906. doi:10.1371/journal.pone.0018906

(24) Morimoto, RI. 2008. Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. *Genes & Development*. **22**:1427-1438. doi:10.1101/gad.1657208

(25) Balch, WE, Morimoto, RI, Dillin, A, and Kelly, JW. 2008. Adapting proteostasis for disease intervention. *Science*. **319**(5865): 916-919. doi:10.1126/science.1141448

(26) Klaips, CL, Jayaraj, GG, and Hartl, FU. 2017. Pathways of cellular proteostasis in aging and disease. *Journal of Cell Biology*. 1-13. doi:10.1083/jcb.201709072

(27) World Health Organization. 2006. Neurological Disorders: Public Health Challenges. *WHO Press*. Geneva, Switzerland. Retrieved from: http://www.who.int/mental_health/neurology/neurodiso/en/

(28) National Eye Institute. Cataracts Defined. https://nei.nih.gov/eyedata/cataract (accessed Apr 2, 2018).

(29) Huntington's Disease Society of America. What is Huntington's Disease? http://hdsa.org/what-is-hd/ (accessed Apr 2, 2018)

(30) ALS Association. Quick Facts About ALS & The ALS Association. http://www.alsa.org/news/media/quick-facts.html (accessed Apr 2, 2018)

(31) Moore, LS, Le, T, and Fan, G. 2013. DNA methylation and its basic function. *Neuropsychopharmacology*. **38**(1): 23-38. doi:10.1038/npp.2012.112

(32) Dana, H, Chalbatani, GM, Mahmoodzahdeh, H, et al. 2017. Molecular mechanisms and biological functions of siRNA. *International Journal of Biomedical Science*. **13**(2): 48-57. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5542916/

(33) Lilienbaum, A. 2013. Relationship between the proteasomal system and autophagy. *International Journal of Biochemistry and Molecular Biology*. **4**(1): 1-26. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3627065/

(34) Hershko, A, and Ciechanover, A. 1998. The ubiquitin system. *Annual Review of Biochemistry*. **67**: 425-479. doi:10.1146/annurev.biochem.67.1.425

(35) Finley, D. 2009. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annual Review of Biochemistry*. **78**: 477-513. doi:10.1146/annurev.biochem.78.081507.101607

(36) Meusser, B, Hirsch, C, Jarosch, E, and Sommer, T. 2005. ERAD: the long road to destruction. *Nature Cell Biology*. **7**(8): 766-772. doi:10.1038/ncb0805-766

(37) Bravo, R, Parra, V, Gatica, D, et al. 2013 Endoplasmic reticulum and the unfolded protein response: dynamics and metabolic integration. *International Review of Cell and Molecular Biology*. **301**: 215-290. doi:10.1016/B978-0-12-407704-1.00005-1

(38) Komatsu, M, Waguri, S, Chiba, T, et al. 2006. Loss of autophagy in the central nervous system causes neurodegeneration in mice. *Nature*. **441**(15): 880-884. doi:10.1038/nature04723

(39) Glick, D, Barth, S, and Macleod, KF. 2010. Autophagy: cellular and molecular mechanisms. *Journal of Pathology*. **221**(1): 3-12. doi:10.1002/path.2697

(40) De Maio, A, Santoro, MG, Tanguay, RM, and Hightower, LE. 2012. Ferruccio Ritossa's scientific legacy 50 years after his discovery of the heat shock response: a new view of biology, a new society, and a new journal. *Cell Stress and Chaperones*. **17**(2): 139-143. doi:10.1007/s12192-012-0320-z

(41) Liberek, K, Lewandowska, A, and Ziętkiewicz, S. 2008. Chaperones in control of protein disaggregation. *The EMBO Journal*. **27**: 328-335. doi:10.1038/sj.emboj.7601970

(42) Ellis, RJ. 1994. Opening and closing the Anfinsen cage. *Current Biology*. **4**(7): 633-634. doi:10.1016/S0960-9822(00)00140-8

(43) Lindquist, S. 2009. Protein folding sculpting evolutionary change. *Cold Spring Harbor Symposia on Quantitative Biology*. **74**: 103-108. doi:10.1101/sqb.2009.74.043

(44) Richter, K, Haslbeck, M, and Buchner, J. 2010. The heat shock response: Life on the verge of death. *Molecular Cell*. **40**: 253-266. doi:10.1016/j.molcel.2010.10.006

(45) Haslbeck, M, and Vierling, E. 2015. A first line of stress defense: Small heat shock proteins and their function in protein homeostasis. *Journal of Molecular Biology*. **427**(7): 1537-1548. doi:10.1016/j.jmb.2015.02.002

(46) van Montfort, RLM, Basha, E, Friedrich, KL, Slingsby, C, and Vierling, E. 2001. Crystal structure and assembly of a eukaryotic small heat shock protein. *Nature Structural Biology*. **8**(12): 1025-1030. doi:10.1038/nsb722

(47) Franzmann, TM, Wühr, M, Richter, K, Walter, S, and Buchner, J. 2005. The activation mechanism of Hsp26 does not require dissociation of the oligomer. *Journal of Molecular Biology*. **350**: 1083-1093. doi:10.1016/j.jmb.2005.05.034

(48) Bakthisaran, R, Tangirala, R, and Rao, CM. 2015. Small heat shock proteins: Role in cellular functions and pathology. *BBA Proteins and Proteomics*. **1854**: 291-319. doi:10.1016/j.bbapap.2014.12.019

(49) Kappé, G, Franck, E, Verschuure, P, et al. 2003. The human genome encodes 10 α-crystallin-related small heat shock proteins: HspB1-10. *Cell Stress & Chaperones*. **8**(1): 53-61. Retrieved from: https://www.ncbi.nlm.nih.gov/pubmed/12820654

(50) Zhang, X, Shi, J, Tian, J, et al. 2014. Expression of one important chaperone protein, heat shock protein 27, in neurodegenerative diseases. *Alzheimer's Research & Therapy*. **6**(9): 78-89. doi:10.1186/s13195-014-0078-x

(51) Arrigo, AP, Simon, S, Gibert, B, et al. 2007. Hsp27 (HspB1) and αB-crystallin (HspB5) as therapeutic targets. *FEBS Letters*. **581**: 3665-3674. doi:10.1016/j.febslet.2007.04.033

(52) Lelj-Garolla, B, and Mauk, AG. 2006. Self-association and chaperone activity of Hsp27 are thermally activated. *Journal of Biological Chemistry*. **281**(12): 8169-8174. doi:10.1074/jbc.M512553200

(53) Armstrong, RN. 1991. Glutathione S-transferases: Reaction mechanism, structure, and function. *Chemical Research in Toxicology*. **4**(2): 131-140. doi:10.1021/tx00020a001

(54) Sherratt, PJ, and Hayes, JD. 2002. Enzyme systems that metabolise drugs and other xenobiotics: 9 Glutathione S-transferases. John Wiley & Sons Ltd. Editor: Ioannides, C. doi:10.1002/0470846305

(55) Sheehan, D, Meade, G, Foley, VM, and Dowd, CA. 2001. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical Journal*. **360**: 1-16. Retrieved from: https://www.ncbi.nlm.nih.gov/pubmed/11695986

(56) Armstrong, R. Glutathione transferase (GST) superfamily. *Enzyme Function Initiative: Bridging Projects*. https://enzymefunction.org/about/groups/bridging-projects/gst-superfamily (accessed Apr 24, 2018)

(57) Harper, S, and Speicher, DW. 2011. Purification of proteins fused to glutathione S-transferase. *Methods in Molecular Biology*. **681**: 259-280. doi:10.1007/978-1-60761-913-0_14

(58) Tudyka, T, and Skerra, A. 1997. Glutathione S-transferase can be used as a C-terminal, enzymatically active dimerization module for a recombinant protease inhibitor, and functionally secreted into the periplasm of *Escherichia coli*. *Protein Science*. **6**: 2180-2187. doi:10.1002/pro.5560061012

(59) Yan, H, Lim, JTE, Contillo, LG, and Krolewski, JJ. 1995. Glutathione S-transferase fusion proteins mimic receptor dimerization in permeabilized cells. *Analytical Biochemistry*. **231**: 455-458. doi:10.1006/abio.1995.0080

(60) Baer, K, Al-Hasani, H, Parvaresch, S, et al. 2001. Dimerization-induced activation of soluble insulin/IGF-1 receptor kinases: An alternative mechanism of activation. *Biochemistry*. **40**: 14268-14278. doi:10.1021/bi015588g

(61) Gazdoiu, S, Yamoah, K, Wu, K, et al. 2005. Proximity-induced activation of human Cdc34 through heterologous dimerization. *PNAS*. **102**(42): 15053-15058. doi:10.1073/pnas.0507646102

(62) Makley, LN, McMenimen, KA, DeVree, BT, et al. 2015. Pharmacological chaperone for α-crystallin partially restores transparency in cataract models. *Science*. **350**: 674-677. doi:10.1126/science.aac9145

(63) Ghosh, JC, Estrada, MR, and Clark, JI. 2005. Interactive domains for chaperone activity in the small heat shock protein, human αB crystallin. *Biochemistry*. **44**: 14854-14869. doi:10.1021/bi0503910

(64) Arbach, H, Butler, C, and McMenimen, KA. 2017. Chaperone activity of human small heat shock protein-GST fusion proteins. *Cell Stress and Chaperones*. **22**: 503-515. doi:10.1007/s12192-017-0764-2

(65) Baranova, EV, Weeks, SD, Beelen, S, et al. 2011. Three-dimensional structure of α-crystallin domain dimers of human small heat shock proteins HspB1 and HspB6. *Journal of Molecular Biology*. **411**(1): 110-122. doi:10.1016/j.jmb.2011.05.024

(66) Blagojce, J, Kelly, MA, Rote, A, et al. 2015. Phosphomimics destabilize Hsp27 oligomeric assemblies and enhance chaperone activity. *Chemistry and Biology*. **22**(2): 186-195. doi:10.1016/j.chembiol.2015.01.001

(67) Bumagina, Z, Gurvits, B, Artemova, N, Muranov, K, and Kurganov, B. 2010. Paradoxical acceleration of dithiothreitol-induced aggregation of insulin in the presence of a chaperone. *International Journal of Molecular Sciences*. **11**(11): 4556-4579. doi:10.3390/ijms11114556

(68) Bhattacharyya, J, and Das, KP. 1999. Molecular chaperone-like properties of an unfolded protein, α_s-Casein. *Journal of Biological Chemistry*. **274**: 15505-15509. doi:10.1074/jbc.274.22.15505

(69) Saha, S, and Das, KP. 2004. Relationship between chaperone activity and oligomeric size of recombinant human αA- and αB-crystallin: A tryptic digestion study. *Proteins: Structure, Function, and Bioinformatics*. **57**(3): 610-617. doi:10.1002/prot.20230