PROTEIN INTERACTIONS *IN VITRO* AND IN THE CELL

BY

DRISHTI GUIN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2019

Urbana, Illinois

Doctoral Committee:

Professor Martin Gruebele, Chair
Professor Brian Freeman
Professor Deborah Leckband
Assistant Professor Diwakar Shukla

ABSTRACT

In recent years, better instrumentation and greater computing power have enabled imaging of elusive biomolecule dynamics in cells, driving many advances in understanding the chemical organization of biological systems. Interactions in the cell affect both biomolecular stability and function. The same protein or nucleic acid can behave differently depending on time in the cell cycle, location in a specific compartment, or stresses acting on the cell.

In chapter 1, the three main cellular interactions, crowding, sticking and quinary structure and the current methods to quantify them both *in vitro* and *in vivo* are described in detail. The chapter concludes with a short discussion on protein evolution in the cell, the factors that drive protein evolution and shape protein interaction networks considering current biophysical evidence. These interactions can significantly affect the free energy ΔG of marginally-stable and low-population proteins and due to epistasis direct the evolutionary pathways in an organism.

Chapters 2 and 3 focus on the novel denaturant dodine that denatures proteins in the ~mM concentration range. The denaturation propensity of dodine is investigated on a complex two-domain enzyme protein in chapter 2. Chapter 3 further explores in detail the denaturation action of dodine on α-helices using both experiments and simulations.

Chapter 4-6 detail the construction of a three-color Fluorescence Relaxation Imaging (FReI) and its utilization in characterizing protein-protein interactions inside mammalian cells. Chapter 5 and 6 examine the binding of the heat inducible, Hsp70 and the constitutively expressed, Hsc70, respectively, to a model substrate protein PGK inside cells and *in vitro*.

The appendices provide a summary of protocols and computer codes used in the analyses of chapters 2-6.

ACKNOWLEDEMENTS

Completing my PhD has been the most challenging and at the same time the most rewarding experience of my life. The list of people without whom this wouldn't have been possible is long and I owe everyone a great debt for their contribution, however small.

First and foremost, I must thank my adviser Prof. Martin Gruebele for his tremendous support and kindness along the way and without whose guidance this would not have come to fruition. It was not only a great honor but also a tremendous opportunity to work for Martin and learn from his great depth of knowledge in the field. I am deeply indebted to Martin for being patient with me and being present whenever help was needed. I realized very quickly that Martin is right most of the time and my biggest achievements have been winning arguments against him even though those moments are few and far apart. I must also thank my committee members, Prof. Brian Freeman, Prof. Deborah Leckband, Prof. Diwakar Shukla and Prof. Renske van der Veen for helpful guidance. A special thanks to Prof. Freeman for spending many hours guiding me and for helpful tips on the Hsp70 project and for teaching me what biology I know.

I would also like to thank all the members of the Gruebele group for their help and support through thick and thin. For getting me up on my feet when I first started, I must thank Dr. Hannah Gelman and Dr. Anna Jean Wirth. Without their time and patience, I wouldn't be standing on the firm ground that I am today. I also owe gratitude to Dr. Shahar Sukenik, Dr. Caitlin Davis and Dr. Lydia Kisley who taught me a lot about science in general. I owe special thanks to Dr. Caitlin Davis for being a constant companion and support at conferences and meetings without whom it wouldn't have been as fun and educational. Finally, I would like to thank our younger graduate students, Mayank Boob, Yuhan Wang, Gopika Gopan and Meredith Rickard and REU students Kori Sye and Brian Bozymski. It is true that one learns the most through teaching and teaching some of the younger members of the lab and watching them grow has been both a pleasure and an honor. Finally, I owe thanks to Dr. Kapil Dave. We started our careers the same year and he has both pushed me to do better and helped me stand on my feet in the hardest of times.

The acknowledgements wouldn't be complete without thanking Liz Ryland, a forever friend and a source of support during my graduate school tenure. Outside the lab I had the opportunity to

serve on the Seminar Committee and my work would not have been successful without the help and support of my co-chair Alex Doran. I also owe thanks to Theresa Struss, Karen Watson, Beth Myler and Katie Trabaris our amazing IMP office secretaries. They taught me life skills and made me the woman I am today. Special thanks to Theresa Struss for being my mother outside of home, not to mention all the candy that kept me alive and fed through the long days and long nights.

I cannot express in words the thanks to my family in India who have been far away yet have supported me in all my endeavors. My darling sister, Noyonika Guin, without whose support and love this would not have been possible. My late grandmother, Mira Guin, who will be with me forever and without whose influence on my life I would not have been where I am today. My thanks wouldn't be complete without a special mention of my boyfriend, Brian Nguyen, who has been a constant support throughout graduate school. I have shared many an intellectual discussion with him from which I have learnt immensely. I wouldn't have been able to get through the lowest moments during my PhD without his patience, love, care and support.

TABLE OF CONTENTS

CHAPTER 1

THE WEAK CHEMICAL INTERACTIONS THAT DRIVE PROTEIN EVOLUTION: CROWDING, STICKING AND QUINARY STRUCTURE IN FOLDING AND FUNCTION[1]

1.1 INTRODUCTION

In this thesis, I will consider a variety of biomolecular interactions, some *in vitro* and some in-cells. They range from hydrophobic disruption of protein structure (the aliphatic tail of dodine) to recognition of hydrophobic stretches (by Hsp70), as well as electrostatic perturbation of the protein or solvent (dodine headgroup). In some cases, the behavior is modulated by crowding (Ficoll effect on Hsp70-protein interaction), in other cases by sticking (lysate effect). In this chapter, I will review what is known about crowding and sticking in cells and *in vitro*, to set the stage for experimental and computational work presented in later chapters.

The cell was discovered in the 17th century by the Dutch scientist van Leeuwenhoek and observed in the simplest eukaryotic organism – a single-celled protozoan. Cells showcase many layers of biological organization, from membrane-bound organelles and phase-separated membrane-less 'liquid droplets' to the cytoskeleton. The structural organization of cells, from small solutes to organelles, is highly dynamic and far removed from the dilute buffers that are used in *in vitro* experiments.

The energetics of the cell cover a wide dynamic range, from interactions barely exceeding the thermal noise k_BT at room temperature, to strong chemical bonds. The protein and DNA backbones are made of stable bonds that are relatively hard to hydrolyze. Even the cellular cytoskeleton is held together by interactions on the order of 40 kJ/mol or 15 k_BT (the interaction free energy of tubulin dimers[1]). Although such specific and relatively strong interactions are easy to detect, weakly interacting (a few k_BT) and thus harder to detect multi-partner networks are more resilient towards malfunction (e.g. due to a mutation), and more flexible to evolve. While hubs in networks may be conserved, their many partners can evolve more independently.[2]

One of the hallmarks of complex networks is that many of their interactions are only on the order of a few k_BT. Weak interactions are numerous in cells, and can add up cooperatively, thus significantly affecting the

spatial and temporal distribution of macromolecules in the cell. To fully understand these weak interactions, they are best studied directly in cells or even in living organisms, or at least in a solvation environment that mimics those aspects of the cell important for the interaction in question. We have only just discovered the tip of the iceberg of how biological organization emerges from and affects dynamics at the molecular level. Moving forward, quantitative in-cell and *in vivo* experiments will become more important to study weak interactions that are sensitive to their solvation environment.

In this chapter, we discuss three concepts that can be used to classify weak interactions inside the cell; 'crowding,' 'sticking' and 'quinary structure.' Crowding is due to the short-range repulsive wall of interaction potentials with macromolecules in the cell (Figure 1.1). The repulsive wall is mainly due to physical forces not subject to evolution, although the size distribution of macromolecules that exclude volume could evolve. Sticking is due to longer-range attractive forces among a combination of macromolecular surfaces and small solutes within the cell. The effective forces can be electrostatic or entropic in nature, and can evolve, for instance by changing the charge or hydrophobicity of an amino acid side chain on a protein surface. Sticking is not necessarily favorable to cell health, but when it evolves into a favorable interaction, the resulting transient structure is referred to as 'quinary structure,' as a continuation of the hierarchical levels of evolved protein organization: primary, secondary, tertiary and quaternary structure.

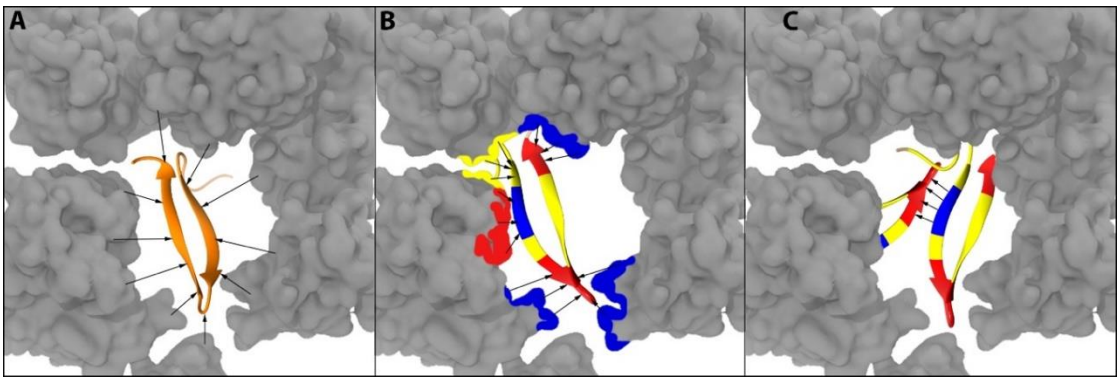


Figure 1.1: Cartoon representation of crowding, sticking and quinary interactions. The cytoplasm scaffold image was provided by Meredith Rickard. (A) Protein (orange) crowded by neighboring biomolecules (grey). (B) Ribbon structure of protein shows charged (negative – blue and positive – red) and hydrophobic patches (yellow). Patches on the surface of the protein interact with neighboring biomolecules via electrostatic and hydrophobic interactions. (C) Quinary functional interaction between two small proteins via surface electrostatic interaction.

A number of literature reviews have discussed in detail the experimental work and the current state-of-the-art theoretical models that describe macromolecular crowding[3,4] and quinary structure.[5–8] We do not seek to recapitulate the comprehensive work of these published reviews. Instead, the focus of this chapter

is to inspect the available data on chemical forces from the point of view of evolutionary pressures that have shaped protein (and other biomolecular) interactomes. Evolution produces complex interaction networks robust to perturbation, but it has to use chemical interactions outlined in figure 1.1 as basic building blocks. Our main goal is to emphasize the importance of collaborative studies in evolution and biophysical chemistry to understand how structural and functional aspects of the cell have evolved. Finally, to conclude we provide an outlook on the future and the scope of further experiments in post-reductionist chemistry of the cell.

1.2 WEAK INTERACTIONS IN THE CELL – A BRIEF HISTORY

In this section we discuss the three weak in-cell interactions (Figure 1.1) that contribute to organization near the k_BT level of thermal energy, namely crowding, sticking and quinary structure. All three of these can impact cell fitness; for example, crowding can enhance protein stability; sticking can lead to aggregation; and quinary structure of two enzymes can improve their substrate processivity. The terms macromolecular crowding and quinary structure were first introduced in the 1980s in the context of the cell. While the effects of crowding have been investigated rigorously *in vitro*, studies probing quinary structure have gained momentum only recently *in vitro* or *in vivo*. On the other hand, non-specific sticking has been known to exist since the earliest studies of protein aggregation.[9] However, its implications in cell have been observed only recently using in-cell NMR.[10]

1.2.1 MACROMOLECULAR CROWDING

The bio-macromolecular concentration inside an *Escherichia coli* cell can reach over 300 mg/mL.[11] The crowding inside eukaryotic cells can be even higher due to the cytoskeletal framework.[12] Macromolecular crowding was defined by Minton in 1983 as excluded volume effects exerted by the solutes in a solution.[13] In simple terms this means that solutes in a solution cannot overlap due to neighboring molecules' short-range repulsion, which is ultimately due to Pauli exclusion of atomic electrons on a length scale of ~ 0.1 nm.[14] About 30-40% of the cellular volume in eukaryotes is occupied by protein and nucleic acid molecules.[15] At these concentrations excluded volume effects can manifest as 1) macromolecular crowding: the volume excluded for one solute molecule by another and 2) macromolecular confinement: molecules confined to a smaller effective volume than the solution volume due to a high concentration of species. This situation can result in phenomena such as jamming, where macromolecular diffusion through interstitial spaces in the cytoplasm is severely hindered.[16]

Both crowding and confinement can have a significant effect on ubiquitous cellular functions such as association, activity, stability and conformation. For example, macromolecular crowding generally stabilizes proteins because the configurational entropy of the unfolded state is more severely reduced than

that of the native state.[17–19] (It is worth noting that unfolded proteins have smaller molar volume than folded proteins *in vitro*, hence proteins denature at high pressure.) Similarly, crowding can also increase association constants by favoring lower entropy, lower energy complexes.[20–23] Although the many features of the cell such as pH, ionic strength, osmolarity and redox potential can be accounted for by using suitable buffers, the main feature of biological macromolecules missing from traditional buffered *in vitro* experiments is their size. The last decade has seen an explosion of experimental data characterizing the effects of macromolecular crowding by using large inert polymers to account for macromolecule size. However, cellular components are far from inert and interact with each other constantly. As a result, simple crowders are not necessarily good mimics of the in-cell environment.[24] In the next section we discuss these interactions as an effect of macromolecules that is complementary to crowding.

1.2.2 STICKING

In 2016, Pielak and coworkers found that the stability of the small protein SH3 remained unchanged in an *E. coli* cell using in-cell NMR.[25] This observation could not be explained by the simplest version of macromolecular crowding that stabilizes the folded state due to excluded volume effects. Moreover, excluded volume effects are purely entropic, but enthalpic contributions are often seen as a consequence of crowding.[4,26]

Macromolecular crowding and its effects have been characterized rigorously *in vitro* using inert synthetic polymers such as Ficoll, dextran or PEG. However, the interior of the cell is far from inert. Proteins and RNA are charged and interact via electrostatics (screened charge or polar interactions), hydrogen bonding, and hydrophobic interactions (which are partly entropic due to enhanced ordering of water molecules near aliphatic or aromatic amino acid side chains). These forces can together be grouped into the non-specific forces that lead to macromolecular sticking inside a cell. In 2017, Oliveberg and coworkers showed that two mammalian proteins that tumble freely in mammalian cells get stuck in the bacterial cytoplasm.[10] On the other hand native bacterial proteins with the same fold tumble freely inside the bacterial cytosol. This is strong evidence that biomolecular surfaces are under evolutionary selection not just for residue-specific function, but also such that surface charge and hydrophobicity are optimized according to organism. For example, a freely tumbling and diffusing molecule can sample multiple potential binding partners in a short time. On the other hand, stickiness could evolve into a useful signaling interaction (see section 2.3). Either way it is clear that weak chemical forces in the cell can modulate stability,[27] activity[28] and diffusion.[29]

1.2.3 QUINARY STRUCTURE

The term ‘quinary’ was first used by Vaĭnshteĭn in 1973 to describe the fifth level of organization of proteins and nucleic acids in natural and synthetic aggregates such as those in viruses, chromosomes, molecular films etc.[30] In 1980, Edelstein also described quinary structure as being the fifth level of organization of protein subunits in helical lattices.[31] The term quinary structure as it is frequently understood today was introduced by McConkey, also as the fifth level of protein organization consisting of functional interactions that are weaker and more transient than quaternary protein structure.[32] Quinary structure can be hard to isolate or purify by harsh *in vitro* techniques. In McConkey’s definition, quinary structure is a likely reason why most cellular proteins evolve slower than expected: protein surfaces are subject to maintaining functional interactions with a few (on the order of 1 to 20) partners, while at the same time avoiding interactions with thousands of other types of macromolecules in the cell.

McConkey also noted that even though hard to replicate *in vitro*, in his day there were already examples of quinary structure formation among cellular constituents. These included i) the ribosome that transiently interacts with many factors and ii) some of the cytoskeletal framework that was destroyed by the separation methods of the time. We now know of many more examples of transiently interacting complexes inside cells. This has allowed us to understand the functional role of quinary structure as well as the range of thermodynamic and kinetic parameters that make an interaction inherently transient. Quinary structure requires a balance of both thermodynamic stability and kinetics. If two states are of similar stability but have a high kinetic barrier, then the system will get trapped in each state for a long time. Similarly, if one state is significantly lower in free energy than the other, then the system will mostly end up in the lower free energy state. When both barriers and stability differences are on the order of a few $k_B T$ on the free energy scale, transient association results. Small free energy differences and barriers make quinary structure highly susceptible to modulation by the cellular environment, and changes in the cellular environment that occur during the cell cycle or stress.

1.3 MACROMOLECULAR CROWDING

Over the years, the effects of macromolecular crowding on protein folding, assembly and other biological processes have been extensively characterized *in vitro* and *in vivo* by theoretical and experimental approaches.[33] Initial work by Zimmerman and Minton led to over two decades of considerable research in the field. A number of reviews have comprehensively described the existing literature.[3,34–38] In this section, we briefly highlight the basic principles of crowding and confinement and discuss *in vitro* crowding agents in experiments referencing earlier reviews where more detail is needed. We discuss in a greater detail some recent developments in the field of theoretical and computational studies that have facilitated all-atom

simulations of the cytoplasm. We finally conclude by detailing recent efforts in characterizing the state of crowding in the cell as well as the consequences of crowding in the pharmaceutical industry.

1.3.1 EXCLUDED VOLUME EFFECT

Excluded volume effects, a direct consequence of crowding by biological macromolecules due to repulsive interactions at short range, are perhaps the best characterized interactions that occur in the cell on the thermal energy scale.[39] Crowding by macromolecules can be visualized in the simplest approximation as hard spheres, nearly close-packed in a fixed volume. The interstitial space between the spheres can be occupied by small solvent or solute molecules such as water and metal ions, but it cannot be occupied by other large spheres present in the packed structure. The same constraints prevent the addition of any more spheres into the volume. This interstitial volume is then said to be excluded for these spheres.

In the above example, the volume available to species of any shape or size is limited simply by the impenetrability of the hard spheres in the volume. For real macromolecules, short-range repulsion has a finite range, which adds further add to the excluded volume (Figure 1.2A). Hence, the introduction of a particle from a free into a crowded environment leads to a reduction in entropy which, when scaled by temperature, is equal to the work done to insert the particle. Therefore, systems are driven to minimize the excluded volume to increase entropy.

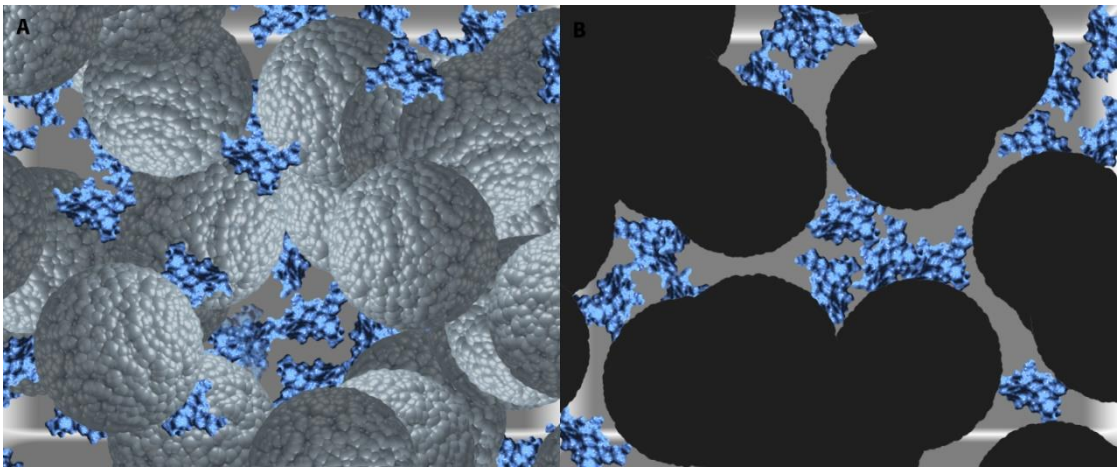


Figure 1.2: A qualitative representation of (a) crowding and (b) confinement. (a) crowding agent represented by grey balls with a model protein in blue. both crowder and protein can freely diffuse in solution. (b) protein in blue is confined to pores by a matrix shown in black.

Rigorous descriptions of crowding due to hard particles in a fluid have been published.[40–43] The effects of crowding on folded biological macromolecules are well described by such simple models that treat macromolecules as hard particles with an effective coarse-grained size and shape. For example, unfolded

proteins, can be treated as random chains which have a greater ability to thread through interstitial spaces as compared to folded proteins generally modeled as hard spheres. Models have been developed to mimic the effect that crowders have on biomolecule stability due to excluded volume.[44–46] Cheung and Thirumalai calculated that the change in melting temperature (T_m) for a protein with crowder volume fraction Φ_C can be estimated by:

$$\Delta T_m \approx 0.84 T_m(\text{Kelvin}) {\varphi_C}^{1.8} \quad [1.1]$$

The volume fraction Φ_C for spherical crowders with radii R comparable to the radius of gyration of the folded protein is

$$\varphi_C = \frac{4\pi}{3}(R)^3 \frac{M}{V}, \quad [1.2]$$

where M is the number of crowders and V is the volume of the smallest cubic box that can be drawn around the protein assuming a random coil denatured state.[44]

1.3.2 CROWDING *VS.* CONFINEMENT

Both crowding and confinement (Figure 1.2) are the consequences of macromolecular volume exclusion. Specifically crowding refers to the amount of free energy required to transfer a macromolecule from a dilute solution into a crowded environment. This is equivalent to the amount of energy expended to create a cavity large enough to accommodate the introduced macromolecule. Hall and Minton[36] showed that for a molecule "X" this free energy can be approximated by

$$\Delta F_X^{crowding} = -RTln\left(\frac{V_{availble,X}}{V_{total}}\right) = RTln(\gamma_X), \quad [1.3]$$

where $V_{available,X}$ is the volume available to a species "X" in a crowded environment, V_{total} is the total volume and $\gamma_X = \frac{V_{total}}{V_{availble,X}}$ is the activity coefficient of the species "X." Eq. [3] simply reflects the entropic cost of changing the available volume around a solute.

On the other hand, confinement refers to the free energy required to transfer a biomolecule from the set of configurations allowed in dilute solution to the set allowed in a bounded volume (such as a rigid cavity). Zhou and Minton[3] showed that for a molecule "X" this free energy change is given by

$$\Delta F_X^{confine} = -RTln(\frac{W_{allowed}}{W_{all}}), \quad [1.4]$$

Where, $W_{allowed}$ is the number of allowed configurations in a crowded solution and W_{all} is the number of all configurational states allowed in a dilute solution. Confinement is a somewhat more general concept than crowding, and reflects the entropic cost of reducing the number of configurations of a macromolecule, not just the cost of reducing the surrounding volume.

Crowding and confinement both limit the number of states macromolecules can occupy. Crowding favors more compact states of a molecule under a given set of solvation conditions. (However, the smaller molar volume of a random coil polypeptide relative to a native state will not generally lower its free energy sufficiently because of its reduced configurational entropy in a smaller free volume.) Confinement doesn't always prefer the most compact state. Rather confinement favors conformations that have shapes complementary to the shapes of the confining cavity. For example, spherical objects fit better in spherical pores whereas rod-like shapes fit better in a more cylindrical pore. The effect of confinement on both protein folding and association has been discussed in detail.[12,47]

1.3.3 *IN VITRO* CROWDING AGENTS

The effects of crowding have been characterized extensively *in vitro* using chemical agents to mimic a crowded environment. This section briefly discusses popular crowding agents and their advantages and disadvantages. Readily available polysaccharides such as Ficoll and dextran, or the poly-ether poly(ethylene glycol) (PEG) are the most popular polymeric crowding agents.[37] Each polymer is available in various molecular masses to mimic the size of different biomacromolecules. Both Ficoll and dextrans interact with proteins mostly via repulsive excluded volume effects. However, a number of studies have shown that for PEG this repulsive interaction is compensated to some extent by attractive interactions with hydrophobic and non-polar side chains on the protein surface.[48–52] For example, PEG seeks out mixed hydrophilic/hydrophobic patches on the protein surface and forms transient but recurring structure around them.[53] Thus even these simple crowders do not necessarily act like structureless spheres.

Crowding studies in general seek to mimic cellular congestion and the effect of congestion on other macromolecules. The intracellular space contains biomolecules such as proteins, DNA and RNA of various shapes and sizes. Most of the crowders described above are used because of the ease of availability and experimental manipulation. These non-biological agents are unlike the crowding biomolecules inside cells and might not, therefore, provide physiologically relevant information.[27,34,54] To probe the effect of biological macromolecular crowding, biomacromolecules such as proteins have also been used as crowders extensively.[20,22,23,28,54–56] Some of the commonly used proteins are bovine serum albumin (BSA),[20,23,55] SubL,[54] hemoglobin,[28] RNase A,[57] β-lactoglobulin[57] and lysozyme.[20] Numerous studies have shown that contrasting effects are observed for non-biological polymeric crowding

agents and biological crowders on a wide variety of reactions such as enzyme activity, protein refolding and denaturation.[28,56,58–60] The difference between a polymeric crowder such as Ficoll and a protein crowder such as BSA is, in most cases, explained as a consequence of the difference between their shapes or sizes. For example, Derham and Harding concluded that in protein crowders, the formation of higher activity oligomeric states due to the excluded volume effect showed an initial increase in enzyme activity.[28] However, in polymer crowders a reduction of activity was seen across all concentrations of crowders. They postulated that this is because the reduction in molecular diffusion is higher for polymers such as dextran or PEG as compared to globular protein crowders. This lower diffusion in polymeric crowders was thought to cancel out any positive contribution of oligomeric states to activity. Similar results were also observed for a small protein SubL on the unfolding of λ-repressor mutant 6-85,[56] and BSA on refolding of egg lysozyme[58,60] *vs.* non-biological crowders, the protein crowders being more effective.

In highly heterogenous environments, diffusion can be complex and is sometimes described as "anomalous diffusion". This occurs when the mean squared displacement, $\langle r^2(t)\rangle$, follows a power law of the form

$$\langle r^2(t)\rangle = 6\mathrm{D}t^{\alpha}, \qquad [1.5]$$

where α is degree of deviation from normal diffusion given by $\alpha = 1$ and D is a constant that does not depend on time. In recent years, several diffusion studies of crowding (and sticking) effects have been conducted *in vitro* by tracking the movement of a fluorescent tracer molecule using Fluorescence Correlation Spectroscopy (FCS).[61–63] These tracer molecules are either labeled with small fluorescent probe molecules such as Alexa488 and fluorescein or with a fluorescent protein such as EGFP. The main thrust of this research area is to characterize the diffusion behavior of macromolecules in crowded solutions. However, the results from these studies are contradictory, with some groups arguing for anomalous diffusion[63] and others for normal diffusion.[62] A recent study using hole-burning[64,65] of fluorescence intensity coupled with whole-cell imaging showed that anomalous diffusion best describes the overall flow of the proteins GFP and FRET-labeled PGK in mammalian cells, although a good semi-quantitative description of the diffusion of folded protein could be obtained by a normal diffusion model.

Even with these limitations, the diffusion behavior in heterogenous, crowded environments can shed light on the behavior of macromolecules inside the cell and the deviations from ideality reported in many crowding studies *in vitro*.

1.3.4 TOWARDS THEORETICAL TOOLS TO CHARACTERIZE CROWDING

Many theoretical models have predicted the effect of crowding based on statistical mechanics. Eq. [1] is an example. These statistical models, in most cases, assume that crowders are inert and do not interact, other

than through purely repulsive forces, with the macromolecules of interest. These theories have been extensively described and reviewed .[3,34,66]

The Zhou and Minton groups have used simple crowding models where both protein and crowder can be modeled by effective hard spheres or rods. The Zhou group used experiments as well theoretical calculations[67–69] based on scaled particle theory (SPT)[40] and Widom's particle insertion method[43] to probe the effect of crowding on protein stability[70] and binding[71] to another protein as well as on membrane proteins.[72] Using the SPT for hard spheres, Zhou predicted the free energy change due to crowding of a polymer chain as [69]

$$\frac{\Delta F_U^{crowd}}{RT} = -ln(1-\varphi) + 3\varphi Y^2\left(1 + \frac{1}{Y\sqrt{2\pi}}\right). \qquad [1.6]$$

The first term is similar to eq. [1], written in terms of the excluded volume fraction φ. Y is the ratio of the radius of gyration of the unfolded chain in dilute solution to the radius of the crowder. For large biomolecules or small crowders, the correction to simple excluded volume is thus bilinear in both excluded volume fraction and ratio of biomolecule to crowder size, and increases the crowding free energy further. The Minton group has also used statistical mechanics-based calculations relying on SPT to predict the effect of crowding.[36,45,46] Recently, Pielak and coworkers used SPT to predict the magnitude of the stabilizing effect of hard-core repulsion on two distinct shapes.[73] Crowders such as sucrose, BSA and lysozyme were modeled as hard sphere crowders using SPT and the magnitude of stabilization on an elongated dumbbell-shaped *vs.* a more compact dimer was calculated. In all their calculations the more compact dimer showed a higher degree of stabilization. Moreover, crowding effects were slightly destabilizing for highly elongated dimers, such as where the monomers are only touching. This reiterates the observation that the crowded cellular milieu prefers a more compact shape than an elongated shape and could be part of the reason why a globular protein like PGK is stabilized in the cytoplasm whereas the more elongated VlsE is destabilized.[24,27]

Moving away from these simple analytical models,[74] the Cheung and Thirumalai groups used molecular simulations to fold a small, fast-folding, model protein WW-domain in a crowded environment.[44] The Cheung group also conducted similar molecular simulations studies in collaboration with experimental groups of Wittung-Stafshede[74,75] and Waxham.[76,77] In 2012, Cheung and Wang used the all-atom cytoplasmic model of Elcock and McGuffee and developed an algorithm that coarse-grains slices of the cytoplasm.[78] They studied the thermodynamic properties of apoazurin in this coarse-grained cytoplasm and compared it to two models where each macromolecule in the cytoplasm is replaced by either (1) hard spheres of 55 Å, the size of Ficoll70, or (2) hard spheres of the same volume as the macromolecule. While

the hard sphere model allows heterogeneity in only the size, the coarse-grained cytoplasm has heterogeneity in both size and shape of crowding macromolecules. As was expected, the size of the crowding agent affects the predicted thermodynamic parameters; the stability of apoazurin was ~5 °C lower in Ficoll70 model than in either the variable-size hard sphere model or the coarse-grained cytoplasm. However, the variable-size hard sphere model could not adequately reflect the extent of structural fluctuations observed in the coarse-grained cytoplasm even though the average values of the thermodynamic parameters are similar in both. This concludes that not just crowder size but also crowder shape contribute to crowding effects.

These methods, while computationally more expensive, can in principle provide more accurate fitting models for experimental observations of kinetics and conformational changes induced by the environment.

1.3.5 CHARACTERIZING CROWDING INSIDE CELLS

Fluorescence is a tool used extensively to measure observables inside living cells, partly due to the ease of tagging cellular components with fluorescent labels, and its relatively non-perturbative nature in cells.[79] Fluorescence techniques such as FCS, Fluorescence Recovery After Photobleaching (FRAP) and Förster Resonance Energy Transfer (FRET) have been used to characterize crowding in cells. FCS has been used widely to measure diffusion to indirectly quantify crowding inside cells, for example via changes in diffusion.[29,80] Weiss and Guigas used FCS to measure the viscoelastic properties of Alexa488 labeled gold colloidal particles (5 nm diameter) in a variety of cell lines from various organisms and at different states of health. They then used viscoelasticity as a measure of crowding and surprisingly, found a lower degree of crowding in the nucleus than the cytoplasm despite the high DNA content. They also found, not surprisingly, that diffusion was size-dependent. For example, the colloidal particles or ~6 nm diameter BSA both show anomalous diffusion whereas much smaller species such as GFP diffused normally.

A newer area of research in the field is the development of molecular crowding sensors that enable direct measurement of crowding inside cells.[81,82] Poolman and Boersma designed FRET-based crowding sensors (Figure 1.3) that report on the crowding-induced conformational changes in the protein.[81,83] The sensors contain an α-helical peptide labeled with the FRET pair mCerulean3 and mCitrine at the N- and C-terminus, respectively. The protein adopts more compact high-FRET conformations in crowded environment and can be genetically encoded in various cell types. Using this sensor, the authors were able to calibrate the crowding in *E. coli* cells and found it to be ~180 mg/g with a volume fraction of 0.13. Poolman, Boersma and Liu later designed a set of sensors that are useful where in-cell calibration of the sensor is prohibited, for example during time-lapse measurements.[83] A set of nine probes was designed with the same FRET pair but varying linker sizes. Six of the nine probes showed efficacy for in-cell

measurements. Other sensors, including PEG-based ones developed by Ebbinghaus and coworkers, use similar FRET-based approaches to measure crowding.[39,82,84]

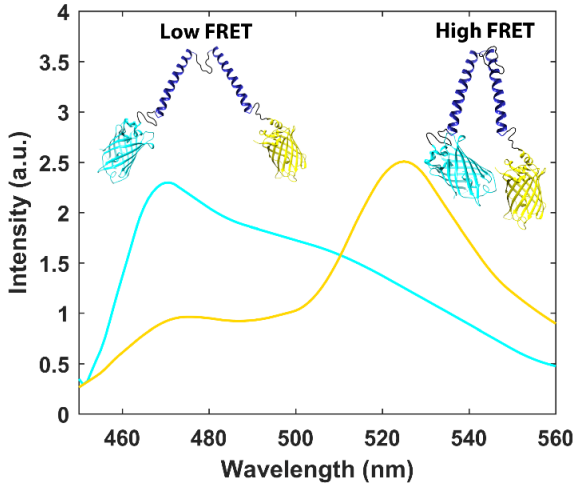


Figure 1.3: An in-cell crowding sensor designed by Poolman and Boersma highlighting the low- and high-fret states. Data was adapted from reference 79. The sensor consists of two helices connected via a flexible linker (inset) labeled with mCerulean at the N-terminus and mCitrine at the C-terminus. In dilute solutions the sensor adopts a low-FRET state where the two fluorescent tags are further apart (left inset), and fluorescence is dominated by mCerulean fluorescence (blue trace). In a crowded environment the two helices are pushed closer together adopting the high-FRET state (right inset). Fluorescence is dominated by mCitrine fluorescence due to mCerulean-mCitrine FRET (yellow trace).

The capability of some sensors to be genetically encoded enabled the measurement of free volume change inside cells due osmotic pressure.[85] These sensors could potentially be used to measure the heterogeneity of crowding in subcellular compartments via localization tags as well as on phase-separated membrane-less organelles. Such studies could yield key information about crowding inside the cell and its effects on modulation of biological processes.

1.3.6 CROWDING OUTSIDE THE CELL

The apparent stabilization of proteins in crowded environments has been harnessed by the pharmaceutical industry in protein- and peptide-based drug therapies. Even though protein drugs hold great promise, they are easily degraded by proteolytic enzymes, cleared by the kidneys, have a short circulating half-life and generate neutralizing antibodies. One of the most widely used methodologies to circumvent these problems is PEGylation, or the addition of PEG chains to proteins. The first evidence of improved drug delivery through PEGylation was shown by Frank Davis and his colleagues in 1970.[86] A detailed review of the use of PEG to improve drug delivery was recently published by Chess and Harris.[87] PEG interacts with proteins via transient interactions at the protein surface that can be specific in nature.[51,53] Gruebele and

Chao showed that when hooked to the protein surface, PEG preferentially interacted with specific patches on the surface of $\lambda_{6\text{-}85}$ irrespective of the attachment site thereby stabilizing particular secondary structures. Furthermore, PEG becomes structured near lysine residues that are surrounded by hydrophobic amino acids (Figure 1.4). This happens as PEG oxygen can hydrogen bond with lysine sidechains while the methylene groups interact with the hydrophobic amino acids. This strategy could be potentially useful in designing drug targets that require stabilization at particular sites or to improve efficacy of PEG to stabilize particular proteins.

Another area of research is to study proteins in gel matrices. Entrapment of biological molecules in gels have been used in many processes such as drug delivery,[88] sensors,[89,90] separation methods,[91] and microfluidics,[92] *etc*. Bulk measurements in gel matrices show stabilization for apomyoglobin[93,94] and lysozyme,[93] among others. However, here too the picture of inert crowding may not be as simple as with PEG: it was recently found that the behavior of a model protein, PGK, at the gel surface was not the same as in the bulk of the matrix.[95] It was also found that the interaction of the gel with PGK, and not degree of confinement, influenced protein properties in gels. Likewise, zwitterionic polymers, originally thought not to interact significantly with proteins, have been shown to affect protein stability and unfolded state compactness via interactions with the protein surface.[96] Gels, undeniably, have potential benefits in many applications relying on enhanced protein stability. However, their interaction with proteins need to be explored beyond simple crowding and confinement effects.

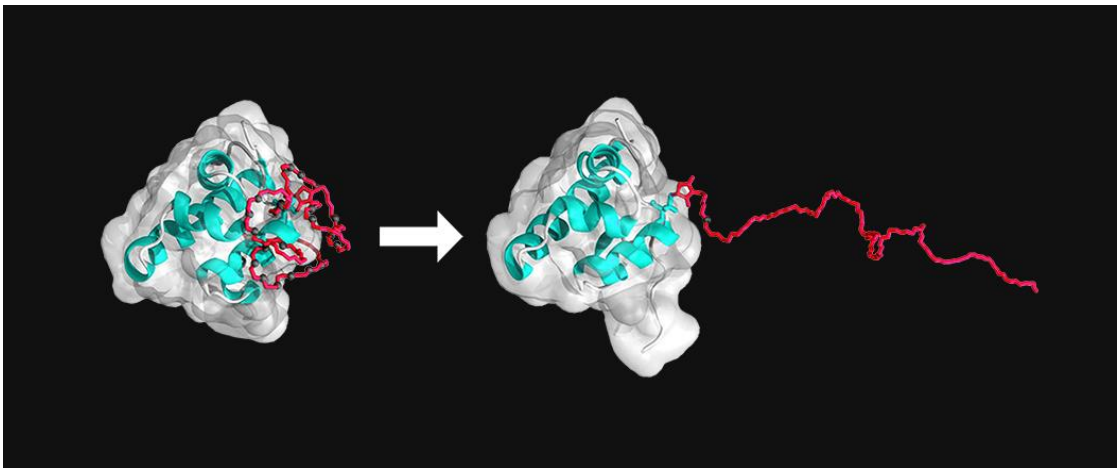


Figure 1.4: The interaction between PEG and $\lambda_{6\text{-}85}$. Figure adapted from reference 51. Figure shows PEG interacting with protein surface on the left and extended into solution on the right.

1.4 STICKING

In the cytoplasm, the packing of macromolecules causes what we simply refer to as crowding; we already alluded in the previous section to limitations of the simple steric picture. Crowding has been characterized

in vitro using inert macromolecules and polymers. However, crowding macromolecules in the cell include proteins and nuclei acids that contain charged, polar, and nonpolar patches. Proteins and RNA surfaces hence are not inert repulsive walls, and interact via electrostatics (charged and polar interactions), hydrogen bonding and hydrophobic interactions. These somewhat longer-range forces, sometimes attractive and sometimes repulsive, are the cause of stickiness of the cytoplasm. In this chapter we refer to these interactions collectively as "sticking" in the cell.

1.4.1 THE THERMODYNAMIC CONSEQUENCES OF STICKING

The change in free energy ΔG^0 of a protein has both entropic and enthalpic contributions according to the Gibbs-Helmholtz equation (N = native, U = unfolded)

$$\Delta G^0_{N\rightarrow U} = \Delta H^0_{N\rightarrow U} - T\Delta S^0_{N\rightarrow U}, \qquad [1.7]$$

and these can be separated by the temperature dependence of the free energy. Crowding manifests as excluded volume effects in the cell. As described above the contribution of excluded volume effects is purely entropic. However, enthalpic effects are often seen in presence of crowders. These enthalpic effects can arise from non-specific forces that include electrostatic attraction and repulsion.

In-cell NMR is a non-perturbative technique to characterize these forces inside the native environment of the cell.[97] One such case was reported for the SH3 domain of the *Drosophila* signal transduction protein using in-cell NMR. Pielak and coworkers studied the stability of SH3 by fluorine labeling of the tryptophan residue in *E. coli* cells and compared it to the stability in polymeric crowding agents.[25] As expected, SH3 was stabilized in Ficoll, dextran and PEG but inside cells no such stabilization was observed. On the contrary both T_m and ΔG^0_u either decreased or were unchanged as compared to buffer. Similar results were obtained in cell lysate where no stabilization was observed. These results point to a destabilizing effect that counteracts the entropic stabilization from crowding. Indeed, when polycationic protein crowders such as BSA and lysozyme were used with the polyanionic SH3, a destabilizing effect was also observed. The degree of destabilization reduced when NaCl was used to screen the charges or the overall positive charges on the protein crowders was decreased by altering the pH. These results indicate that weak non-specific attractive interactions can effectively counteract stabilizing entropic effects due to crowding.

Similar results were also observed for wild-type GB1 protein, where attractive interaction due to mutating a surface aspartic acid to lysine destabilized GB1 in *E. coli* cells.[98,99] Moreover, while this mutation was innocuous in buffers, the average effect seen in the cell is 10-fold larger. GB1 and *E. coli* proteins (on average) are slightly negatively charged, hence mutating a negatively charged aspartic acid to a positive

lysine changes the overall charge of the protein by +2 and therefore increases attractive interaction with other cellular proteins. Such charge reversal mutations have also proved useful to differentiate non-specific from functional interactions using in-cell NMR for other proteins.[100]

The stickiness of the cytoplasm can therefore negatively affect protein stability and function. These effects abound in the cell where the bio-macromolecular concentration is very high and where these macromolecules are not inert. Attractive interactions with other biomolecules can destabilize proteins because the number of favorable interactions can increase as the protein unfolds. As we will discuss later, stickiness in the cell caused by a random mutation could however evolve into new favorable quinary structure.

1.4.2 EVIDENCE OF STICKINESS *IN VITRO*

Several *in vitro* studies point towards the existence of these weak hydrophobic and electrostatic interactions and hydrogen bonding that cause sticking even outside the cell. The effects of sticking have been characterized by studying protein stability in buffers with globular proteins as crowders such as BSA and lysozyme.[97,101,102] In this scenario sticking is due to either attractive or repulsive interactions between charged biomolecules. Attractive interactions are destabilizing as they favor exposure of more surface which leads to unfolding. For example, negatively charged $SOD1^{I35A}$ (-0.5 e) and GB1 (-4 e) are destabilized in positively charged lysozyme (+8.5 e). Repulsive interactions stabilize the native state by reinforcing hard-core repulsion due to excluded volume. For example a negatively charged SOD1 dimer (-5 e) is slightly stabilized in BSA (-8.5 e) due to non-specific electrostatic repulsion.[102] However, in many cases the stabilization due to the net charge on the interacting proteins is easily overcome by non-specific localized attractive interactions between protein surfaces. Lysozyme (+8.5 e), BSA (-8.5 e) and an anionic lysate destabilize chymotrypsin inhibitor 2 (CI2).[97,103,104] Similar effects have also been seen for ubiquitin in BSA and lysozyme where the stability is either decreased or unchanged and for $SOD1^{I35A}$ in BSA where the stability is largely unaffected.[105]

Protein self-association due to sticking has been a major problem in the biopharmaceutical industry.[106,107] Aggregated proteins in pharmaceutical applications can have far-reaching consequences on patient health. Aggregated protein in the blood stream can persist and cause an immune response leading to the patient becoming immune to the drug or in worst-case scenarios acquiring an autoimmune disease.[108–110] The formation of weak transient non-specific complexes due to sticking precedes the formation of more stable protein aggregates. These protein self-associated transient complexes are difficult to isolate as stable intermediates and must be studied using a method that is sensitive to low-population states undergoing very fast assembly-disassembly kinetics *in situ*.

Paramagnetic relaxation enhancement (PRE) is one such technique. For example, the human growth hormone (hGH) is involved in many regulatory processes. GH deficiency (GHD) causes slow muscular development and stunted growth and is generally treated using hormone replacement therapy.[111] One of the main challenges in hGH therapy is the propensity of hGH to form soluble dimers, trimers and higher oligomers as well as insoluble aggregates.[112] Consequently the aggregation of hGH has been extensively studied.[112–114] Using PRE, Led and coworkers showed that hGH forms transient weakly associated complexes that give way to longer-lived aggregates as the concentration increases.[115,116] In order for aggregation to occur, multiple sites on the protein surface interact via weak non-specific interactions with other hGH molecules (K_d = 0.9 mM). Similar ultra-weak self-association ($K_d \geq 15$ mM) was also observed for an *E. coli* protein histidine-containing phosphocarrier protein (HPr).[117] These metastable oligomeric species could act as nucleation events for the formation of higher order aggregates such as amyloid fibrils or viral capsids. Transient oligomeric species are in rapid equilibrium with the monomer and can either be assimilated into a higher order aggregate or dissolve back into the solution. PRE measurements are a useful tool to characterize these early metastable intermediates that form due to sticking.

1.4.3 *IN VIVO* METHODS TO CHARACTERIZE STICKINESS

In-cell spectroscopy has gained considerable momentum in the last decade driven by the need to study biomolecules in their native environment. While *in vitro* experiments have been indispensable to our understanding of biomolecule structure and function, a large part of the interactome relies on weak interactions that are disrupted *in vitro* unless a native-like solvation environment can be painstakingly reproduced. In this section we discuss the two main methods to characterize sticking *in vivo*: 1) Fluorescence microscopy and 2) In-cell NMR.

Fluorescence microscopy is a time-honored technique to study biomolecule dynamics *in vivo*. The ease of availability of fluorescent labels as well as robust and flexible tagging methods have made fluorescence microscopy the method of choice for studying many different processes *in vivo*. These include but are not limited to diffusion,[29,80,118] binding[119,120] and stability[27,85,121,122] of biomolecules. Moreover, the availability of different fluorescence techniques such as Förster Resonance Energy Transfer (FRET), fluorescence lifetime imaging (FLIM),[123] fluorescence correlation spectroscopy (FCS),[61,124] or fluorescence recovery after photobleaching (FRAP)[125] can identify different aspects of the tagged biomolecule dynamics. An important outcome of the stickiness of the cytoplasm is the slowdown of biomolecules as they diffuse through the cytoplasm. This can be easily visualized by fluorescently labeling biomolecules and then tracking them as they diffuse through the cytoplasm.[126] Gruebele and Guo measured the diffusion of a GFP-labeled protein, phosphoglycerate kinase (PGK), in both the folded and unfolded state inside cells and saw anomalous diffusion.[29] They also observed that unfolded PGK

diffused slower than folded PGK in the cell. It was concluded that this slow-down could not be explained by the larger hydrodynamic radius of the unfolded protein chain alone, but required sticking of exposed hydrophobic patches to other cellular constituents.

While fluorescence techniques have been extensively diversified for in-cell applications, they are somewhat limited when conformational dynamics of biomolecules needs to be monitored. Most of these techniques also involve using fluorescent proteins that are bulky and could perturb the tagged biomolecule.[127] In the last decade, in-cell NMR has been developed to complement fluorescence microscopy.[100,128–130] Isotopically labeled proteins for in-cell NMR can be both over-expressed or microinjected directly in the cell. Generally, NMR line broadening is a signature of interaction between biomolecules and can be used as such in-cells to characterize both functional and non-specific interactions. Non-specific interactions are easy to visualize when exclusively eukaryotic proteins are studied in prokaryotic organisms such as *E. coli*. In this case most functional partners for the protein of interest are absent and only non-specific interactions occur.[100] In-cell NMR and mutagenesis have been used to predict the role of specific surface charges to non-specific electrostatic interactions as explained in section 4.1.[97–99,101] Gierasch, Wang and Zhuravleva showed that these non-specific interactions are not modulated by a single variable, but are a function of several factors like overall charge, distribution of hydrophobic patches and conformational flexibility.[99]

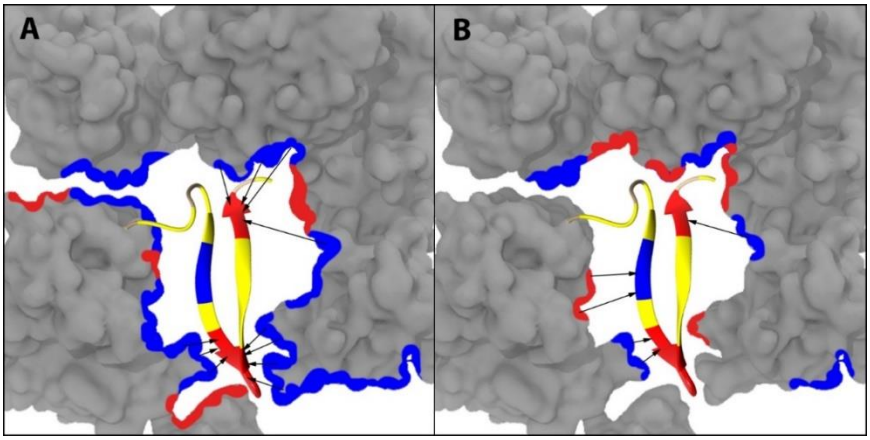


Figure 1.5: Cartoon representation of sticking in a (A) bacterial vs. (B) mammalian cytoplasm (grey). Arrow lengths signify interaction strength, longer arrows show weaker interactions and vice versa. Cytoplasm scaffold image was provided by Meredith Rickard. (A) A mammalian protein in a bacterial cytoplasm. Many bacterial proteins are negatively charged and proteins from other organisms such as mammals may get stuck and show slowed diffusion due to attractive electrostatic interactions. (B) A mammalian protein in a mammalian cytoplasm. A mammalian cytoplasm has a distribution of charges that is evolved to be compatible with native mammalian proteins. The cartoon shows a smaller number of negatively charged patches that reduces sticking.

In most cases, sticking has a negative consequence for protein stability or function. For example, ProtL,[131] SOD1,[102] and GB1-D40K[98] are destabilized in-cells as compared to buffer. This is evidence that entropic stabilization from crowding can be overcome by destabilization due to sticking. Furthermore, there is also evidence that sticking can lead to disruption of functional interactions. Oliveberg and coworkers showed that two mammalian proteins HAH1 and SOD1 diffused freely in mammalian cells but seemed to get stuck in the bacterial cytoplasm (Figure 1.5).[10] A freely diffusing protein can form functional interactions with its partners. A slowdown such as the one observed for SOD1 and HAH1 in *E. coli* can reduce the sampling of such interactions.

In-cell measurements are just starting to scratch the surface of the widespread non-specific interactions that are abundant in the cell. The usability of in-cell NMR is somewhat limited to studying eukaryote-specific proteins in prokaryotes. Eukaryotic proteins that are conserved in prokaryotes pose a challenge because very strong functional interactions can broaden lines beyond detection. However, in-cell NMR has been successful in eukaryotic systems by mutating out residues in the binding pocket to reduce broadening due to strong interactions. For example, Shirakawa and coworkers were able to record in-cell NMR spectra of a ubiquitin derivative (Ub-3A), whose affinity to cytosolic proteins was reduced by making three point mutations in the binding interface.[132] While wild-type ubiquitin showed peak broadening due to interactions with endogenous proteins *in vivo*, Ub-3A showed well-resolved peaks. Hydrogen exchange in combination with NMR also showed that binding to proteins in the cell destabilize wild-type ubiquitin *in vivo* compared to *in vitro*. Interestingly, even though Ub-3A has lower binding affinity, it is still destabilized *in vivo* due to sticking.

1.4.4 TOWARDS CREATING A STICKY CYTOPLASM USING THEORETICAL MODELS

A new area of theoretical advancement is the explicit coarse-grained or even atomistic treatment of protein-protein interactions in crowded environments. In 1996, Field and Bicout published the first research study on modeling the cytoplasm of an *E. coli* cell.[133] The Field version of the cytoplasm consisted of three particles, ribosomes, proteins and tRNA, modeled as spheres. The interactions between these particles consisted of short-range Lennard-Jones and long-range electrostatic terms. In 2008, Ellison and coworkers developed another model of the bacterial cytoplasm which also represented proteins as spheres.[134] This model was an important step forward from the model of Field and Bicout in that the model included >100 proteins at physiologically relevant concentrations. However, due to the limitations of this model it could not reproduce the *in vivo* diffusion rate of GFP accurately. Contrary to the 10-fold decrease in diffusion *in vivo*[135] this model only produced a 2-fold reduction in diffusion rate. The authors hypothesized that since the model's prediction was based solely on steric repulsion due to excluded volume between

macromolecules, there are perhaps other effects inside cells that need to be accounted for to get more accurate results. As we will see later, sticking due to attractive potentials is one such effect.

In recent years, other more realistic models for crowding and sticking have been developed. In 2010, Elcock and McGuffee developed an all-atom model of the *E. coli* cytoplasm (Figure 1.6). The model consists of the 50 most abundant macromolecules of the *E. coli* cytoplasm, 45 of which are proteins. Moreover, this model considers two of the most common types of interactions in the cell *i.e.* hydrophobic and electrostatic. With these additional considerations Elcock and McGuffee were able to provide a quantitative rationalization of the destabilization of CRABP *in vivo* and $\lambda_{6\text{-}85}$'s unchanged stability *in vivo* (*vs. in vitro*). In these cases, sticking of the unfolded protein counteracts the crowding-enhanced stability of the native state. However, this all-atom cytoplasmic model is very computationally intensive and still limited in its uses for simulating cellular phenomena.

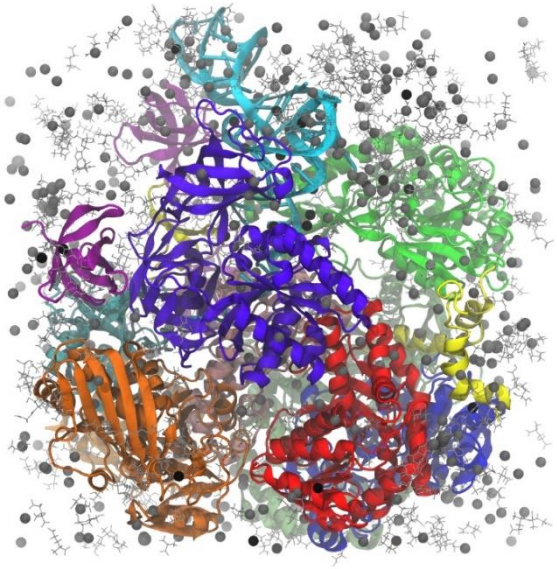


Figure 1.6: A snapshot showing proteins packed into a theoretical model of the bacterial cytoplasm. Image was provided by Dr. Taras Pogorelov and Meredith Rickard. The model consists of the most abundant cytosolic proteins, metabolites, ions and water molecules in an E. coli cell. Such a model allows for both crowding and sticking via protein surface charges.

By 2015, Feig and coworkers presented another atomistic model of a cytoplasm based on *M. genitalium*.[136] This model of the cytoplasm is essentially complete and consists of all the components necessary for protein translation, folding and degradation in addition to the metabolic core functions. Using this methodology Feig and coworkers were able to construct complete metabolic pathways in a cytoplasmic subsection. With the increase in computational power in leaps and bounds over the past decade, the

drawbacks of these all-atom simulations owing to their time-intensive nature could be resolved in the near future and allow for the simulation of biological reactions in a highly detailed model of the cell.

1.5 QUINARY STRUCTURE

The term 'quinary' was used by McConkey to describe transient functional assemblies present in the cell.[32] Proteins show four main levels of organization: primary (amide bond formation, $-\Delta G = 8\text{-}16$ kJ/mol in solvent), secondary (hydrogen bonds during α-helix and β-sheet formation, $-\Delta G < 1$ kJ/mol in solvent), tertiary (disulfide bond formation, hydrophobic core, salt bridges etc., $-\Delta G = 0\text{-}0.1$ kJ/mol/residue in solvent) and quaternary (hydrogen bonding, electrostatics and hydrophobic interactions in protein oligomerization, $-\Delta G = 40\text{-}60$ kJ/mol). Quinary structure is the 5th level of protein structural organization, where proteins interact weakly ($K_d > 1$ μM) and form short-lived functional complexes in the cell.[85] Quinary structure is characterized by low thermodynamic stability and a low kinetic barrier, but unlike sticking, it has useful functional consequences that improve cell health. The electrostatic, hydrophobic and other interactions that underlie quinary structure formation are referred to here simply as 'quinary interactions.'

In this section we briefly describe the advances in quinary structure determination and our current knowledge of the existence of quinary structure formation in biological systems. We briefly describe the concept of the metabolon. The metabolon is one the most well-characterized instances of quinary structure formation in cells. In the last decade technical advancements have led to the discovery of many more examples of quinary structure. We describe in detail two such cases where quinary structure may play key roles in regulating biological function: 1) phase separation in living systems and 2) formation of encounter complexes.

1.5.1 WHAT IS AND ISN'T QUINARY STRUCTURE?

For an interaction to be quinary it must satisfy three conditions: 1) low stability of the complex, 2) rapid kinetics of dissociation/association and 3) interaction should confer some functionality. For a very high kinetic barrier the interaction wouldn't be transient, *i.e.* the complex may get trapped in one state for a long time. Similarly, for a highly stable state, the system will more often end up in that state with the interaction being more long-lived than transient. Both above-mentioned scenarios lead to tightly bound stable complexes instead of transient quinary structure. And if there is no significant function, the interaction is merely 'sticking'. These features make quinary structure highly susceptible to disruption by *in vitro* biochemical separation methods.

In evolutionary terms, it is possible that sticking evolves into functional quinary structure, which then evolves into more specific, stronger interactions. The difference between sticking and quinary structure is not the energy scale, but that the former is frequently a destabilizing force without any functional contribution, whereas the latter endows functionality. One possible reason that most quinary structure has not evolved greater specificity is that proteins are involved in networks with multiple binding partners. A protein's surface is finite in extent, and can only accommodate so many binding partners: either a few strong ones (larger surface area occupied by each binding interaction) or more weaker ones (smaller surface area occupied by each of the binding partners). In terms of information theory, the surface of a protein can only encode a certain number of bits of information, and these must be divided up among multiple binding interactions in a zero-sum game. If we take an amino acid surface area $A_A \approx 10$ Å^3 and a protein surface area A_P, and assume that either +, -, polar or non-polar can be encoded on each amino acid patch (four states), then the total information is $\sim(A_P/A_A)^4$, or in terms of an information entropy (proportional to number of bits because $\ln[x] = \ln_2[x]/\ln[2]$),

$$S_{quinary} \sim 4R ln[\frac{A_P}{A_A}]. \quad [1.8]$$

Quinary structure, due to its low stability, is highly susceptible to the environment. Cellular properties, such as whole proteome pI values, are highly organism-specific. For example, anionic proteins are more abundant in *E. coli* at physiological pH.[137–139] Since quinary structure is highly sensitive to the local environment, it is possible that they are also organism-specific. In-cell NMR studies have proved useful in probing the energy scale of weak interactions in cells.[140] However, many of these studies rely on the examination of eukaryotic proteins in prokaryotes to reduce peak broadening due to strong interactions that may occur in the native eukaryotic cytosol (Figure 1.5).[100] Due to its organism specificity, this strategy could disrupt quinary structure associated with the protein of interest that are present in the eukaryotic cytosol. For this reason, in-cell NMR studies of eukaryotic proteins in *E. coli* have been classified as sticking and described in Section 4.3.

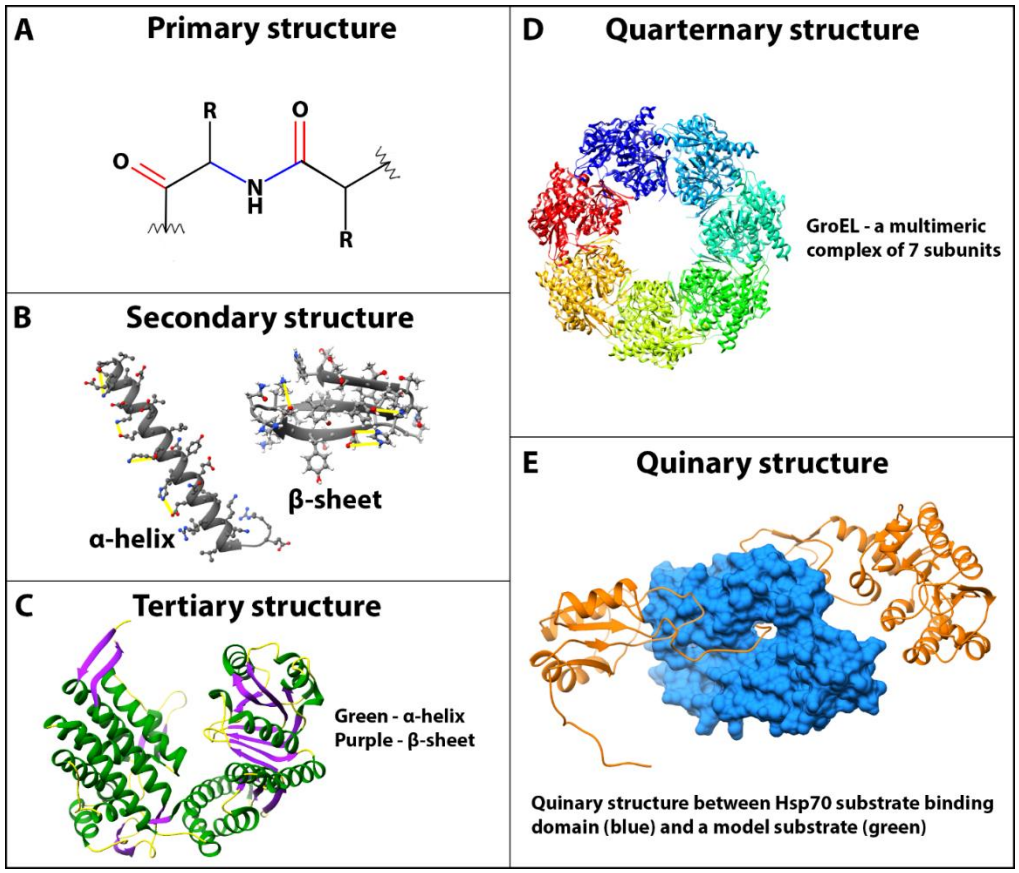


Figure 1.7: The different levels of organization of in protein folding. (A) Primary structure consists of backbone amide bonds. (B) Secondary structure shows two example folds, α-helices and β-sheets held together by hydrogen bonds shown in yellow. (C) Tertiary structure shows a fully folded GroEL monomer (PDB ID: 1SS8) consisting of both α-helices and β-sheets. (D) GroEL (PDB ID: 1SS8) – a multimeric complex made of 7 monomers, an example of quaternary structure. (E) Quinary structure (weak binding) between Hsp70 substrate binding domain (blue, PDB ID: 2KHO) and a model substrate phosphoglycerate kinase (orange, PDB ID: 3PGK).

It is important to differentiate between sticking and quinary structure since both have similar energetics. We also do not refer to strong, long-lived interactions, such as those underlying the hemoglobin tetramer or GroEL formation, as quinary, but rather as quaternary (Figure 1.7).

1.5.2 THE METABOLON

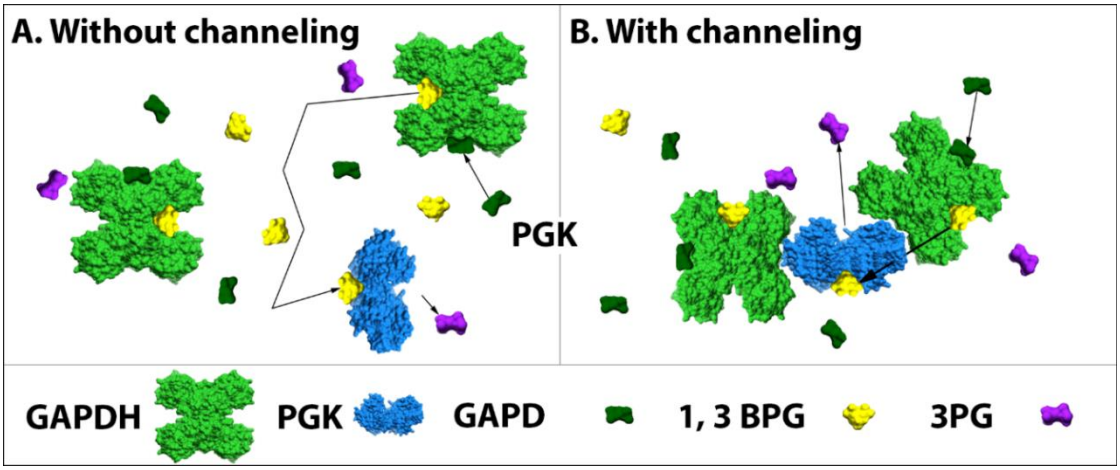


Figure 1.8: Cartoon showing substrate channeling between GAPDH and PGK during glycolysis. GAPDH (green) and PGK (light blue) were accessed by PDB IDs 1IHY and 3PGK. GAPD (dark green), 1,3 BPG (yellow) and 3PG (purple) were accessed by PubChem numbers 729, 683 and 439183. All structures were rendered using UCSF Chimera. (A) Without channeling the substrate 1,3 BPG must diffuse after being produced by GAPDH before being bound to PGK. (B) During channeling the substrate 1,3 BPG is channeled between both enzymes.

Perhaps one of the best representative examples of quinary structure formation is the metabolon. The word metabolon was coined by Paul A. Srere in 1985.[141] Quoting the 1985 communication by Srere, a metabolon is a 'supramolecular complex of sequential metabolic enzymes and cellular structural elements'. This supramolecular complex is formed of many enzymes associating into quinary structure and increases reaction efficiency through substrate-channeling (Figure 1.8).[142–144] Metabolic pathways generate many intermediates, a majority of which have no specific function other than to be fed into the next reaction in the sequence. During substrate-channeling, substrates are prevented from escaping into the bulk cytoplasm by efficiently channeling them to the next processing enzyme in the supramolecular complex. This dramatically accelerates reaction rates by 1) avoiding the time delay for the enzyme or substrate to diffuse in the cytoplasm to encounter one another, and 2) allowing the enzyme to compete for relatively low copy number substrate molecules before they decay into side products.

Metabolons exist in many metabolic pathways, including fatty acid oxidation,[142] amino acid metabolism,[145] glycolysis,[146] lipid biosynthesis[142] and the tricarboxylic acid (TCA) cycle.[144,147] The TCA cycle, also known as the Krebs cycle, is an essential metabolic pathway that generates the energy rich molecule ATP through aerobic respiration in cells.[148] The Krebs metabolon has been studied in great detail and consists of eight enzymes forming a supramolecular complex via quinary interactions.[143,144,147,149] Theoretical modeling showed that on association the charge patters on the enzyme surfaces rearrange to create continuous positively charged zones.[143] This allows substrate-

channeling of negatively charged substrates from one enzyme to another across the positively charged surface. An analytical equation was developed to characterize the effect of substrate-channeling on the kinetics of a bi-enzyme complex. Assume a simple coupled reaction scheme with enzyme E_1 and E_2:

$$S \xrightarrow{E_1,\ V_0} I \xrightarrow{E_2,\ K_m,\ V_{max}} P \qquad [1.9]$$

where, the substrate S is converted to intermediate I by enzyme 1 that operates at the constant velocity V_0 which is then converted to product P by enzyme 2 with Michaelis constant K_m and maximum attainable velocity V_{max}. For this bi-enzyme system following pseudo first-order kinetics the time required for the intermediate I to build up to sufficient levels so as to maintain steady state flux is called the transient time, τ. The variable τ can be used to characterize the effect of efficient channeling of substrate by the metabolon as substrate-channeling leads to dramatic decrease in the transient time. By building on older models by Easterby[150] and Ovádi,[151] Elcock and coworkers formulated, using an analytical approach, the following general equation for the calculation of τ:[152]

$$\tau = \frac{K_m(1-p_c p_r)}{V_{max}} \qquad [1.10]$$

where, p_c is the channeling probability that describes whether the intermediate I is successfully transferred to the next enzyme and p_r is the probability that complex formation successfully leads to formation of product P rather than dissociating to reform the intermediate. Using this formula, the transient time for two representative enzymes, citrate synthase (CS) and mitochondrial malate dehydrogenase (mMDH), in the Krebs metabolon was determined to be 0.03 secs and 2.5 secs with and without channeling, a nearly two orders of magnitude difference in efficiency.[143] In the above example the importance of substrate-channeling is evident in the ~100-fold reduction in τ. This is particularly important in the case of CS/mMDH because the forward mMDH reaction has an unfavorable equilibrium constant and their intermediate, oxaloacetate, occurs at concentrations that are too low to sustain experimental reaction rates.[153,154]

Surprisingly, the quinary structures involved in substrate channeling have rarely been imaged directly in cells. One recently developed technique uses cell volume modulation in response to osmotic pressure to modulate enzyme association.[85] Using osmotic pressure Gruebele and coworkers were able to characterize the K_d of two weakly associating enzymes, $GAPDH_4$ and PGK, both members of the glycolysis metabolon. The K_d for $GAPDH_4$-PGK oligomerization was calculated to be ~14 μM in U-2 OS cells. Moreover, the authors showed that two GAPDH tetramers associated with one PGK. It is known that

GAPDH copy numbers in-cells is at least twice as high as PGK.[155] Therefore, the formation of the $GAPDH_4$-PGK-$GAPDH_4$ ternary complex may help GAPDH compete for cellular levels of PGK and in the efficient transfer of bis-phosphoglycerate substrate from GAPDH tetramer to PGK during glycolysis (Figure 1.8). Other studies have also shown substrate channeling in the glycolysis metabolon. Molecular dynamics simulations by Barton, Minteer, Sigman and coworkers demonstrated substrate channeling between another enzyme hexokinase (HK) and G6PDH (glucose-6-phosphate dehydrogenase) in the glycolytic pathway.[156] The interaction between G6PDH and HK precede the ATP generation step involving PGK and GAPDH. Taken together, these results potentially emphasize the channeling of multiple substrates between enzymes during glycolysis.

The benefits of substrate channeling are clearly visible in many other biological processes. For example, during protein synthesis, multi-synthetase complexes provide amino acid substrates to the ribosome in the form of aminoacyl-tRNAs. These aminoacyl-tRNAs are channeled directly from aminoacyl-tRNA synthetases to elongation factor to the ribosome without diffusing into the bulk cytoplasm.[157,158] Such quinary structure has resisted characterization as it is disrupted by *in vitro* purification methods.[158] The many regulatory functions of the aminoacyl-tRNA synthetase complexes have been discussed in detail in published reviews.[158]

Shakhnovich and coworkers showed that there is evolutionary pressure to develop interaction networks that support substrate channeling and select against non-functional interactions.[159] The authors investigated in detail the reasons behind the gene dosage toxicity (GDT). In simple terms GDT is a phenomenon where overexpression of certain genes has toxic effects. Using the enzyme DHFR (dihydrofolate reductase) the authors showed that imbalance of protein-protein interactions due to overexpression is the result of toxicity during DHFR overexpression. One of the key findings revealed by these results was that evolutionary selection shapes protein-protein interactions to facilitate the formation of metabolons that support efficient substrate channeling. These interaction networks are highly organism specific such that when *E. coli* DHFR was replaced by a foreign DHFR, promiscuous mis-interactions increased indicating that there is selection pressure against such promiscuity.

Quinary structure is thus an important biological organizing principle ensuring that reactions in the cell proceed efficiently. Weak interactions allow supramolecular complexes to be assembled and disassembled quickly in response to cellular signals and act as effective biological switches. With the advent of better *in vivo* imaging techniques, the observation of metabolon formation in many other biological compartments may surface in the future.

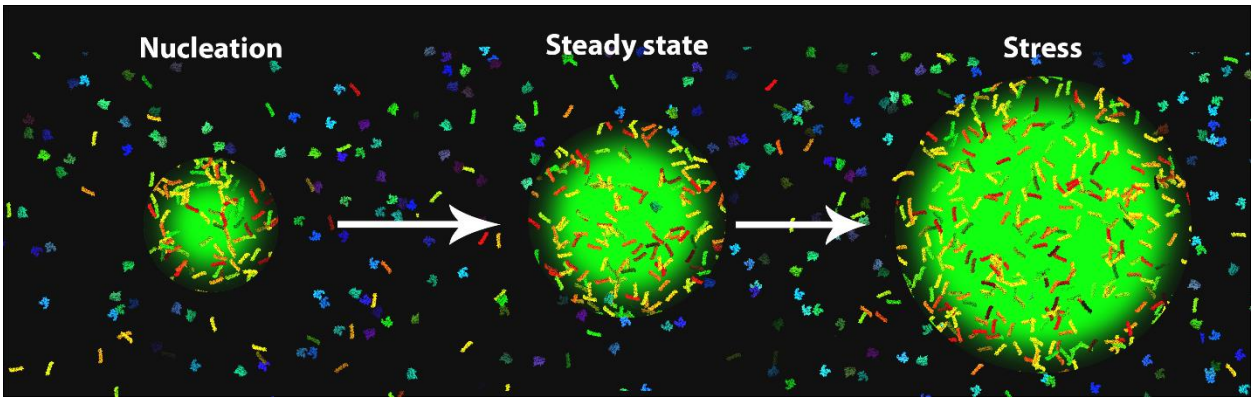


Figure 1.9: Cartoon representation showing the three steps in the formation of phase separated droplets in the cell. During nucleation a set of constituent proteins are recruited and form a small droplet. The droplet grows and reaches steady state. However, if stress is applied the droplet growth increases as a stress response mechanism to protect or shield its constituents from unfolding/misfolding.

1.5.3 THE ROLE OF QUINARY STRUCTURE IN CELLULAR ORGANIZATION

A single quinary interaction may not confer a lot of additional functionality or stability to the cell, but in large numbers they can add up and result in robust networks of interactions. Weaker interactions allow a protein to interact with more partners by using fewer bits of the recognizable information stored on a protein's surface. Such weakly interacting networks are more likely to be robust against deletion of any one component or interaction.[2]

In extreme cases, the highly crowded cellular environment can even lead to phase separation of biomolecules due to quinary interactions, similar to that observed in saturated solutions (Figure 1.9). This creates microenvironments, also referred to as membraneless organelles or 'liquid droplets,' with defined functionality and specific composition: examples include Cajal bodies, stress granules, nucleolus, P-bodies and paraspeckles.[160] The composition of these membraneless organelles typically ranges from a few to several hundred protein species or RNA molecules.[160]

These microenvironments are considered quinary structure due to two main reasons: 1) The interacting molecules are highly dynamic, showing liquid-like properties where molecules dynamically exchange with the surrounding environment and many components are recruited only transiently in response to certain stimuli and 2) they are functional and improve cell health.[161–163] At the center of this phenomenon are multivalent interactions between groups of binding partners often involving highly charged disordered proteins (IDPs) and RNA molecules.[164,165] Many of these quinary interactions occur in the low-complexity regions (LCRs) in IDPs that consequently have been shown to be modulators of phase separation in cells.[166] Quinary interactions has also been implicated during stress-induced phase separation[166,167] and reversible aggregation of endogenous proteins.[168]

The biochemical environment in these phase-separated regions has important functional implications. Phase separation can affect reaction kinetics and specificity by substantially increasing the local concentration of the reactants. For example the rate of mRNA processing is significantly reduced when key components fail to concentrate within the histone locus body or Cajal bodies in zebrafish.[169–171] Phase separation could also inhibit activity by sequestering molecules in these regions or act as an on/off switch where functions can be rapidly regulated through the formation and dissolution of the condensed phase.[160] For example key components of the protein synthesis machinery are sequestered inside stress granules during heat shock in *Saccharomyces cerevisiae*.[167] Finally, by either releasing molecules from the condensate or recruiting molecules into the condensate, phase separation can help to maintain stable levels of molecules in the bulk phase despite fluctuations in expression.[160]

The eukaryotic cytoskeleton also exhibits many instances of quinary structure formation with K_d in the ~μM regime. In yeast, *Saccharomyces cerevisiae*, evidence suggests that actin and microtubule cytoskeletons may function together during mitosis and in mating cells.[172] Barnes and coworkers found that a component of the actin cytoskeleton, coronin 1p (Crn1p), provides a functional link between the actin and microtubular cytoskeleton in yeast.[173] Crn1p interacts only weakly with microtubules with a K_d of 15-20 μM. However, in the presence of actin, the K_d for Crn1p and microtubule association increases 10-fold. This indicates that quinary structure formation (K_d – 2 μM) via Crn1p crosslinks actin filaments and microtubules in mitotic and mating cells. Another example of quinary structure is evident for the cofactor, dynactin, of the cytoplasmic dynein-1 motor that transports cargos along the microtubular cytoskeleton.[174] The microtubule binding domain of a dynactin subunit interacts with microtubules with a K_d of 10 μM.[175] These cytoskeletal quinary structures abound in the eukaryotic cytoplasm and provide structure and shape to the cell.

1.5.4 QUINARY STRUCTURE AND ENCOUNTER COMPLEXES

Phase separation is not the only phenomenon in biological systems that is driven by quinary structure. Biological systems need regulatory switches for a large number of processes. These switches must be highly sensitive and specific and signaling must be performed with a high fidelity to ensure proper function. Most signaling pathways consists of the formation and dissolution of multicomponent complexes consisting of proteins, nucleic acids and other small molecules.[176] These complex must not only associate with high specificity but also dissociate when the signal is turned off. A tightly bound complex that does not dissociate cannot be turned off.

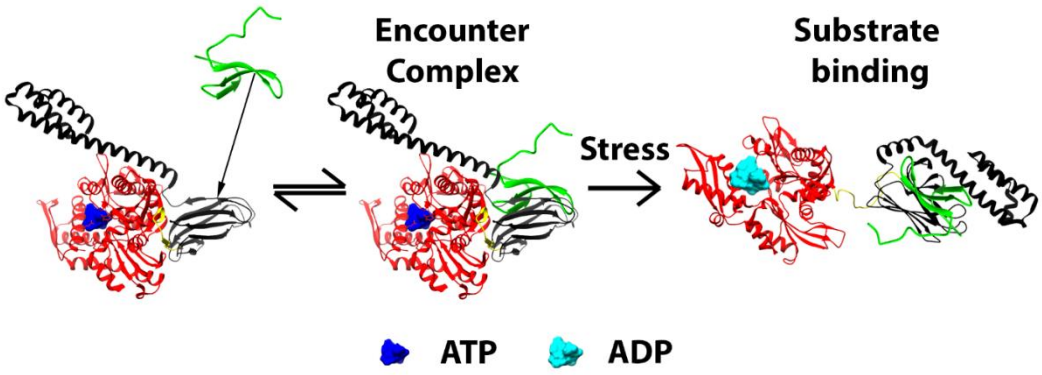


Figure 1.10: Formation of encounter complexes leading to binding in the Hsp70 system. Red and black show the N-terminal and C-terminal nucleotide and substrate binding domain respectively. In the ATP bound state Hsp70 forms transient encounter complexes with potential substrate proteins. During stress ATP is replaced by ADP which facilitates conversion of the encounter complexes to productive stably bound substrate-chaperone complex. Hsp70 molecules were rendered using UCSF Chimera and accessed using PDB IDs (ATP bound state – 4B9Q and ADP bound state – 2KHO).

The cell solves this problem using encounter complexes (Figure 1.10).[177] In simple terms, biomolecules diffuse freely and form weak transient encounter complexes via collisions with a high k_{on} and k_{off} rate. These weak complexes can then reorient, reposition or undergo conformational changes to give rise to stronger and more specific interactions with a low K_d. These weak transient complexes are functional and signaling pathways that lead to their formation are possibly formed through evolutionary selection under which non-functional interactions evolve into productive interactions. In addition to evolved functionality, the K_d for these interactions are in the range of 10s of μM[178] and hence we group them under quinary structure formation for the purpose of this chapter. In effect, quinary structure can be a short-lived precursor for stronger binding.

Such a mechanism is utilized by the molecular chaperone 70 kDa heat shock protein (Hsp70). Hsp70 is maintains cellular proteostasis by binding unfolded, misfolded or nascent chain peptides and preventing further unfolding (Figure 1.10).[179,180] In a normal cell, Hsp70 is in the ATP bound state where it binds substrates with a high k_{on} and k_{off} rate.[181] Substrate binding then induces ATP hydrolysis which in turn leads to a conformational change in the C-terminal domain of Hsp70. This conformational change results in a more tightly bound Hsp70-substrate complex.

Paramagnetic relaxation enhancement (PRE) is a useful method to detect low-population (<10%) encounter complexes. To do this, paramagnetic labels are generally introduced outside the specific interaction site. Clore and coworkers showed that the encounter complexes formed due to quinary interactions are important

intermediates that increase the rate of formation of specific functional interactions by reducing the dimensionality of the search process.[178,182] In the bacterial signaling system that catalyzes phosphorylation reactions (phosphotransferase system) the authors were able to show evidence of the formation of these encounter complexes using PRE. They observed these rare, fast-exchanging complexes for the N-terminal domain of enzyme I (EIN), $IIA^{Mannitol}$ and $IIA^{Mannose}$ with the phosphocarrier protein (HPr) with $K_d \approx$ 10-50 μM.[178] These encounter complexes are formed due to longer-range electrostatic and shorter-range van der Waals attraction and have been discussed in detail in published reviews.[183,184] In the case of the bacterial phosphotransferase system, the negatively charged residues on EIN, $IIA^{Mannitol}$ and $IIA^{Mannose}$ interact with the positively charged surfaces of HPr. In the case of EIN-HPr, a small population of a ternary encounter complex (HPr-EIN-HPr) was also observed.[185] This second type of encounter complex occurs predominantly when the active site of Enzyme I is occupied, and possibly helps in efficiently reloading the enzyme active site the moment it frees up, as well as competing for the cellular pool of HPr.

A similar ternary complex was also observed for protein-DNA binding.[182] In the nucleus, DNA is present at mM base pair concentration. Under nucleus-like experimental conditions involving sub-millimolar free DNA, Clore and Iwahara showed that the association of protein and DNA occurs via a ternary encounter complex where free DNA associates with a DNA-protein bound complex rather than the association of free protein with free DNA. The formation of the ternary complex can accelerate target recognition rate in protein-DNA interactions resulting in translocation rates that are up to three orders of magnitude faster than the *in vitro* protein-DNA dissociation rate for a transcription factor, HOXD9 homeodomain. This phenomenon simultaneously explains the highly dynamic nature of protein-DNA interactions observed *in vivo*, as well as the long half-life of the complex measured by traditional *in vitro* biochemical methods.[186,187]

Therefore, quinary interactions underlying encounter complex formation plays important biological roles in living systems. The two roles discussed in detail in this section relate to the efficient formation of a specific enzyme complex and the efficient reloading of the substrate at the active site of this complex. Both these functions are important for improving enzymatic turnover *in vivo*, however, this scratches only the surface of quinary structure formation and its biological implications. There is also evidence that formation of these weak complexes is involved in enhancing electron transfer for plastocyanin[188] and cytochrome c.[189,190] Future studies are required to fully appreciate the functional diversity and landscape of these short-lived functional encounter complexes.

1.5.5 CHARACTERIZING QUINARY STRUCTURE FORMATION *IN VIVO* AND *IN VITRO*

Because quinary structure is transient and weakly bound, on the order of a few k_BT in the energy scale, routine purification methods can disrupt it and therefore, it must be characterized *in vivo* where possible.

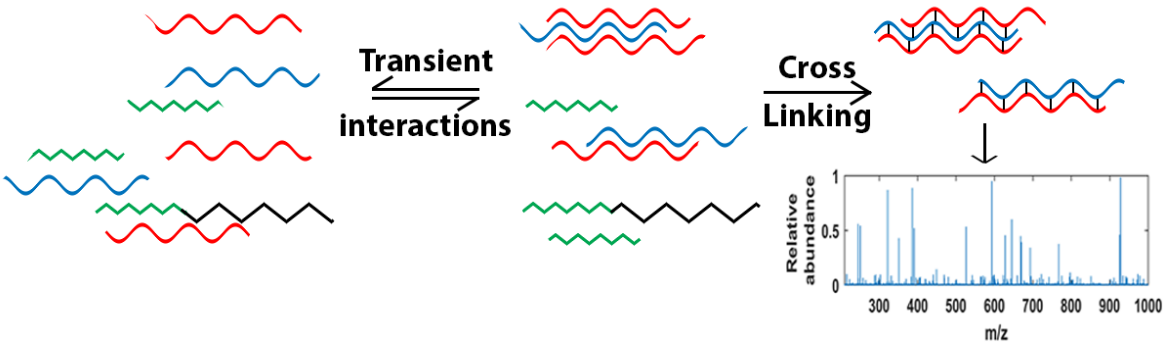


Figure 1.11: Chemical cross-linking mass spectrometry. Transiently interaction species are cross-linked to form more stable species that can be then isolated in vitro. These fragments can then be characterized using mass spectrometry giving rise to discrete peaks corresponding to complex molecular weight on the x-axis and relative abundance on the y-axis.

This is not to say that quinary structure formation cannot be observed *in vitro*. For example, using NMR Shekhtman and coworkers showed that ribosome-protein quinary structure plays a key role in enzymatic activity of thymidylate synthase (TS).[191] Addition of ribosomes *in vitro* enhances enzymatic activity 20-fold, in good agreement with the ~10-fold increase observed *in vivo*.[192] Interestingly, the authors also found that ribosome-specific quinary interactions can decrease the activity of another enzyme, dihydrofolate reductase (DHFR). By modulating the enzyme activity via quinary structure formation, the ribosome plays an important role in metabolism by acting as the hub where enzymes and metabolites are concentrated.[142] Ribsome-mediated quinary structure formation with the 30S subunit of the ribosome has also been shown with mRNA, and for aminoacyl-tRNA synthetases such as with LysRS as described in section 5.2.[193] These interactions are significantly weakened when antibiotics that inhibit the 30S subunit are added.

In-cell NMR was originally established in bacteria, but eukaryotic systems have also recently been used, including yeast, human cell lines and *Xenopus laevis* oocytes.[100,128,132,194,195] NMR experiments have the potential to probe both sticking (see Section 4.3) and quinary structure which we describe in detail in this section. NMR experiments in yeast have been successfully used to demonstrate biomolecule interactions under physiological expression conditions in physiologically relevant cellular compartments. This is possible in yeast due to a wide variety of genetic tools facilitated by very well understood yeast

genetics.[195] Moreover, a very stable yeast cell wall reduces leakage problems[196] and spectra can be obtained over long periods of time (~6 hours) with a high signal-to-noise ratio.

In addition to above-mentioned enzymes and mRNA, chemical cross linking and mass spectrometry revealed >800 proteins that can potentially bind to mRNA or ribosomes in eukaryotic cells (Figure 1.11).[191,197–199] Experiments in yeast using NMR by Shekhtman and coworkers showed that transient interactions with RNA play a key role in deciding biochemistry, like protein activity and localization, of ubiquitin and β-galactosidase.[195,199] The resulting quinary structure is further modulated by growth medium and RNA levels in cells. For example, in yeast grown in methanol as the carbon source a well-resolved $^{1}H–^{15}N$ HSQC spectrum of ubiquitin is obtained. This is because the presence of a large amount of preprocessed mRNA and large ribosomal subunit leads to a reduction in RNA-protein quinary structure formation. In sharp contrast, quinary structure formation is enhanced, and the ubiquitin spectrum is broadened beyond detection in yeast grown in mixed dextrose-methanol medium. Moreover, in the mixed medium both proteins are sequestered in their inactive state in vesicles. Ubiquitin interferes with dextrose metabolism and its inactivation can allow the cell to explore more efficient metabolic pathways,[200] also known as catabolic inactivation.[201] RNA-protein quinary interactions can therefore act as important regulators of many in-cell processes.

Even with some disadvantages (e.g. high protein density required),[196] NMR in human cell lines[100,132] and *Xenopus laevis* oocytes[202,203] is a powerful techniques to probe quinary structure. Proteins are constantly interacting with other proteins in the cellular cytoplasm. These interactions lead to peak broadening and can be easily visualized by NMR. Formation of quinary structure has been observed for both ubiquitin in HeLa cells and *Xenopus laevis* oocytes and profilin1 in human HEK293T cells.[100,132,203] The contribution of these interactions can be measured by introducing appropriate mutations and comparing their NMR spectra. For example, mutations in parts of the protein that interact strongly with other protein partners will lead to larger differences in the spectrum than those that interact weakly. This was successfully demonstrated for both ubiquitin and profilin1 by introducing mutations in the binding pocket and on the surface respectively, which then yielded well-resolved NMR spectra.[100,132]

Because the stabilization due to quinary structure formation is on the order of a few k_BT, they are highly susceptible to changes in the local environment. Gruebele and coworkers showed that different environments in the cell show different degrees of stabilization due to quinary structure formation.[121] A model protein phosphoglycerate kinase (PGK) is stabilized in mammalian U2OS cells compared to *in vitro*.[122] This is partially due to macromolecular crowding[18], however PGK stability in-cell is also

modulated to some extent by its cellular localization.[121] PGK in the nucleus is more stable than that in the cytoplasm and endoplasmic reticulum (ER). The degree of crowding in the cytoplasm and the nucleus is similar according to diffusion measurements.[80] Since the local environment in the nucleus is very different from that in the cytoplasm, the nature of sticking or quinary interactions are possibly also different in both environments.[204]

Weak modulations of the cellular environment such as cell volume changes can be used to characterize the K_d for quinary structure formation. Such weak perturbations only modify quinary structure in the cell without destroying the cell, thus, allowing the detection of quinary structure in the native cellular environment. Cell volume changes by osmotic pressure modulation was used by Gruebele and coworkers to determine the K_d for $GAPDH_4$-PGK binding, described in detail in section 5.2 and for mCherry-AcGFP1 oligomerization to be ~2 µM, an order of magnitude lower than *in vitro*.[85] PRE-measurements described in detail in section 5.4 have also proved very useful to probe quinary structure and the formation of metastable encounter complexes.

1.6 EVOLUTION

We now turn to the connection between physico-chemical interactions in the cell, such as crowding and productive (quinary) or disruptive (sticking) interactions, and evolution. Evolution of proteins has been investigated in great depth and many comprehensive reviews exist that critically analyze the developments in the field.

A recent article by Spitzer, Pielak and Poolman on the emergence of life sheds light on some interesting concepts that drive biological evolution.[205] The authors describe how, among other things, crowding plays an important role in driving evolution in biological systems. Evolution cannot occur in uncrowded systems where surfaces are far from each other and do not interact. In dilute conditions, non-covalent molecular forces cannot maintain cellular organization because thermal disordering effects overcome attractive ordering forces. Indeed evolution of biomolecular surfaces under crowded or confined conditions has been shown to be a likely prerequisite for the transition from the inanimate to the living.[137]

Since the focus of this chapter is cellular forces at protein surfaces, we briefly touch upon how evolution has shaped surface interactions. Biomolecular surfaces are constantly evolving under constraints to improve fitness and function. The potential benefits of combining protein biophysical chemistry with evolutionary biology are many, and as such the inclusion of evolutionary biology in the study of protein biophysics already greatly benefits our understanding of protein function.[206,207] We discuss two main methods of analyzing evolutionary data: epistasis mapping and ancestral sequence reconstruction. Finally, we conclude

the chapter by outlining what future collaborative efforts between evolutionary biology and biophysical chemistry could bring to the table in terms of our understanding of protein interactions in the cell.

1.6.1 HOW PROTEIN SURFACES EVOLVE – CYTOCHROME *C* AND HEAT SHOCK PROTEINS

Evolution means constant change in biology, but some proteins are peculiar because they have stayed effectively unchanged over millions of years. The structure of cytochrome *c* was solved in 1971 and was used to characterize the correlation between evolution and function.[208,209] Cytochrome *c* is a small protein (104 residues in vertebrates), present in the mitochondria of every eukaryotic organism. Dickerson estimated its evolutionary rate.[209] During random mutational drift, if *m* is the fraction of mutations in a polypeptide chain, then the fraction *n* of the polypeptide sequence that actually changes is:

$$n = (1 - e^{-m}). \qquad [1.11]$$

n takes into account that repeated mutations at the same locus do not reduce sequence identity. The calculated rate was then used to approximate the time required for a 1% change in sequence between two divergent lines of evolution in a million years (MY). For a small fibrinopeptide, whose function is to be excised out of fibrinogen and converted to fibrin in a blood clot, this time was 1.1 MY, a rapid rate of protein evolution. Presumably for fibrinogen, any change in the sequence is permissible if it still allows for successful excision of the peptide. The fibrinogen mutation rate is therefore close to the actual DNA mutation rate (Figure 1.12). Conversely, cytochrome *c* interacts with several large macromolecules and hence utilizes more of its overall surface for function. The time scale for cytochrome *c* is 20 MY and large portions of the cytochrome *c* surface have highly conserved charged and aromatic residues. There is also a general correlation between protein size and evolutionary rate.[209]

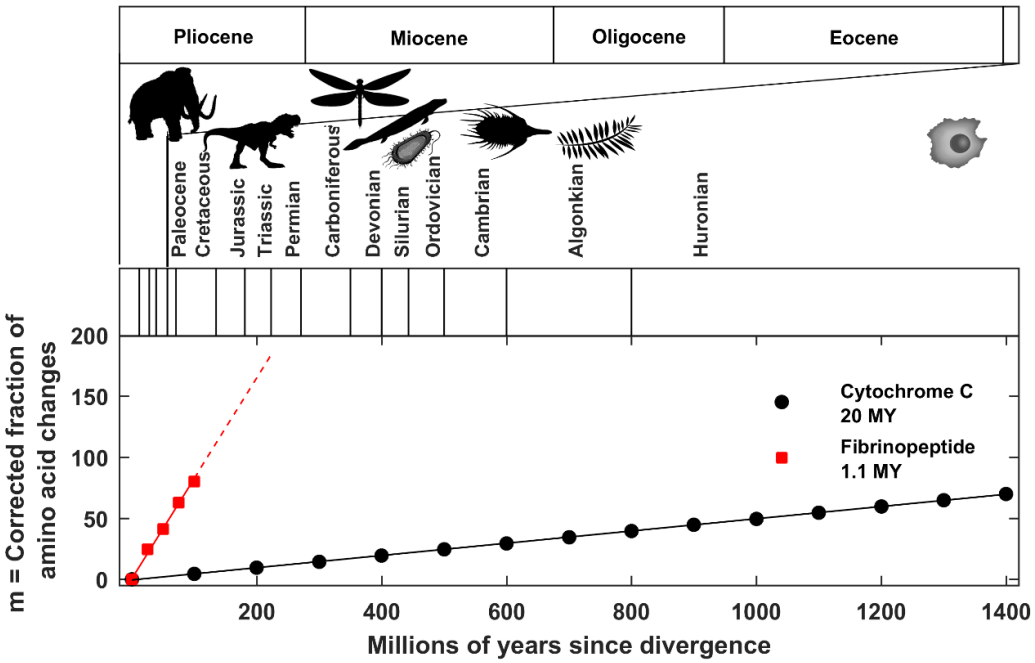


Figure 1.12: The evolution of Cytochrome c vs. fibrinopeptides. Figure adapted from reference 256. Cytochrome c evolves very slowly taking almost 20 MY for a 1% change in residues whereas fibrinopeptides evolve at the rate of occurrence of mutations, 1.1 MY. This shows that residues in Cytochrome c are highly conserved.

The evolutionary role and importance of the family of heat shock proteins has been discussed and probed in great detail.[210–212] They have many interaction partners and are evolutionary capacitors.[213] Not surprisingly, therefore, they are also highly conserved and well-known for their stress response function.[214,215] Among them, the 90 kDa heat shock protein (Hsp90) plays a fundamental role in the expression of genetic variation.[216–218] Its function is special because in addition to stress response, it can also expose or suppress genetic variation. Macro-scale changes in genetics through evolution are generally gradual, but all living species must survive sudden environmental changes and maintain robust developmental systems that do not change abruptly during environmental stasis. This is generally done by storing a certain amount of unexpressed genetic variation that is typically hidden during developmental homeostasis, for example, in species hybrids.[219] Hsp90 can expose such variation by destabilizing and stabilizing transcription factors associated with promoter regions responsible for morphological remodeling of metamorphosis.[218] Likewise, in cell-cycle control Hsp90 supports both activators and inhibitors of the same function to control the process output downstream.[220] Hsp90, therefore acts as a capacitor, that through controlled exposure of cryptic alleles could account for the rapid morphological changes evident in fossil records. Another 70 kDa heat shock protein (Hsp70) that is also highly conserved maintains cellular

stasis during stress and has been referred to as an evolution facilitator, for example by protecting less stable protein mutants and facilitating more sequence variation.[214,221]

These two examples of protein evolution teach us something very important about the combination of evolutionary biology and biophysical chemistry. Cytochrome *c* shows us how important it is to consider evolution from the perspective of protein function. It also points to conclusions that could be made only by combining results derived separately from evolution and studies of protein function. Heat shock proteins highlight that evolutionary changes occur both on macro- and micro-scales. Interestingly, biological systems have evolved not only fitness with the help of genetic variation, but have also endowed the products of evolution itself, proteins, with the latent power of evolutionary change. In the future heat shock proteins are prime candidates to tell us a more complete story of evolutionary change than what is visible only through genetic data.

Lastly, although evolution is the force for change, it is important to probe systems, such as heat shock proteins, that have resisted that change. By correlating systems that change frequently and those that remain static, we can begin to reconstruct the elusive protein evolutionary pathways that lead from now into the past.

1.6.2 WHAT DRIVES EVOLUTION OF PROTEIN SURFACES?

The cell is an environment where proteins must navigate crowded spaces so their surfaces can not only make functional interactions, but also avoid debilitating non-functional interactions with the majority of surfaces they encounter. As discussed in 5.3, only so many bits of information can be encoded on a protein's surface, and must be used up in a compromise between interaction strength (stronger = generally more surface), number of interactions encoded (more = less surface per interaction), and avoiding undesirable sticking (although this could evolve into quinary structure). Proteins evolve constantly and have been doing so since the Hadean eon (~4 billion years ago).[222,223] They have weathered extreme heat and extreme cold but have maintained function.[224] Interestingly, the principles that drive the complex evolutionary behavior in biological systems are fairly basic and follows two main rules: 1) maintain or improve organism fitness and 2) keep intact or improve function.

Perhaps the two driving forces in protein evolution that are most evident at first glance are folding and stability. Since misfolding and aggregation can have a considerable negative effect on organism viability,[225] there is selection pressure to evolve thermodynamically stable[226] and /or aggregation-resistant protein sequences.[227] Several different adaptations reduce aggregation propensity for proteins.[227] For example, protein sequences limit hydrophobicity and maintain an overall low net charge.

Hydrophobic patches longer than 5 residues are represented significantly less than what would be predicted from a statistically independent distribution.[228] Moreover, disordered proteins have charge distributions to avoid aggregation, and globular protein fold to bury hydrophobic residues in the core and hence reduce the propensity of these hydrophobic surfaces from coming into contact and aggregating with other hydrophobic surfaces. Since aggregation is concentration-dependent, protein sequences have further evolved such that the aggregation propensity is just below their solubility limit *in vivo*.[229]

Early studies have also shown that different proteins from the same organism can evolve at vastly different rates.[230] Proteins with multiple partners evolve more slowly.[2,231] This is because a greater proportion of the protein is directly involved in function and the protein is only able to accept a reduced subset of mutations that are at either net neutral or positive for all the interacting partners. Any other evolutionary changes, for such a system must occur via coevolution where changes to one protein lead to an additional selection pressure for reciprocal changes in the interacting partners. This is in part explained by the neutral theory of evolution,[232] where the rate of protein sequence evolution is given by,

$$k = \mu p, \qquad [1.12]$$

where k is the rate of protein sequence evolution, μ is the rate of mutation and p is approximated by the proportion of mutants that are neutral because beneficial mutations are considered too rare to affect the rate of evolution.

Another factor driving protein evolution is the pressure to reduce non-specific protein-protein interactions that compete with productive specific interactions. Interestingly, even though the cell is a sea of sticky surfaces where the need for specificity is paramount, protein interactions follow a scale-free network topology. In a scale-free topology, the minimum free energy gap (ΔE) between the weakest specific interaction and the most competitive non-specific interaction decreases in a power law fashion,[233,234] such that

$$\Delta E \sim N^{-\gamma} \qquad [1.13]$$

where, N is the number of interfaces and γ is the scaling factor that determines how quickly the energy gap drops to 0 as the number of interfaces increase and correspondingly how quickly binding specificity is lost. Even though γ is small (0.13-0.19) the gap reduction is significant for typical proteome sizes, such that, for a typical yeast proteome a gap as small as ~2.5k_BT is already reached with ~1000 interfaces. Moreover, for a proteome with N protein types there are aN specific interactions and bN^2 non-specific ones (where a = 1 to 20 and b are constants), and hence there is significant evolutionary pressure to decrease deleterious non-

specific interactions. Since increasing the proteome size radically increases non-specific interactions and decreases ΔE, this limits the number of proteins that can function effectively in the cell, and organisms must therefore survive with a limited number of proteins. Increasing protein size provides one way out by increasing the dynamic range of interactions (weak to strong) available but has its own problems: increased misfolding and aggregation requires increased chaperoning, i.e. yet more proteins that act as caretakers.

Indeed, even though the evolutionary distance is large, the number of proteins remains similar between simple multicellular organisms and humans.[235] Given these constraints and consistent with scale-free topologies, binding in biological systems is optimized by favoring networks where a few proteins interact with a large number of partners, while most proteins interact with just a few other partners.[233,236]

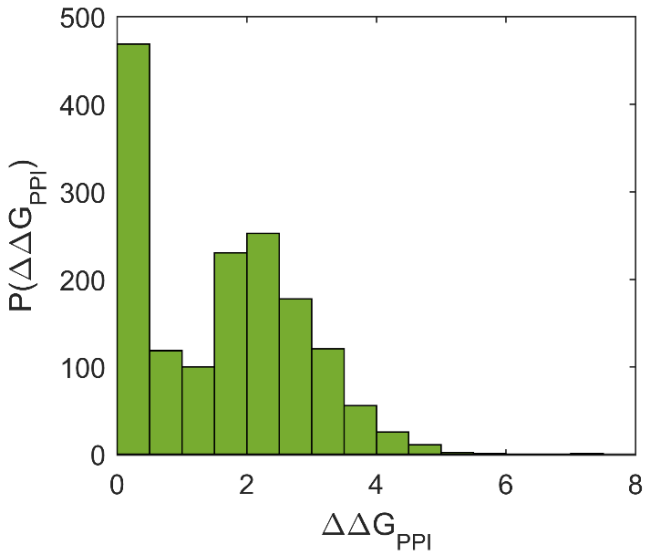


Figure 1.13: Bar plot showing relative abundance of proteins ($P(\Delta\Delta G_{PPI})$) as a function of stabilization due to PPIs ($\Delta\Delta G_{PPI}$). Figure was adapted from reference 231. The population at $\Delta\Delta G_{PPI}=0$ shows no stabilization due to PPIs.

The pressure to reduce non-specific interactions has also been associated with protein abundance and surface hydrophobicity or hydrophilicity.[237] Protein abundance is negatively correlated with the number of its functional interaction partners as well as the number of non-specific interactions.[238] Functional interactions in a proteome generally involve hydrophobic interactions.[239,240] Consequently, hydrophobicity also decreases with protein abundance to decrease non-specific interactions. In agreement with this observation more abundant proteins in the *E. coli* cytoplasm are less hydrophobic.[241] Conversely, abundant hydrophilic IDPs (disordered proteins) make up to 30% of the eukaryotic proteome.[242] Thus, surface properties of a protein play a crucial role in determining protein evolutionary rates.[238,243]

Further reviews that discuss protein evolution in more detail have been published.[244–246] We discuss the biophysical aspects of protein interaction networks in more detail in the next section.

1.6.3 EVOLUTIONARY ADVANTAGES OF PROTEIN INTERACTION NETWORKS

The hydrophobic core of proteins is highly conserved and mutations in the core can quickly disrupt protein structure and stability.[247] Although surface mutations contribute only weakly to overall stability,[248] they also exhibit a surprising level of conservation.[249] This is due to protein-protein interactions (PPIs). In the previous section, we discussed how protein interaction networks follow scale-free topologies. These protein interaction networks (PINs) benefit the organism in many ways.

In spite of the obvious disadvantages due to non-specific interactions, highlighted in section 6.2, networks are of great importance to thriving biological systems. They facilitate signaling and ensure that the cell is robust to random failure of a few network components. Additionally, in higher organisms the number of interactions between proteins increases, and consequently larger multi-protein complexes are observed as compared to their ancestors.[250] Such multi-protein complexes are favored over larger size of individual proteins because larger proteins are more expensive to fold, as noted in 6.2.[251] Smaller individual proteins are easier to fold and are less prone to aggregation, thus improving overall fitness of the organism. Protein interactions can also give rise to allostery and cooperativity, which results in a more efficient on/off switch.

PPIs further reduce toxic aggregation by reducing the effective monomer concentration in the cytosol; monomeric proteins bound to partners are not available for aggregation and are therefore removed from the aggregation-prone pool of monomers. This additional stabilization of the folded state due to suppression of aggregation was calculated[213] and contributes to the overall stability such that,

$$\Delta G = \Delta G_{folded} + \Delta\Delta G_{PPI} \qquad [1.14]$$

$$\Delta\Delta G_{PPI} = k_B T ln\left(\frac{C_A}{(1+e^{\Delta G/k_B T})U_A}\right), \qquad [1.15]$$

where C_A is the total concentration of any protein, U_A is the concentration of all unusable states of A, comprised of all unfolded peptides and insoluble oligomers and ΔG is the thermodynamic stability of the free monomeric state. In yeast, the stabilization due to PPIs was found to be on average about ~2k_BT and in some cases as high as 5-6k_BT (Figure 1.13).[213] Therefore, each protein in a PIN effectively stabilizes its interaction partners to some extent. On the other hand, 'foreign' proteins such as noted in the in-cell NMR experiments (section 4.3) experience sticking instead of effective stabilization.

Stabilization from PPIs can also offset small amounts of destabilizing variations in protein sequences, which often leads to a higher degree of functional diversification.[252,253] This phenomenon has been termed evolutionary capacitance.[213] We discuss the concept of capacitance further in section 6.5. Indeed, the contribution of $\Delta\Delta G_{PPI}$ becomes more important for proteins whose effective population is low (due to genetic drift) and for proteins with a low inherent stability (ΔG_{folded}), such as IDPs (disordered proteins).

The relationship between protein fitness and evolution has been investigated in detail. The Shakhnovich group studied the effect of mutations on the folding free energy and showed that protein abundance is negatively correlated with evolutionary rate using simulations.[254,255] Since destabilizing effects of deleterious mutations are multiplied by protein abundance, more abundant proteins lead to larger amounts of toxic misfolded structures as a result of lethal mutations. Protein abundance and fitness hence plan an important role in determining evolutionary rates: higher abundance leads to higher stabilities and slower evolutionary rates since average mutations are more deleterious. These relationships between the protein biophysical landscape and evolutionary rate have been discussed in detail in already published reviews.[256,257]

1.6.4 CURRENT METHODS IN PROTEIN EVOLUTIONARY BIOLOGY

In this section we briefly touch upon two methods that have been used to study evolution of biological systems, and that are related to protein evolution and cell health and can be applied to studies of protein surface evolution. Detailed description of the methods discussed below are beyond the scope of this chapter but have been discussed elsewhere.[258,259]

1.6.4.1 EPISTASIS MODELING

The term ‘epistatic’ was first used in biology by Bateson in 1909 to describe masking of a gene by another gene.[260] From the point of view of proteins, epistasis refers to the modification of a mutation’s phenotype by another mutation.[261,262] For example, a deleterious mutation could be masked by a mutation at a remote site; or a mutation’s enhanced activity could only arise in the context of a subsequent mutation at a remote site, so that what started out as genetic drift later becomes enhanced fitness (Figure 1.14). As a result, any mutation event in an evolutionary pathway is contingent on the context of many past events. Epistasis is responsible for evolutionary benefits that occur due to a potentiating mutation in an ancestor, and thus cannot be evolved easily by stochastic single point mutations.[263] Thus, epistasis is a symmetry-breaking event that gets frozen in as time progresses. The genetic definition used by Bateson is the same basic concept but defines these changes in the genotype of the organism instead of the phenotype.

The recognition of pairwise epistasis goes back ~100 years,[260] however, the importance of higher-order epistasis has only recently been emphasized.[264–266] Pairwise epistasis is the difference in the effects of two mutations introduced together *vs.* separately. More generally higher-order epistasis refers to when a set of $n + 1$ mutations are introduced together *vs.* separately in *n-tuples* (e.g. a triplet relative to the three pairs). Such higher-order epistasis is evident in many biological systems.[266]

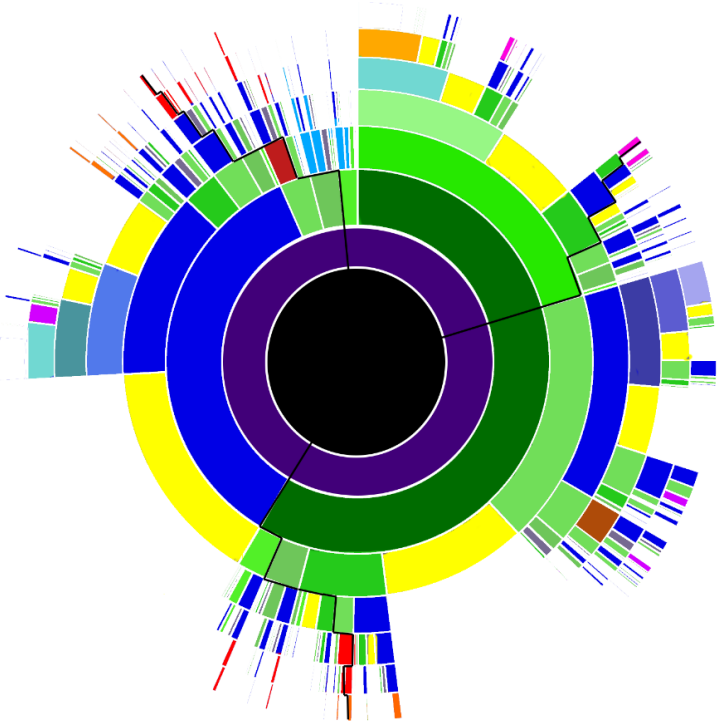


Figure 1.14: A sunburst plot showing epistasis. All proteins start from the black ancestor. Each corner represents a possible mutation and color differences show mutation accessibility from the previous mutation blue (easiest) to red (hardest). For example, it's easiest to go from blue to blue and hardest to go from blue to red. These affects add up, for example if starting at blue traveling along blue is the easiest. The three example black paths show three different evolutionary trajectories starting from the common black ancestor. Epistasis means that every mutation is to some extent determined by the path taken by the previous mutations and the landscape color of the previous mutation.

Historically epistasis has been studied either in simple model organism such as *E. coli, S. cerevisiae* or *C. elegans*, or by computationally modeling epistasis in model systems.[258] For example, Plotkin and coworkers modeled how epistasis affects accumulation of mutations over time.[262] They studied the stability of lysine-arginine-ornithine-binding periplasmic protein (*argT*) from *Salmonella typhimurium* as a function of mutations. They showed that at any given time, mutations that became successfully incorporated were contingent on previous mutations and typically would have been deleterious if introduced at an earlier time. Additionally, once a mutation is fixed, any reversal becomes increasingly detrimental to fitness. This is because a mutation, M, fixed at any given time interacts with many other mutations that

occur at later times and deletion of M negatively affects all other mutations that were contingent on M. This phenomenon has been referred to as an "evolutionary Stokes shift,"[267] in analogy to light excitation and fluorescence: once an excited state is relaxing, blue fluorescence at the same wavelength as the absorption is increasingly unlikely to be emitted as relaxation progresses. Thornton and Harms, similarly, used an error prone polymerase to generate mutants of an ancestral protein of the vertebrate glucocorticoid receptor (GR).[268] They showed that the evolution of the GR protein from its ancestor occurred through a series of epistatic changes that are otherwise improbable and non-deterministic. Such an event is so rare, that if evolution were to reoccur from the start the GR protein would evolve either from a completely different ancestor or not at all; more generally the vertebrate endocrine system would be substantially different.

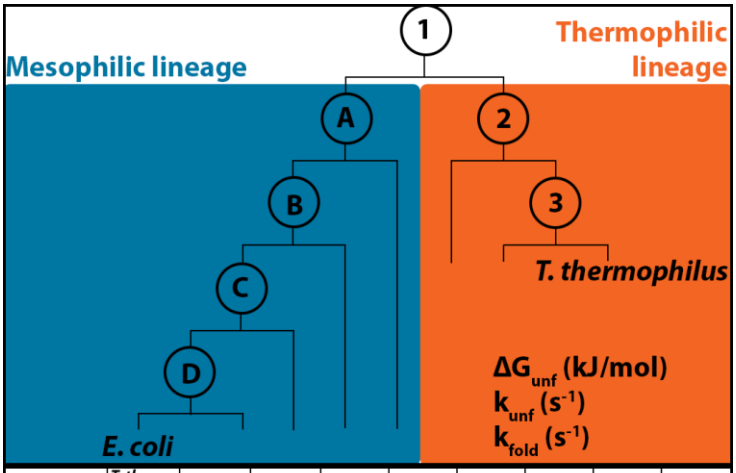
Evidence of epistasis has been also been shown experimentally in model organisms. For example, epistasis was shown to be responsible for evolving a citrate metabolism gene in *E. coli* over ~30000 generations.[263] A potentiating mutation that occurred at ~20000 generations led to significantly greater tendency for the evolution of the citrate gene in later clones. Similarly Kryazhimskiy, Desai and coworkers characterized epistasis, adaptability and fitness in *S. cerevisiae* over 500 generations.[269] They found that the adaptability of a genotype is negatively correlated with fitness; lower fitness leads to higher rate of adaptation, and differences in adaptability are almost entirely correlated with the overall fitness, rather than with individual genetic mutations underlying that fitness.[269]

Epistasis modeling can be a powerful tool to predict the evolutionary pathway undertaken by modern proteins. However, Harms and Sailer showed that even with current tools, evolution remains difficult to back-track due to epistasis.[265] Since proteins occur in an ensemble of conformations, any single mutation exerts its effect on each conformation in a slightly different way. Calculating this requires either the knowledge of how each conformation is affected by a mutation, which is currently impossible to measure, or the calculation of all higher-order epistases, which is currently computationally too difficult. Simply calculating the average ensemble effect of the mutation on the entire population, a mean field approach, leads to large uncertainties in prediction. For now, epistasis can predict phenotypes with relatively high accuracy and help connect the dots between how modern proteins came to be from ancestral ones.

1.6.4.2 ANCESTRAL SEQUENCE RECONSTRUCTION

With the advent of efficient sequencing techniques by Frederick Sanger in 1955,[270] Emile Zuckerkandl and Linus Pauling postulated that such sequences could be used to reconstruct the sequences of ancestral proteins. Ancestral sequence reconstruction was thus born. In simple terms, it is the extrapolation back in time of genetic sequences from current proteins to their common ancestors. It relies on sufficiently realistic models of evolution to predict ancestral states and immense progress has been made with the improvement

in computational power and the development of more efficient algorithms. Details of some common methods and algorithms used in ancestral sequence reconstruction have been published.[259,271] Among the available prediction methods, Bayesian inference is believed to be able to estimate ancestral sequences with a high accuracy and has been widely used.[272]



	T. therm-ophilus	3	2	1	A	B	C	D	E. coli
ΔG_{unf}	53.6	52.3	43.9	42.7	39.3	36.0	38.9	38.9	40.6
k_{unf}	$4x10^{-6}$	$2.1x10^{-4}$	$1.1x10^{-3}$	$1.6x10^{-3}$	$2.4x10^{-4}$	$1.1x10^{-4}$	$1.7x10^{-4}$	$7.3x10^{-6}$	$1.7x10^{-5}$
k_{fold}	0.6	4.1	14.1	10.4	9.7	0.77	2.8	.13	.74

Figure 1.15: Thermodynamic and kinetic evolution of modern RNase H thermophilic and mesophilic homologs. Data for figure was adapted from reference 255. Overall stability is reflected by ΔG_{unf}, folding rate by k_{fold} and unfolding rate by k_{unf}.

Once ancestral protein sequences have been predicted, they can be experimentally reconstructed in two ways. First, via step-by-step site-directed mutagenesis of specific residues in the modern protein, recapitulating in reverse the path from ancestral to modern protein.[273,274] This approach is limited to proteins for which structure-function relationships are well understood and only works under the assumption that the mutations themselves do not significantly affect the function or folding of the protein. Alternatively, the entire sequence of the ancestral sequence can be assembled in its entirety *de novo* provided that the ancestral sequence is known or can be inferred.[275] This method does not require prior knowledge of the structure-function relationship. Of course, site-directed mutagenesis can also be used to reconstruct the full set of mutations of an ancestral protein at once.[276]

Benner and coworkers reconstructed the ancestor of the yeast protein that consumes and metabolizes ethanol. Since most organisms cannot metabolize alcohol, this lends yeast a significant survival advantage over other competing organisms.[277] The ancestor of this enzyme specialized not in alcohol consumption but in alcohol production. The alcohol production was a consequence of recycling NADH during anaerobic

glycolysis. The alcohol is eventually lost to the environment. Similarly, sequence reconstruction also showed that ancient enzymes exhibit a relatively slow evolution of protein structure even as the amino acid sequence varies.[278] This suggests that the evolution of non-promiscuous activity of highly specialized enzymes and enzyme complexes may been completed in the era of the last universal common ancestor.

More recently, ancestral sequence reconstruction was used to investigate evolution of the folding pathway of a model protein, RNaseH.[279,280] Marqusee and coworkers used sequence reconstruction to probe evolution of the thermodynamic and kinetic stability of RNase H (Figure 1.15) from a common ancestor into the mesophilic and thermophilic branches of modern day RNase H. They observed a 90 to 400-fold decrease in the protein unfolding rate (k_{unf}) as well as 10 to 20-fold decrease in protein folding rate (k_{fold}) for mesophiles and thermophiles over the ancestor. The large decrease in unfolding rate shows evolutionary pressure in favor of increasing kinetic stability. Increasing kinetic stability allows the protein to be more resistant to unfolding and subsequent aggregation or misfolding. Thus, reduction of unfolding rate is an evolved protein property. However, the smaller decrease in the folding rates of RNase H could indicate neutral drift. Therefore, evolution does not seek to evolve fast folding proteins beyond a certain limit. Good enough is often good enough.

While both thermophilic and mesophilic proteins are stabilized overall, thermophile RNase H stability is much higher. This is possible because RNase H is a three-state folder allowing for changes in the rate-limiting step, from native to the intermediate state, to only affect kinetic stability without much change to the overall thermodynamic stability.[281]

This can be easily visualized by the equation where for a two-state folder:

$$\Delta G_{unf} = -RT\ln\left(k_{unf}/k_{fold}\right) \quad [1.16]$$

whereas, for a three-state folder it is:

$$\Delta G_{unf} = \Delta G_{intermediate} + -RT\ln\left(k_{unf,rate-limiting}/k_{fold,rate-limiting}\right) \quad [1.17]$$

1.7 WHAT IS YET TO COME: LOOKING AT PROTEIN BIOPHYSICS FROM AN EVOLUTIONARY PERSPECTIVE

Biological systems are highly evolved, robust and efficient molecular machines, not just structurally and functionally but also in terms of interactions. Promiscuous or generalist proteins were more common in our ancestors. They provided a wider array of lesser catalytic activities, more useful in a rapidly changing fitness landscape such as the Paleoarchean era. Specificity evolved as enhancement of certain functional traits,

better adopted to stable niches, improved fitness. However, not all proteins became specific. This is because of two reasons: 1) Specificity is difficult to maintain; two highly specific partners must co-evolve to keep function intact. 2) If all proteins became highly specialized, then the biomolecule load due to protein variety N in a cell would be enormous. As described before, non-specific interactions that scale as N^2 would then become highly detrimental to the organism's survival. Thus more proteins that are 'perfect ' is not necessarily better than fewer proteins that are 'good enough,' and so generalists still abound in modern day species.[282,283] Over a third of modern *E. coli* enzymes exhibit promiscuity.[283] However, promiscuity does remain a source for new specific functions in the future, should the environment change again.[284]

Quinary interactions likely evolved from non-functional sticking, very likely assisted by epistasis. A later mutation that made a pre-existing sticky interaction weakly functional would have been very beneficial to organism fitness. Moreover, for the few proteins that have many binding partners, like chaperones, promiscuity is an inherent trait. In *E. coli* the promiscuous interactions that have survived evolution have weakly-binding K_d in the μM regime,[282] a hallmark of quinary structure formation. Promiscuous interactions and quinary structure were therefore shaping the organization of even the earliest cells.

Quinary structure formation is strongly influenced by the crowded cellular environment and purification methods often disrupt such interactions. *In vitro* experiments in simple buffers, therefore, only have limited access to probing quinary structure and in-cell studies or carefully constructed more complex *in vitro* systems are needed. With technical advancements, instruments to image quinary structure inside cells are fortunately becoming more routine.

Much remains to be gained by collaborative efforts between evolutionary biology and biophysical chemistry. The integration of the knowledge of protein structure, function and evolution will provide a complete picture of how proteins interact inside cells.

CHAPTER 2

DODINE AS A TRANSPARENT PROTEIN DENATURANT FOR CIRCULAR DICHROISM AND INFRARED STUDIES[2]

2.1 INTRODUCTION[2]

Protein-denaturant interactions have been used to characterize protein stability *in vitro*. In chapters 2 and 3 I probe these interactions in more detail by using an alternative denaturant, dodine. Dodine interactions with proteins enable protein denaturation at much lower, millimolar, concentrations. These "low concentration" denaturants do not interfere with spectroscopy unlike more traditional "Molar" denaturants.

Protein denaturation by urea or guanidine hydrochloride is a time-honored technique to study protein stability. The denaturant concentration at the unfolding midpoint (50% folded population for a two-state folding protein) is related to the stability of the native state, and the slope of the denaturation curve provides useful information about the solvent-accessible surface area of folded vs. unfolded state.[285–287]

These denaturants are ideal for fluorescence detection, but for other detection techniques, such as circular dichroism or infrared spectroscopy, they pose a problem. Their absorption coefficients in important UV regions (<210 nm for circular dichroism spectroscopy) or IR regions (1500-1700 cm-1 near the protein amide I/1' bands) are large, and the required molar concentrations obscure protein bands.[288,289] Deuteration as a work-around is not fully successful in dealing with this problem.

Other denaturation techniques work at much lower concentration, but have their own drawbacks. For example, detergents can act at concentrations comparable to the protein concentration, but they often produce entirely non-cooperative unfolding curves.[290,291] That precludes the comparison of thermodynamic stability among protein mutants for example. Nonetheless, continuous unfolding can be useful in some circumstances. The problem of denatured state residual structure is now at the forefront of protein folding and misfolding studies,[292] and continuous tuning of denatured state structure is a useful tool for such studies.[293]

We hypothesize that a molecule combining small denaturants (urea, guanidine) and long-chain detergents into a single unit might provide both a cooperative transition at a very low denaturant concentration, and

[2]This chapter reprinted with permission from Drishti Guin, Kori Sye, Kapil Dave and Martin Gruebele. Dodine as a transparent protein denaturant for circular dichroism and infrared studies. Protein Science, 25, 1061-1068, 2013. Copyright 2016 The Protein Society.

some continuous tunability of denatured states. Based on early work with molecules such as decylbiguanidine[294] we recently tested dodine (dodecylguanidine), an alkylated guanidine, as a denaturant. Dodine is easily available commercially because of its use as a fungicide (Figure 2.1A). We showed that dodine denatures two mini-proteins cooperatively at sub-mM concentrations, at least 1000x lower than the active concentration range of urea or guanidine.[295] It was less effective for the β-sheet model protein than for the helical bundle model protein.

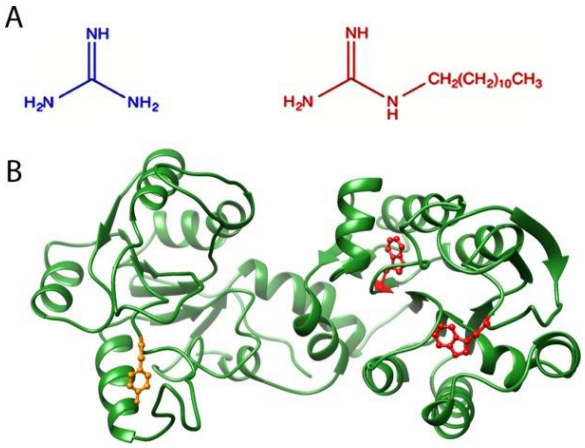


Figure 2.1: (A) Structures of neutral forms of guanidine (blue) and dodine (red). (B) Ribbon structure of wild type phosphoglycerate kinase (PGK) from yeast. Red residues are two tryptophan present in the C-terminal domain of the wild type protein; orange residue is additional tryptophan present in the N-terminal domain of the PGK Y122W mutant.

Here we test dodine on a much larger protein, the ATP synthase phosphoglycerate kinase (PGK, Figure 2.1B). This mixed α/β protein has been studied extensively by fluorescence,[296,297] circular dichroism,[296–298] infrared absorption,[299] NMR[300] and FRET.[18,301,302] Several of these biophysical probes would benefit from a cooperative denaturant that acts at sub-millimolar (mM) concentrations to reduce spectral interference. We found that dodine at just 0.25 mM concentration can induce cooperative denaturation of PGK, if the PGK concentration is in the micromolar range. As it did for mini-proteins, dodine has an effect on PGK similar to over 1000x higher concentrations of guanidine, although its denatured state has more residual structure than in guanidine. Sub-mM dodine also facilitates thermal and pressure denaturation of PGK, comparable to 0.35-0.7 M guanidine. Unlike the latter, it produces excellent CD spectra of denatured PGK, analyzable to wavelengths down to 200 nm with low noise, and nearly background-free infrared spectra over the entire amide I' band. Its main drawback compared to urea or guanidine is the lower reversibility of unfolding curves on a time scale < hours. Dodine could indeed prove to be a versatile denaturant for many biophysical studies, where IR and UV transparency

is required, some continuous tunability of the unfolded ensemble is acceptable or even desirable, and slow reversibility can be tolerated.

2.2 METHODS

2.2.1 CHOICE OF PROTEINS, EXPRESSION AND PURIFICATION

Temperature and denaturant unfolding studies of wild type yeast PGK and its well-behaved Y122W mutant[297,303] were conducted using commercially available guanidinium hydrochloride (guanidine) and n-dodecylguanidinium acetate (dodine) (both Sigma-Aldrich, St. Louis, MO). Conformational changes upon thermal, pressure and chemical denaturation were monitored by circular dichroism (sensitive to overall secondary structure) and tryptophan fluorescence peak shift (a more local probe). Both tryptophan residues in the wild type protein are located in the C-terminal domain (W308 and W333). The mutant tryptophan at position 122 serves as an additional probe for the N-terminal domain.[297] All titrations were performed in 20 mM Tris buffer. Since the solubility of the commercially available dodine acetate is low in pure water, dodine was pre-dissolved in small amounts of ethanol (15 mM unless stated otherwise) and then diluted down to experimental concentrations (sub-mM concentrations). Dodine concentrations were below the critical micelle concentration (CMC = 8.7 mM) for all experiments.[295]

Both PGK mutants were expressed as described previously.[304] Each protein was isolated by affinity chromatography. The lysis buffer was 150 mM NaCl, 50 mM Na_2HPO_4, and 20 mM imidazole (Sigma-Aldrich, St. Louis, MO); pH 8.0. Protein was purified on a 5-mL HisTrap HP column on an AKTA pure FPLC system (GE Healthcare Life Sciences) and eluted with 150 mM NaCl, 50 mM Na_2HPO_4 and 500 mM imidazole; pH 8.0. Protein purity was assessed using SDS-PAGE gel electrophoresis and protein was dialyzed into 20 mM Tris buffer; pH 6.7. 20 mM Tris was chosen because dodine is less soluble in phosphate and acetate buffers.

2.2.2 SPECTROSCOPIC PROBES OF DENATURATION

Circular dichroism was measured using a spectropolarimeter with Peltier temperature control (JASCO Inc., Easton, MD). Unless otherwise noted, all spectra were recorded from 250 to 200 nm at a scan rate of 50 nm/min at 1 nm resolution and averaging 5 accumulations. Thermal melts were done using a 2 mm path length cuvette. Unless otherwise noted, protein concentration in circular dichroism experiments was approximately 2 μM.

Tryptophan fluorescence measurements were taken on fluorimeter equipped with Peltier temperature control (Agilent Technologies, Santa Clara, CA). The excitation wavelength was 280 nm, and emission spectra were collected from 290 to 450 nm. Samples were measured in 400 μL cuvettes at 5 μM

concentrations, unless otherwise noted. The reproducibility of dodine-induced unfolding and dodine-assisted thermal unfolding as detected by fluorescence was checked on the course of several months with different samples (Figure 2.18).

Infrared (IR) spectra were taken using a Nicolet 6700 IR spectrometer (Thermo Scientific, Waltham, MA). PGK from an ammonium sulfate suspension (Sigma-Aldrich, St. Louis, MO) was pelleted by centrifuging at 10,000 rpm for 10 minutes. The supernatant was removed, and the pellet was re-dissolved in 20 mM Tris buffer made with 99% deuterium oxide (Sigma-Aldrich, St. Louis, MO) to avoid the strong absorption of water in the amide I′ region. The protein concentration was approximately 150 μM for IR measurements. For the dodine FTIR, dodine was diluted from a 50 mM dodine stock in ethanol. For the guanidine FTIR, guanidine was diluted from a 5 M stock in deuterium oxide. The samples were placed between barium fluoride windows and spectra were collected from 400 to 4000 cm^{-1}. FTIR spectra were smoothed by resampling points where water vapor absorbance was small.

2.2.3 REVERSIBILITY AND HYSTERESIS MEASUREMENTS

Both thermal titration and denaturant titration (isothermal) reversibility was monitored using tryptophan fluorescence at 2 μM protein concentration. For thermal reversibility a downward temperature melt was done on the fluorimeter with a 20 minute wait time between temperatures to allow proper equilibration.

For isothermal reversibility an aliquot of protein was allowed to equilibrate at room temperature with denaturant for 30 minutes. The aliquot was then diluted in buffer with protein to keep protein concentration constant and further equilibrated for 20 minutes. Gradual recovery in λ_{peak} was measured using tryptophan fluorescence.

Hysteresis was monitored using tryptophan fluorescence by cycling at 1 minute intervals for the required time period.

2.2.4 DATA ANALYSIS

Data was analyzed using Matlab (Mathworks Inc., Natick, MA). Circular dichroism spectra were analyzed by the change of the circular dichroism signal at 222 nm for α-helices[305], as well as singular value decomposition (SVD). SVD yielded only two major basis functions (Figures 2.13 and 2.14) in dodine, confirming the quasi-two-state folding as detected by CD. Tryptophan fluorescence was analyzed by tracking the shift in fluorescence spectral peak wavelength (λ_{peak}) as a function of temperature or denaturant concentration. Melting temperatures (T_m) and denaturation midpoints (C_m) were calculated using a two-state thermodynamic fit when possible. Native and denatured state baselines were assumed to be linear in the perturbing variable X (temperature, pressure or concentration) so that the signal could be modeled as:

$$S_i = m_i(X - X_m) + b_i \quad [2.1]$$

where, *i* denotes either native (N) or denatured (D) state, m is the linear slope and *b* the linear intercept. The total signal is then modeled as:

$$S(X) = S_N(X)f_N(X) + S_D(X)f_D(X) \quad [2.2]$$

where the native and denatured protein fractions *f* are given by

$$\Delta G_U \approx \delta g_1(X - X_m), \text{ and } f_N = \frac{1}{1+e^{-\frac{\Delta G_U}{RT}}}, f_N = \frac{e^{-\frac{\Delta G_U}{RT}}}{1+e^{-\frac{\Delta G_U}{RT}}} \quad [2.3]$$

Fits to this equation allow us to extract a midpoint value (X_m) for the perturbation of interest, such as T_m (melting temperature), P_m (melting pressure) or C_m (midpoint denaturant concentration). Models with more floating parameters (e.g. a heat capacity model for unfolding[306]) did not result in smaller fitting uncertainties. SVD analysis of the fluorescence yielded more accurate *m*-values and δg_1 values than wavelength shift, so those are reported in SM.

2.2.5 SINGULAR VALUE DECOMPOSITION (SVD) ANALYSIS

In SVD, a sequence of CD or fluorescence spectra taken at different denaturant concentration, temperature, or pressure, is arranged into a matrix (wavelength in the column dimension, changes in temperature or pressure or denaturant in the row dimension, or *vice-versa*). This matrix of raw data A is decomposed uniquely into three matrices:

$$A = USV^T \quad [2.4]$$

The matrix V^T contains a set of orthogonal basis functions that can be linearly combined to represent the original spectra in *A* as a function of denaturant concentration, pressure or temperature. The matrix *S* is a diagonal matrix that contains a set of singular values, one for each function in V^T; the larger the singular value, the more important the function for reconstructing the original spectra. Each column in the matrix *U* contains the amplitude for each basis vector in V^T. *U* shows how during a denaturation the functions in V^T contribute to the observed signal in *A*. Roughly, the first column in V^T tells us how the intensity of the spectrum changed during denaturation, while the second column in V^T tells us how the first derivative of the intensity (wavelength shift) changed for the overall spectrum. If the protein is well-approximated by a two-state transition, higher singular values (3, 4 …) will be small compared to 1 and 2.

2.2.6 DODINE BINDING TO PROTEIN MODEL

The dodine binding to protein was modeled as a simple denaturant binding reaction given by

$$P + nD \overset{K_{eq}}{\Longleftrightarrow} PD_n \quad [2.5]$$

Where, P is the protein and D is denaturant. Assuming either no denaturant or all denaturant is bound, P is the native protein and PD_n is denatured protein, so that n would be the total number of binding sites on the protein.[307] The equilibrium constant of the above reaction can be written as

$$K_{eq} = \frac{[x]}{[P_o - x][D_o - nx]^n} \quad [2.6]$$

Where, P_o and D_o are total protein and denaturant concentration respectively. At mid-point of titration with denaturant

$$[PD_n] = [P] = \frac{P_o}{2}, \text{ or } x = P_0/2 \quad [2.7]$$

Such that K_{eq} is reduced to

$$K_{eq} = \frac{1}{\left[D_o - \frac{nP_o}{2}\right]^n} = e^{\frac{-\Delta G}{RT}} \quad [2.8]$$

$$D_o = \frac{nP_o}{2} + e^{\left(\frac{\Delta G}{nRT}\right)} \quad [2.9]$$

A fit of D_o vs. P_o then predicts values of the effective free energy of binding as well as total binding sites *n*. The actual binding consists of a succession of equilibria, whose equilibrium constants may increase (cooperative) or decrease (anti-cooperative) as more dodine molecules bind to protein. However, with only 4 data points in Figure 2.1, the simple effective model fit to the equations shown above was enough.

2.3 RESULTS

2.3.1 DENATURANTS, PGK WILD TYPE *VS.* Y122W MUTANT, AND DENATURATION MEASUREMENTS

We extend the use of dodine (Figure 2.1A) as a mM denaturant to a larger and more complex two-domain enzyme, PGK (Figure 2.1B). Dodine was obtained as the dodinium acetate salt, and guanidine as the

guanidinium chloride salt. We refer to them as “dodine” and “guanidine” for simplicity. Experiments were carried out in tris buffer (see Methods).

Wild type PGK has a melting temperature of T_m = 54 ˚C, and its Y122W mutant has T_m = 48 ˚C.[308] Previous results have shown that PGK has a cooperative denaturation midpoint $C_m \approx 0.77$ M in aqueous guanidine solution, very similar to the PGK Y122W mutant with $C_m \approx 0.75$ M in guanidine.[297,309–311] Wild type W308 and W333 probe the C-terminal domain, while the complementary Y122W probes the N-terminal domain of the two-domain enzyme.

Another commonly used nitrogen-based denaturant, urea, has also been characterized as a denaturant for PGK, with a $C_m \approx 2.3$ M.[312] Guanidine has been shown to be a more effective destabilizing agent for α-helices,[313] although guanidine is only slightly more effective than urea for β-sheet rich proteins.[314] This could explain why urea is a weaker denaturant for a mixed α/β fold such as PGK. Based on our experience with mini-proteins, we expect that dodine will act at concentrations roughly 1000 times lower than guanidine, and 2000 times lower than urea.[295]

Chemical or thermal denaturation was monitored using circular dichroism and fluorescence peak wavelength shift. Pressure denaturation was monitored by fluorescence peak wavelength shift. All results discussed below for the wild type were replicated with the Y122W mutant. No significant differences in unfolding behavior were found between the two mutants besides a slightly lower melting transition of PGK Y122W, as expected from its lower stability. This supports the notion that both N- and C-terminal domains show similar effects upon dodine denaturation.

2.3.2 DODINE *VS.* GUANIDINE COOPERATIVE UNFOLDING TRANSITION

In dodine both head group and alkyl chain contribute to the unfolding transition.[292,293] The C_m was determined by monitoring tryptophan peak fluorescence wavelength λ_{peak} as a function of denaturant concentration. The measured C_m for PGK with guanidine is 0.84 M (Figure 2.2A). Dodine is an effective denaturant at over 1000-fold lower concentration than guanidine: its C_m ranges from 0.28 to 0.44 mM depending on protein concentration (Figure 2.2B). This concentration dependence is characteristic of denaturants that act on the same concentration scale as the target protein.[290,315] The cooperativity of the transition with dodine increases with decreasing protein concentration.

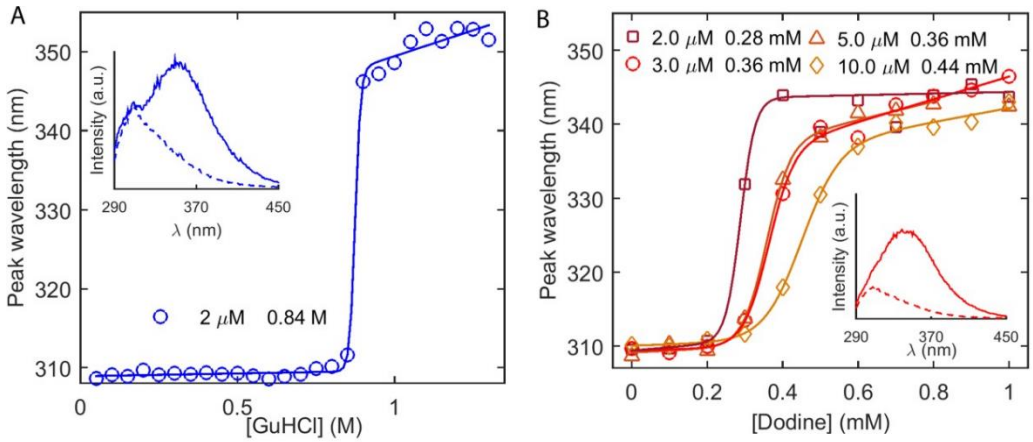


Figure 2.2: Fluorescence denaturation curves of PGK with (A) guanidine (blue) and (B) dodine (red). Markers show experimental data and solid lines are fits to experimental data. The insets show emission spectra at 0 M denaturant (dashed) and final denaturant concentration (solid). (A) The concentration midpoint (C_m) in guanidine is 0.84 M for 2 μM protein. The total wavelength center of mass shift is $\Delta\lambda_{peak}$ ~39 nm. Inset also shows evolution of double peak on denaturation. (B) Denaturation via dodine is protein concentration dependent. The inset for dodine shows absence of double peak. $\Delta\lambda_{peak}$ ~30-34 nm.

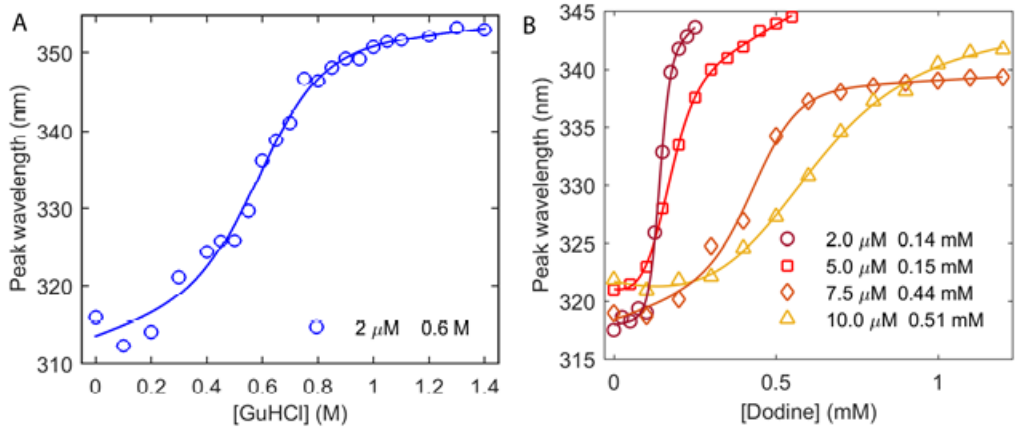


Figure 2.3: Fluorescence denaturation curves of PGK Y122W with (A) guanidine and (B) dodine. PGK Y122W has a folded peak wavelength of 315 nm, about 5 nm higher than the wild type. Markers show experimental data and solid lines are fits to experimental data. (A) The concentration midpoint (C_m) in guanidine is 0.6 M for 2 μM protein. The total wavelength center of mass shift is $\Delta\lambda_{peak} \leq 35$ nm. (B) Denaturation via dodine is protein concentration dependent. $\Delta\lambda_{peak} \geq 23$ nm.

Upon denaturation with guanidine, the PGK fluorescence spectrum evolves from a peak at 310 nm with a shoulder at 325 nm to a double peak with a major peak at 350 nm (Figure 2.2A, inset). The native state fluorescence at 310 nm is characteristic of the tyrosine residues in PGK, while the shoulder at 325 nm is due to the two tryptophans. The two tryptophan sidechains that are buried in the folded state become solvent exposed on unfolding, hence giving rise to the large red shifted emission peak at 350 nm. Dodine produces a similar outcome, with a more pronounced peak at 345 nm (Figure 2.2B, inset). Similar to previous

findings[295], guanidine results in a slightly larger red-shift of λ_{peak} than dodine (39±1 *vs.* 32±2 nm in Figure 2.2). This could be because guanidine denatures the protein more extensively, or it is possible that dodine alkyl chains solvate the denatured hydrophobic core, and their lower dielectric constant (compared to water) reduces the red-shift upon solvation. The issue of solvation/binding of protein by dodine is discussed in more detail later. The mutant PGKY122W closely mimics the unfolding behavior of the wild type protein, except with a slightly lower stability (Figure 2.3).

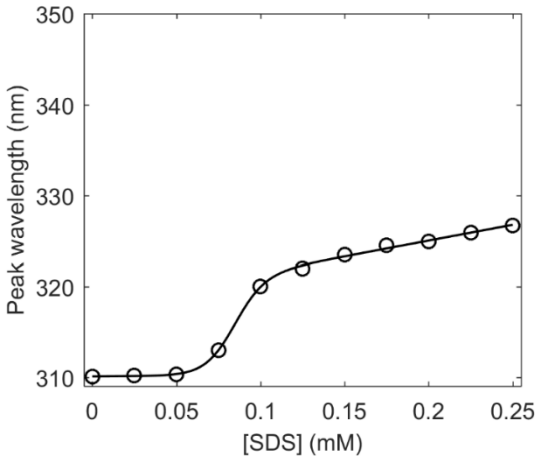


Figure 2.4: Isothermal control experiments with SDS and PGK wild type. Note the highly expanded wavelength scale to show detail near 310 to 325 nm. While dodine induces a cooperative transition to 345 nm (Figure 2.2 and 2.3), SDS shows a much small cooperative transition with C_m ~0.08 mM, followed by a gradual structure loss. Symbols represent experimental data; line is fit to experimental data

A control experiment shows that SDS does not result in as strong a cooperative transition as dodine (Figure 2.4). A small cooperative transition in the signal with C_m = 0.08 mM [290,316] is followed by a gradual loss of structure as a linear function of SDS concentration.

We found that while the isothermal denaturation of PGK with guanidine is reversible to a large degree, with dodine it is largely irreversible on a time scale of ≤ 1h (Figure 2.5).

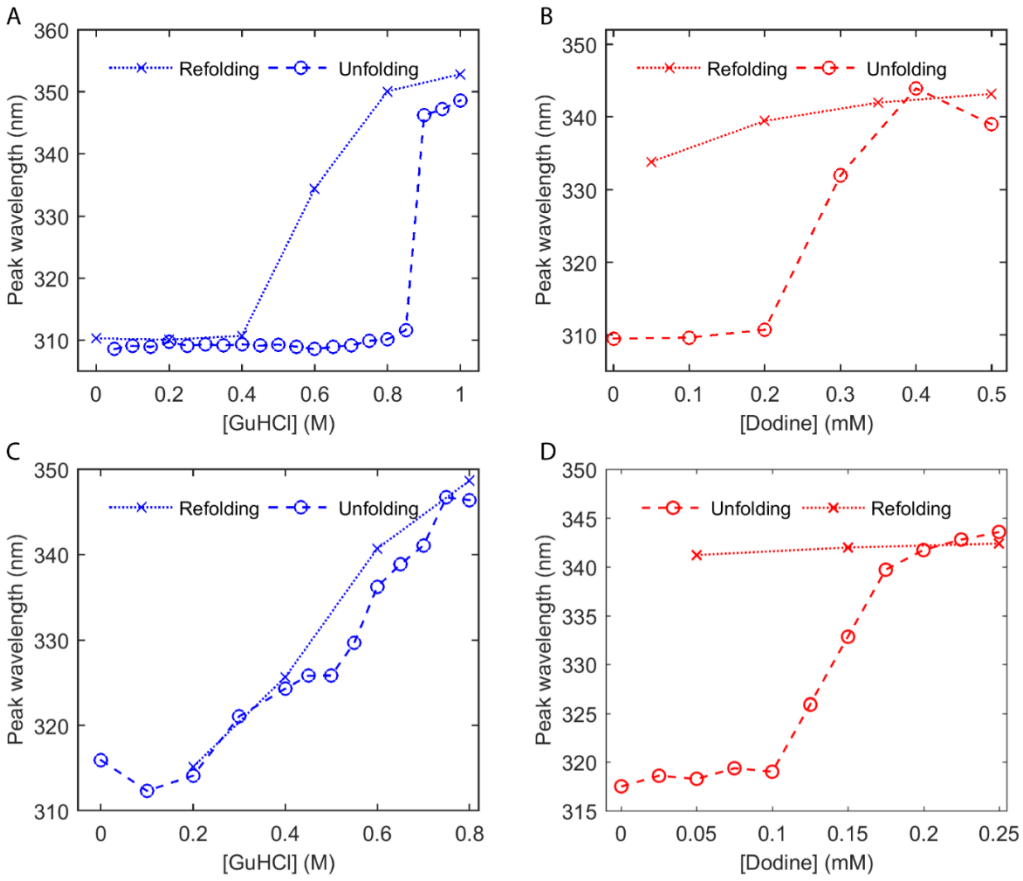


Figure 2.5: Isothermal reversibility of (A) PGK wild type with guanidine, (B) PGK wild type with dodine, (C) PGK Y122W with guanidine and (D) PGK Y122W with dodine. Both guanidine titrations are reversible while dodine titrations are irreversible.

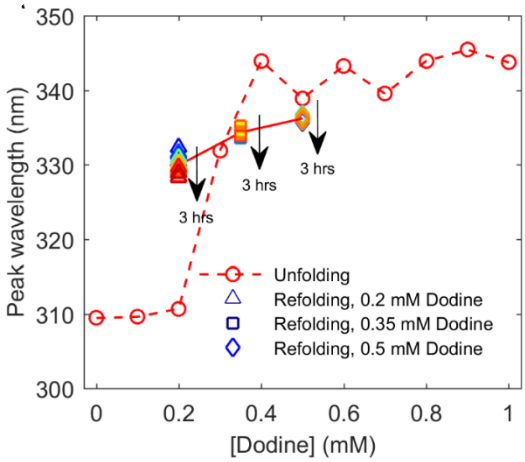


Figure 2.6: Hysteresis of isothermal refolding of PGK wild type with dodine. Bluer colors show earlier times and black arrows total experiment time. Refolding was monitored over 3 hours on decreasing dodine concentration.

Reversibility was further checked by conducting hysteresis measurements over a time scale of 1 hour (Figure 2.6). A small hysteresis of the peak wavelength occurs over the time scale at the lowest dodine concentration (0.2 mM, upward triangle). For higher concentrations recovery in peak wavelength showed very little change (0.35 mM, squares).

2.3.3 LOW CONCENTRATION OF DODINE FACILITATES PGK THERMAL AND PRESSURE DENATURATION

To further extend our understanding of dodine as a potential denaturant for larger, more complex proteins, we investigated whether dodine destabilized PGK by lowering its melting temperature or pressure. To compare the effect of dodine with guanidine, denaturation was conducted in 1) buffer without denaturant, 2) 0.35 to 0.7 M guanidine and 3) 0.05 to 0.2 mM dodine. The denaturation was monitored by circular dichroism spectroscopy and fluorescence spectroscopy (Figure 2.7).

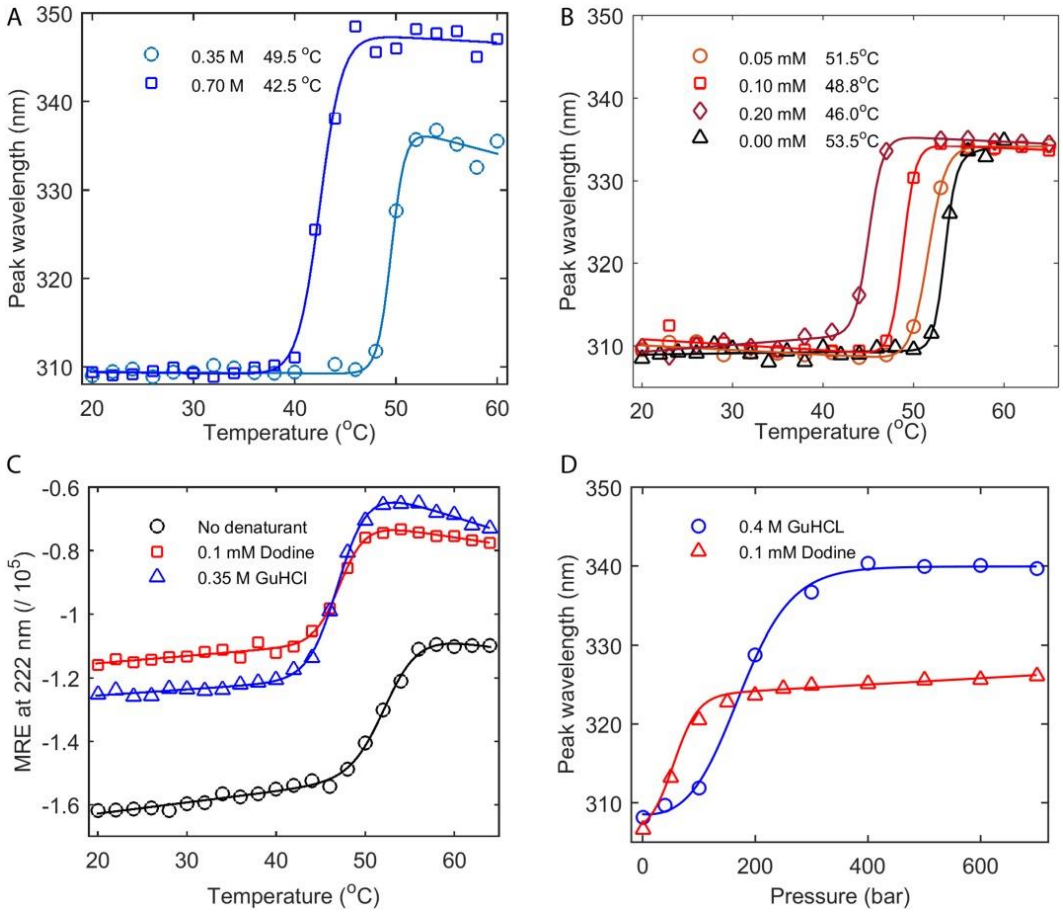


Figure 2.7: Thermal and pressure melts of PGK assisted by guanidine or dodine. (A) Thermal melt assisted by guanidine (blue) and monitored by Trp fluorescence. 0.35 *M* guanidine reduces the melting temperature from 54 °C to 49.5 °C. 0.7 *M* guanidine reduces the melting temperature a further 7 °C to 42.5 °C. (B) Thermal melt assisted by dodine (red) and monitored by Trp fluorescence. A temperature drop

Figure 2.7 (cont.)

equivalent to 0.35 *M* guanidine requires only 0.1 m*M* dodine. (C) Thermal denaturation monitored by circular dichroism (mean residue ellipticity at 222 nm). 0.1 m*M* dodine (red) produces a similar 7 °C drop in melting temperature as 0.35 *M* guanidine (blue). (D) Pressure denaturation monitored by Trp fluorescence. 0.1 m*M* dodine (red) leads to earlier unfolding than 0.4 *M* guanidine (blue), but a smaller wavelength shift. All solid lines are approximate two-state fits to data, and markers show experimental data.

Figure 2.7A shows that 0.35 M guanidine-assisted thermal denaturation of PGK produced a 5 °C drop in the melting temperature of PGK. Dodine produced a comparable drop in T_m at only 0.1 mM (Figure 2.7B). The cooperativity of the thermal unfolding transition is preserved for concentrations of dodine between 0.05 and 0.2 mM. Hence dodine is an effective assistant of thermal denaturation at concentrations over 1000-fold lower than those of guanidine. Again, guanidine yielded a ≈ 5-10 nm larger red shift than dodine.

Circular dichroism detection (Figure 2.7C) yielded results similar to the ones seen using tryptophan fluorescence spectroscopy. 0.1 mM dodine is able to sufficiently destabilize the protein producing a melting curve similar to 0.35 M guanidine. The loss of ellipticity during guanidine assisted thermal denaturation is larger (47%) than that by dodine (34%). This agrees with the fluorescence-detected isothermal and thermal melt data shown in Figures 2.2 and 2.7A-B: guanidine results in slightly larger change in the fluorescence and CD spectra of denatured PGK than dodine.

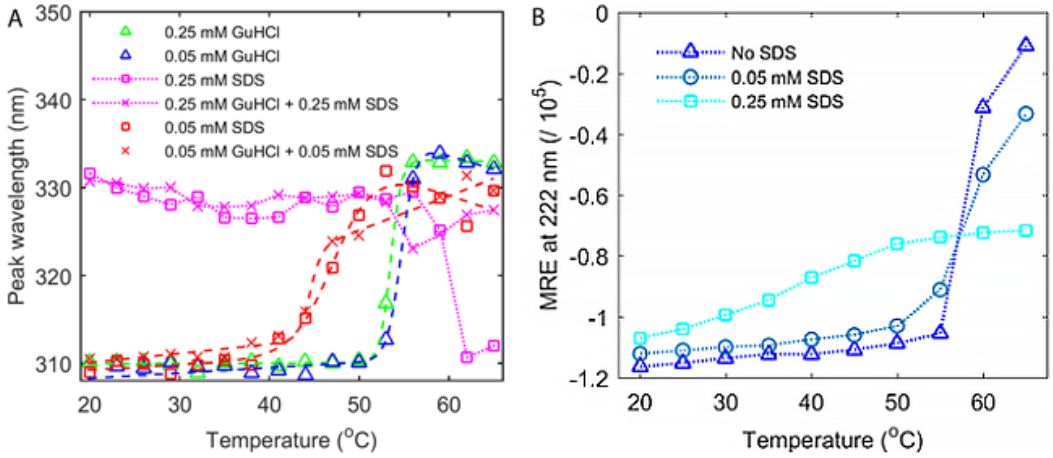


Figure 2.8: Control experiments with SDS-assisted temperature denaturation up to 60 °C monitored by (A) wavelength peak shift of tryptophan fluorescence and (B) circular dichroism. At low concentration, SDS thermal unfolding curves are cooperative and show significant destabilization of PGK in (A). This conflicts with the observation in (B), where SDS at 0.05 mM resembles PGK unfolding in no denaturant, albeit the loss of structure with SDS on unfolding is lower as seen by the smaller change in MRE (~80000 for SDS vs. ~110000 for no denaturant). At higher SDS (0.25 mM) the structure loss is significantly lower (cyan squares, (B)) and the unfolding transition is non-cooperative. Also apparent is that an equimolar mixture of SDS and guanidine does not show the same effect as dodine as seen with wavelength peak shift of tryptophan fluorescence. Markers and dotted lines represent experimental data. Dashed lines are two state fits to experimental data.

Control experiments (Figure 2.8) show that both SDS alone and a mixture of SDS and guanidine do not reproduce the effect dodine has on protein thermal stability at mM concentration. This further proves that denaturation by dodine is due to cooperative action of both its head group and its alkyl tail. While 0.05 mM SDS produced a cooperative transition, 0.25 mM SDS showed no cooperativity and in SDS thermal melts the protein is much less extensively denatured ($\lambda_{peak} \approx 330$ nm) than thermal melts assisted by dodine or even thermal melts of protein without denaturant. SDS thermal melts showed signs of aggregation at higher temperature, an effect that was absent with both dodine and guanidine at all temperatures. This interaction is discussed further in the next section.

Figure 2.7D summarizes dodine- *vs.* guanidine-assisted pressure denaturation detected by tyrosine and tryptophan fluorescence. 0.4 M guanidine results in a P_m of about 266 bar. In contrast, 0.1 mM dodine induces unfolding below 100 bar, but with a smaller red-shift in tryptophan fluorescence. The smaller red shift could be due to formation of an intermediate, or it could be due to solvation of the tryptophan sidechain by dodine during pressure denaturation, resulting in a reduced dielectric constant near the tryptophan. Dodine hence facilitates PGK denaturation with pressure at sub-mM concentrations, but the denatured state has more structure than the guanidine denatured state, or is preferentially solvated by dodine instead of water. The mutant PGKY122W closely mimics the thermal unfolding behavior of the wild type protein, except with a slightly lower stability (Figure 2.9).

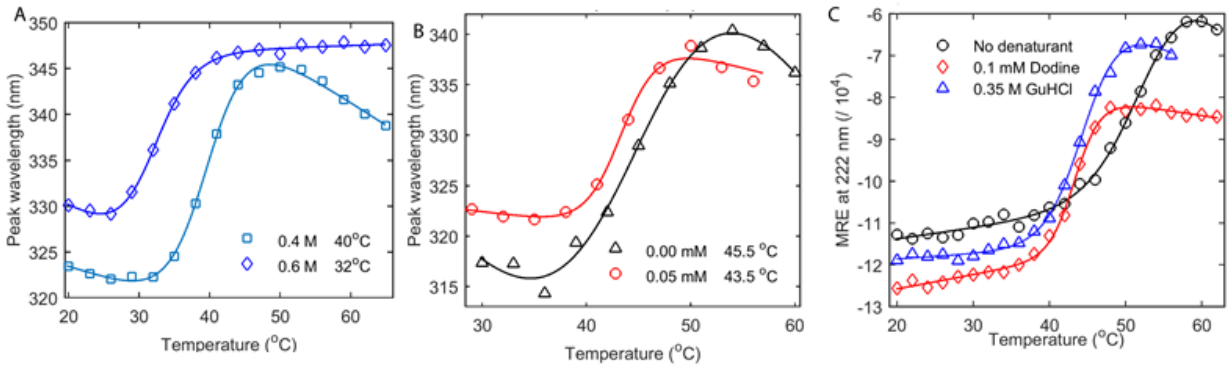


Figure 2.9: Thermal melts of PGK Y122W. Solid lines are fits to data and markers show experimental data. (A) Thermal melt assisted by guanidine and monitored by tryptophan fluorescence. Some cold denaturation of the mutant is evident at both guanidine concentrations because the baseline of 315 nm is not reached (T_m=45°C in 0 M guanidine in (B), T_m=40°C in 0.4 M guanidine, T_m=32°C in 0.6 M guanidine). (B) Thermal melt assisted by dodine and monitored by tryptophan fluorescence. Slight cold denatration of the mutant PGK also occurs in 0.05 mM dodine (T_m=43°C in 0.05 mM dodine). (C) Thermal melts monitored by circular dichroism. 0.1 mM dodine and 0.35 M guanidine cause the same decrease of stability as monitored by CD.

We believe that a substantial fraction of the missing red-shift is explained by preferential dodine binding to the protein and to the exposed tryptophans, just as guanidine is thought to bind to protein side chains and/or backbone.[317,318] To show that this hypothesis is at least consistent with the data, we took the simple 1-step binding model

$$P + nD \overset{K_{eq}}{\Longleftrightarrow} PD_n,\ K_{eq} = e^{\frac{-\Delta G}{RT}} \qquad [2.10]$$

and solved it analytically for the relationship between D_0 and P_0, the total concentrations of dodine and protein at the unfolding midpoint in Figure 2.2B. The methods section describes details of the model. Figure 2.10 shows midpoint of denaturation as a function of protein concentration. On fitting this to a simplified model of denaturant binding to protein (see above), a total of n~38±5 binding sites were found to fit the observed slope roughly. Even though the model does not account for (anti)cooperativity, it is able to predict the approximate number of dodine molecules that bind to the protein. Since the number of dodine per protein (~200) is not in large excess of the bound number (~40) the concentration mid-point of denaturation is a function of protein concentration. This does not occur with the weaker denaturants urea or guanidine, where there are >200,000 times more denaturant molecules than protein present, far more than would be required to completely solvate the protein surface with a full layer.

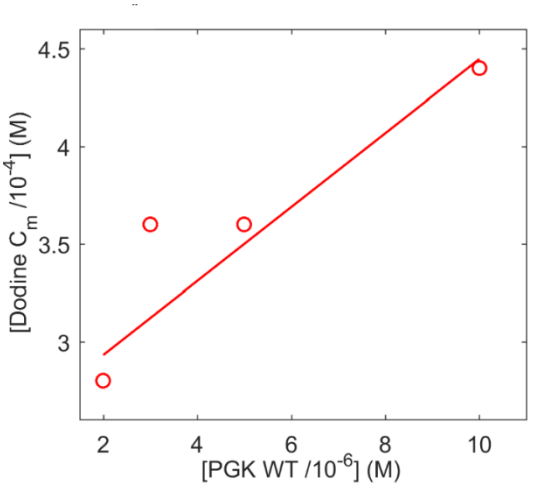


Figure 2.10: Dodine binding C_m as a function of protein concentration. Circles represent experimental data whereas solid line represent fit to experimental data. Fits yielded a value of ~38 for number of binding sites.

This simple model, without multiple binding steps, cooperativity, or anti-cooperativity of dodine binding (i.e. constant ΔG), was sufficient to reproduce the trend in Figure 2.2B (Figure 2.10). Higher protein concentration requires higher dodine concentration to push the equilibrium so that enough dodines are

available to solvate and denature the protein. According to the fit in Figure 2.10, the ≈200:1 excess of dodine over protein observed at the unfolding midpoint is required such that ≈30 to 40 dodines solvate and denature each protein. Given the simplicity of the model, this should be taken as a rough estimate only.

We also tested the reversibility of thermal denaturation of PGK with dodine and guanidine. Thermal denaturation with guanidine as the assisting destabilizer is reversible, whereas with dodine it is not when the temperature is raised above 50 °C (Figure 2.11). The same trend in reversibility was observed for isothermal denaturation of PGK (Figure 2.5) also described above.

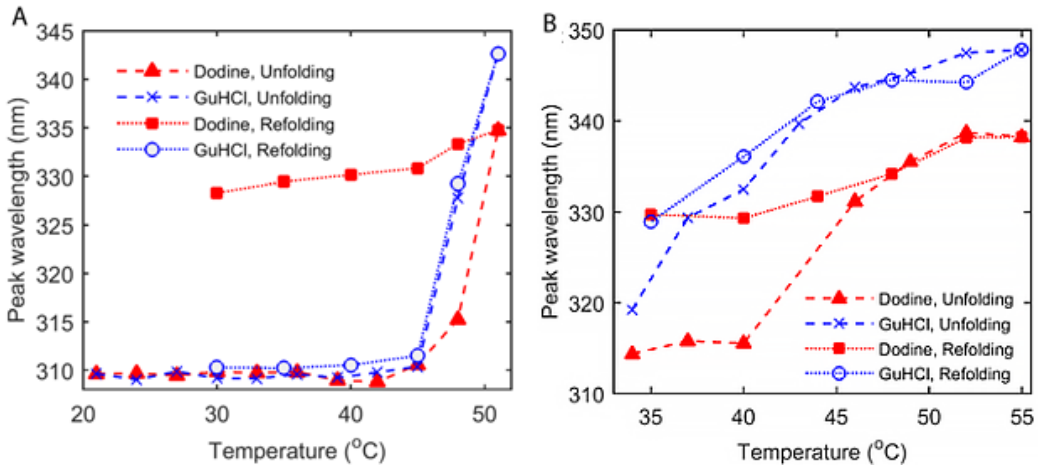


Figure 2.11: Thermal reversibility of (A) PGK wild type with guanidine and dodine and (B) PGK Y122W with guanidine and dodine. Like isothermal reversibility both guanidine thermal melts are mostly reversible while dodine thermal melts are mostly irreversible.

Hysteresis measurements showed that no reversibility was observed over a time scale of 1 hour showed that thermal denaturation of PGK in dodine is irreversible on those timescales (Figure 2.12).

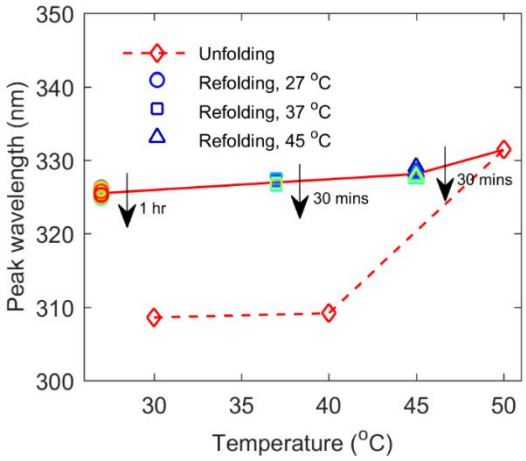


Figure 2.12: Hysteresis of thermal refolding of PGK wild type with dodine (0.1 mM). Bluer colors show earlier times and black arrows total experiment time. Very little recovery was observed at any temperature over time scales of 1 hr (27 °C) or 30 minutes (37 °C and 45 °C). Refolding may be possible, but a very long equilibration time, on the order of days, may be required.

2.3.4 INTERACTION OF SDS WITH PHOSPHOGLYCERATE KINASE

Thermal denaturation of PGK assisted by SDS as destabilizing agent shows an interesting feature. At low SDS = 0.05 mM a cooperative transition is visible as temperature increases (Figure 2.7A). The destabilization evident with this SDS concentration in Figure 2.8A, however, conflicts with the circular dichroism reported melting transition (Figure 2.8B). With 0.05 mM SDS CD the melting temperature reported by overall loss of α-helical structure correlates well with PGK in the absence of SDS, showing no significant destabilization. This can be explained by reported observations present in current literature on SDS mechanism of protein denaturation. The initial peak shift in tryptophan fluorescence may be due to an initial expansion of protein to form a partially denatured hyperfluorescent state that prematurely exposes the tryptophan residues.[319,320] Tryptophan fluorescence, hence is sensitive to folding intermediates that may form. Overall loss of structure is lower in the presence of SDS as shown by both peak shift and change in CD MRE (~80000 for 0.05 mM vs. ~110000 for no SDS). This may be attributed to some residual structure due to protein/SDS mixed micelles, if protein favors micelle formation at sub-CMC detergent concentration. All the experiments here were carried out well below the SDS critical micelle concentration in absence of protein, but protein may nucleate micelles.

It is interesting to note that the (partly) unfolded protein at 0.25 mM SDS seems to retain more structure as seen from the low overall final peak wavelength of 330 nm (Figure 2.8A). This observation was further reinforced by the CD measurements where the loss in MRE with 0.25 mM SDS was half of the overall loss without denaturant (Figure 2.8B). Furthermore, while fluorescence peak shift stays constant over the temperature range, CD shows a non-cooperative linear structure loss with increase in temperature bolstering

the argument that tryptophan fluorescence is perhaps not able to accurately describe global unfolding of PGK in SDS. It has been shown using SAXS that membrane proteins sometimes embed in dodecyl maltoside micelles.[291] It is possible that the protein PGK becomes embedded in SDS micelles in an intermediate state that initially shows a shift to a higher wavelength, mimicking unfolding. The protein then unfolds in this hydrophobic micellar pocket. This would lead to the disappearance of the tryptophan fluorescence shift expected on unfolding as the residues are no longer solvent exposed but are trapped in the hydrophobic micellar pocket.

It is also possible that the protein is stabilized by SDS in an intermediate state (λ_{peak} = 330 nm) that is resistant to unfolding but aggregates at higher temperatures as was observed for the 0.25 mM SDS case. [291]

2.3.5 M-VALUE DETERMINATION FROM SINGULAR VALUE DECOMPOSITION (SVD) ANALYSIS

Peak wavelength shift nicely defines the midpoint of denaturation transitions, but it is not even approximately linearly related to folded and unfolded protein concentrations. We use this probe in the above discussed results because it is a good intuitive measure of T_m, C_m and P_m. However, the calculation of *m*-values requires a more linear probe. (The *m*-value is the slope of the unfolding curve at the transition midpoint; a higher slope indicates more cooperativity of the transition.) Singular value decomposition (SVD) of the fluorescence spectra provides a more linear measure of the contribution of various states to the fluorescence signal, and was used to extract *m*-values.

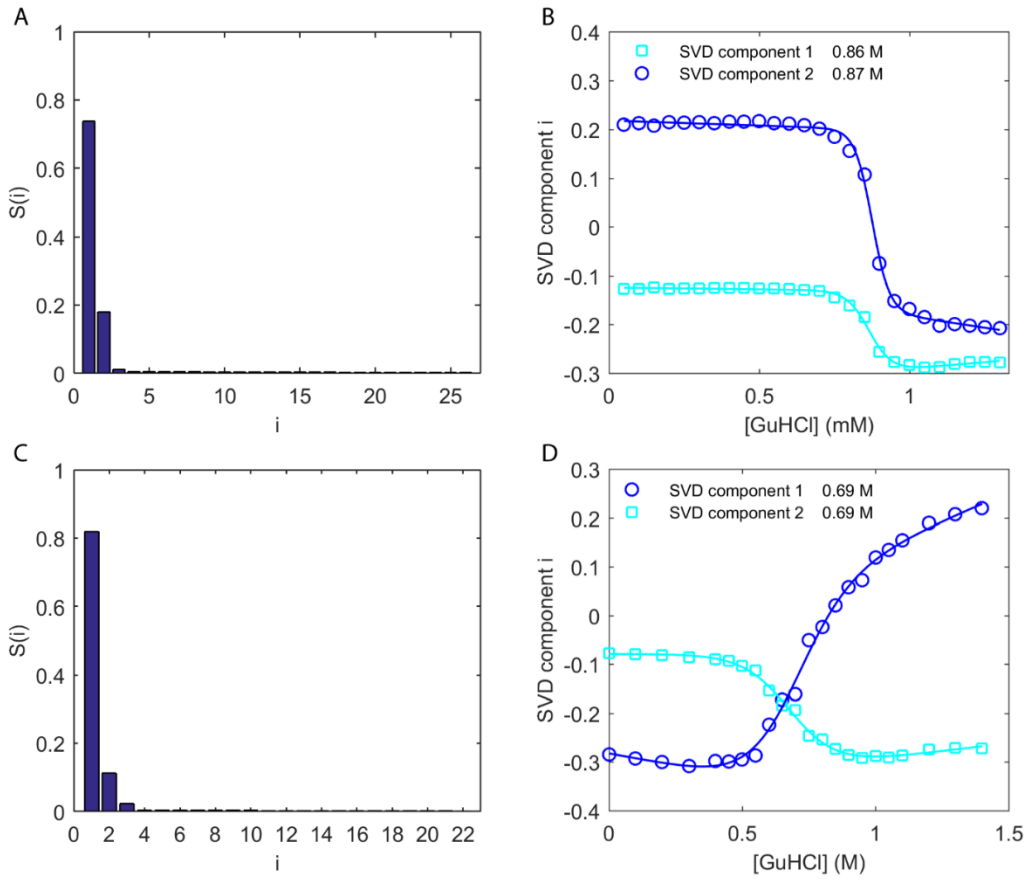


Figure 2.13: Singular value decomposition (SVD) of PGK wild type and PGK Y122W isothermal titration with guanidine. (A) and (C) are histograms showing weight of each basis component obtained from SVD of PGK wild type and PGK Y122W respectively. (B) and (D) show evolution of SVD basis vectors 1 and 2 for PGK wild type and PGK Y122W with temperature. The titration of PGK is well defined by consideration of SVD componenets 1 and 2 for both wt and mutant. The C_m of titration obtained by both components is in close agreement with analyzed fluorescence data.

Figure 2.13 and 2.14 show the SVD analysis for the isothermal and thermal denaturation of both wildtype and mutant PGK.

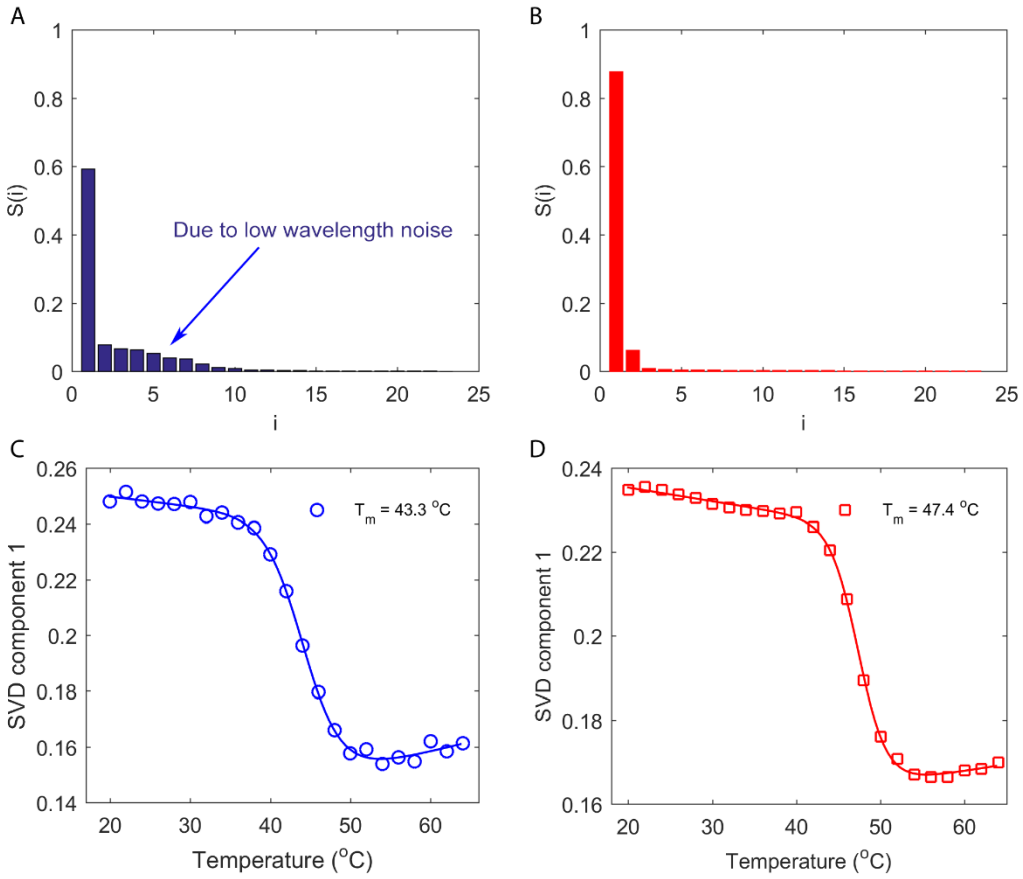


Figure 2.14: Singular value decomposition (SVD) of PGK wild type circular dichroism probed thermal titration with guanidine and dodine. Histograms showing weight of each basis component obtained from SVD of PGK wild type with (A) guanidine and (B) dodine and evolution of SVD basis vector 1 with (C) guanidine and (D) dodine with temperature. Vector components higher than the second in (A) arise from large noise below 210 nm with guanidine, and are absent in (B) with dodine, another indication that the dodine denaturation produces cleaner spectra for data analysis.

It is apparent even in SVD-analyzed experiments that the width of transition is larger for the mutant (Figure 2.13). This may be due to the additional tryptophan in the N-terminal domain for the Y122W mutant. It has been shown that the two domains do not unfold completely in concert, with the C-terminal unfolding before the N-terminal domain.[321] The broadening of the transition could be caused by both domains unfolding at slightly different concentrations of guanidine, making the transition appear less sharp.

The *m*-values (Table 2.1) obtained from SVD analysis of the isothermal titration of both mutants with guanidine are smaller for the mutant (~29 kJ/mol/M), as expected from the decrease in sharpness of the transition. The values obtained for the wild type (~60 kJ/mol/M) are larger than that reported in literature (~40 kJ/mol/M).[296] This could be due to differences in solvation conditions (e.g. buffer used).

Table 2.1: *m*-values and δg_1 for PGK wild type and PGK Y122W with guanidine obtained from singular value decomposition (SVD) of fluorescence-detected thermal and isothermal guanidine titration, as well as circular dichroism-detected thermal titration. These values are more accurate than what would be obtained from peak wavelength shift data.

	Tryptophan fluorescence				Circular dichroism
	Isothermal guanidine titration *m*-value (kJ/mol/M)		Thermal titration δg_1 (kJ/mol/K)		Thermal titration δg_1 (kJ/mol/K)
	SVD component 1	SVD component 2	SVD component 1	SVD component 2	SVD component 1
PGK wild type	61	67	1.8	2.0	1.4
PGK Y122 W	29	24	1.1	1.1	1.1

2.3.6 DODINE IS TRANSPARENT IN THE ULTRAVIOLET AND INFRARED REGIONS AT ITS WORKING CONCENTRATION

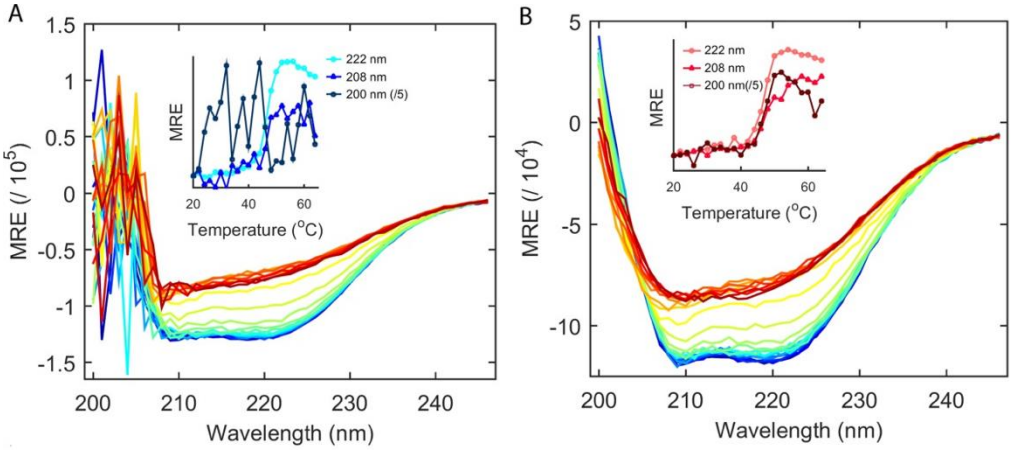


Figure 2.15: Circular dichroism in dodine (red) and guanidine (blue) solutions. (A) Circular dichroism thermal denaturation with protein in 0.35 *M* guanidine. Bluer colors show lower temperatures. Inset shows large noise due to guanidine absorption for the MRE at 200 nm (dark blue) and 208 nm (blue) compared to 222 nm (cyan). (B) Circular dichroism thermal denaturation with dodine. Traces show no absorption due to dodine. Inset shows low noise in MRE at all wavelengths 200 nm (dark red), 208 nm (red) and 222 nm (orange).

Figure 2.15 shows the circular dichroism spectra of PGK wild type taken with 0.35 M guanidine and 0.1 mM dodine. The strong absorbance of guanidine in the <210 nm region (Figure 2.15A) makes it difficult to observe the temperature dependence of the circular dichroism signal around 208 nm (Figure 2.15A, inset), an important signature of α-helical structural element of the protein. A much lower concentration of dodine is required to induce the same amount of destabilization that allows the collection of high-quality circular dichroism spectra down to 200 nm (Figure 2.15B).

The mutant PGKY122W shows similar behavior when circular dichroism spectra were collected with 0.35 M guanidine and 0.1 mM dodine (Figure 2.16).

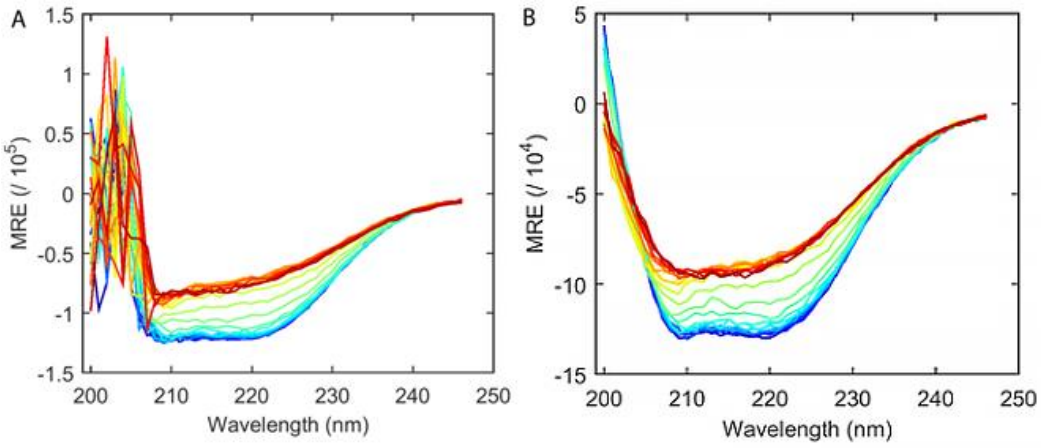


Figure 2.16: CD spectroscopy of PGK Y122W with guanidine (A) and dodine (B). Bluer colors show lower temperatures in the range 20 to 60 °C. (A) Circular dichroism thermal denaturation in 0.35 M guanidine. (B) Circular dichroism thermal denaturation with 0.1 mM dodine. While guanidine absorption saturates the signal below 208 nm and makes data acquisition impossible, whereas a high-quality CD spectrum is acquired down to 200 nm in dodine.

Figure 2.17 compares infrared spectra of PGK wild type in the amide 1' with and without GuDCl or dodine denaturants.

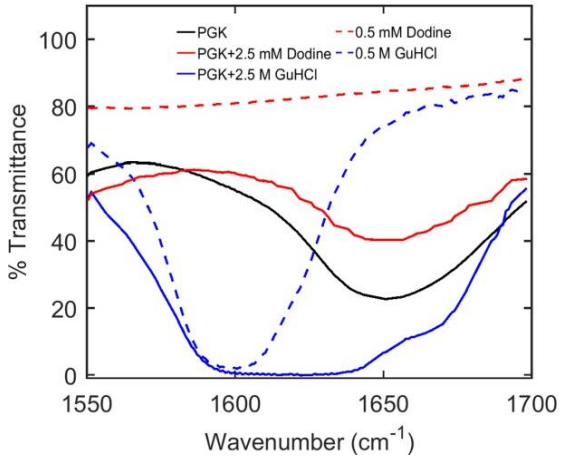


Figure 2.17: Fourier transform infrared spectroscopy in dodine (red) and guanidine (blue) solutions. FTIR spectra of protein (black) with guanidine (blue) and dodine (red), smoothed over 5 cm^{-1}. Dashed curves show absorption peaks of denaturant without protein. Saturation of the amide I′ peak can be seen in the blue curve due to guanidine.

While protein spectra taken in 0.5 mM dodine clearly exhibits the full shifted amide I' peak with different intensities, spectra taken with GuDCl show obscuration of the amide I' peak due to strong guanidine absorption that saturates transmission in the 1590 to 1640 cm^{-1} spectral range. Dodine's transparency in both ultraviolet and infrared regions could be an important tool for secondary structure determination where a destabilization with a denaturant is required without introducing noise in the spectra.

2.3.7 SMALL CHANGES IN DODINE CONCENTRATION CAN AFFECT ACCURACY

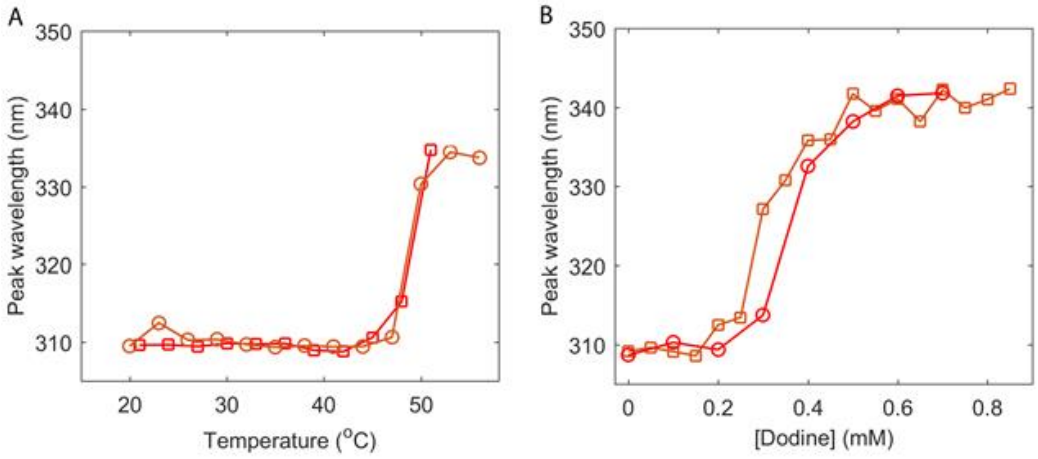


Figure 2.18: Reproducibility of dodine denaturation. Since dodine experiments are very sensitive to small fluctuations in dodine concentration that may be a result of change in volume or evaporation experimental fluctuations need to be minimized as far as possible. (A) Good agreement between two sets of dodine assisted thermal denaturation of PGK wild type. (B) Small fluctuations in C_m due to fluctuations in concentration in the order of 0.05 mM, also with wild type PGK.

Figure 2.18 shows two sets of data, utilizing different protein preparations, but the same concentrated dodine stock solution. Both thermal and isothermal titrations with dodine are sensitive to small changes in dodine concentration. Since dodine is used from an ethanol stock, small changes in dilution can be introduced. During our experiments, only one concentrated dodine stock was utilized, so the relative results are accurate. Figure 2.18A shows good agreement in temperature unfolding data of PGK wild type. Figure 2.18B shows a small shift of C_m due to experimental error. Since dodine acts at sub-millimolar concentrations, special care has to be taken to make sure its concentration is measured accurately.

2.4 DISCUSSION

Dodine can be used to destabilize proteins for thermal or pressure denaturation at concentrations over 1000-fold lower than the widely used nitrogen-based denaturants. Dodine retains cooperativity when its concentration is titrated, although cooperativity is greatest when the protein:dodine concentration ratio is the smallest possible. The unfolding by guanidine is more extensive as seen by the overall peak shift

(fluorescence) or spectral magnitude change (circular dichroism) upon unfolding. In contrast, dodine provides more rapid transition upon pressure denaturation. This could potentially be a useful tool for detecting pressure denaturation at low applied pressure, if it holds up during future studies of other proteins. It remains to be seen whether the smaller tryptophan fluorescence wavelength shift is caused mainly by more structured denatured states, or mainly by preferential dodine solvation of the tryptophan sidechain. Although dodine yields well-reproducible denaturation curves, it is not as reversible as guanidine, which generally achieves 80% reversibility for PGK over a short period of tim*e*. This is typically the case for amphiphilic denaturants. However, dodine can be used to obtain high quality CD and IR spectra in wavelength regions not accessible with guanidine.

CHAPTER 3
DODINE AS A KOSMO-CHAOTROPIC AGENT[3]

3.1 INTRODUCTION

As we saw in chapter 2 for the dodine-PGK interaction, dodine can be an effective denaturant at millimolar concentrations. PGK is a large multidomain protein. A smaller protein would allow us to probe, in greater detail, dodine-protein interactions by combining both simulations and experiments. In this chapter we do so by using the small helical protein domain Protein B and examining its interaction with dodine experimentally and by analyzing molecular dynamics simulations. Simulations were carried out by S. Mittal, Experiments by D. Guin, and analysis was performed by both together.

Dodine, commonly used as a fungicide,[322,323] contains a dodecyl tail and a guanidinium headgroup giving it both denaturant and detergent properties. Earlier studies with dodine as a protein denaturant showed that dodine acts at millimolar concentrations.[295,324] Unlike urea and guanidinium-Cl that act at molar concentrations,[285,325,326] dodine's millimolar range of action is especially useful when high denaturant concentration interferes with spectroscopic probes such as circular dichroism (CD), ultraviolet (UV) and infrared (IR) absorption.[327] Detergents like sodium dodecyl sulfate (SDS) also unfold proteins at millimolar concentration, but dodine acts much more cooperatively:[290] susceptible proteins unfold fairly suddenly, not gradually, when dodine is added.

Dodine exhibits some unusual behavior as a denaturant. For example, it denatures the α-helical protein λ-repressor fragment and the mixed α, β protein phosphoglycerate kinase, but it has no discernible effect on the β-sheet protein Fip35 WW domain.[295,324] Thus the question remains how dodine interacts microscopically with proteins, sometimes acting to destabilize, sometimes not. It seems plausible that the Janus-like denaturant headgroup and long aliphatic tail of dodine could interact very differently with the polypeptide backbone and hydrophobic amino acid sidechains of the core thereby facilitating this unusual behavior.

[3]This chapter reprinted with permission from Drishti Guin, Shriyaa Mittal, Brian Bozymski, Diwakar Shukla and Martin Gruebele. Dodine as a kosmo-chaotropic agent. J. Phys. Chem. Lett., 10, 2600-2605, 2019. DG and SM contributed equally to this work.

Here we combine experimental spectroscopy and all-atom molecular dynamics (MD) simulation to dissect how different parts of dodine interact with a protein. We probe the interaction of dodine with the small three α-helix bundle protein B (PrB). PrB is the albumin-binding domain from the bacterial surface protein PAB.[328] PrB was chosen for two main reasons. Firstly, its small size facilitates the characterization of the dodine-protein interaction by experiment and MD simulation. Secondly, previous studies have shown that dodine denatures α-helical proteins more readily than β-sheet proteins.[295] Since PrB is predominantly α-helical, it serves as a good model to probe the specific interaction of dodine with helices and a small hydrophobic core. We use intrinsic tryptophan fluorescence[329] and far-UV CD[305,330] to probe the integrity of local tertiary structure and global secondary structure. Since PrB lacks any intrinsic tryptophan residues that report on protein unfolding, we engineered a mutant PrB21 with a tryptophan (W) in helix 1 and a tyrosine (Y) in helix 2 (Figure 3.1A, inset). In the folded state, longer-wavelength W fluorescence is quenched by Y, which fluoresces at shorter wavelength.[331] When the two residues move apart during unfolding, W fluorescence is restored.

3.2 METHODS

3.2.1 FLUORESCENCE AND CIRCULAR DICHROISM MEASUREMENTS

The PrB21 protein was synthesized by WatsonBio, LLC. and was stored lyophilized at -20 °C. The protein sequence was the same as used previously for simulations by Shaw *et. al.*[332] except with one point mutation L12W. The final sequence is:

LKNAIEDAIA EWKKAGITSD FYFNAINKAK TVEEVNALVN EILKAHA

For experiments a small amount of protein was dissolved in water and absorbance at 280 nm (A_{280}) was measured using a UV spectrophotometer (Shimadzu UV-1800). The concentration of the dissolved protein was measured using Beer Lambert's law using A_{280} assuming an extinction coefficient of 6990 M^{-1} cm^{-1}. The extinction coefficient was estimated using ExPASy server tool ProtParam.

Tryptophan fluorescence measurements were performed on a FP-8300 spectrofluorometer equipped with a Peltier temperature controller (JASCO). Tryptophan was excited at 285 nm and fluorescence spectra was collected from 280 to 450 nm. All experiments were performed with 5 μM protein in 20 mM tris buffer (pH 7.0) in a 300 μL cuvette. Tris buffer was used because of dodine precipitated in other tested buffers (phosphate and acetate). Dodine was purchased as the acetate salt from Sigma-Aldrich 45466. Because dodine solubility in water is very low, a stock was prepared at 24.87 mM in 100% ethanol. For experiments small amounts of dodine stock were diluted in water at the desired concentration. For experiments with

GuHCl (Sigma-Aldrich), a 10 M GuHCl stock was prepared in water. Since 10 M GuHCl precipitates in water at room temperature, the stock was warmed to ensure solubility before each experiment.

Circular dichroism was measured using a J-715 spectropolarimeter with Peltier temperature control (JASCO). Unless otherwise noted, all spectra were recorded from 240 to 200 nm at a scan rate of 50 nm min^{-1} at 1 nm resolution by averaging five accumulations. All experiments were performed in 2 µm cuvettes with 5 µM protein concentration.

Above 1 mM, dodine formed micelles at higher temperatures and an unfolding curve could not be extracted from fluorescence data due to aggregation.

3.2.2 DATA ANALYSIS AND CURVE FITTING

Data was analyzed using MATLAB (MathWorks). Tryptophan fluorescence spectrum was analyzed by calculating the peak wavelength center of mass (λ_{Peak}) change as a function of perturbant (temperature and denaturant). λ_{Peak} was estimated by fitting the maximum of the spectrum to a polynomial of degree 2.

Melting temperature (T_m) and denaturation midpoint (C_m) were calculated using a two-state sigmoidal fit to the experimental data. Native and denatured state baselines were assumed to be linear with respect to the perturbing variable X (temperature or denaturant concentration) so that the signal S can be estimated by:

$$S_i = m_i(X - X_m) + b_i \quad [3.1]$$

Where, i is either the native (N) or the denatured (D) state, m is the slope and b is the intercept. X_m is the corresponding midpoint of the curve with the perturbant X. The total signal $S(X)$ can then be estimated as a sigmoidal function with respect to perturbant X:

$$S(X) = S_N(X)f_N(X) + S_D(X)f_D(X) \quad [3.2]$$

Where, f_N and f_D are the populations of the native and denatured states respectively

$$K_{eq} = e^{-\frac{\Delta G_{N\rightarrow D}}{RT}}, \quad [3.3]$$

$$\Delta G_{N\rightarrow D} \approx \delta g_1(X - X_m), \quad [3.4]$$

$$f_N = \frac{K_{eq}}{1+K_{eq}}, \quad [3.5]$$

$$and f_D = \frac{1}{1+K_{eq}} \quad [3.6]$$

Fits to equation 3.2 yield the midpoint value (X_m), either for temperature (T_m) or denaturant (C_m).

3.2.3 MOLECULAR DYNAMICS SIMULATIONS

The three-dimensional (3D) coordinates of the PrB was used as a starting structure for MD simulation. PyMol Mutagenesis Wizard was used to mutate the protein sequence to match experimental sequence[333] except an additional mutation at position 39, V39K. Hence, MD simulations results reported are for a PrB21 V39K mutant of PrB (mentioned above). We performed an additional 5 μs simulations with PrB sequence (consistent with that used in experiments for this study). Results from these MD simulations are shown in appendix B. These figures provide a comparison between the observations reported for PrB21 and its V39K mutant. Our conclusions are consistent with the MD simulation results from simulations performed on PrB21 V39K mutant.

The tleap program in AmberTools14 was used to build the MD systems.[334] Protein chain termini were capped with neutral acetyl and methylamide groups. The protein was solvated in a box containing TIP3P water molecules in a periodic box of size 56.83 × 53.062 × 51.585 $Å^3$.[335] A salt (NaCl) concentration of 0.15 M was used to neutralize the MD system leading to an MD system with 14,023 atoms. The standard protonation states were used for the titratable groups.

The structure-data file for a dodine molecule was obtained from the PubChem Compound Database.[336] ACPYPE script was used to generate the dodine molecule topology using General Amber Force Field of AMBER14.[337] A previous study used a similar method for the dodine parameters.[338] The solvated protein MD system was used as an input to Packmol which randomly arranged dodine molecules around the PrB molecule.[339] An MD system with 10 dodine molecules (0.16% dodine/water ratio and concentration 0.087 M) consisted of 19,844 atoms. Another MD system with 20 dodine molecules (0.31% dodine/water ratio and concentration 0.175 M) consisted of 20,378 atoms.

The MD systems were energy minimized for 20,000 steps using the conjugate gradient method, heated from 0 to 300 K and equilibrated for 10 ns. The MD simulations were performed in constant NPT conditions and 1 atm. The temperature was controlled using a Berendsen thermostat, and the pressure was maintained using a Berendsen barostat.[340] Long range electrostatic interaction was treated with the Particle Mesh Ewald method,[341] and bonds involving hydrogens were constrained using the SHAKE algorithm.[342]

The nonbonded distance cutoff was set to 10 Å, and an integration step of 2 fs was used. All simulations were performed using the AMBER FF14SB force field.[343]

The solvated protein MD system (with no dodine) was simulated for 500 ns at 300 K to compare with the dodine simulations. The 0.087 M dodine MD system was simulated at 320 K, 340 K and 360 K. And, the 0.175 M dodine MD system was simulated at 360 K. Each of their production run simulations comprised of 5 replicas – each of which is 500 ns long. The solvated protein MD system (with no dodine) was also simulated for 100 ns at 400 K to obtain an unfolded structural ensemble of PrB. The largest radius of gyration from this ensemble was used as a starting point for 500 ns simulation of the unfolded protein at 300 K temperature to observe relaxation of the protein from unfolded structure obtained at the higher temperature. Total accumulated simulation time was ~11.1 μs and snapshots were saved at a frequency of 100 ps.

MDTraj 1.9.1 and CPPTRAJ V14.25 were used for the analysis of MD simulations.[344,345] Analysis and graphing scripts are written in the Python programming language. VMD was used to visualize MD simulation results.[346]

3.2.4 CONTACT COEFFICIENT

The contact coefficient calculations are adapted from the method described by Stumpe and Grubmüller to quantify frequency of interaction between urea and an amino acid.[347] Results for contact coefficient are shown where a contact is defined as atoms within 0.35 nm. Similar calculations were performed with a cutoff of 0.4 nm and similar distributions were observed. The contact coefficient was normalized to the maximum value separately for the head and the tail groups, respectively. The coefficients were then subtracted so that positive values preferentially interact with the head group and vice versa. Values closer to zero show no preference for either the head or the tail group. Hydrophobicity values are based on Kyte-Dolittle hydrophobicity scale.[348]

3.2.5 HELICAL CONTENT

The helical content of the 3 helices are calculated as defined in the NAMD 2.11 manual.[349] The python implementation is taken from https://github.com/amoffett/alpha_helical_content as used in ref. 19. The individual helices for the analysis in this work are defined as residues 4-15 (helix 1), residues 19-27 (helix 2), and residues 31-45 (helix 3) shown in Appendix B.1.

3.3 RESULTS

3.3.1 DODINE DISRUPTS TERTIARY STRUCTURE AND MILDLY STABILIZES PRB21 BY INDUCING HELIX FORMATION

PrB21 denatures cooperatively as a function of dodine concentration, with a midpoint concentration C_m = 0.26±0.02 mM (Figure 3.1A) that is over 10^4 times smaller than for guanidinium-Cl (Figure 3.2A). We stay at ≤1 mM dodine to avoid micelle formation (see SI for further Methods details). Dodine addition increases the fluorescence wavelength, indicative of reduced quenching of W by Y and increased exposure of the W sidechain to a more polar solvent environment (W Stokes-shifts in aqueous solution).[350] Either of these events indicates loss of local tertiary structure of PrB21. However, λ_{Peak} reaches only 331±1 nm in dodine, *vs.* a Stokes shift to 350 nm when PrB is denatured with guanidinium-Cl (Figure 3.2A).[295,324] Thus the endpoint in Figure 3.1A is not a fully unfolded state.

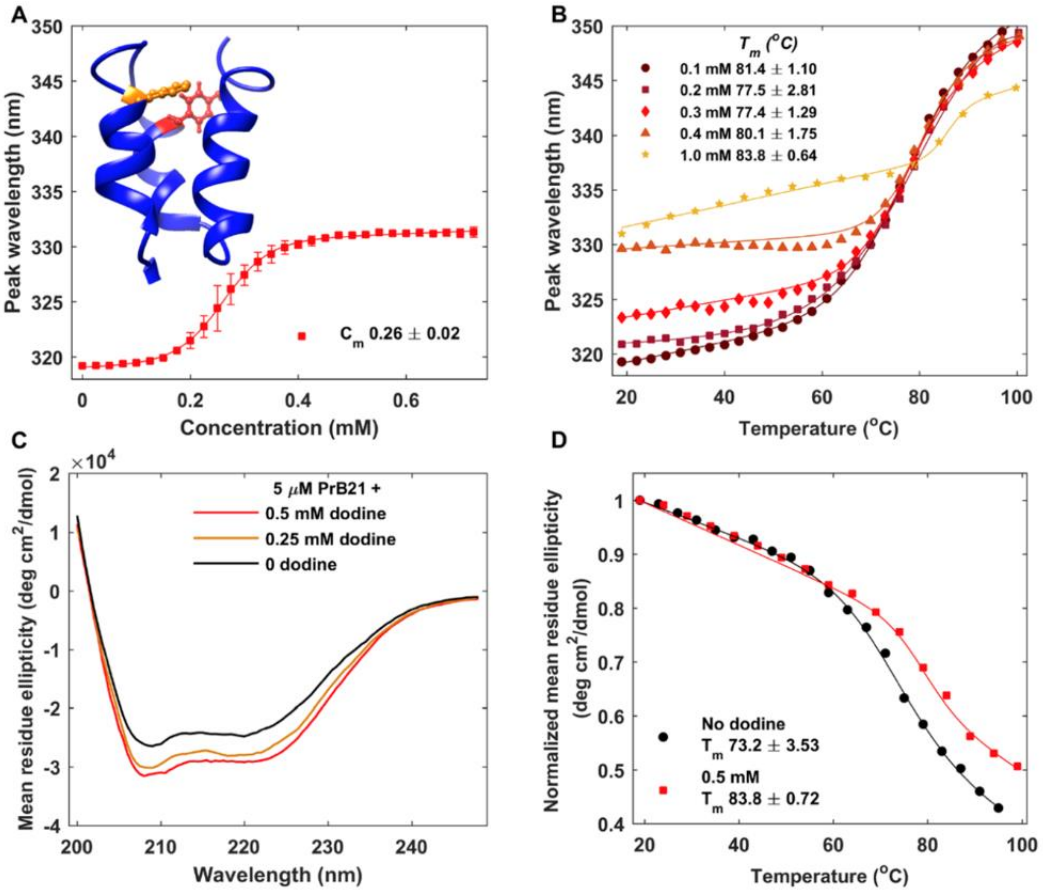


Figure 3.1: The effect of dodine on PrB21 stability. Solid markers show experimental data and solid lines show model fits in A, B and D. Solid lines in C show experimental data. Unfolding was monitored by change in λ_{Peak} as a function of denaturant concentration. Secondary structure was monitored by mean residue ellipticity of PrB21 at various dodine concentrations. Errors bars are reported by the standard error of the mean for two repeats. (A) Denaturation of PrB21 with dodine at 20 °C. Inset shows ribbon structure of PrB21

Figure 3.1 (cont.)

with the tryptophan residue on helix 1 highlighted in orange and tyrosine residue on helix 2 highlighted in red. (B) Dodine-assisted thermal denaturation of PrB21. (C) Circular dichroism spectra of native PrB21 and PrB21 in dodine. (D) Thermal denaturation of PrB21 without denaturant (black) and 0.5 mM dodine (red) monitored by mean residue ellipticity change as a function of temperature.

Unlike with PGK discussed in chapter 2 denaturation of PrB21 in both guanidinium-Cl and dodine at 20 °C is largely independent of protein concentration (Figure 3.3) and reversible (Figure 3.4) when monitored by fluorescence peak wavelength (λ_{Peak}) shift in buffer.

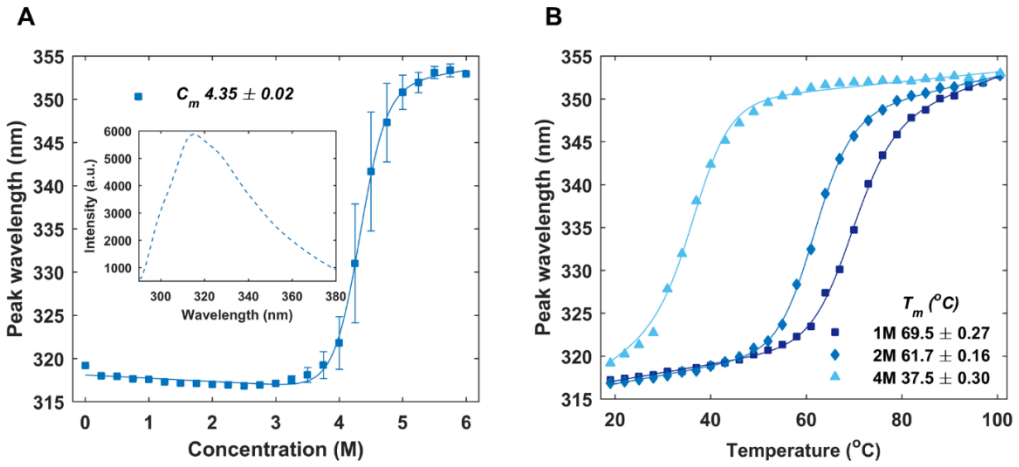


Figure 3.2: Denaturation of PrB21 assisted by GuHCl. Solid markers and dashed lines show experimental data and solid lines show fits to experimental data. Unfolding was monitored by change in λ_{Peak} as a function of perturbant (denaturant concentration or temperature). Errors bars are reported by the standard deviation for two repeats. (A) Isothermal denaturation of PrB21 with GuHCl. The C_m for denaturation of PrB21 with GuHCl fit to 4.35 M. The inset shows spectrum of native PrB21 shows a peak at 315 nm typical of tyrosine fluorescence and a shoulder at 325 nm typical of tryptophan fluorescence in a buried hydrophobic environment. On unfolding the λ_{Peak} shifts ~35 nm. (B) Thermal denaturation of PrB21 in tris assisted by 1 M (dark blue), 2 M (blue) and 4 M (cyan) GuHCl. GuHCl destabilizes PrB21 towards thermal denaturation. PrB21 without denaturant unfolds with a T_m of 78.9 °C (Figure 3.1B). In 1 M, 2 M and 4 M GuHCl PrB21 unfolds with a T_m of 69.5, 61.7 and 37.5 °C, respectively.

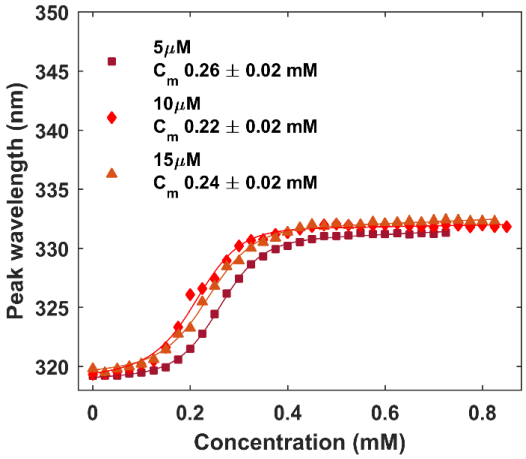


Figure 3.3: PrB21 denaturation is independent of protein concentration. Solid markers and dashed lines show experimental data and solid lines show fits to experimental data. Unfolding was monitored by change in λ_{Peak} as a function of denaturant concentration. Titration of 5 (maroon), 10 (red) and 15 μM (orange) PrB21 with dodine. All denaturation C_ms fit to ~0.25 mM and dependence of C_m on protein concentration was observed.

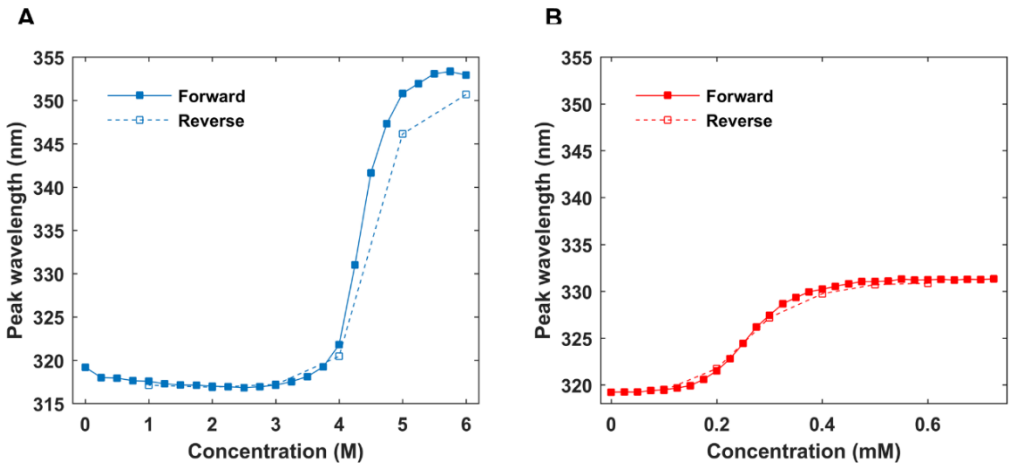


Figure 3.4: Reversibility of isothermal denaturation of PrB21 with (A) GuHCl and (B) dodine. Solid markers and solid lines are forward melts and open markers and dashed lines show reverse melts. Unfolding was monitored by change in λ_{Peak} as a function of denaturant. Denaturation of PrB21 with both GuHCl and dodine are completely reversible.

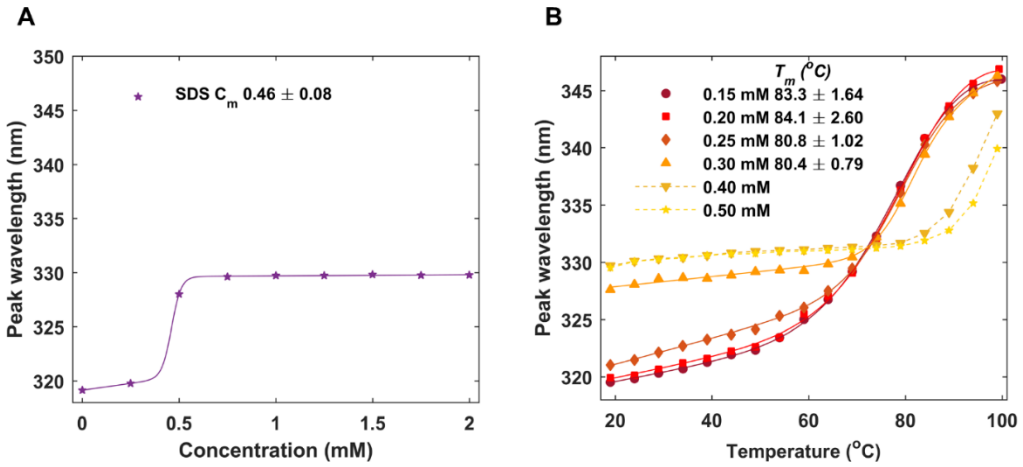


Figure 3.5: SDS-assisted isothermal and thermal denaturation of PrB21. Solid markers and dashed lines show experimental data and solid lines show fits to experimental data. Unfolding was monitored by change in λ_{Peak} as a function of dodine concentration or temperature. (A) Isothermal denaturation of PrB21 with SDS. The C_m fit to 0.46 mM and the final denatured state was partially unfolded with λ_{Peak} ~330 nm. (B) Thermal denaturation of PrB21 with 0.15 mM (maroon), 0.20 mM (red), 0.25 mM (dark orange), 0.30 mM (orange), 0.40 mM (dark yellow) and 0.50 mM (yellow) SDS. Between 0.15 mM – 0.30 mM SDS PrB21 unfolded with a T_m of 82±2 °C and no change in T_m was observed.

Is further unfolding possible? Thermal denaturation in Figure 3.1B answers this question. PrB21 unfolds cooperatively with a melting temperature (T_m) of 79 °C in the absence of any denaturant to $\lambda_{Peak} \approx 350$ nm (see methods for fit details). As expected, addition of guanidinium-Cl reduces T_m of PrB21 (Figure 3.2B). Dodine has no such effect: up to 0.4 mM dodine, T_m remains at 79 °C. At the highest concentration of dodine (1 mM), we even observe stabilization against thermal denaturation ($T_m \approx 84$ °C).

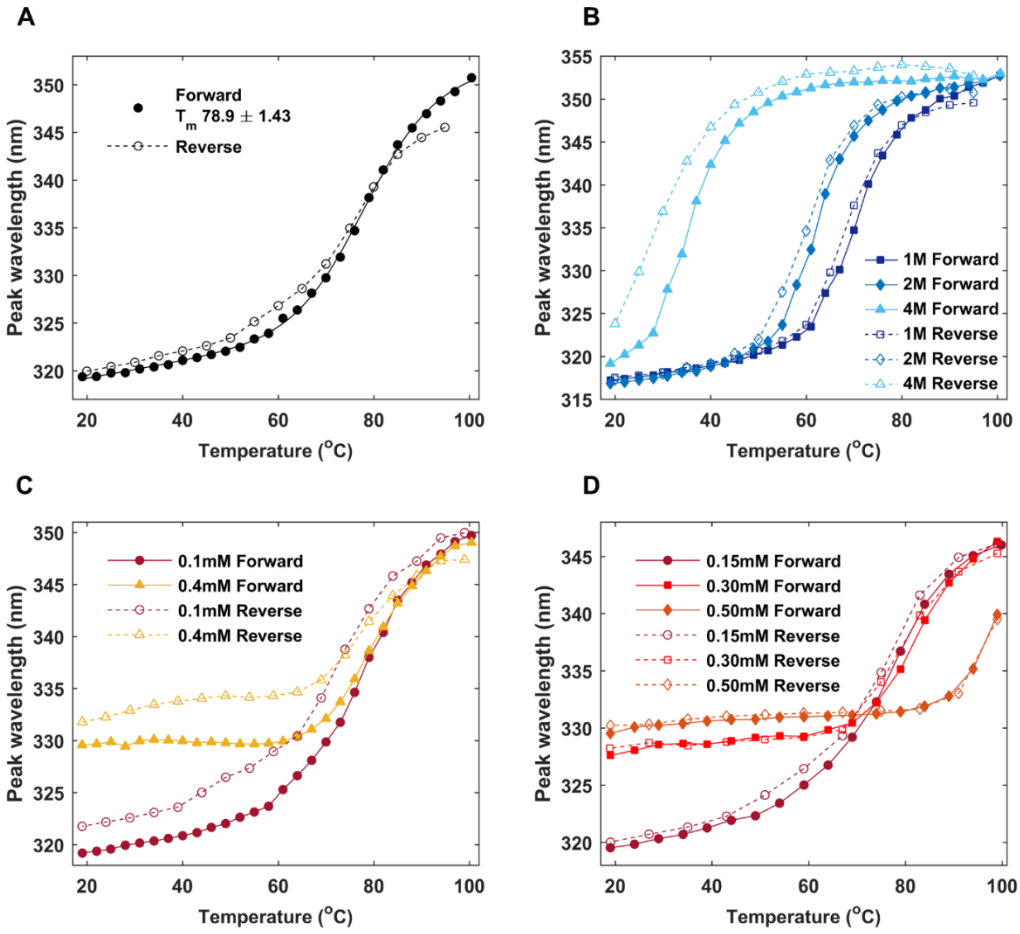


Figure 3.6: Reversibility of isothermal denaturation of PrB21 with (A) no denaturant, (B) GuHCl, (C) dodine and (D) SDS. Solid markers and solid lines are forward melts and open markers and dashed lines show reverse melts. Unfolding was monitored by change in λ_{Peak} as a function of denaturant. All thermal melts of PrB21 with and without denaturant are reversible to a large degree.

Instead of reducing T_m, addition of dodine raises the native state fluorescence baseline in Figure 3.1B to a longer wavelength, in accord with Figure 3.1A. Thus, dodine denaturation in Figure 3.1A and thermal denaturation in Figure 3.1B are distinct events. Dodine cooperatively attacks the tertiary structure of PrB21 without affecting the thermal stability of the residual tertiary structure. A similar effect was also observed with SDS (Figures 3.5 and 3.6). Higher SDS concentrations (0.4 mM and 0.5 mM) stabilize PrB21 such that unfolding is not completed at all temperatures tested and the curves are shifted to a higher T_m (Figure 3.5).

In contrast to denaturing local tertiary structure, dodine increases global helical content when probed by CD (Figure 3.1C). In addition, dodine stabilizes helical structure against thermal denaturation (Figure 3.1D). We hypothesize that the gain in helical content when dodine interacts with the amphiphilic helices

of PrB21 leads to the formation of a 'super-helical' state which is responsible for thermal stabilization of PrB21 in Figure 3.1B.

Thus, dodine denaturation differs from guanidinium-Cl denaturation in four basic aspects: (1) dodine acts at 10,000-fold less concentration; (2) although dodine disrupts local tertiary structure around our W-Y probe pair, it has no effect on thermal denaturation of residual tertiary structure; (3) dodine increases helical secondary structure; and (4) dodine stabilizes secondary structure against thermal denaturation. We further investigate the reason behind this behavior by using MD simulations to provide atomic-level structural insight.

3.3.2 SIMULATIONS SHOW AN INCREASE IN PRB21 HELICAL CONTENT AND RADIUS OF GYRATION WITH TEMPERATURE IN 0.087 M DODINE

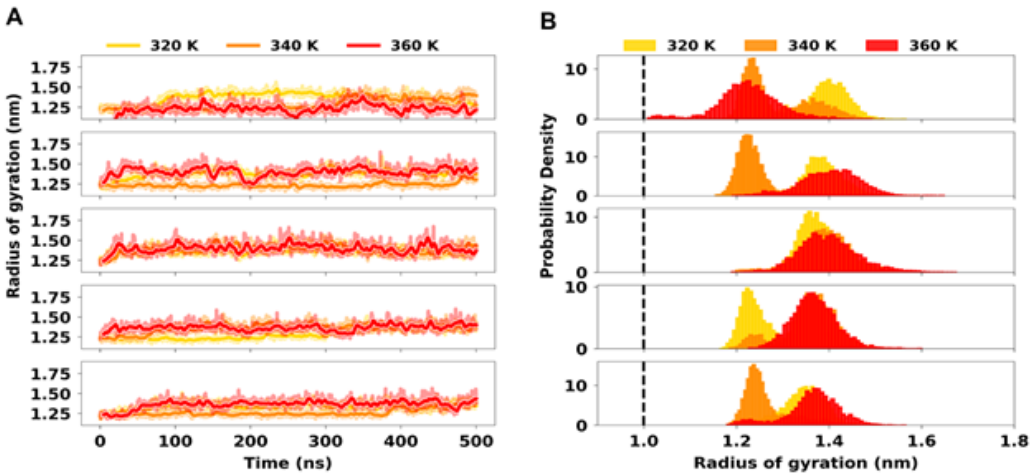


Figure 3.7: (A) Radius of gyration of the protein plotted as a function of time for 500 ns simulations in 5 replicas and (B) probability distributions for the radius of gyration in simulations with 0.087 M dodine at 320 K (yellow), 340 K (orange), and 360 K (red). The black dashed line indicates the crystal structure (native state) values.

We first examined the effect of temperature on the non-native state by conducting molecular dynamics simulations (see methods) with a dodine concentration of 0.087 M (10 dodine molecules and one protein in the simulation box) at three different temperatures, 320, 340 and 360 K starting from folded PrB21. The chosen dodine concentration keeps the simulated dodine:protein ratio similar to experiment and micelle formation is not an issue with 10 dodine molecules on the time scale of the simulation. The probability distributions of radius of gyration, W-Y distance and helical content at each temperature were calculated from five replicas (Figures 3.7-3.9).

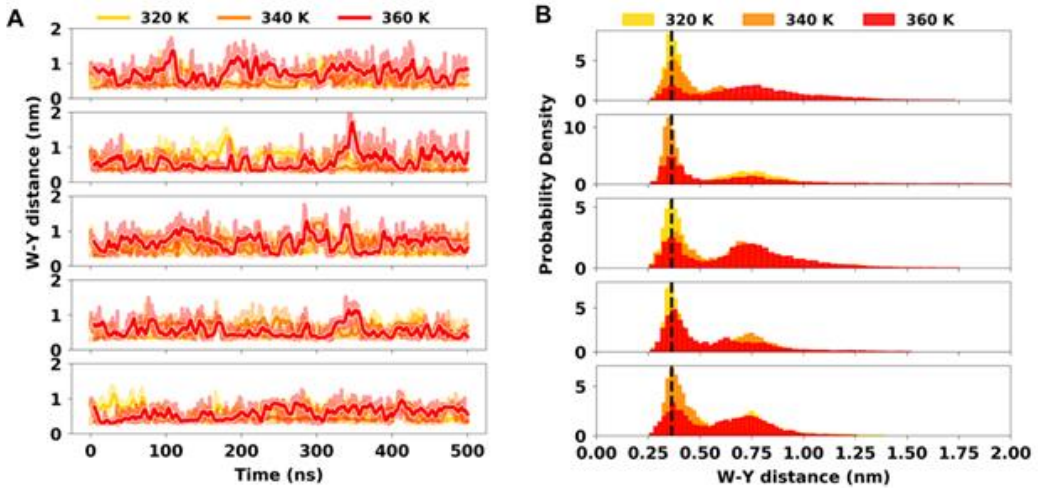


Figure 3.8: (A) Trp-Tyr (W-Y) distance plotted as a function of time for 500 ns simulations in 5 replicas and (B) probability distributions for the Trp-Tyr (W-Y) distance in simulations with 0.087 M dodine at 320 K (yellow), 340 K (orange), and 360 K (red). The black dashed line indicates the crystal structure (native state) values.

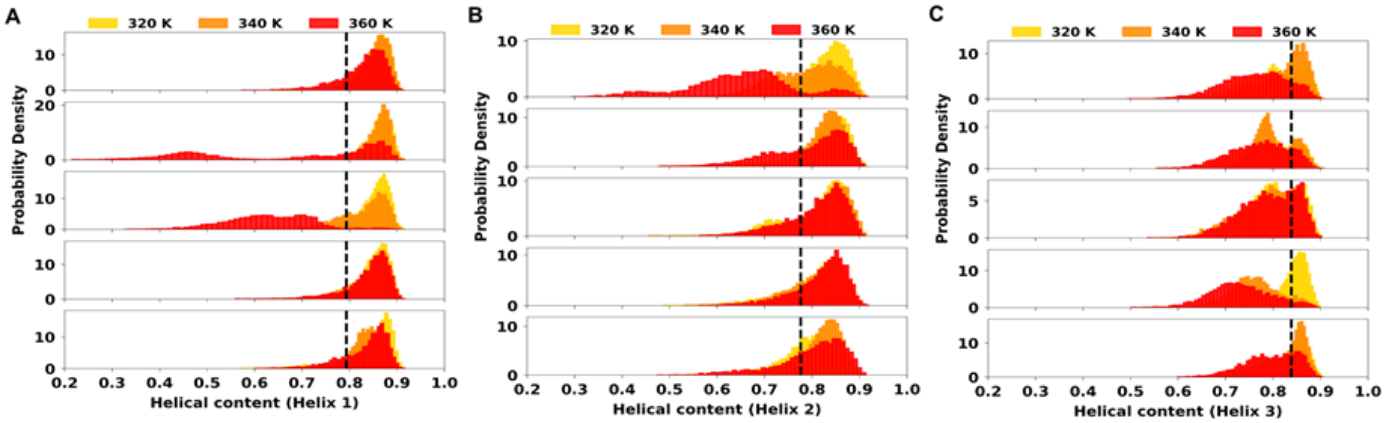


Figure 3.9: Probability distributions for helical content of helixes 1, 2 and 3 in simulations with 0.087 M dodine at 320 K (yellow), 340 K (orange), and 360 K (red) for five replicas. The black dashed line indicates the crystal structure (native state) values.

While R_g is informative of the global shape of the protein, W-Y distance is a proxy for the experimentally observed fluorescence, and helical content is a proxy for the experimental CD spectrum.

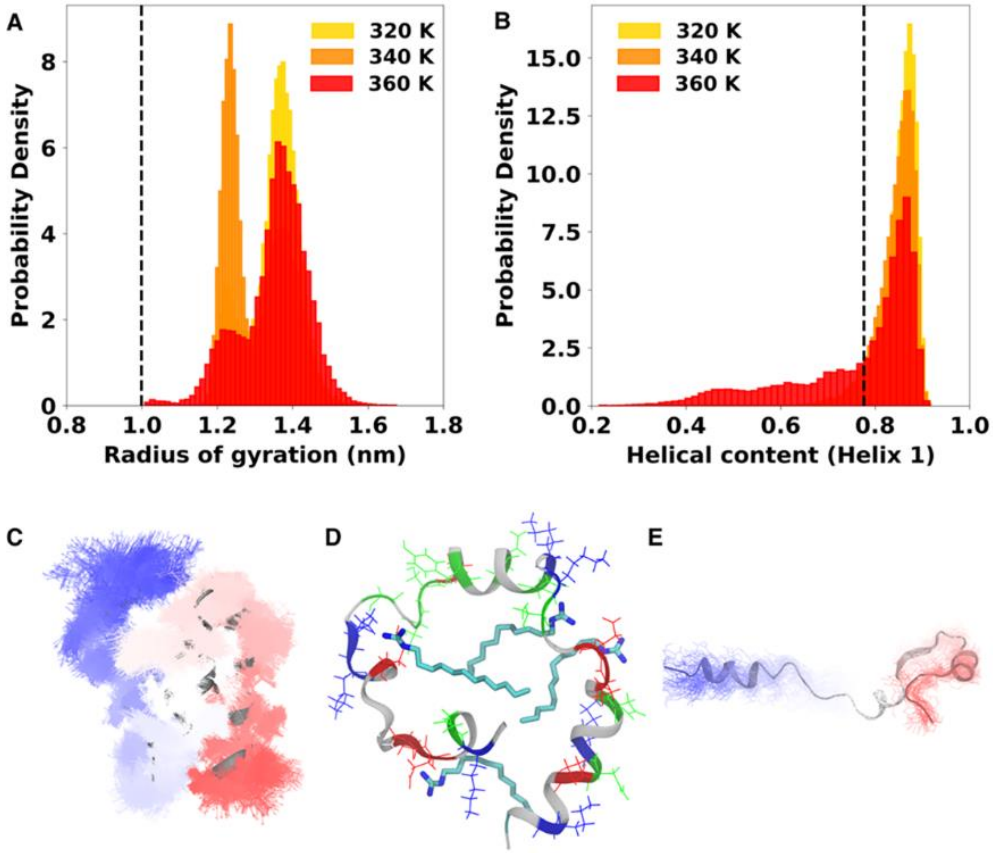


Figure 3.10: The structural effect of temperature and dodine on PrB21. Combined probability distribution for (A) radius of gyration and (B) helical content of helix 1 of PrB21 in simulations with dodine (0.087 M dodine, 10:1 dodine:PrB21 ratio) at 320 K (yellow), 340 K (orange), and 360 K (red). The black dashed line indicates the crystal structure (native state) values. (C) An overlay of ~500 structures obtained from a 500 ns simulation of PrB21 without dodine at 300 K. The first frame is shown in gray as cartoon representation. (D) A representative dodine-expanded form of PrB21 highlighting three dodine molecules in the protein core. (E) An overlay of ~50 structures obtained from the first 5 ns of a PrB21 refolding simulation. The first frame is shown in gray as cartoon representation. In panels C and E the protein is shown in line representation where residues 1 to 47 are colored red to blue. In panel D the residues are colored according to hydrophobicity and dodine molecules' carbon atoms are shown in cyan and nitrogen atoms are shown in blue.

The distribution of radius of gyration (R_g) in dodine (Figure 3.10A) is bimodal, where both ensembles are expanded compared to the native state (R_g ~1.2 to 1.4 nm, Figures 3.10A and 3.7B). Like most materials, proteins usually expand when temperature is increased, as is observed when PrB21 R_g increases from 1.2 nm at 340 K to 1.4 nm at 360 K.[351,352] Interestingly, at 340 K a significant fraction of the population adopts a more compact state than at 320 K or 360 K in four out of five replicas. Compaction of PrB21 as the temperature increases from 320 K to 340 K is counterintuitive. This is also seen in the W-Y distance, since the state when these residues are farther apart (~0.75 nm) is less populated at 340 K as compared to 320 K and 360 K (Figures 3.8B and 3.11).

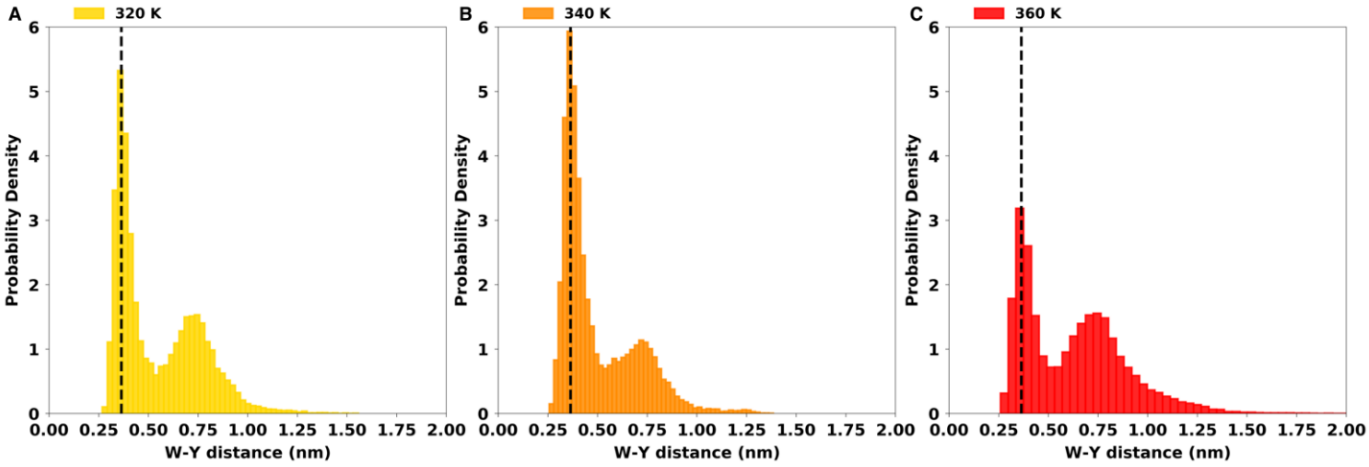


Figure 3.11: Combined probability distribution for Trp-Tyr (W-Y) distance with 0.087 M dodine at (A) 320 K (yellow), (B) 340 K (orange), and (C) 360 K (red). The black dashed line indicates the crystal structure (native state) values.

It is well-known that the strength of hydrophobic interactions increases with temperature because it becomes more favorable to exclude water between hydrophobic surfaces, and thus increase solvent entropy.[353] It is possible that dodine shifts the balance of protein configurational entropy and hydrophobicity in favor of the latter for PrB21 at 340 K. However, as temperature is increased to 360 K, protein configurational entropy dominates hydrophobic collapse. Therefore, the interaction of the PrB21 hydrophobic core with dodine hydrophobic tails could overcome complete thermal unfolding at an optimal temperature.

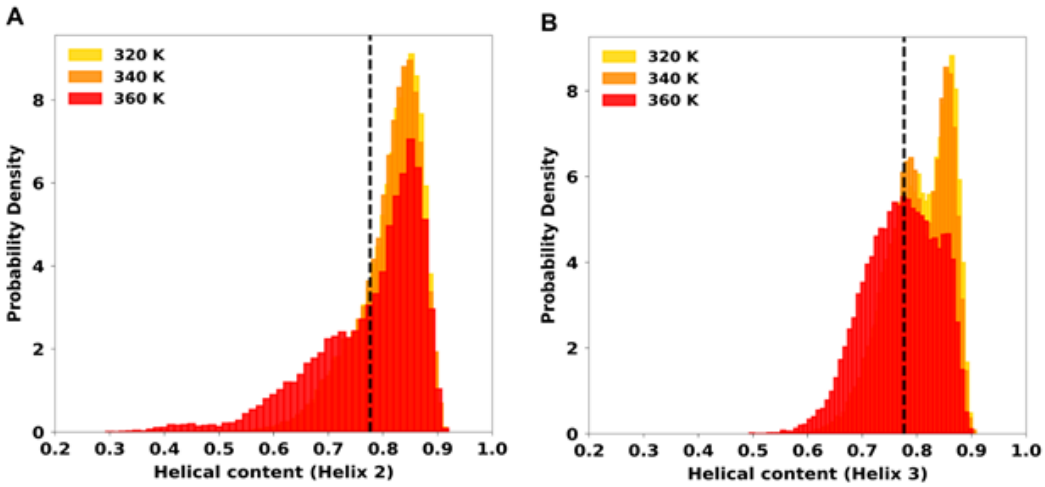


Figure 3.12: Combined probability distribution for helical content of (A) helix 2 and (B) 3 with 0.087 M dodine at 320 K (yellow), 340 K (orange), and 360 K (red). The black dashed line indicates the crystal structure (native state) values.

To test the 'super-helical state' hypothesis proposed based on CD experiments, we examined helical content in the simulations. Helical content of PrB21 increases in dodine at all three simulation temperatures (Figures 3.10B and 3.12A) for helix 1 and 2. Helical content for helix 3 increases up to 340 K and then decreases to the native level at 360 K (Figure 3.12B). At 360 K the helical content distribution is significantly broadened as compared to 320 and 340 K for both helix 1 and 2. This broadening indicates the formation of multiple low-helical states. These low-helical states and the decrease in helicity for helix 3 are consequences of increasing temperature because helical structure is temperature sensitive and even stable helices are denatured at sufficiently high temperature.[354]

3.3.3 INCREASING DODINE CONCENTRATION ACCELERATES THE FORMATION OF THE DODINE-INDUCED SUPER-HELICAL STATE OF PRB21

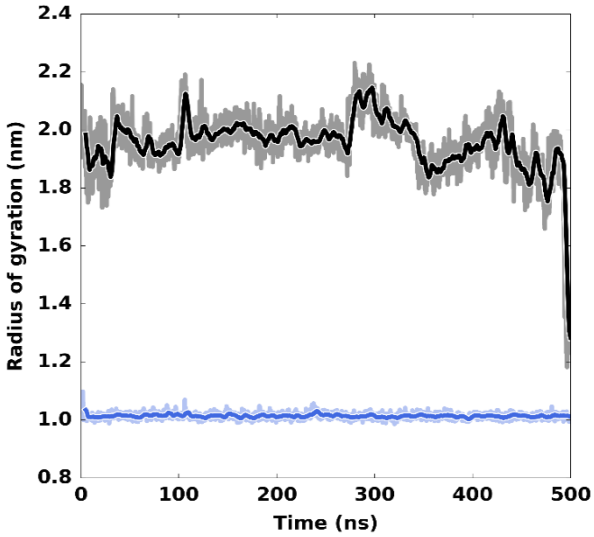


Figure 3.13: Radius of gyration of the protein is plotted as a function of time for 500 ns simulations. Apo protein simulations starting from folded structure (blue) and unfolded structure (black).

We compared the dodine-induced super-helical state to the folded and unfolded PrB21 conformations in absence of dodine. Simulations of PrB21 at 300 K starting with the folded structure yielded a native R_g of ≈1 nm (Figure 3.10C). Relaxation of the unfolded structure yielded an R_g of ≈2 nm, but still had high helical content in the simulation (Figures 3.10E and 3.13). In the presence of dodine, PrB21 forms an expanded but super-helical state (Figures 3.10D and 3.7) which is far from either the native state or from an unfolded coil. In addition to increased helical content, this expanded state has an increased W-Y distance and dodine tails are intercalated into the protein core, unzipping the helices and expanding the PrB21 tertiary structure (Figure 3.10D). We propose that these three simulated states (native, expanded and unfolded) are analogous to the '320 nm fluorescence', '330 nm fluorescence' and '350 nm fluorescence' states in Figure 3.1 as they

match the trends in W-Y distance and helical content as a function of dodine concentration and temperature inferred from experiment.

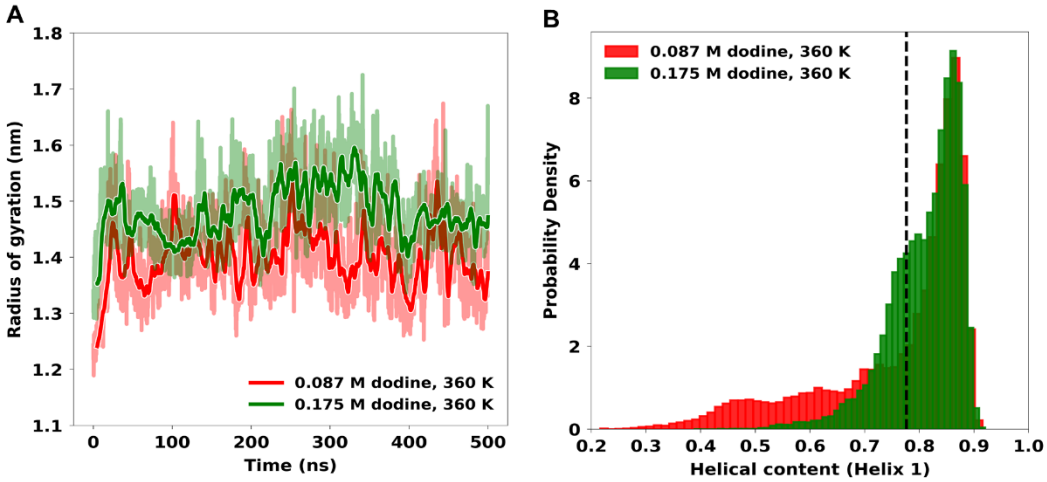


Figure 3.14. MD simulations of PrB21 with 0.087 M and 0.175 M dodine at 360 K. (A) Radius of gyration of the protein as a function of time, and (B) Probability distribution for the helical content of helix 1 with 0.087 M dodine (red) and 0.175 M dodine (green) at 360 K. The dark traces in the left panel show the data averaged over a 5 ns window and the lighter traces show data recorded at each timestep in our simulations. The black dashed line indicates the crystal structure (native state) values.

We also compared simulations at a constant temperature (360 K) but at two different dodine concentrations. We compared 10 and 20 dodine molecules, corresponding to ~0.087 M and ~0.175 M to study the effect of higher dodine concentration on PrB21 (maintaining a similar protein:dodine ratio as in experiment). The rate of formation of the non-native super-helical state is dodine concentration-dependent in our simulations. In all five replicas, we observed formation of the super-helical state in < 5 ns with 0.175 M dodine *vs.* ~10 to 20 ns with 0.087 M dodine (Figures 3.14A and 3.15).

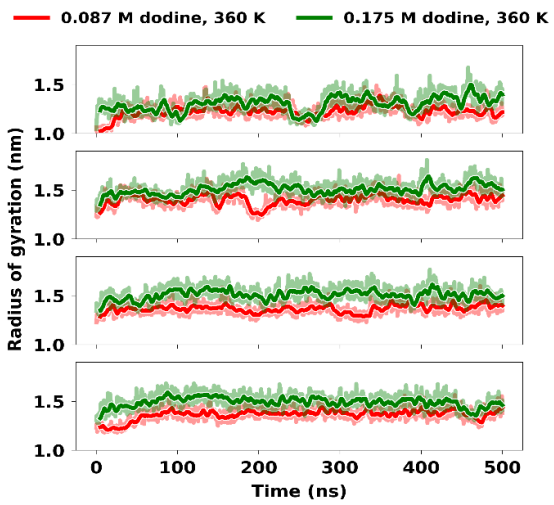


Figure 3.15: Radius of gyration of the protein as a function of time with 0.087 M dodine (red) and 0.175 M dodine (green) at 360 K for four simulation replicas. Replica one is shown in Figure 3.14A.

As expected, R_g increases with increasing dodine concentration (Figure 3.16).

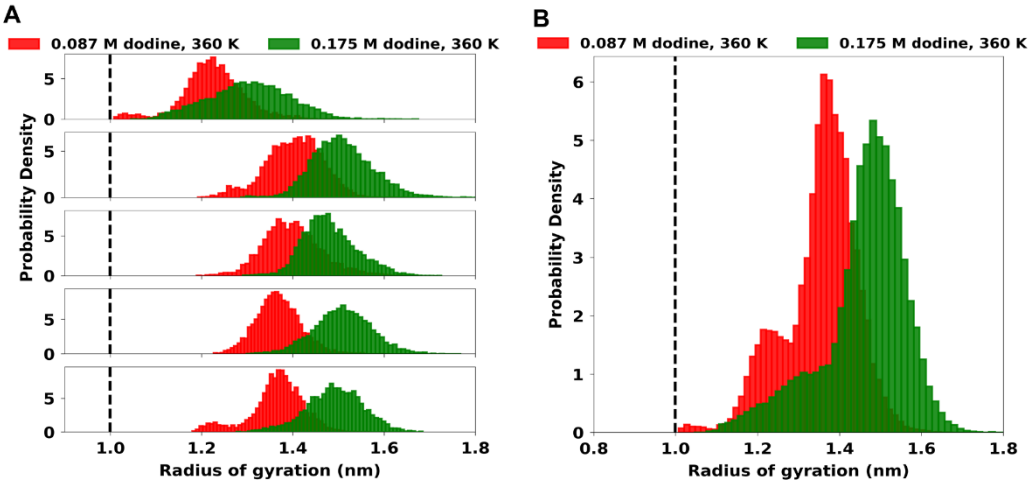


Figure 3.16: (A) Probability distribution for five replicas and (B) combined probability distribution of radius of gyration with 0.087 M dodine (red) and 0.175 M dodine (green) at 360 K. The black dashed line indicates the crystal structure (native state) values.

The probability distribution of helical content for each replica (Figure 3.17) was combined to estimate the overall probability distribution for helix 1, 2 and 3 as a function of dodine concentration. As previously discussed for 0.087 M dodine at 360 K, the helical content increased for helix 1 and 2 and remained unchanged for helix 3 in 0.175 M dodine (Figures 3.14B and 3.18). In Figure 3.10B, we also observed the formation of low-helical states due to thermal denaturation of helices at 360 K with 0.087 M dodine. The population of this low-helical state decreases when the concentration of dodine is increased to 0.175 M

(Figures 3.14B and 3.18A). With 0.175 M dodine the helical content at 360 K mimics that at 320 K with 10 dodine molecules (Figure 3.10B). This means that the addition of dodine stabilizes helical content against thermal denaturation. These simulations help explain why experimentally PrB21 is not destabilized against thermal denaturation by dodine.

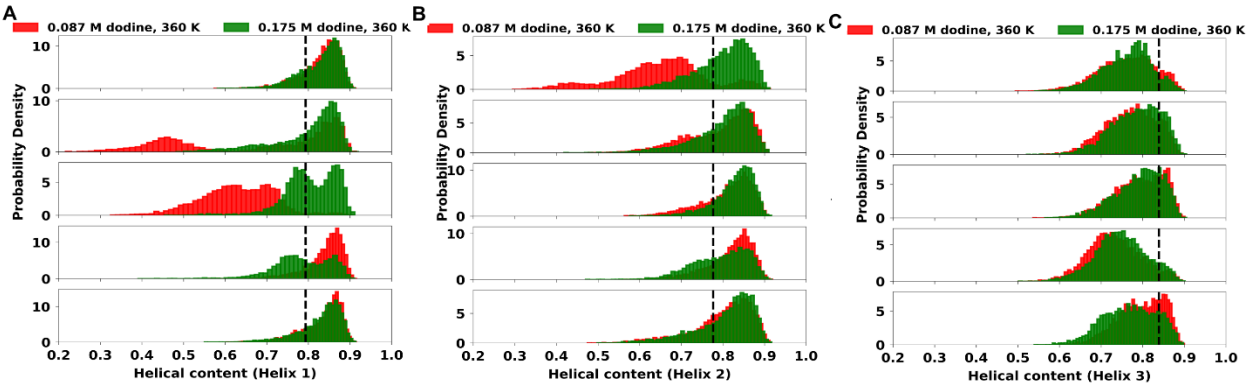


Figure 3.17: Probability distribution for helical content with 0.087 M dodine (red) and 0.175 M dodine (green) at 360 K for five simulation replicas. The black dashed line indicates the crystal structure (native state) values.

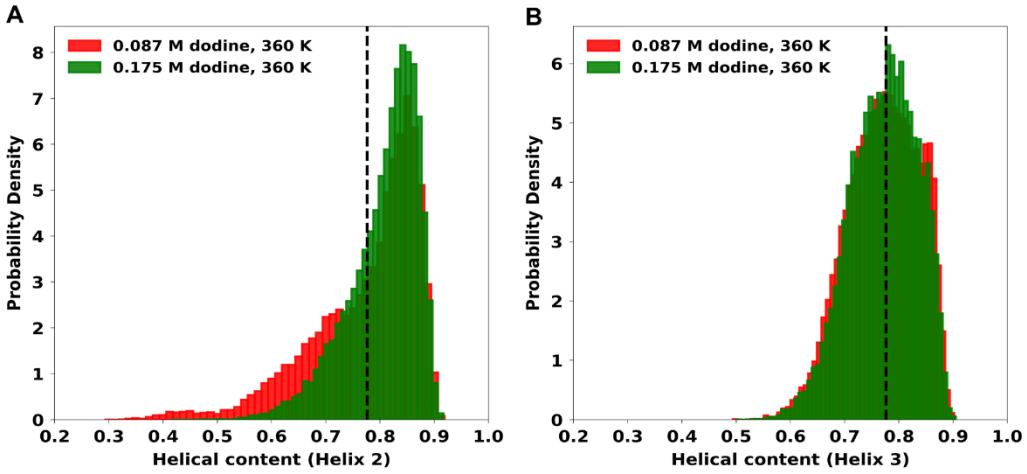


Figure 3.18: Combined probability distribution for helical content of helix 2 and 3 with 0.087 M dodine (red) and 0.175 M dodine (green) at 360 K. Probability distribution for helix 1 is shown in Figure 3.14B. The black dashed line indicates the crystal structure (native state) values.

3.3.4 DODINE INTERACTION WITH PRB21 CORRELATES WELL WITH RESIDUE HYDROPATHY

Next, we turned to a more detailed structural analysis of how dodine interacts with the solvent-exposed protein surface. Different amino acid side chains may interact differently with the denaturant head or detergent tail[347] of dodine to stabilize the super-helical but expanded state. Figure 3.19 shows the contact-

coefficient$_{\text{head-tail}}$ (CC_{HT}), for each residue with either the tail or head of the dodine molecule (Figure 3.19A). Positive values (up to +1) imply that the residue prefers interaction with the guanidinium-Cl head group and negative values (up to -1) imply preferred interaction with the hydrophobic dodecyl tail. Values close to zero show no preference for either the head or the tail.

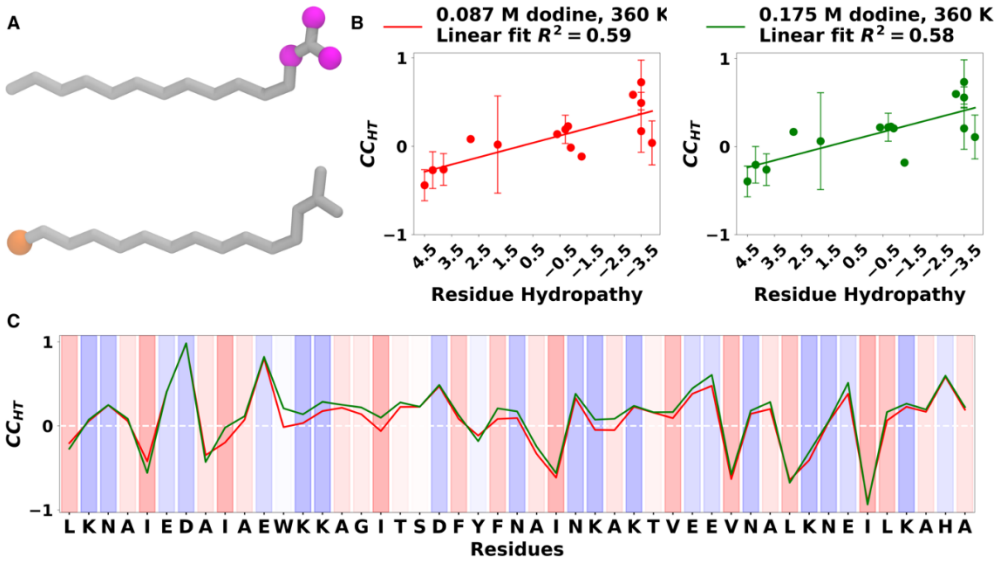


Figure 3.19: Contact-coefficient$_{\text{head-tail}}$ values for each residue of PrB21. (A) Highlighted atoms show atoms included to calculate head group interactions (pink) and tail interactions (orange). (B) Contact-coefficient$_{\text{head-tail}}$ (CC_{HT}) as a function of residue hydropathy for all residue (backbone+sidechain) of PrB21 in 0.087 M dodine (red) and 0.175 M dodine (greem) at 360 K. Positive values indicate preference for guanidinium-Cl head while negative values indicate preference for the hydrophobic tail. (C) Same plot as a function of sequence. The background colors represent residue hydrophobicity with blue colors being more polar and redder more hydrophobic.

The interaction of residue sidechains and backbones with dodine is shown for 0.087 M and 0.175 M dodine at 360 K as a function of residue hydropathy (Figures 3.19B and 3.20A), and as a function of protein sequence (Figure 3.19C). About a quarter of all residues interact preferentially with either the head ($CC_{HT} \geq 0.25$) or the tail ($CC_{HT} \leq$-0.25, Table 3.1). Moreover, these interaction values correlate well with the amino acid hydropathy. More than ~90% of all amino acids that preferentially interact with the head are polar and those that interact with the tail are hydrophobic (Table 3.1). Interactions with the head group increased at higher dodine concentration, while those with the tail remained mostly unchanged with both dodine concentration and temperature. A similar result was obtained when only residue backbones were considered (Figure 3.20C).

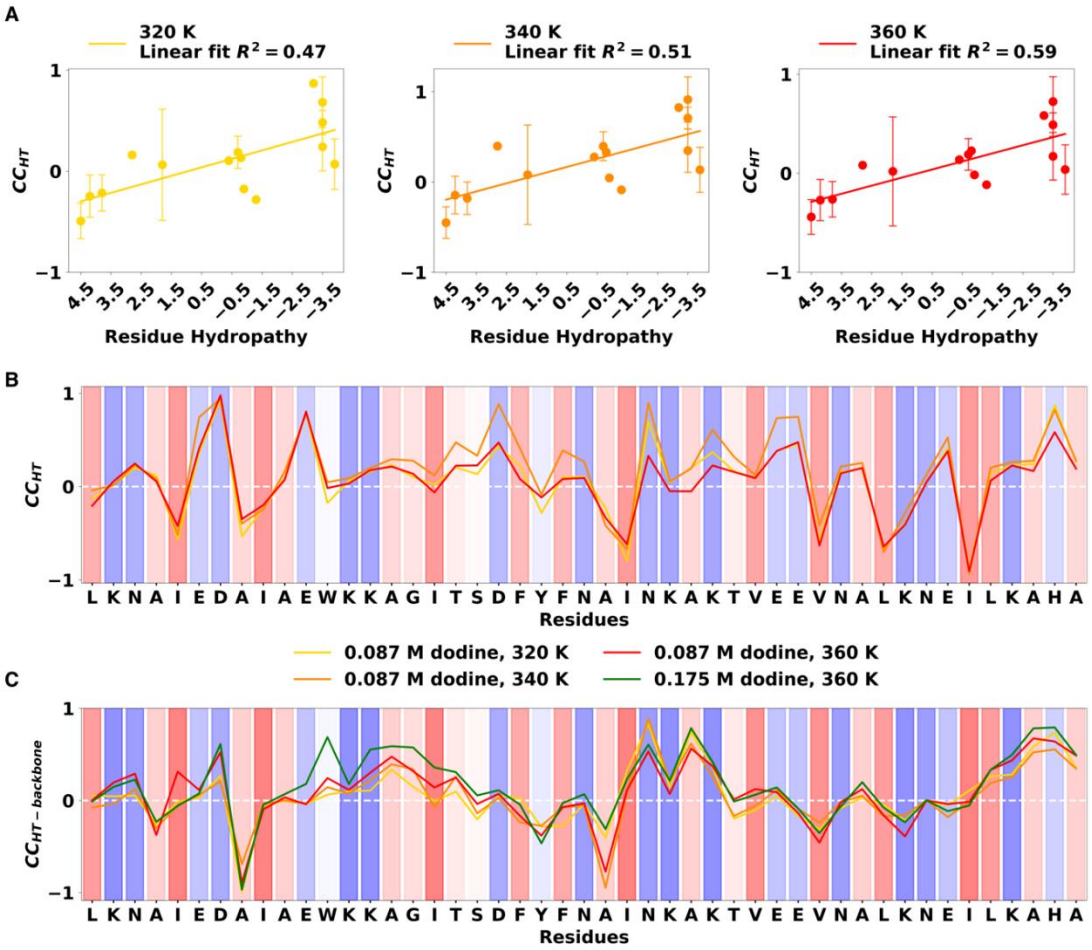


Figure 3.20: (A) Contact-coefficient$_{head\text{-}tail}$ (CC_{HT}) as a function of residue hydropathy for all residue (backbone+sidechain) of PrB21 in 0.087 M dodine at 320 K (yellow), 340 K (orange) and 360 K (red). Positive values indicate preference for guanidinium head while negative values indicate preference for the hydrophobic tail. (B) Same plot as a function of sequence. The background colors represent residue hydrophobicity with blue colors being more polar and redder more hydrophobic. (C) Contact-coefficient$_{head\text{-}tail}$ (CC_{HT}) values as a function of sequence for backbone atoms only.

Thus, the dodine tail preferentially interacts with hydrophobic sidechains in helices, stabilizing amphipathic helices in PrB21. The charged head group preferentially interacts with negatively charged sidechains and with the backbone, leading to tertiary structure denaturation. Polar residues such as threonine (T) or serine (S) show little propensity either way. The increased dodine head interaction with protein backbone-only at higher dodine concentration (Figure 3.20C) signals an increase in guanidinium-Cl-like denaturant properties at higher dodine concentrations. In contrast, no significant concentration- (Figure 3.19C) or temperature-dependence (Figure 3.20B) of CC_{HT} for backbone+sidechains was observed.

Table 3.1: Statistics of amino-acid residues involved in interactions with dodine molecules. Residues with $CC_{HT}>=0$ have preferable interactions with the guanidinium head group, and with the hydrophobic dodecyl tail otherwise.

System	Atom type	% (CC_{HT} >= 0.25)	% (CC_{HT} <= -0.25)	% Polar residues with CC_{HT} >= 0.25	% Hydrophobic residues with CC_{HT} <= -0.25
10 dodine 320 K	All	21	19	100	89
	backbone	15	13	-	-
10 dodine 340 K	All	21	19	100	100
	backbone	15	11	-	-
10 dodine 360 K	All	19	17	100	100
	backbone	26	13	-	-
20 dodine 360 K	All	28	19	92	100
	backbone	32	11	-	-

3.4 DISCUSSION

The exploration of novel denaturants such as dodine has recently gained more interest because traditional denaturants interfere with common spectroscopic techniques. Here, we propose a possible mode of interaction between dodine and a model helical protein, PrB21. Dodine interacts with the protein via both the guanidinium-Cl head group (attacks backbone and charged side chains) and the tail (attacks non-polar side chains). At low temperature, this interaction induces an expanded super-helical state of PrB21. Increasing the dodine concentration does not affect the thermal stability of this non-native PrB21 state because the aliphatic tail counteracts the denaturant effect of the guanidinium-Cl headgroup. Only at high temperature does the protein finally unfold completely. A control with SDS supports this model because SDS, which lacks the denaturing guanidinium-Cl group, stabilizes PrB21 even more against thermal denaturation (Figure 3.5B).

Dodine has shown quite a range of denaturation behaviors with different proteins, consistent with our explanation in terms of dodine's stabilization of secondary structure and insertion into the hydrophobic core. The small β-sheet protein Fip35 WW domain (34 residues) is not denatured by dodine. Unlike PrB21, WW has no significant hydrophobic core for dodine to insert into; likely dodine would have to be used above its critical micelle concentration (6-8 mM) to be effective.[295] The larger α-helical protein $\lambda_{6\text{-}85}$ is denatured by dodine, but no super-helical state forms. As the hydrophobic core gets larger, the decrease in stability due to loss of tertiary structure may not be sufficiently overcome by stabilizing secondary structure elements. In this case dodine would act more like a traditional denaturant. The guanidinium-Cl head group is imperative for the destabilizing denaturant-like properties of dodine, but is able to act at much lower concentration thanks to assistance from the hydrophobic tail interacting with the larger core. This is indeed

reinforced by dodine's ability to effectively denature phosphoglycerate kinase, a much larger (415 residue) protein with two large hydrophobic cores. The Janus-like behavior of headgroup and tail make dodine a simultaneous secondary structure stabilizer and tertiary structure destabilizer or a 'kosmo-chaotrope,' dependent on protein size.

CHAPTER 4

THE DESIGN AND CONSTRUCTION OF A THREE-COLOR FAST RELAXATION IMAGING (FReI) SETUP

4.1 INTRODUCTION

The principle behind Fast Relaxation Imaging (FReI) has already been described in detail in previously published works.[122,355] A FReI instrument is able to measure biomolecule dynamics *in vitro*, in-cell and *in vivo* by suddenly perturbing the environment to induce a reaction, and then utilizing Förster Resonance Energy Transfer (FRET) or another spectroscopic technique to observe the reaction as it equilibrates. Generally, two interacting biomolecules are first labeled with a donor fluorophore (mEGFP or AcGFP1) and an acceptor fluorophore (mCherry). The system is perturbed using temperature and allowed to come to equilibrium. Binding or interaction can be characterized by tracking the change in FRET efficiency over time as a function of temperature. This technique can also be applied to track unfolding or folding of a singular protein by labeling either termini with the donor and acceptor fluorophore instead of a two binding or interacting biomolecules. The concept of FRET is described in more detail in the following sections as well as the schematic and details of the FReI system.

4.1.1 FÖRSTER RESONANCE ENERGY TRANSFER (FRET)

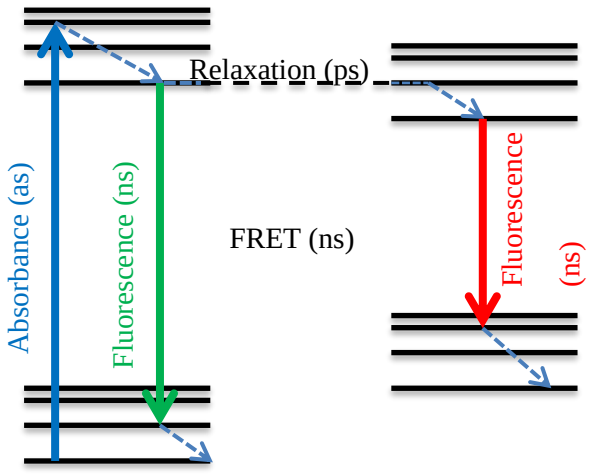


Figure 4.1: Jablonski diagram showing energy transfer during FRET.

Förster Resonance Energy Transfer or FRET, named after the German scientist Theodor Förster, is a phenomenon by which energy is transferred non-radiatively between two chromophores by dipole-dipole coupling. This transfer of energy occurs between a donor excited state to an acceptor exited state in turn leading to acceptor fluorescence. The Jablonski diagram (Figure 4.1) shown above highlights this process.

The efficiency of this energy transfer is given by:

$$E = \left[1 + \left(\frac{r}{R_0}\right)^6\right]^{-1} \quad [4.1]$$

Where, r is the distance between the two fluorophores and R_0 is the Förster distance given by:

$$(R_0)^6 = \frac{2.07}{128\pi^5 N_A}\frac{\kappa^2 Q_D}{n^4}\int F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda \quad [4.2]$$

Where, Q_D is the fluorescence quantum yield of the donor in the absence of the acceptor, κ is the dipole orientation factor, n is the refractive index of the medium, N_A is Avogadro's number, and the integral term is the spectral overlap that roughly measures what fraction of the donor energy is transferred to the acceptor.

Since FRET depends on the distance between the two fluorophores it can measure the spatial arrangement of fluorophores. Both binding/unbinding and folding/unfolding lead to distance changes between fluorophores that can be measure as a FRET efficiency change. R_0 is the fluorophore separation distance at which the energy transfer efficiency is 50 %. R_0 is 3-6 nm for standard pairs of fluorophores, resulting in a 4-5 fold FRET efficiency change over the 2-3 nm distance changes relevant for global protein structural changes or coarse changes in protein-protein distances.[356]

4.2 FAST RELAXATION IMAGING (FREI)

FReI relies on FRET to measure protein-protein binding and protein unfolding. The instrument consists of a traditional epi-fluorescence microscope, a laser heating setup and an imaging setup. On my instrument, the system of interest in the cell or in a buffer is perturbed using programmed temperature jumps (described in more detail in later sections), although fast volume, pH or other perturbations are also possible. Each temperature jump consists of a sharp increase in temperature on timescales faster than the folding/binding times of the system, followed by a short period where the temperature is held constant and the system is allowed to settle to the new equilibrium. This enables the capturing of both kinetics and thermodynamics of the system.

FRET efficiency is imaged by capturing both donor and acceptor fluorescence on a CMOS sensor by splitting the light into two or more channels using dichroic mirrors to collect several color channels as a function of time after each temperature jump. Finally, the data is processed on a local computer using MATLAB. A full list of parts is given in the appendix C (C.1 and C.2) following the main chapters.

4.2.1 THE MICROSCOPE

The FReI instrument's main body consists of a commercial epi-fluorescence microscope. A standard microscope provides various convenient advantages. The stage allows us to use the microscope's motorized control for imaging as well as the flexibility of easily switching between multiple objectives using the built-in turret. We chose the Axio Observer A1 (Figure 4.2, Carl Zeiss Inc.) which allows us imaging in both the transmittance mode and reflectance mode (for fluorescence). The microscope is equipped with 2.5x, 10x, 40x and 63x objectives for imaging.

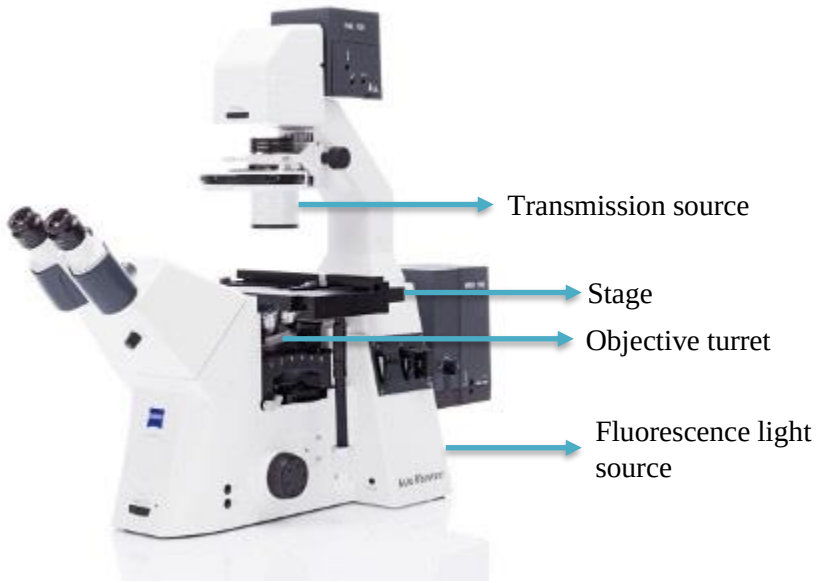


Figure 4.2: The Carl Zeiss Axio Observer A1. Image is sourced from zeiss.com.

We chose a white light reflectance source to be able to excite multiple fluorophores without switching light sources. The UHP-T-LED white (Prizmatix, Inc) was chosen for both its high brightness as well as a full spectrum of wavelengths. Since the acceptor fluorophore is only indirectly excited by the donor energy transfer, the donor molecule excitation light source needs to deliver enough intensity, so the acceptor signal is over the background noise. Additionally, the LED emits collimated light, which reduces artifacts from off-axis excitation. The UHP-T-LED provides >4 watt of collimated white light power, enough to get detectable signal from FRET. The LED also contains wavelengths from 400-700 nm (Figure 4.3). The LED

is controlled using an analog BNC input, 0-5 V and a TTL trigger using the instrument control software LabVIEW (National Instruments).

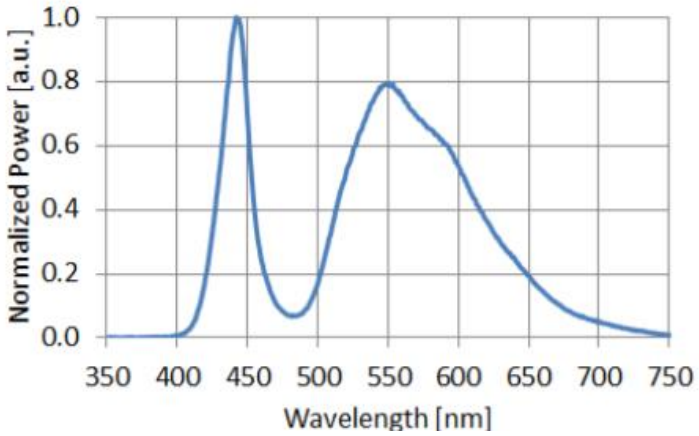


Figure 4.3: The normalized power output as a function of wavelength for the UHP-T-LED white. The graph was sourced from the specs sheet provided by Prizmatix Inc. Note that for GFP excitation at ~470-480 nm, a more optimal 'white' excitation source could be used.

The LED light is passed through a filter cube equipped with an excitation filter, an emission filter and a dichroic mirror (Figure 4.4).

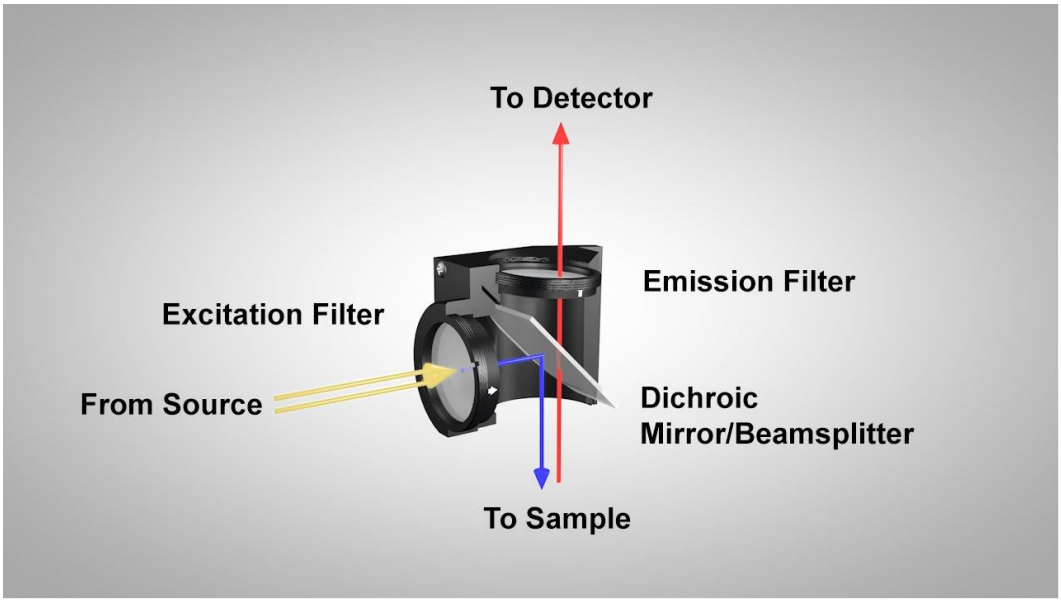


Figure 4.4: A filter cube schematic showing the excitation and emission filter and a dichroic beamsplitter. The image is sourced from thorlabs.com.

The microscope is equipped with a turret than can hold up to 6 filter sets at a time. Custom filter sets were purchased from Chroma Inc. for GFP and mCherry excitation/emission (Table 4.1). Both excitation filters are bandpass filters while the emission filters were chosen to be long-pass filters to allow for maximum emission fluorescence to be collected at the detector. The GFP filters are designed to collect both GFP and mCherry emission while the mCherry filters are purely for mCherry excitation and only directs mCherry emission towards the detector.

Table 4.1: Filter set part numbers for EGFP and mCherry purchased from Chroma Inc.

	Excitation filter	**Dichroic beamsplitter**	**Emission filter**
EGFP	ET470/40x	T495lpxt	ET500lp
mCherry	ET580/25x	T600lpxr	AT610lp

4.2.2 HEATING SETUP

Since the cells are in an aqueous medium and most of our buffers are aqueous, we used a 2 µm, 1 W CW fiber laser from AdValue Photonics. From the infrared absorbance spectrum of water, we determined that the absorptivity of water at 2 µm is $10^4\ m^{-1}$. This absorptivity of water at 2 µm is enough to heat the sample up to 4 °C while being invisible to the imaging optics at the timescales at which our proteins of interest fold or bind (~2 seconds).

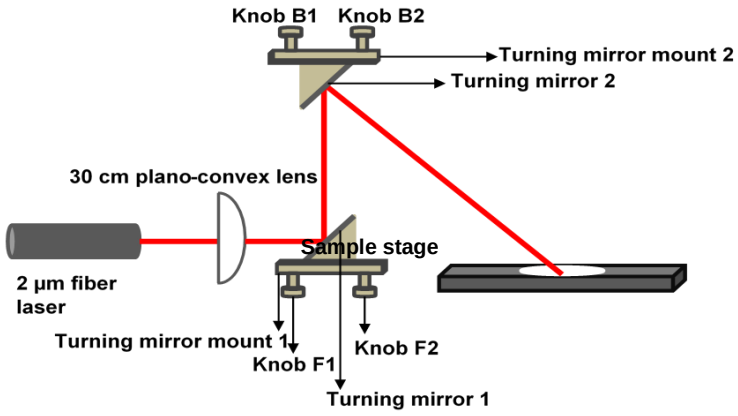


Figure 4.5: A schematic of the laser heating setup for the FReI instrument. Beam path of the IR laser beam.

The collimated 2 mm laser beam is first focused down to ~0.2 mm using a 30 cm focusing lens to increase the power density of the laser beam at the sample (Figure 4.5). We direct the laser beam towards the sample

at a 45 ° angle using turning mirrors to avoid blocking the transmission light path with a vertically mounted laser.

The laser is equipped with an analog voltage input that accepts voltage input from 0 – 1.2 V for a maximum power output of 1 W at 1.2 V. The voltage input is supplied through BK precision 9130 power supply (BK Precision Corp.) that is controlled remotely via LabVIEW (National Instruments Ltd.).

The spatial alignment of the laser beam at the sample is controlled through the front (knob F1 and F2) and back (knob B1 and B2) vernier screws on the two turning mirror mounts (Figure 4.4). The following procedure outlines the alignment process. Note than one must be familiar with the laser operation safety manual and laser goggles before aligning the laser

1. Switch the objective to 10x and turn on the laser and LED controller.
2. Prepare a ~5 µM mCherry sample with 6 µm green fluorescent polystyrene beads (Phosphorex, Inc). Generally, 1 µL of a 1:20 dilution of the bead solution is added to 500 µL of 5 µM mCherry.
3. Adhere a 120 µM double-sided spacer (Grace Biolabs Ltd.) on a cleaned glass slide. Pipette 300 µL of the mCherry sample in the middle of the spacer and make sure the liquid spreads and covers the circular area of the spacer. Gently lay down a #1.5 cover slip on the spacer and make sure its firmly stuck on the top sticky side of the spacer (Figure 4.6).

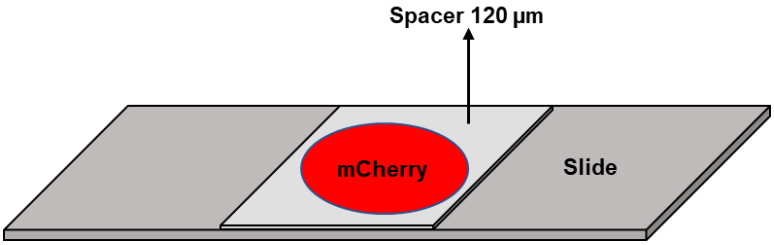


Figure 4.6: Sample slide preparation for laser beam alignment.

4. Once the slide is ready place the slide on the stage cover slip side down and let the beads settle for about 5-10 minutes. Excite the green fluorescent beads using the EGFP filter set mentioned in the section above and focus on the sample.
5. Turn the laser on and roughly align the laser to the objective using IR paper (Thorlabs Inc.).
6. Switch to the mCherry filter set and heat the sample with a 1 second 0.7 V pulse sent to the IR laser.

7. Analyze the recorded video, the area where the laser hits temporarily turns dark. Move the laser beam using the 4 knobs and repeat step 6. Continue moving the beam until the heating are is in the center.
8. Switch objective to 40x and repeat steps 6 and 7.
9. Switch objective to 63x and repeat steps 6 and 7.

Figure 4.7 shows the aligned beam using a 10x objective. The heating temporarily bleaches the part of the mCherry sample where the laser is aimed and leads to a dark spot allowing us to be able to locate and visualize the laser spot on the sample.

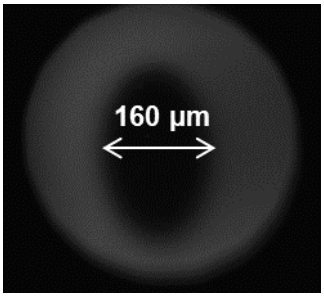


Figure 4.7: Beam profile of the laser captured with a 40x objective lens.

4.2.3 THE IMAGING SETUP

Instead of the traditional two-color imaging the goal was to build a three-color imaging setup capable of capturing fluorescence from up to three colors at the same time. While two-color FRET is able to investigate interactions of two biomolecules, a third degree of freedom is needed to really understand the network of interactions between components. Most interactions in the cell are complex and involve more than two components. In order to investigate more complex interactions a three-color FReI imaging setup is needed. However, we wanted to retain the capability to image two-color FRET without complex optics switching and the system would then need to be able to image in both two- and three-color mode with a simple switch.

The fluorescence emission from the sample is directed out of the camera port through a telescoping lens (15 cm focal length) in order to collimate the beam. A collimated beam avoids issues related to the distance constraints of a focusing beam. The light is first turned 90° using a turning mirror M1, and then collimated through the lens is placed at 15 cm from the camera port to achieve telescoping (Figure 4.8). The beam is then split into red and green light using a T600lpxr dichroic mirror mounted on a custom dichroic holder. The holder is mounted on a rotation mount RP01 (Thorlabs Inc.) on a 1.4”-20 tapped aluminum optical breadboard (MB810, Thorlabs Inc.). The green light is passed through a second dichroic mirror T495lpxt that then splits the light into green and blue colors. The second dichroic is held by a quick release filter

mount (SFH2, Thorlabs Inc.) for easy removal and switching between two-color and three-color imaging. Both dichroic mirrors are oriented such that they are at 45 ° to the beam. The red channel is directed towards a turning mirror M2, green to turning mirror M3 and blue to turning mirror M4 that directs the beams towards a knife edged prism (MRAK25, Thorlabs Inc.). The prism then combines the beams onto a manual iris vari-focal C-mount lens (Azure 1060ZM, Azure Photonics) attached to a CMOS camera (Lt225, Lumenera Corp.). The camera was chosen for its wide detector size (2048 X 1088 pixels) which allows imaging of all three channels side by side at the same time. Such a configuration avoids using multiple cameras which would then require time-syncing the three colors. The Lt225 comes with a CMOS sensor with a 2/3" optical format providing a resolution of 2048 x 1088 using 5.5 μm^2 pixels and provides up to 170 fps at full resolution, data is generally collected at 60 fps for most of our in-cell and *in vitro* applications.

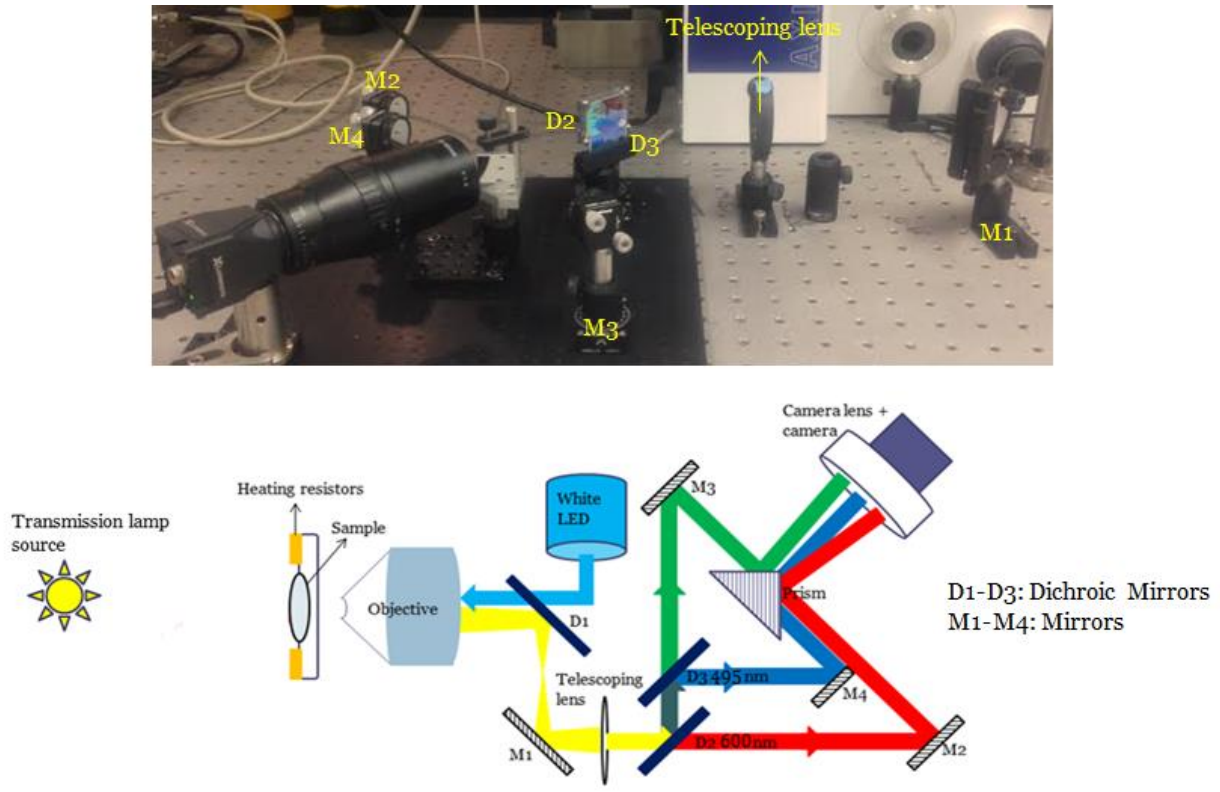


Figure 4.8: Three-color FReI imaging setup. Top image shows the imaging setup on the microscope and bottom shows the beam path through the splitting optics imaged on the camera.

The camera transmits images via a USB 3.0 port to the computer. The three channels are first aligned using the transmission light source by turning the knobs on the turning mirrors until the channels look well-separated and visually aligned (Figure 4.9). The camera lens magnification needs to be then set such that that images are captured at 1 μm per pixel. In order to calibrate this magnification a USAF target (Newport

Inc.) was utilized (Figure 4.9). Since each bar width is constant the width divided by the number of pixels contained in each bar yields the magnification. The magnification for the setup was calculated to be ~0.8 μm/pixel with the 63x objective.

The channel brightness artifact in figure 4.9 is due to variable power outputs from the white light bulb at different wavelengths. The transmission source emits more power at the redder wavelengths than the bluer wavelengths.

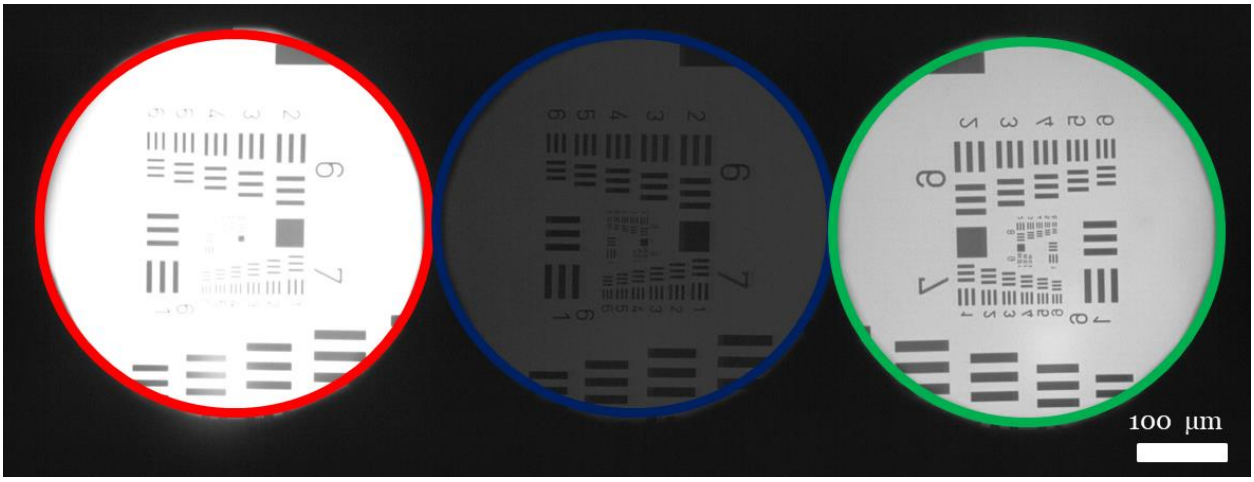


Figure 4.9: USAF target imaged with a 40x objective. Channels are red, blue and green in that order.

4.3 JUMP SHAPING AND CALIBRATION

In order to heat the sample on timescales faster than the folding/binding time of the protein of interest (~2 seconds) a voltage waveform is shaped such that it leads to a step-shaped increase in temperature. The step-shaped increase in temperature perturbs the equilibrium of the system to a new state following which the temperature is held constant allowing the new state to come to equilibrium. This allows for the observation of both folding/binding kinetics and thermodynamics. A detailed procedure for shaping such a waveform has been described in the published thesis of Dr. Hannah Gelman [357].

The sample is first prepared similarly to the method outlined under laser alignment step 2 and 3. The beads allow for the focusing on the surface of the cover slip in contact with the buffer, also the surface where the cells are adhered and therefore, reduce any errors that may arise from slight differences in heating at different points along the 120 μm depth of the cover slip. The sample is then excited with the white light LED using the mCherry filter set.

A continuous waveform is shaped that yields ~10 steps of ~4 °C jumps with 8 seconds of imaging between steps. The size of each jump can be calibrated from the intensity change using the mCherry quantum yield change of 2.1 % per degree temperature change [358]. The 8 bit images are stored in a matrix where each pixel intensity is stored as a number between 0 – 255. Between each jump is a 2 second wait during which the laser actively heats but no video is recorded. At this time each matrix containing pixel-by-pixel intensity data is written as a binary file to the computer hard drive. Each voltage waveform consists of a series of voltage values sent to the laser at 10 ms time intervals for ~100 seconds. The voltage waveform consists of two parts, an initial spike that lasts for about 100 ms and heats the sample quickly, followed by a slow changing phase to hold the temperature constant after the initial spike (Figure 4.10). The fast phase (spike) needs to be regularly calibrated and shaped because it drifts depending on environmental conditions (e.g. type of slides used) as well as due to slight drifts of the laser beam position.

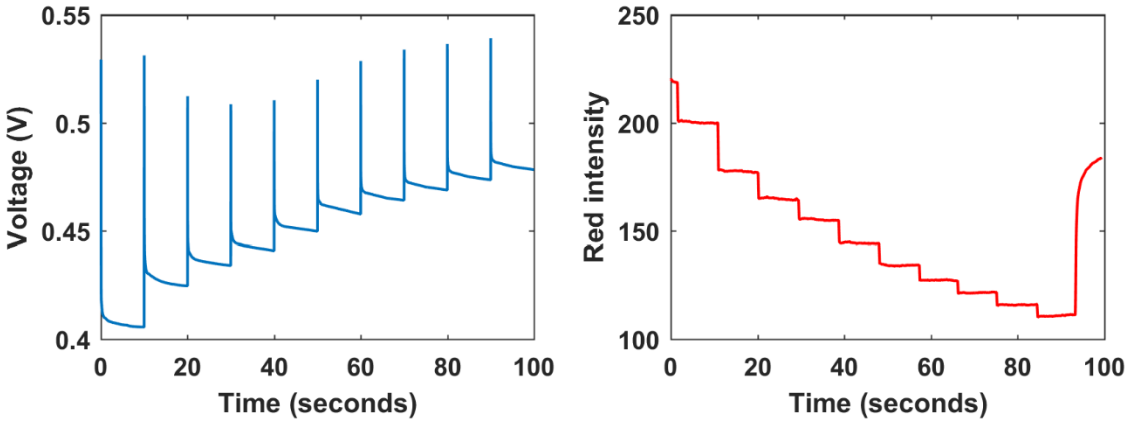


Figure 4.10: (A) Voltage waveform for programmed temperature jump and (B) intensity change due to heating.

In order to shape the initial fast changing phase of the voltage waveform, it is easiest to start with an old voltage profile that the user can use as a framework for the new jump. Generate and analyze a video of the jump by using mCherry fluorescence as a proxy for temperature. The voltage values can then be increased or decreased depending on whether a higher or a lower temperature increase is desired. A good way to do this is to fit about 20-30 points using an exponential function with 1 or 2 terms. The function can then be modified to increase or decrease the voltage values. This has the additional benefit of generating a smoothed waveform that has fewer breaks and yields a more continuous temperature change. Generate and analyze a video with the jump after every fitting step. After the first 20-30 points have been fit to satisfaction and yield the required temperature change, fit the next 20-30 points that contain at least 5-10 points from the last portion of the waveform that was shaped earlier. This ensures that while shaping, continuity is

maintained between the separately fitted parts of the final jump. Continue the process until the full jump is shaped and yield 10 step-shaped temperature jumps.

In figure 4.10(B) one can observe that the final mCherry red intensity does not return to its original state. This is due to irreversible bleaching from both temperature and photo-bleaching.

CHAPTER 5

HEAT SHOCK-INDUCED CHAPERONING BY HSP70 IS ENABLED IN-CELL

5.1 INTRODUCTION

In this chapter we describe how the FReI instrument described in the previous chapter was used to compare Hsp70-substrate interactions *in vitro* (with and without ATP and co-chaperon Hsp40) and in-cell. Our hypothesis was that the in-cell environment may be more conducive to proper ATP hydrolysis-dependent heat shock than *in vitro*. This question was motivated because there are many denaturant-based assays for Hsp70 function in the literature but these assays only capture binding with unnatural substrates. No studies currently show binding of Hsp70 to a native protein during heat shock.

The cell is a crowded environment (300-400 mg/mL of macromolecules)[11] containing many surfaces capable of weak interactions. Quinary structure (transient interactions evolved for function) and crowding (excluded volume) are features indispensable to the proper function of many cellular proteins [18,101]. On the other hand, non-specific sticking can destabilize proteins, reduce effective binding constants by competing with productive binding, or reduce the number of encounter complexes by reducing diffusion rates in the cell compared to *in vitro* [10,27].

These properties of the cytoplasm are difficult to mimic *in vitro*. The absence of an in-cell like environment is not a problem when folding of very stable proteins or very strong protein-protein interactions (nanomolar dissociation constant K_d) are studied in buffer, but it could have important consequences for the folding of marginally stable proteins or weak (micromolar dissociation constant K_d) protein-protein interactions, such as chaperoning.

Members of the 70 kDa family of heat shock proteins (Hsp70s) have long been implicated in the maintenance of in-cell protein homeostasis by binding substrates during early stages of folding or cellular stress [359–361]. Hsp70s are known to be promiscuous [362–364]. Given recent evidence that promiscuity and stress response could be tuned via quinary structure [85,166,365], it is possible that crowding and weak in-cell interactions could be important for proper Hsp70-substrate binding under heat shock. Hsp70 substrate affinity is foremost controlled by the binding and consecutive hydrolysis of ATP at the N-terminal

nucleotide binding domain (NBD) [366]. The NBD found in the Hsp70 family shares structural similarity to actin and some sugar kinases, underscoring the importance of ATP hydrolysis machinery in the cell [367,368]. ATP hydrolysis is induced by substrate binding to the C-terminal substrate binding domain (SBD) followed by domain rearrangement allowing up to two orders of magnitude increase in substrate affinity [360,369].

Hsp70 has been shown to bind short peptides [370], intrinsically disordered proteins (such as RCMLA and carboxamidomethylated ribonuclease A) [359,371,372], chemically denatured proteins (such as β-galactosidase) [373], and obligate substrates (such as luciferase) [373] *in vitro*. Few examples of binding upon heat shock exist, such as NCA-SNase and staphylococcal β-lactamase exist [359,374]. These two marginally stable proteins unfold at 37 °C, and Hsp70 could only bind them in their unfolded state. Such unfolded substrates do not provide insight on how Hsp70 binding proceeds with normally folded globular proteins during heat shock.

More recent *in vitro* studies have shown binding of bacterial Hsp70 (DnaK) to full length native substrates upon force-unfolding [180]. In these studies, DnaK was not only able to stabilize unfolded proteins to potentially prevent misfolding, but also bound near-native protein conformations to stabilize them against unfolding [375]. Significant advances have also been made in our understanding of the physical interactions between substrate and chaperone [179,364,376,377]. However no study has compared *in vitro* and in-cell substrate binding of a normally folded full-length protein during heat shock [359], even though heat shock is at the core of the concept of a "Heat shock protein."

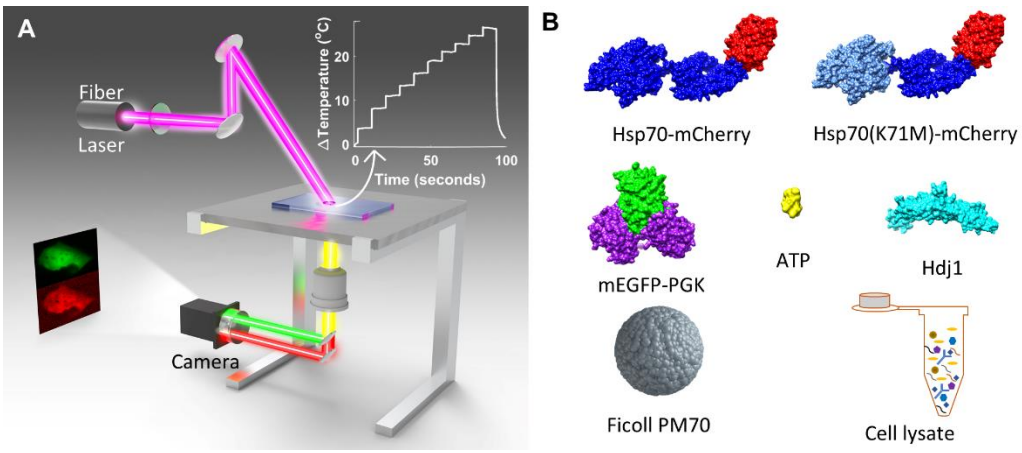


Figure 5.1: Experimental design and molecular structures. (a) Schematic of epifluorescence microscope equipped with a 2000 nm wavelength infrared fiber laser (Advalue Photonics) for fast programmed temperature ramping. (b) Structures of proteins and molecules used in experiments and Figures 5.4,5.6-5.7. Structures were accessed by either PDB IDs Hsp70 (2KHO), PGK (3PGK), Hsp40 (2QLD), mCherry (2H5Q)

Figure 5.1 (cont.)

and mEGFP (3EVP) or PubChem ID for ATP (5957). All structures including Hsp70-mCherry were rendered using USCF Chimera 1.12.

To elicit a chaperoning response that mimics the cellular heat shock response, we thermally denatured PGK [378] in mammalian U-2 OS cells. Figure 5.1A shows a schematic of the epifluorescence microscope equipped with a 2000 nm infrared wavelength fiber laser that is used to heat the aqueous medium in and around the cell to induce a heat shock. We also show that sticking overwhelms any productive binding of Hsp70 to its substrate phosphoglycerate kinase (PGK) *in vitro*.

Since we subject cells to sudden heat-stress in our experiments, we studied the human cytoplasmic heat-inducible Hsp70 isoform, also known as Hsp72 or HSPA1A (Appendix C, C.2), which we refer to as Hsp70 throughout. To study binding, we engineered a FRET pair consisting of mHsp70 (Hsp70 labeled by the red acceptor mCherry at the C-terminus) and ePGK (PGK labeled by the green donor mEGFP at the N-terminus) (Figure 5.1B). mCherry and mEGFP are known not to interact directly with one another below the mM level [85]. The intensity of green and red fluorescence was imaged onto a CMOS sensor as a function of temperature. Binding was detected by change in FRET efficiency between ePGK and mHsp70. Both protein concentrations were in the few micromolar range in-cell, comparable to physiological concentrations of these proteins in mammalian cells.

This chapter shows in five steps that the cellular environment is needed for an effective chaperoning response. 1) I first show that fluorescently labeling PGK and Hsp70 does not significantly disrupt their stability and function, 2) I show that PGK is an Hsp70 substrate, i.e. its refolding is enhanced in a standard denaturant assay, 3) I show that Hsp70 upon heat shock *in vitro* does not bind to PGK any differently than an ATPase activity-deficient mutant, and that crowding or co-factors alone do not rescue this deficiency (Figure 5.1B), 4) I show that an ATPase-dependent heat shock response does occur in mammalian cells, and 5) finally I show that the in-cell response has an onset even below the melting temperature of PGK.

There is a precedent for such “pre-emptive holdase” activity for the Hsp70 homolog DnaK, where the chaperone binds to full-length native proteins in force-unfolding experiments [180]. Our experiments establish such a mechanism for Hsp70 and its substrate PGK in cells.

5.2 METHODS

5.2.1 PROTEIN EXPRESSION AND PURIFICATION

Protein expression was carried out according to previously established protocol [95]. Fusion protein sequences were cloned into pDream 2.1/MCS vector by Genscript Corp. and used for dual expression in E.

coli and mammalian cells unless stated otherwise. Sequences of relevant Hsp70 (HSPA1A), Hsp40 (Hdj1) and PGK mutants are listed in Appendix D, D.1. Hsp70 was cloned with a C-terminal mCherry (mHsp70) and mEGFP (mEGFP-Hsp70) fusion protein with a 6×His-tag for purification. Hsp40 sequence with 6×His tag was cloned into pET15b for expression in *E. coli*. All PGK sequences were also cloned into pDream as N-terminal mEGFP fusion proteins with 6xHis-tags for purification purposes (ePGKs). PGK mutants (PGK0-PGK3) with varying stabilities were designed in-house using PCR site-directed mutagenesis. The goal of the fluorescent labeling was not to obtain a maximum FRET signal when Hsp70 and PGK interact, but to obtain enough signal while not disrupting the function of the proteins.

For protein expression, the plasmid of interest was chemically transformed into BL21-CodonPlus(DE3)-RIPL cells (Agilent) according to manufacturer protocol. Transformed cells were grown in 1L Lennox LB broth (Fisher Biosciences) at 37 °C for 4 hours. The cells were then induced with 1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG, Inalco) and allowed to express protein for 12-16 h at 18 °C. The cells were pelleted by centrifugation (Beckman Coulter Avanti J-E, 5000 rpm, 25 minutes, 10 °C) and resuspended in lysis buffer (500 mM NaCl, 50 mM Na2PO4, 20 mM imidazole, pH 7.4). The cell suspension was then supplemented with protease inhibitors (1 mM PMSF, Sigma) and 20 µL DNaseI (New England Biolabs)/10 mL suspension. Cells were then broken by sonication (Qsonica, 70% intensity, 6 s pulse/min, 6 min total process time). The cell lysate was re-centrifuged (10000 rpm, 20 min, 10 °C) to remove debris and filtered once through a 0.45 µm filter (Millipore) and twice through 0.22 µm filter (Millipore).

Each protein was isolated by affinity chromatography. Proteins were purified on a 5-mL HisTrap HP column by FPLC (GE Healthcare Life Sciences AKTA) and eluted with 500 mM NaCl, 50 mM Na_2HPO_4 and 500 mM imidazole; pH 7.4. Protein purity was assessed using SDS-PAGE gel electrophoresis, mass spectroscopy and protein was dialyzed into appropriate storage buffers: 1) PGK storage buffer: 20 mM phosphate, pH 7 and 2) Hsp70/Hsp40 storage buffer: 100 mM Tris-HCl, 0.1 mM EDTA, 100 mM NaCl, pH 6.9 with KOH. For long term storage, glycerol was added to a final concentration of 30% volume/volume, flash frozen in liquid N_2 and stored at -70 °C.

5.2.2 MHSP70 AND MHSP70K71M ATPASE ACTIVITY ASSAY

Prior to all measurements the glycerol from the frozen stocks were removed by spin filtration buffer exchange. ATPase assays were conducted with the ATP sensor protein PercevalHR[379]. The plasmid for bacterial expression, pRsetB-PercevalHR (Addgene plasmid # 49081), was obtained from Addgene and was a gift from Gary Yellen. The protein was expressed as described in the main text and dialyzed in 5 mM MOPS, 300 mM NaCl, pH 7.3. Absorbance measurements were made in K1 buffer with 2 mM total nucleotide (2 mM ATP, 2 mM ADP or 1 mM each of ATP and ADP) and 5 µM PercevalHR. ATPase

assays were conducted also in K1 buffer with 5 µM mHsp70 or mHsp70K71M, 2.5 µM Hsp40, 10 µM PercevalHR and 1 mM ATP. The PercevalHR emission intensity was monitored at 520 nm for 500 nm excitation for 2 hours at 15 second intervals at 37 °C.

5.2.3 STOPPED FLOW PGK REFOLDING ASSAY

To monitor PGK unfolding and refolding, a FRET-labeled PGK1 (FRET-PGK1) was designed with an N-terminal AcGFP1 and a C-terminal mCherry label and cloned into a pDream vector by Genscript [122].

Stopped flow measurements were conducted with an Olis RSM 1000 instrument. FRET-PGK1 was excited at 475 nm and emission spectra were collected between 493 to 720 nm at 31 scans/sec scan speed. Donor and acceptor intensity were calculated by measuring area under the curve between 535-585 nm and 585-635 nm, respectively. FRET efficiency was calculated as described in the methods section of the main text. Stopped flow unfolding of FRET-PGK1 was measured by rapid mixing of 82 µM FRET-PGK1 and 0.6 M GuHCl in a 1:5 ratio. The final concentration after mixing for unfolding experiments was 13.6 µM FRET-PGK1 and 0.5 M GuHCl. Stopped flow refolding of FRET-PGK1 was conducted by rapidly diluting 82 µM FRET-PGK1 in 0.5 M GuHCl in buffer with or without Hsp70/Hsp40 in a 1:5 ratio. The final concentration after mixing for refolding experiments was 13.6 µM FRET-PGK1 and 0.083 M GuHCl. Chaperone free refolding buffer contained 0.5 mM ATP and 0.5 mM DTT in K1 buffer (25 mM HEPES, 50 mM KCl and 10 mM MgCl2; pH 7.6 with KOH). Chaperone assisted refolding was conducted in K1 buffer with 0.5 mM ATP, 0.5 mM DTT, 24 µM Hsp70 and 24 µM Hsp40.

5.2.4 U-2 OS CYTOPLASMIC LYSATE PREPARATION

Cells were grown to 70% confluence in T-75 culture flasks (Thermo Fisher Scientific). Prior to lysis, growth medium was removed, and cell monolayer was washed with 10 mL PBS (Corning). Cells were trypsinized (Corning) with 1 mL trypsin/flask and collected in 15 mL conical tubes. The cells were then pelleted by centrifuging for 5 minutes at 300×g (NuAire Awel C48). The supernatant was aspirated, and the cells were washed with 10 mL ice-cold PBS. The pellet was spun down again for 5 minutes at 300×g after which the PBS wash was aspirated. A 1 mL aliquot of lysis buffer was prepared by mixing cold Pierce lysis buffer (Thermo Fisher Scientific, 87787) was supplemented with 10 µL 100× protease inhibitor cocktail (Thermo Fisher Scientific, 78442). The pellet was resuspended in 200 µL ice-cold lysis buffer and incubated on ice for 30 minutes. The lysed cells were centrifuged again for 5 minutes at 13000 g on a tabletop centrifuge (Eppendorf 5415D) and the supernatant was collected. The protein concentration of the lysate was estimated by a BCA assay (Thermo Fisher Scientific, 23225). Aliquots of cell lysate were stored at -20 °C.

5.2.5 *IN VITRO* CIRCULAR DICHROISM AND FLUORIMETER MELTS

Prior to all measurements the glycerol from the frozen stocks were removed by spin filtration buffer exchange. Tryptophan fluorescence measurements and *in vitro* FRET binding experiments were conducted on an FP8300 spectrofluorometer equipped with Peltier temperature control (JASCO). For *in vitro* PGK characterization, tryptophan was excited at 295 nm, and emission spectra were collected from 290 to 450 nm. Samples were measured in 300 µL cuvettes at 5 µM concentrations, unless otherwise noted.

Circular dichroism (CD) was measured using a J-715 spectropolarimeter with Peltier temperature control (JASCO). Unless otherwise noted, all spectra were recorded from 250 to 200 nm at a scan rate of 50 nm/min at 1 nm resolution and averaging five accumulations. Thermal melts were done using a 2 mm path length cuvette. Unless otherwise noted, protein concentration in circular dichroism experiments was 5 µM. Protein unfolding was monitored by measuring the change in mean residue ellipticity at 222 nm.

In vitro FRET-PGK1 unfolding with denaturant guanidinium hydrochloride (GuHCl) was measured by conducting isothermal titrations at 20 °C of 5 µM FRET-PGK1 with GuHCl between 0-0.8 M. AcGFP1 was excited at 475 nm and emission spectra were collected from 470 to 700 nm in 300 µL cuvettes.

For all FRET measurements mEGFP was excited at 485 nm and emission spectra were collected from 480 to 700 nm in 300 µL cuvettes. Unfolding and aggregation of Hsp70 was monitored by the change in FRET efficiency *vs.* temperature by melting an equimolar mixture of mHsp70 (1 µM) and mEGFP-Hsp70 (1 µM) with and without 2 mM ATP and 10 mM DTT in K1 buffer.

For *in vitro* FRET binding experiments ePGKn (n=0-3) was heated with mHsp70 with 10 mM DTT and 2 mM ATP in K1 buffer. For measurements with 1:1 mHsp70:ePGK, 1 µM ePGK was mixed with 1 µM mHsp70, for 1:5 mHsp70:ePGK, 2.5 µM ePGK was mixed with 0.5 µM mHsp70 and for 5:1 mHsp70:ePGK, 5 µM mHsp70 was mixed with 1 µM ePGK. Control measurements were conducted with either 5:1 mHsp70:mEGFP (5 µM mHsp70 and 1 µM mEGFP) or 5:1 mCherry:ePGKs (5 µM mCherry and 1 µM ePGKs). ATP was maintained in all *in vitro* experiments by addition of an ATP regenerating mix. The mix consists of 50 mM creatine phosphate, 0.2 mg/mL of creatine phosphokinase in appropriate buffer. Regeneration of ATP has been shown to increase refolding of denatured luciferase by Hsp70 [380]. For crowding Ficoll70 (Sigma) was added to a final concentration of 300 mg/mL along with 1 µM Hsp40. For measurements in cell lysate and Ficoll70, 0.6 mg/mL cell lysate was added in addition to 300 mg/mL Ficoll70 and 1 µM Hsp40. Temperature melts with Ficoll showed signs of aggregation at higher temperatures.

5.2.6 B-GALACTOSIDASE HSP70 REFOLDING ASSAY

The refolding assay was performed according to previously described protocol [373]. Briefly 1 μM β-galactosidase was denatured by 10-fold dilution into denaturation buffer (K1 buffer supplemented with 5 mM BME) with 3 M GuHCl. For a folded control measurement β-galactosidase was diluted into denaturation buffer without GuHCl. Both denatured and folded β-galactosidase were then incubated at 30 °C for 30 minutes. Refolding was then initiated by 125-fold dilution of denatured β-galactosidase into refolding buffer (K1 supplemented with 10 mM DTT and 2 mM ATP) supplemented with BSA (3.2 μM) or chaperone (1.6 μM Hsp70 or 3.2 μM Hsp40) at 4 °C. For a folded control, folded β-galactosidase was also diluted 125-fold into refolding buffer with 3.2 μM BSA. At each time-point 10 μL of the appropriate mixture was removed and added to 10 μL 0.8 mg/mL ONPG substrate. The ONPG was then incubated at 37 °C for 15 minutes. The chromogenic reaction was arrested by adding 80 μL 0.5 M sodium carbonate. The extent of chromogenic reaction was estimated by measuring absorbance at 412 nm in a UV-Vis spectrophotometer (Shimadzu UV-1800) in a quartz cuvette. Refolding activity of Hsp70 was estimated by calculating the percentage recovery of GuHCl denatured β-galactosidase with respect to folded control.

5.2.7 FRET BINDING EXPERIMENTS IN CELLS

U-2 OS cells (ATCC) were grown in DMEM (Corning) supplemented with 10% FBS, 5% penicillin streptomycin (Fisher) and 5% sodium pyruvate (Fisher). At ~70% confluence cells were transferred on to pre-cleaned glass coverslips (18 mm × 18 mm, #1.5) in a 35 mm falcon dish (MatTek Corp.) by trypsinization. They were then co-transfected with the appropriate plasmid(s) with Lipofectamine 2000 (Fisher) (5 μL Lipofectamine for every 2 μg plasmid) in DMEM without penicillin. 4 hours after transfection cells were washed with PBS and allowed to grow for 18-26 hours in DMEM with penicillin. For co-transfection plasmid concentrations were optimized for 1:1 expression of both proteins. For mHsp70 and ePGK roughly 2 μg mHsp70 and 1 μg ePGK was used for co-transfecting 3 35 mm falcon dishes. For mCherry/ePGK and mHsp70/mEGFP controls roughly 2 μg mHsp70/ePGK and 1 μg mCherry/mEGFP was used for co-transfecting 3 35 mm falcon dishes.

18-26 hours after transfection coverslips were washed with PBS (Corning) and adhered to a slide using 120 μm thick spacers (Grace Bio-Labs) for imaging. Cells were imaged in Opti-MEM (Fisher) supplemented with 15% FBS. mEGFP in the cells was excited by a white LED by passing the light through a Chroma ET470/40x bandpass filter and mCherry was excited through a ET580/25x bandpass filter [95]. Both mEGFP and mCherry emission was monitored on a Lt225 (Lumenera Corp.) camera equipped with a CMOS sensor.

Prior to the experiment, each fluorescent protein was individually excited to obtain intensities at room temperature. Hsp70 and PGK were assumed to be in the unbound state at room temperature. Initial concentrations of the two proteins was calculated from their individual fluorescence intensities using a predetermined *in vitro* calibration. Only cells that expressed roughly equimolar quantities of mEGFP and mCherry labeled proteins were chosen for experiments. Cells were heated using fast temperature jumps between 19-46 °C and imaged at 60 fps using LabVIEW (National Instruments) [95].

5.2.8 DATA ANALYSIS

All data was analyzed using MATLAB (MathWorks). *In vitro* tryptophan fluorescence measurements were analyzed by monitoring wavelength peak shift. CD spectra were analyzed by calculating the change in the CD mean residue ellipticity (MRE) at 222 nm [305].

Both *in vitro* and in-cell FRET binding measurements were monitored using FRET efficiency (E_{FRET}). The E_{FRET} was calculated according to:

$$E_{FRET} = \frac{mCherry\ (acceptor)\ intensity}{mEGFP\ (donor)\ intensity + mCherry\ intensity} \quad [5.1]$$

In vitro donor and acceptor intensity were calculated by measuring area under the curve between 495 – 585 and 585 – 700 nm, respectively. In-cell binding was detected via FRET from mEGFP to mCherry and quantified as described previously in [122]. FRET efficiency was estimated at fourteen equally spaced temperatures from 20-46 °C by interpolation and normalized to the first temperature point. The traces were then corrected for bleaching by fitting a line through the first three data points and subtracting the fitted slope. The corrected traces were then averaged.

Melting temperature (T_m), concentration midpoint (C_m) or binding temperature (T_0) were calculated using a two-state sigmoidal fit to the experimental data. If binding is driven by unfolding, sigmoidal curves are also appropriate for the bimolecular process. Native and denatured state baselines were assumed to be linear as a function of perturbant (Y), which can be either temperature or concentration so that the signal S can be estimated by:

$$S_i = m_i(Y - Y_X) + b_i\,, \quad [5.2]$$

where i is either the native or substrate-chaperone unbound state (N) or the denatured or substrate-chaperone aggregated state (D), m is the slope and b is the intercept. Y_X is the corresponding midpoint of

the curve (T_m, C_m or T_0) with the temperature T or concentration C. The total signal $S(T)$ can then be estimated as a sigmoidal function with respect to temperature T:

$$S(T) = S_N(Y)f_N(Y) + S_D(Y)f_D(Y)\,, \quad [5.3]$$

where, f_N and f_D are the populations of the N and D states, respectively.

$$K_{eq} = e^{-\frac{\Delta G_{N\rightarrow D}}{RT}}, \quad [5.4]$$

$$\Delta G_{N\rightarrow D} \approx \delta g_1(Y - Y_X), \quad [5.5]$$

$$f_N = \frac{K_{eq}}{1+K_{eq}}, \text{ and } f_D = \frac{1}{1+K_{eq}}$$

Fits to equation 5.2 yield the melting temperature (T_m), concentration midpoint (C_m) or binding temperature (T_0). For a three-state sigmoidal fit the total signal $S(T)$ can then be estimated with respect to temperature T as:

$$S(T) = S_U(T)f_U(T) + S_B(T)f_B(T) + S_A(T)f_A(T)\,, \quad [5.6]$$

where *U* is the substrate-chaperone unbound state, *B* is the substrate-chaperone bound state and *A* is the substrate-chaperone aggregated state. S_U, S_B and S_A can be modeled by equation 5.1. f_U, f_B, and f_A are the populations of the unbound, bound and aggregated states, respectively.

$$K_{eq1} = e^{-\frac{\Delta G_{U\rightarrow B}}{RT}}, \quad [5.7]$$

$$K_{eq2} = e^{-\frac{\Delta G_{B\rightarrow A}}{RT}}, \quad [5.8]$$

$$\Delta G_{U\rightarrow B} \approx \delta g_1(T - T_{01}), \text{ and } \Delta G_{B\rightarrow A} \approx \delta g_1'(T - T_{02}),$$

where T_{01} is the binding temperature midpoint for substrate and chaperone and T_{02} is the aggregation temperature midpoint where both chaperone and substrate are unfolded above 50 °C and aggregate.

$$f_U = \frac{K_{eq1}}{1+K_{eq1}+K_{eq1}K_{eq2}}, \quad [5.9]$$

$$f_B = \frac{K_{eq1}K_{eq2}}{1+K_{eq1}+K_{eq1}K_{eq2}}, \quad [5.10]$$

$$f_A = \frac{1}{1+K_{eq1}+K_{eq1}K_{eq2}} \quad [5.11]$$

The kinetic data for mHsp70 ATPase activity was fit using the equation:

$$\frac{(1+a)exp^{-kt}}{2} + \frac{1-a}{2} \quad [5.12]$$

where, k is the reaction rate. The mHsp70 fits yielded a value of -0.6623 for a which was fixed at this value for fitting the kinetic data for mHsp70K71M. The data for mHsp70K71M was then fit to the equation:

$$\frac{(1-0.6623)exp^{-kt}}{2} + \frac{1+0.6623}{2} \quad [5.13]$$

5.3 RESULTS

5.3.1 MCHERRY LABELING ONLY SLIGHTLY REDUCES HSP70 ACTIVITY AND STABILITY *IN VITRO*

Although C-terminal fluorescent protein fusions of Hsp70 have been shown to be functional previously [381], fluorescent protein tags such as mCherry could perturb stability and function of the tagged Hsp70 [127]. To confirm that labeled mHsp70 retains its refolding activity, we conducted standard refolding assays with β-galactosidase [373] (Figure 5.2A). Wild-type Hsp70 (wt-Hsp70) with co-chaperone Hsp40 recovered about 55% of unfolded β-galactosidase, while mHsp70 with Hsp40 recovered about 35%.

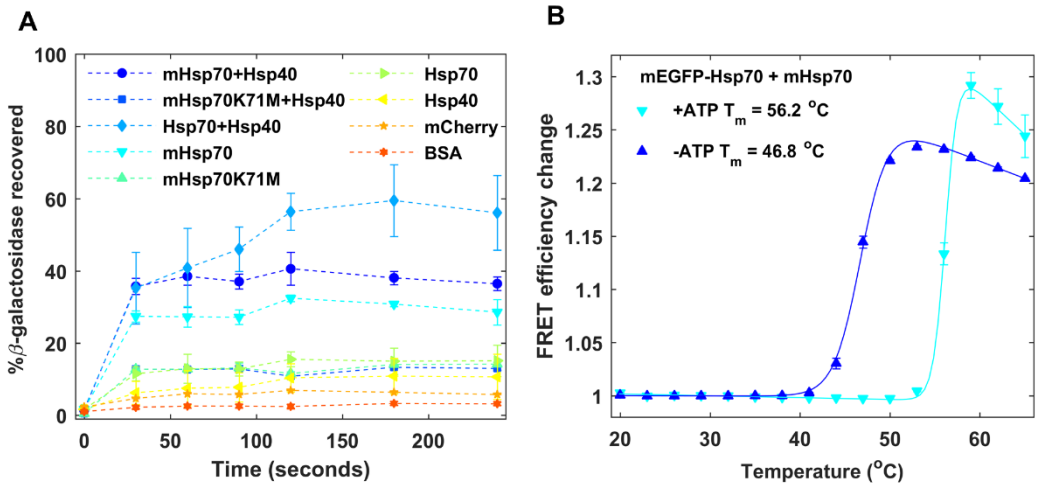


Figure 5.2: *In vitro* characterization of mHsp70. Filled markers and dashed lines show experimental data and solid lines are fits to experimental data. Error bars are standard error from mean for two measurement repeats. (a) β-galactosidase assay to assess refolding activity of mHsp70 compared to wt-Hsp70 and mHsp70K71M. (b) Unfolding of mHsp70 probed by FRET as a function of temperature with and without ATP. FRET was monitored between mEGFP-Hsp70 and mHsp70. Errors in T_m reflect the standard deviation of T_m values obtained from individually fitting two measurements.

Hsp70 function is modulated by ATPase activity and mutants lacking ATPase activity should show diminished substrate recovery. A lysine at position 71 (K71) in the nucleotide binding domain is essential for ATP hydrolysis and any mutations at that position abrogate ATPase activity [382]. The ATPase activity was measured by monitoring the emission of PercevalHR at 520 for 500 nm excitation (Figure 5.3). The absorbance spectrum of the protein PercevalHR shifts according to the nucleotide present (Figure 5.3A). With 2 mM ATP the absorbance peak at 500 nm increased while with 2 mM ADP the absorbance peak at 420 nm increased. As the ATP is hydrolyzed the absorbance at 500 nm decreases and hence the emission at 520 nm for the 500 nm excitation also decreases. PercevalHR was used to measure the rate of ATP hydrolysis by monitoring the emission intensity at 520 nm for 500 nm excitation (Figure 5.3B). The curves were fit using 1st order reaction kinetics and yielded an ATP hydrolysis rate of 6.31 X 10^{-1} sec^{-1} for mHsp70 and 3.69 X 10^{-2} sec^{-1} mHsp70K71M.

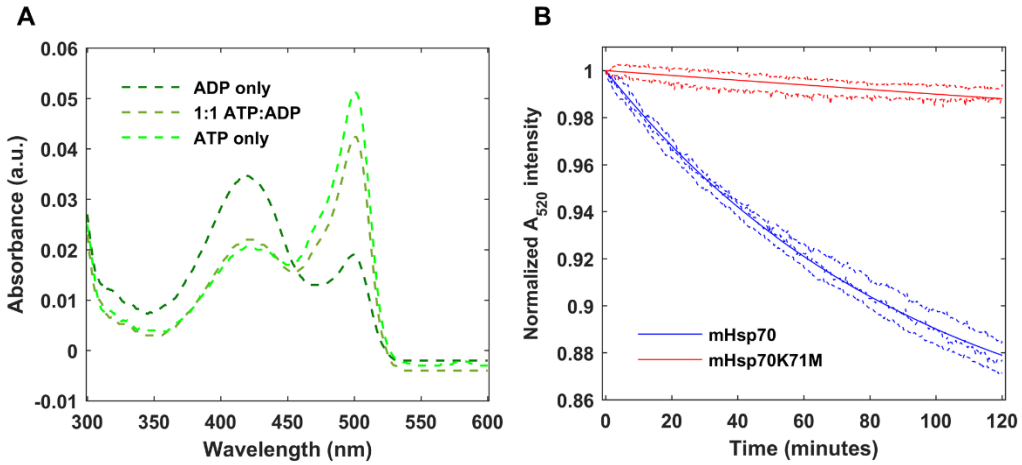


Figure 5.3: *In vitro* measurement of mHsp70 and mHsp70K71M ATPase activity using PercevalHR. Dotted lines and dashed lines show experimental data and solid lines show fits to experimental data. (A) Absorbance spectra of Perceval-HR with 2 mM ADP (dark green), 1 mM ATP + 1 mM ADP (green) and 2 mM ATP (light green). (B) ATPase activity of mHsp70 (blue) and mHsp70K71M (red) monitored by Perceval-HR emission intensity at 520 nm for 500 nm excitation.

As expected, the rate of ATP hydrolysis by the mutant was decreased by up to an order of magnitude and as a result mHsp70K71M showed much lower (<15%) β-galactosidase recovery, whether Hsp40 was present or not (Figure 5.2B). mCherry, Hsp40 and the model crowder protein BSA showed no significant chaperoning activity on their own.

Surprisingly, mHsp70 without the co-chaperone Hsp40 recovered about ~28% of unfolded β-galactosidase, refolding almost twice as much β-galactosidase as wt-Hsp70 by itself (~15% recovery). We speculate that

mCherry linked to Hsp70 helps with substrate recruitment due to crowding or hydrophobic surface patches on the mCherry surface. In effect, mCherry acts as a low-performance co-chaperone substitute.

To confirm that the stability of the labeled construct is the same as wt-Hsp70, thermal stability of mHsp70 was characterized in a temperature-controlled fluorimeter and compared to published results. The thermal stability of Hsp70 has been previously characterized *in vitro* where Hsp70 unfolded with a melting temperature (T_m) of 42 °C without ATP and 60 °C with Mg^{2+}/ATP [383]. This unfolding both with and without ATP was accompanied by a 25% structure loss (by circular dichroism (CD)). Using FPLC gel-filtration the authors also showed that this unfolding transition led to self-aggregation and formation of oligomeric species.

We monitored Hsp70 unfolding and aggregation by tracking FRET efficiency (E_{FRET}) change of an equimolar mixture of donor-Hsp70 (mEGFP-Hsp70) and acceptor-Hsp70 (mHsp70) *in vitro*. Both mCherry and mEGFP are stable in the temperature range tested and do not unfold until ≥70 °C (Figure 5.4). Thermal melts of mHsp70 and mEGFP-Hsp70 showed the expected increase in E_{FRET} upon unfolding due to self-aggregation.

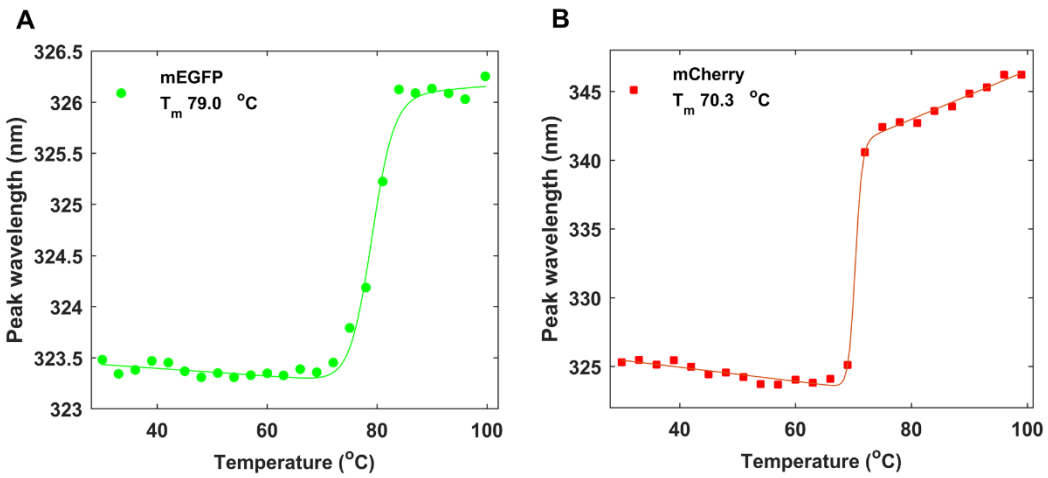


Figure 5.4: *In vitro* thermal unfolding of mEGFP and mCherry monitored by tryptophan fluorescence peak shift. Errors shown reflect the 1 standard deviation precision of the fit. Systematic errors are not accounted for. Filled markers show experimental data and solid lines are fits to experimental data. (A) Thermal denaturation of mEGFP fit to a T_m of 79 °C. (B) Thermal denaturation of mCherry fit to a T_m of 70.3 °C.

The E_{FRET} curves fit to a T_m of 45 °C without ATP, and 55 °C with ATP (Figure 5.2B) similar to previously published results. mHsp70 unfolding was also characterized by CD and tryptophan fluorescence, and showed similar unfolding behavior with the different spectroscopic probes (Table 5.1). These changes in

FRET at high temperature are not to be confused with productive binding of Hsp70 to substrates. See Appendix D, D.3 for additional melting curves.

Table 5.1: Transition temperature mid-points for Hsp70 and mHsp70 calculated from two-step sigmoidal fits to the experimental data. Errors shown reflect the 1 standard deviation precision of the fit. Systematic errors are not accounted for.

Protein	Nucleotide	Nucleotide concentration (mM)	Transition temperature (°C)		
			CD	Trp SVD	FRET
Hsp40	NA	0	56.2 ± 0.3	NA	NA
Hsp70	NA	0	45.8 ± 0.9	47.3 ± 0.2	NA
Hsp70	ATP	20	NA	52.4 ± 0.1	NA
Hsp70-mCherry	NA	0	45.8 ± 0.5	44.4 ± 0.3	45.5 ± 0.1
Hsp70-mCherry	ATP	20	NA	53.5 ± 0.2	55.2 ± 0.1
Hsp70-mCherry	ADP	20	NA	53.5 ± 0.4	NA

5.3.2 UPON REFOLDING FROM DENATURANT, PGK IS A SUBSTRATE OF HSP70 *IN VITRO*

We designed and studied four different mutants of ePGK (Table 5.2) with successively lower thermal denaturation midpoints T_m to map out the relationship between substrate stability and chaperone binding. The thermal denaturation of ePGK mutants was probed by tryptophan fluorescence peak shift or CD mean residue ellipticity (Figure 5.5).

Table 5.2: Melting temperatures and point mutations for PGKs tested. T_m was estimated from two-step sigmoidal fits to the experimental data. Errors shown reflect the 1 standard deviation precision of the fit. Systematic errors are not accounted for.

Protein	Mutations (Y122W+)	Melting T_m (°C)		In-cell binding T_0 (°C)
		CD	Trp peak shift	FRET
mEGFP-PGK0	P111T	50.4 ± 0.3	48.4 ± 0.1	>46
mEGFP-PGK1	W308F, W333F	43.2 ± 0.5	45.8 ± 0.3	>41.0 ± 0.4
mEGFP-PGK2	P111T, P204H	43.3 ± 0.4	43.3 ± 0.1	39.1 ± 0.1
mEGFP-PGK3	W308F, P204H	41.0 ± 0.4	40.5 ± 0.3	37.1 ± 0.2

Mutant ePGK0 has a T_m (48.4 °C) well above the viable temperature range for mammalian cells. It serves as a control where little binding is expected below 48 °C. ePGK1 (T_m = 45.8 °C), ePGK2 (T_m = 43.3 °C) and ePGK3 (T_m = 40.5 °C), although based on very different mutations, all have a T_m that lies in the range where unfolding is accessible in the cell. All mutants have T_ms above physiological temperature for the U-2 OS human cell line.

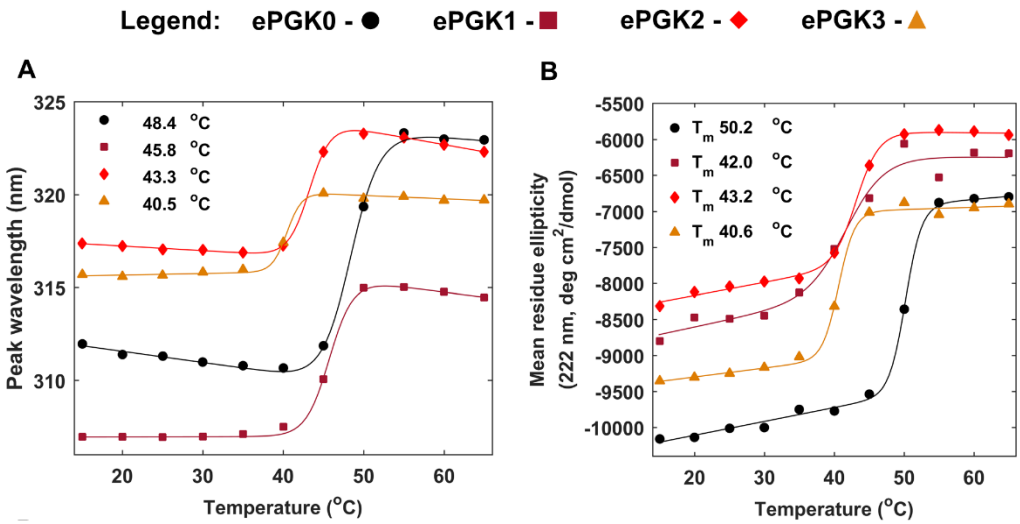


Figure 5.5: *In vitro* characterization of ePGKs. Filled markers show experimental data and solid lines are fits to experimental data. Errors shown reflect the 1 standard deviation precision of the fit. Systematic errors are not accounted for. (A) Unfolding of ePGK0 through ePGK3 monitored by tryptophan fluorescence peak shift *vs.* temperature. (B) *In vitro* thermal unfolding of ePGK(0-3) monitored by circular dichroism. Thermal unfolding was monitored by change in mean residue ellipticity for ePGK0-3 at 222 nm. All unfolding curves were fit using a two-state model.

E. coli PGK is a known substrate of the *E. coli* Hsp70, DnaK [378]. To show that our eukaryotic PGK homolog is an *in vitro* substrate of human Hsp70, we performed stopped-flow PGK refolding assays. PGK1 was labeled with AcGFP1 at the N-terminus and mCherry at the C-terminus [121] such that refolding can be monitored by the change in FRET efficiency between AcGFP1 and mCherry as the two termini move closer together. The refolding rate of initially GuHCl-denatured FRET-PGK1 in buffer and buffer supplemented with Hsp70/Hsp40 was compared. Optimal GuHCl concentrations for refolding experiments were determined by performing isothermal titrations of FRET-PGK1 with GuHCl, yielding a mid-point concentration (C_m) of 0.31 M (Figure 5.6A) in agreement with previously reported values [311].

Unfolding of FRET-PGK1 in 0.5 M GuHCl showed a fast phase with $\tau_1 \approx 5$ secs and a slow phase with $\tau_2 \approx 190$ secs (Figure 5.6B). 40% of the unfolded PGK refolded in the dead time (~1 ms) of the instrument. Addition of Hsp70 and Hsp40 improved folding efficiency of FRET-PGK1 from 58% without Hsp70 to

72% with Hsp70. Thus, PGK1 refolding from denaturant is chaperoned by Hsp70 and eukaryotic PGK is an Hsp70 substrate.

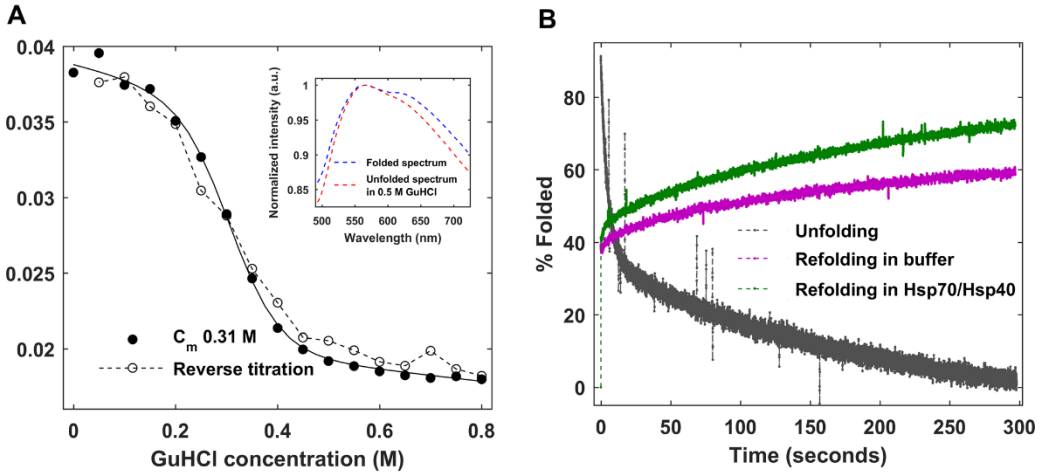


Figure 5.6: *In vitro* FRET-PGK1 refolding assay with Hsp70. Markers and dashed lines show experimental data and solid lines show fits to experimental data. (A) Isothermal denaturation of FRET-PGK1 with GuHCl. GuHCl titration of FRET-PGK1 (solid circles) fit to a C_m of 0.31 M. The unfolding of FRET-PGK1 in GuHCl is reversible to a large extent (open circles). Inset shows spectrum of folded FRET-PGK1 (blue) and unfolded FRET-PGK1 in 0.5 M GuHCl (red) recorded using the RSM 1000 stopped flow. (B) Unfolding of FRET-PGK1 in 0.5 M GuHCl (grey) proceeded via a fast phase with τ_1 = 5.2 secs and a slow phase with τ_2 = 187.2 secs.

5.3.3 MHSP70 IS NOT AN ATP-DEPENDENT HEAT SHOCK CHAPERONE FOR PGK *IN VITRO*

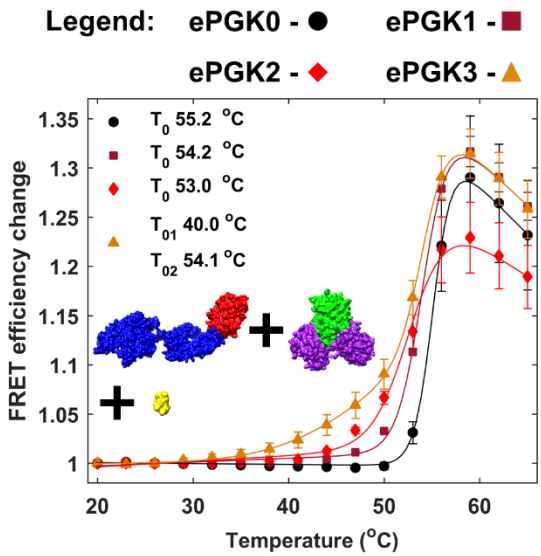


Figure 5.7: *In vitro* binding of 5:1 mHsp70 and ePGKs with ATP. Colors are ePGK0 (black), ePGK1 (dark-red), ePGK2 (red) and ePGK3 (orange). Filled markers show experimental data and solid lines are fits

Figure 5.7 (cont.)

to experimental data. Error bars are standard error from mean for two measurement repeats. Errors in T_m reflect the standard deviation of T_m values obtained from individually fitting three measurements.

To study chaperoning of ePGK0-3 by mHsp70 under heat shock *in vitro*, thermal scans were FRET-detected in buffer K1 with excess ATP (see Methods). 5:1, 1:1 and 1:5 concentration ratios of substrate and chaperone were used. In order to differentiate between melting and association, we refer to the protein thermal denaturation midpoint as T_m, and the association midpoint as T_0. Figure 5.7 shows the interaction of 5× excess mHsp70 (5 µM) with ePGK0-3 (1 µM). In ATP buffer, E_{FRET} *vs.* temperature shows a cooperative transition with an onset of 49 °C and a temperature midpoint (T_0) of 55 °C using a two-state sigmoidal fit (see Methods). This T_0 is similar to the T_m of mHsp70 with ATP in Figure 5.2B, and high temperature association is thus aggregation of unfolding ePGK0 with unfolded mHsp70, and not productive chaperone-substrate binding.

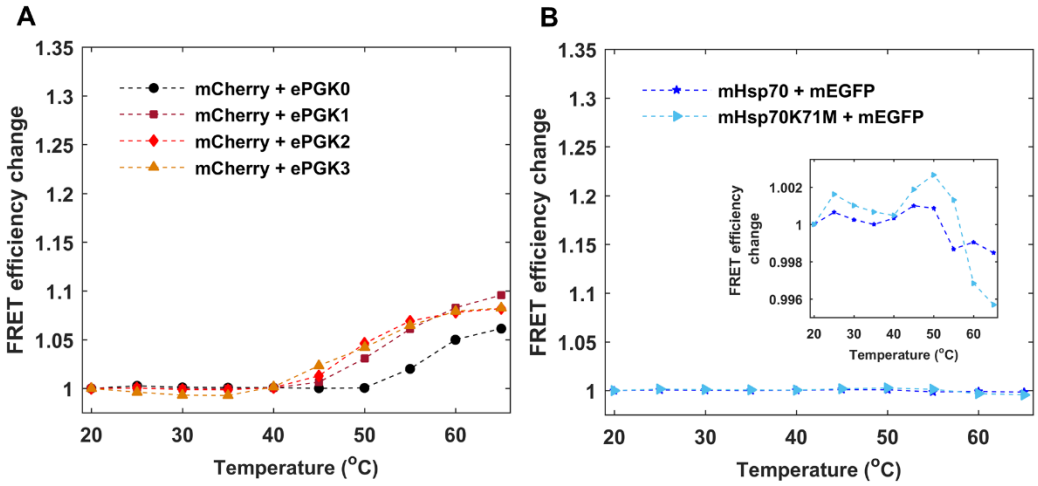


Figure 5.8: *In vitro* fluorometer controls with 5:1 mCherry:ePGKs and mHsp70:mEGFP with ATP. Filled markers and dashed lines show experimental data. Change in FRET efficiency *vs.* temperature for (A) mCherry binding to ePGKs(0-3). A small change in FRET efficiency, 1%, was observed due to non-specific interaction between mCherry and unfolded ePGKs (B) mHsp70 binding to mEGFP. Inset shows the enlarged y-axis for mHsp70/mHsp70K71M non-specific interaction with mEGFP. The change in FRET efficiency for non-specific interaction of either unfolded mHsp70 or unfolded mHsp70K71M with mEGFP is <0.5%.

A transition with a lower onset at 35 °C was clearly observed in Figure 5.7 for ePGK3 (35 °C). Control experiments with mHsp70+mEGFP as well as mCherry+ePGK0-3 showed no such increase in FRET efficiency E_{FRET} (Figure 5.8). Thus, the association is not of mEGFP with mHsp70 or mCherry with PGK. The E_{FRET} curve for ePGK3 and mHsp70 could be fit using a three-state model with two T_0s (T_{01} and T_{02}). T_{01} = 40 °C agrees with the T_m for ePGK3. This initial transition is due to association of mHsp70 and ePGK3

as it unfolds. T_{02} fit to 54 °C and corresponds to the aggregation of unfolded mHsp70 and ePGK3 as discussed above for ePGK0. The binding curves for ePGK1 and ePGK2 could not be fitted reliably by a three-state model but also showed some early onset.

We also tested binding at other mHsp70:ePGK ratios, most importantly for comparison with our in-cell mHsp70:PGK ratio of ≈ 1:1. At 1:1 or 1:5 mHsp70:ePGKs (Figure 5.9), no binding is detectable at all *in vitro*.

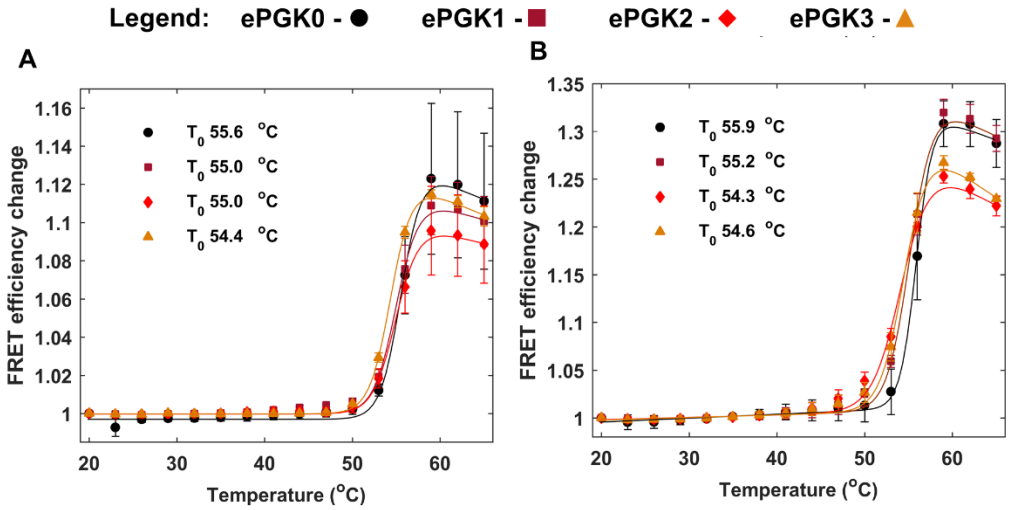


Figure 5.9: *In vitro* binding of 1:5 and 1:1 mHsp70:ePGKs with ATP. Filled markers show experimental data and solid lines are fits to experimental data. Error bars are standard error from mean for three and two measurement repeats for panel A and B, respectively. Fit errors reflect the 1 standard deviation precision of the fit. (A) Thermal unfolding of 1:5 mHsp70:ePGKs with ATP. The binding curves for ePGK0- (black), ePGK1- (dark red), ePGK2- (red) and ePGK3-mHsp70 (orange) were fit using a two-state model and yielded a binding T_0 of 55.6 °C, 55.0 °C, 55.0 °C and 54.4 °C, respectively. (B) Thermal unfolding of 1:1 mHsp70:ePGKs with ATP. The binding curves for ePGK0- (black), ePGK1- (dark red), ePGK2- (red) and ePGK3-mHsp70 (orange) were fit using a two-state model and yielded a binding T_0 of 55.9 °C, 55.2 °C, 54.3 °C and 54.6 °C, respectively.

We hypothesized that the earlier onset of Hsp70-PGK3 interaction at the 5:1 ratio might signal productive binding, which requires ATP-dependence. Such substrate binding would be abolished for the ATPase-deficient mutant mHsp70K71M. However, 5:1 mHsp70K71M:ePGK3 binding yielded curves like those observed with mHsp70 (Figure 5.10A). We conclude that our hypothesis of productive *in vitro* binding is falsified and that mHsp70 does not act as an ATP-dependent heat shock chaperone for ePGK3. As the substrate protein unfolds, it simply sticks *in vitro* to the Hsp70 chaperone in an ATPase activity-independent manner.

Any productive binding of mHsp70 to ePGK3 should show an E_{FRET} change above the background (~ 5%) observed for ATP-independent interaction. We next checked whether addition of 10 µM co-chaperone Hsp40 would increase E_{FRET} above the background. The addition of Hsp40 did not significantly increase mHsp70-ePGK3 binding, and mHsp70K71M again showed a similar curve with Hsp40. We tested various concentrations of co-chaperone Hsp40 to 5:1 mHsp70:ePGK3, but no significant increase over the background could be observed at any of the concentrations tested (Figure 5.11). Thus, the co-chaperone Hsp40 *in vitro* does not rescue ATP-independent sticking of PGK and Hsp70 to become ATP-controlled productive binding.

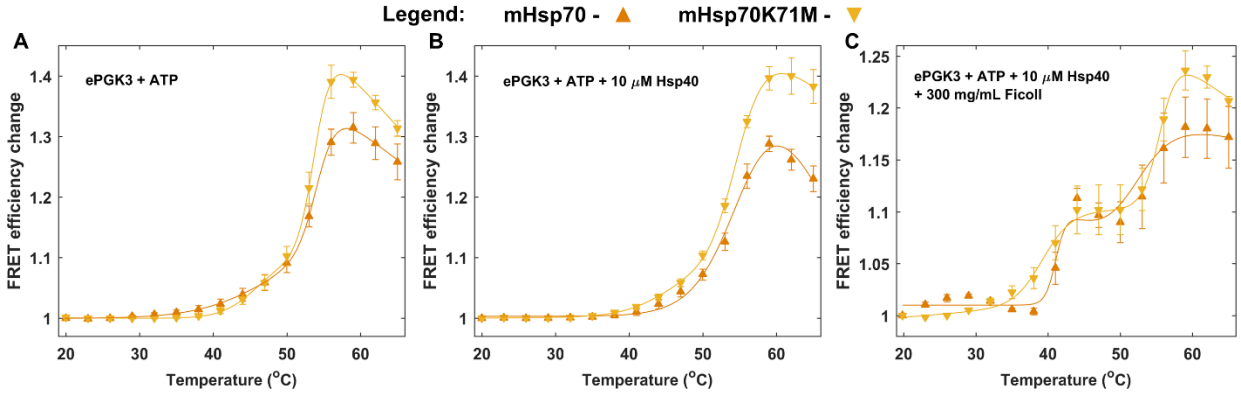


Figure 5.10: *In vitro* binding of 5:1 mHsp70 and ePGK3 with (A) ATP, (B) ATP and 10 µM Hsp40 and (C) ATP, 10 µM Hsp40 and 300 mg/mL Ficoll70. Filled markers show experimental data and solid lines are fits to experimental data. Error bars are standard error from mean for two measurement repeats. Both mHsp70 and mHsp70K71M show similar binding curves.

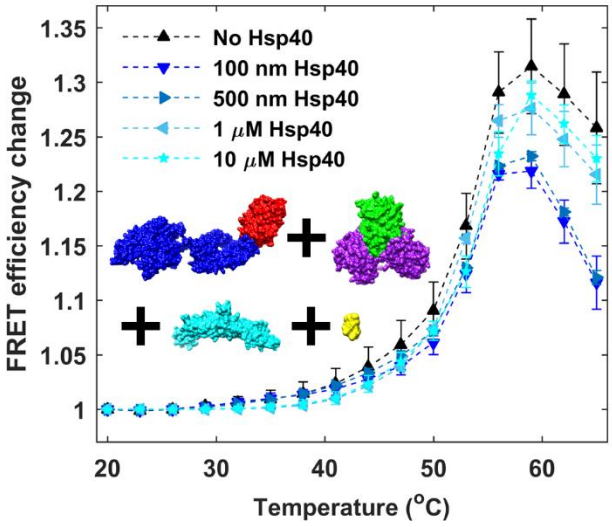


Figure 5.11: *In vitro* binding of 5:1 mHsp70 and ePGK3 with ATP with Hsp40. Filled markers and dashed lines show experimental data. Error bars are standard error from mean for two measurement repeats.

Figure 5.11 (cont.)

mHsp70-ePGK3 binding was conducted at multiple Hsp40 concentrations. Binding curves at all concentrations show similar change in FRET efficiency *vs.* temperature and no significant improvement in binding could be observed at any of the concentrations tested.

Since the cellular environment is extremely crowded, we also conducted experiments *in vitro* with an artificial crowding agent, 300 mg/mL Ficoll70, at an excluded volume similar to the cytoplasm. We again tested 1:1, 1:5 and 5:1 mHsp70:ePGKs (Figure 5.12).

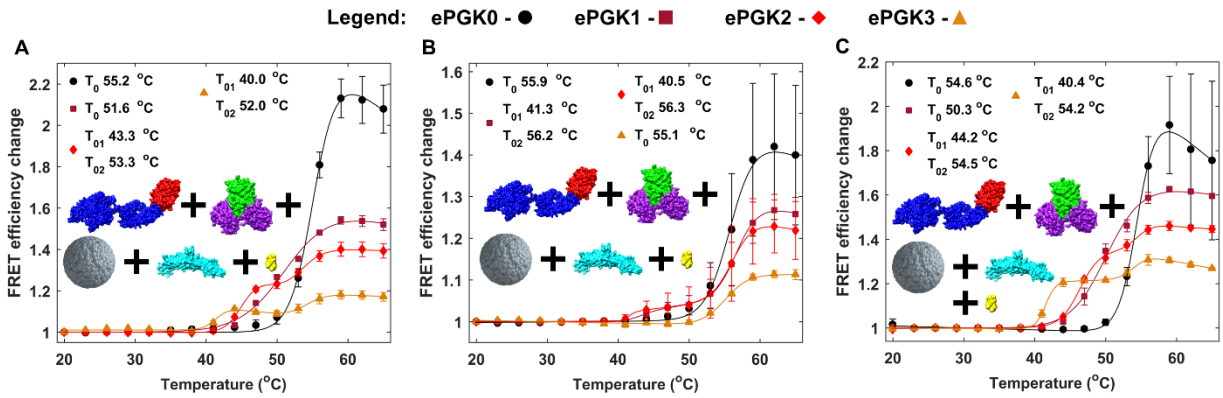


Figure 5.12: *In vitro* binding of mHsp70 and ePGKs with Ficoll70. Filled markers show experimental data and solid lines are fits to experimental data. Error bars are standard error from mean for two measurement repeats. Fit errors reflect the 1 standard deviation precision of the fit. Binding of (A) 1:1, (B) 1:5 and (C) 5:1 mHsp70:ePGKs (ePGK0 (black), ePGK1 (dark red), ePGK2 (red) and ePGK3 (orange)) in buffer with Ficoll70, ATP and Hsp40 (inset).

Ficoll70 resulted in two well-resolved cooperative transitions for ePGK2 and 3. We discuss 1:1 binding in more detail here because this condition is the closest to our in-cell experiments. The binding curves for ePGK2 and 3 were fit using a three-state model and both yielded a T_{02} of ~54 °C (aggregation of unfolded mHsp70 and unfolded ePGK). T_{01} fit to 43.2 °C for ePGK2 and 40.0 °C for ePGK3, both in good agreement with the T_m for ePGK2 and ePGK3. Therefore, this early transition is due to the association of mHsp70 to ePGK as it unfolds. ATPase-deficient mHsp70K71M:ePGK controls in Ficoll70 (Figure 5.10C) again showed the same change in E_{FRET}, with similar T_{01}. Thus, binding in a crowded *in vitro* environment is still dominated by ATP-independent sticking of unfolded substrate to the chaperone.

To mimic the cell more closely, we also performed experiments in 0.6 mg/mL cell lysate plus buffer containing 300 mg/ml of the crowding agent Ficoll70 (Figure 5.13) with 1:1 mHsp70:ePGK. This significantly reduced sticking of ePGK3 compared to just Ficoll70, from ~10% to ~5%. Sticking was abolished completely for ePGK2. We hypothesize that the combination of crowding and weak interactions

of lysate osmolytes and macromolecules with Hsp70 and PGK reduces sticking and could promote proper heat shock function of Hsp70 in cells.

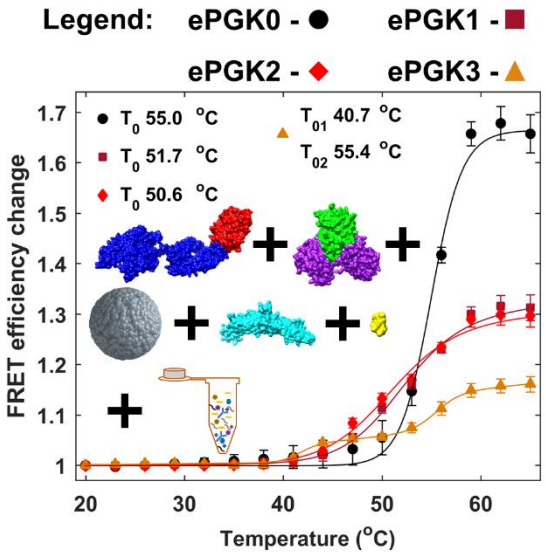


Figure 5.13: *In vitro* binding of 1:1 mHsp70 and ePGKs with cell lysate. Filled markers show experimental data and solid lines are fits to experimental data. Error bars are standard error from mean for two measurement repeats. Fit errors reflect the 1 standard deviation precision of the fit. Binding of ePGK0 (black), ePGK1 (dark red), ePGK2 (red) and ePGK3 (orange) in buffer with Ficoll70, cell lysate, ATP and Hsp40 with mHsp70.

5.3.4 MHSP70-EPGK BINDING UNDER IN-CELL HEAT SHOCK IS SPECIFIC AND ATP-DEPENDENT

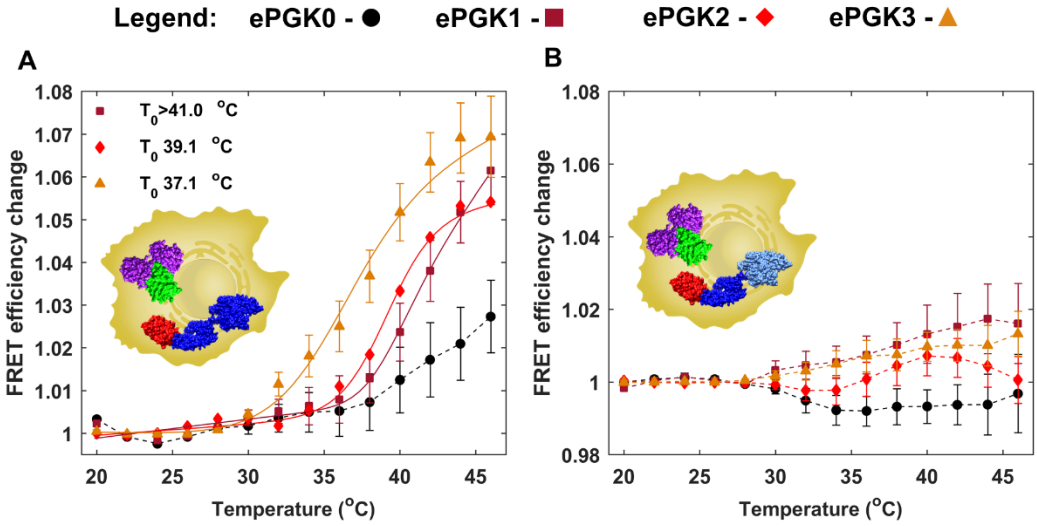


Figure 5.14: In-cell binding of mHsp70 and ePGKs. Filled markers and dashed lines show experimental data and solid lines are fits to experimental data. Error bars are standard error from mean for 15-30 cells. Fit

Figure 5.14 (cont.)

errors reflect the 1 standard deviation precision of the fit. FRET efficiency average for (A) mHsp70- and (B) mHsp70K71M-PGK binding *vs.* temperature for ePGK0 (black), ePGK1 (dark red), ePGK2 (red) and ePGK3 (orange).

For in-cell experiments, ePGK and either mHsp70 or mHsp70K71M are co-expressed in U-2 OS cells (see Methods). Only cells with 1:1±0.1 substrate:chaperone (based on ratiometric fluorescence intensity calibration) were chosen for imaging. ePGK was thermally denatured by a programmed infrared laser-induced temperature ramp (Figure 5.1A. inset). The average of the E_{FRET} curves for mHsp70-ePGK binding for 15-30 cells was calculated for ePGK0-3 (Figure 5.14A). As expected for ePGK0 with T_m = 48 °C, only the onset of binding was observed. In contrast, ePGK2 and ePGK3 showed fully resolved cooperative binding curves with an overall E_{FRET} change of ~8% on average. The T_0 for ePGK2-mHsp70 and ePGK3-mHsp70 binding fit to 39.1 °C and 37.1 °C, respectively. The fitted T_0 values are about 3 °C lower than the corresponding ePGK *in vitro* thermal denaturation T_m values, as we further discuss below.

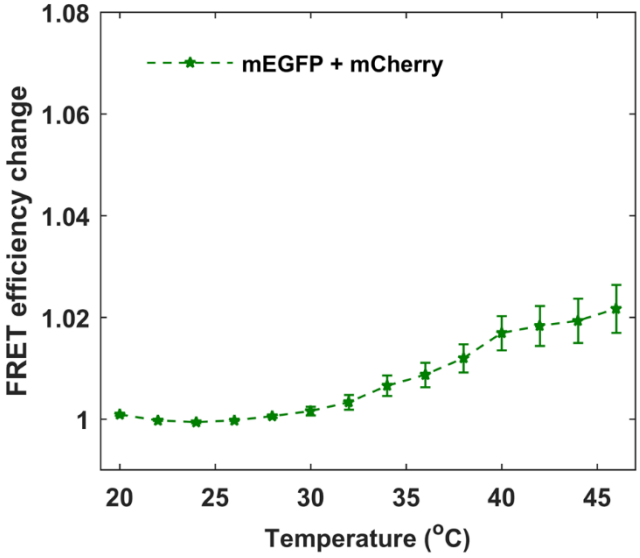


Figure 5.15: In-cell binding of mEGFP-mCherry. The average in-cell binding FRET efficiency change for mEGFP and mCherry is only 2%, similar to the change seen for the mHsp70K71M mutant in Figure 5.7B, and 4x smaller than the change seen for mHsp70K71 with ePGK in Figure 5.14A.

In contrast to our *in vitro* binding results, average E_{FRET} binding curves for the ATPase-inactive mHsp70K71M mutant showed only a small signal change, ≤ 2% (Figure 5.14B). A control in cells transfected with mEGFP and mCherry (Figure 5.15) shows a similar 2% change, so the mHsp70K71M result is likely due to a weak interaction of the fluorescent labels. The signal observed with mHsp70 in cell is at least 4x times higher than with mHsp70K71M. Therefore, the E_{FRET} change observed with mHsp70 is due to productive substrate binding regulated by ATPase activity. In-cell, mHsp70 is indeed a heat shock

chaperone for ePGK. All raw cell traces for the average traces above as well as additional controls are shown in Appendix D, D.9.

5.4 DISCUSSION

Our results show that in-cell, mHsp70 binds ePGK1-3 in a cooperative, ATP-dependent manner. *In vitro*, no ATP-dependent binding is observed upon heat shock. For in-cell binding, the observed T_0 is ~3 °C lower than the PGK T_m *in vitro*. We previously showed that the stability of FRET-labeled PGK1 (fPGK1) is higher in U-2 OS cells by ~2 °C than *in vitro* [118,122]. Thus, the Hsp70-PGK binding temperature T_0 is actually likely to be ~5 °C lower than the in-cell T_m of PGK.

There are two plausible explanations for why mHsp70 already binds ePGK below its in-cell unfolding temperature: mHsp70 either acts as an unfoldase, or as a preemptive holdase for PGK. Unfoldases lower the stability of a protein they bind by shifting the unfolding equilibrium [384], while we coin the term 'preemptive holdase' for a chaperone that recognizes and binds its substrate below T_m without promoting unfolding. Because wt-Hsp70 is overexpressed during the over half hour duration of our previous in-cell fPGK1 thermal denaturation experiments, and a higher fPGK1 T_m is observed in-cell than *in vitro*, these chaperones do not appear to act as unfoldases for PGK in the cell [385].

Instead, Hsp70 may act as a preemptive holdase in the cell. Our proposed mechanism is as follows. It is known that Hsp70 recognizes short (~7 residues) hydrophobic motifs [386]. Increased thermal fluctuations of the native state preceding the thermal unfolding transition have been demonstrated for PGK and other proteins: for example, tryptophan near the protein surface exists in a highly fluctuating environment prior to unfolding [320,387,388]. We therefore hypothesize that prior to unfolding, native state fluctuations of substrate proteins such as PGK expose on their surface short hydrophobic motifs that Hsp70 recognizes and binds preemptively, before full unfolding occurs. Binding of a human Hsp70 isoform, Hsc70 and bacterial Hsp70, DnaK, to dynamically exposed states was recently demonstrated for two substrate proteins using NMR [389] to support this notion.

PGK is known to have seven such hydrophobic motifs that could potentially bind Hsp70, as shown in figure 5.16. These binding sites were predicted using the limbo website, http://limbo.switchlab.org/limbo-analysis [390]. It has been observed previously that DnaK can bind near-native states of maltose binding protein and prevent unfolding with force [180]. Preemptive holdase activity in the cell similarly would enable Hsp70 to bind and hold substrate proteins prior to unfolding, stabilizing substrates against unfolding. We speculate that preemptive holdase activity may have evolved to preempt formation of misfolded or unfolded states of important substrates before they can have a chance to aggregate in the cell.

Figure 5.16: Ribbon structure of PGK. Hydrophobic patches are highlighted in green. These patches are potential binding sites for Hsp70 even when PGK is in its native state.

In contrast, *in vitro* binding experiments in buffer showed equal binding signals of ePGK with both mHsp70 and mHsp70K71M, pointing to an ATP-independent sticking between mHsp70 and unfolded ePGK. We also corroborated that the addition of Hsp40 did not increase binding efficiency *in vitro*. With crowding agent Ficoll70, equal sticking of ePGK2 and ePGK3 to both mHsp70 and mHsp70K71M was observed. Thus, *in vitro* binding, crowded or not, is not under ATP control and not pre-emptive. However, the addition of cell lysate reduced the propensity for ATP-independent sticking in Ficoll70. We therefore conclude that weak interactions with cytoplasmic components, and crowding combine to promote the productive chaperoning interaction of mHsp70 with ePGK under thermal stress.

Chaperoning is likely not the only quinary structure formation process that is significantly modulated by the cytosol. Other weak but functional protein-protein interactions are also likely to elude *in vitro* assays such as pull-downs or titrations, and may require investigation in the native environment of the cell to obtain quantitative K_d values and other binding parameters [6,7,85,98].

CHAPTER 6

CHARACTERIZING IN-CELL AND *IN VITRO* BINDING FOR THE CONSTITUTIVELY EXPRESSED CYTOPLASMIC HSP70 – HSC70

6.1 INTRODUCTION

In the last chapter, we characterized the binding of cytoplasmic heat inducible Hsp70 to a model substrate protein PGK using FReI. The human cell expresses many different homologs of the 70 kDa heat shock protein including the heat inducible Hsp70.[391,392] The constitutively expressed cytoplasmic Hsp70, also known as heat shock cognate (Hsc) proteins or Hsc70 or Hsp73 or HspA8, shares 85% sequence identity with Hsp70 (Emboss Pairwise Sequence Alignment Tool, Needleman-Wunsch algorithm).[393] We refer to the protein as Hsc70 throughout. Unlike Hsp70, Hsc70 performs protein quality control functions in the cell under normal conditions.[394,395] Hsc70 also contains an N-terminal ATPase domain and a C-terminal substrate binding domain with a similar mode of action while binding substrate proteins and performing cellular quality control.[396]

In the healthy, unstressed cell, differences appear between Hsp70 and Hsc70 function.[397] However, some evidence has shown that Hsc70 and Hsp70 co-localize during stress in different cellular compartments and could be performing similar functions by binding [398,399]. In order to compare binding of Hsp70 and Hsc70 we engineered fluorescently labeled fusion Hsc70 constructs similar to that described in chapter 5 and examined substrate binding in-cell. We used the PGK mutants as substrates labeled with mEGFP as described in chapter 5.

6.2 METHODS

6.2.1 PROTEIN EXPRESSION AND PURIFICATION

Protein expression was carried out according to previously established protocol [95]. Fusion protein sequences were cloned into pDream 2.1/MCS vector by Genscript Corp. and used for dual expression in E. coli and mammalian cells unless stated otherwise. Sequence of Hsc70 (HSPA8) is listed in Appendix E, E.1. Hsc70 was cloned with a C-terminal mCherry (mHsc70) and mEGFP (mEGFP-Hsc70) fusion protein with a 6×His-tag for purification. All PGK mutants were cloned as described in chapter 5. All proteins were purified according to the protocol listen in chapter 5.

6.2.2 *IN VITRO* CIRCULAR DICHROISM AND FLUORIMETER MELTS

CD and fluorometer melts were carried out following protocols listed in chapter 5. For all FRET measurements, mEGFP was excited at 485 nm and emission spectra were collected from 480 to 700 nm in 300 µL cuvettes. Unfolding and aggregation of Hsc70 was monitored by the change in FRET efficiency *vs.* temperature by melting an equimolar mixture of mHsc70 (1 µM) and mEGFP-Hsc70 (1 µM) with and without 2 mM ATP and 10 mM DTT in K1 buffer.

For *in vitro* FRET binding experiments 2.5 µM ePGKn (n=0-3) was heated with 0.5 µM mHsc70 with 10 mM DTT and 2 mM ATP in K1. ATP was maintained in all *in vitro* experiments by addition of an ATP regenerating mix. The mix consists of 50 mM creatine phosphate, 0.2 mg/mL of creatine phosphokinase in appropriate buffer. Regeneration of ATP has been shown to increase refolding of denatured luciferase by Hsp70 [380].

6.2.3 FRET BINDING EXPERIMENTS IN CELLS

For detailed methods refer to chapter 5. For mHsc70 and ePGK transfection, roughly 2 µg mHsc70 and 1 µg ePGK was used for co-transfecting 3 35 mm falcon dishes. For mCherry/ePGK and mHsc70/mEGFP controls roughly 2 µg mHsc70/ePGK and 1 µg mCherry/mEGFP was used for co-transfecting 3 35 mm falcon dishes.

6.2.4 DATA ANALYSIS

All data was analyzed using MATLAB (MathWorks). All two- and three- state fits in this chapter are fit using equations listed in chapter 5.

6.3 RESULTS

6.3.1 MCHERRY LABELING DOES NOT AFFECT HSC70 STABILITY

I first wanted to confirm that fluorescently labeled Hsc70 exhibits the same thermal stability as wild type Hsc70. Like Hsp70, Hsc70 was labeled at the C-terminal with mCherry. We call this fusion construct mHsc70. The thermal stability of mHsc70 was compared to that of wildtype Hsc70 (Figure 6.1).

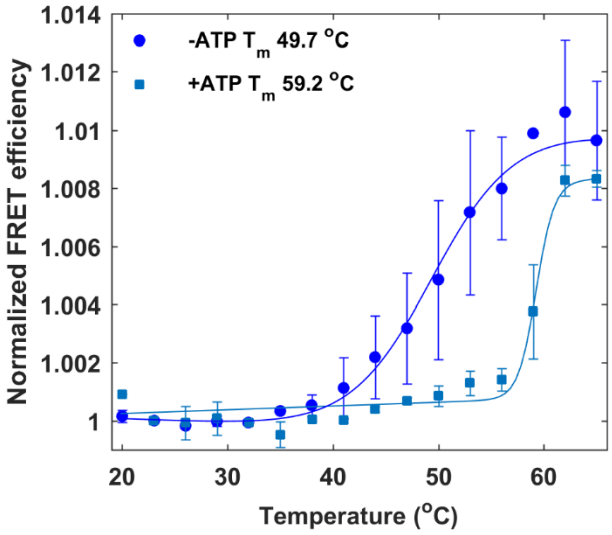


Figure 6.1: mHsc70 thermal unfolding monitored by change in FRET efficiency as a function of temperature. With ATP mHsc70 unfolds with a T_m of 59.2 °C and 49.7 °C without.

mHsc70 thermal stability was characterized by thermally unfolding mHsc70 in a temperature-controlled fluorometer and monitored by FRET efficiency normalized to room temperature (20 °C). Like mHsp70, mHsc70 also shows ~10 °C thermal stabilization in the presence of ATP. The T_m of mHsc70 was compared to wildtype Hsc70 and showed similar stabilization on the addition of ATP (Table 6.1). The thermal unfolding of Hsc70 followed by tryptophan peak wavelength shift and thermal unfolding probed by CD is shown in Appendix E, E.2.

Table 6.1: Transition temperature mid-points for Hsc70 and mHsc70 calculated from two-step sigmoidal fits to the experimental data. Errors shown reflect the 1 standard deviation precision of the fit. Systematic errors are not accounted for.

Protein	Nucleotide	Nucleotide concentration (mM)	Transition temperature (°C)		
			CD	Trp SVD	FRET
Hsc70	NA	0	44.2 ± 0.4	41.4 ± 0.9	NA
Hsc70	ATP	20	NA	53.2 ± 2.0	NA
Hsc70-mCherry	NA	0	44.0 ± 0.5	41.3 ± 0.9	49.7 ± 0.9
Hsc70-mCherry	ATP	20	NA	NA	59.2 ± 0.9

6.3.2 *IN VITRO* MHSC70-PGK BINDING IS EITHER NON-SPECIFIC OR ABOLISHED

Using ePGKs0-3 characterized in chapter 5, preliminary experiments to probe mHsc70-ePGK binding showed no temperature dependence with 1:5 mHsc70:PGK (Figure 6.2). Like in chapter 5, in order to differentiate between melting and association, we refer to the protein thermal denaturation midpoint as T_m, and the association midpoint as T_0.

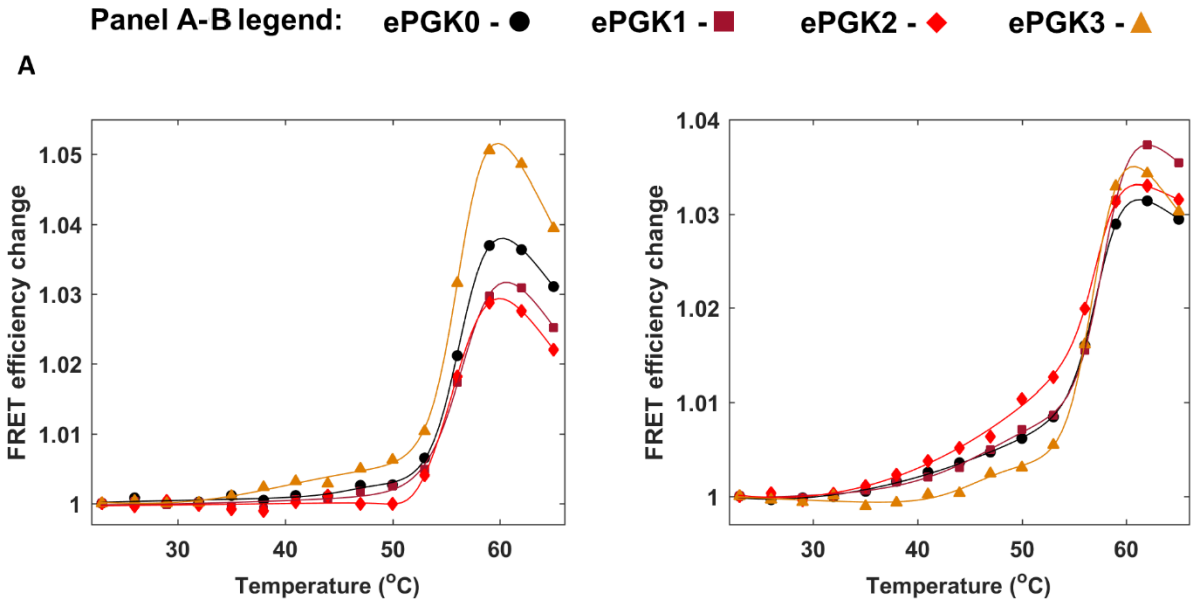


Figure 6.2: mHsc70-ePGK binding in vitro (A) without Hdj1 and (B) with Hdj1. Filled markers show experimental data and solid lines show 2-state fits to experimental data. All curves fit to a T_0 of ~54 ± 1 °C.

The large change in FRET efficiency starting at 50 °C is due to the aggregation of unfolded mHsc70 and ePGKs as described in chapter 5. Other than this aggregation transition, no binding is observed for ePGKs 0-2 between 25-50 °C in the absence of Hdj1 (Figure 6.2A). A small transition is observed for ePGK3 in the absence of Hdj1. To further inspect the nature of this transition, Hdj1 was added to the fluorometer melts to study Hdj1-assisted mHsc70-ePGK binding. On the addition of Hdj1, a small amount of binding is visible for all ePGK mutants before the 50 °C aggregation transition (Figure 6.2B). However, no temperature dependence was visible even for mHsc70-ePGK binding with Hdj1.

These findings point to a small amount of binding *in vitro* for mHsc70. However, more experiments are needed with the ATPase deficient mHsc70K71M mutant to further characterize if this binding is non-specific in nature like that observed with mHsp70.

6.3.3 IN-CELL MHSC70, LIKE MHSP70, SHOWS COOPERATIVE BINDS PGK COOPERATIVELY AS A FUNCTION OF TEMPERATURE

For in-cell experiments, ePGK and either mHsc70 or mHsc70K71M are co-expressed in U-2 OS cells (see Methods). The average of the E_{FRET} curves for mHsc70-ePGK binding for 15-30 cells was calculated for ePGK0-3 (Figure 6.3A).

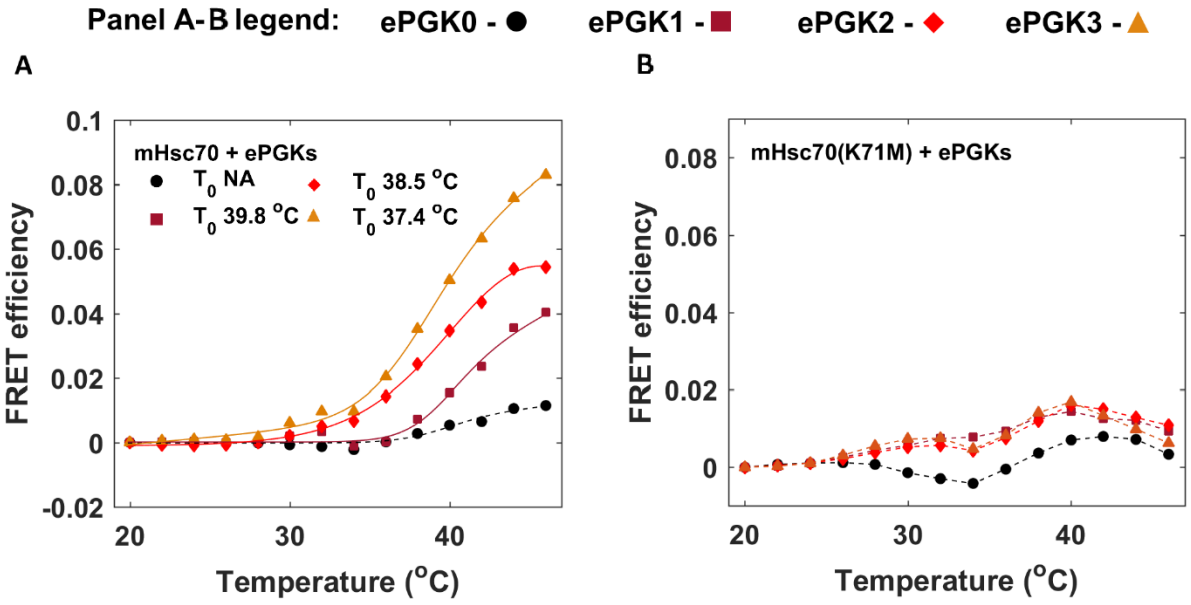


Figure 6.3: Binding of (A) mHsc70 and (B) mHsc70K71M to ePGK0 (black), ePGK1 (dark red), ePGK2 (red) and ePGK3 (orange). Filled markers and dashed lines show experimental data and solid lines show fits to experimental data. Data is averages from 15 – 30 cells.

As expected for ePGK0 with T_m = 48 °C, only the onset of binding was observed. For ePGK1, a partially complete binding curve was observed that fit to a T_0 of 39.8 °C. Both ePGK2 and 3 showed fully resolved binding curves with binding T_0s 38.5 and 37.4 °C. Like Hsp70, Hsc70 also binds PGK prior to unfolding with $T_0 < T_m$ for each individual PGK mutant. However, binding with the ATPase activity deficient mutant mHsc70K71M showed no cooperative binding with any of the PGK mutants in-cell (Figure 6.3B). Thus, the observed binding in figure 6.3A between Hsc70 and PGK is the result of cooperative chaperoning activity of mHsc70.

6.4 DISCUSSION

Our results discussed above combined with our observations from chapter 5 show that both Hsp70 and Hsc70 bind substrates during heat stress prior to unfolding. In the last chapter, we referred to this mode of substrate binding as preemptive holdase activity. Both Hsc70 and Hsp70 therefore, preemptively bind substrates in the cell during heat stress. This indicates that both Hsp70 and Hsc70 occur in the cell during

heat stress and bind unfolding substrates. Some studies have already shown that differences between Hsp70 and Hsc70 function is minimal during heat stress.[398,399] Our results agree with the findings that Hsc70 also performs stress-response during heat stress.

An interesting future direction might be to conduct three-color FRET studies by labeling both Hsc70 and Hsp70 with acceptors and PGK with a donor protein that is capable of exciting both acceptor proteins. This would allow us to probe if both Hsc70 and Hsp70 bind at the same time or if Hsc70 binding precedes Hsp70 binding. Such a mechanism where Hsc70 binding occurs first could happen since Hsp70 upregulation occurs after heat stress and can take up to 10 minutes. Hsc70 could then bind first and transfer substrate to Hsp70 when it becomes available in the cell.

Such a parallel three-color FRET scheme is possible by using mTAGBFP as the donor and mEGFP and LSS-mKate as an acceptor (Figure 6.4).

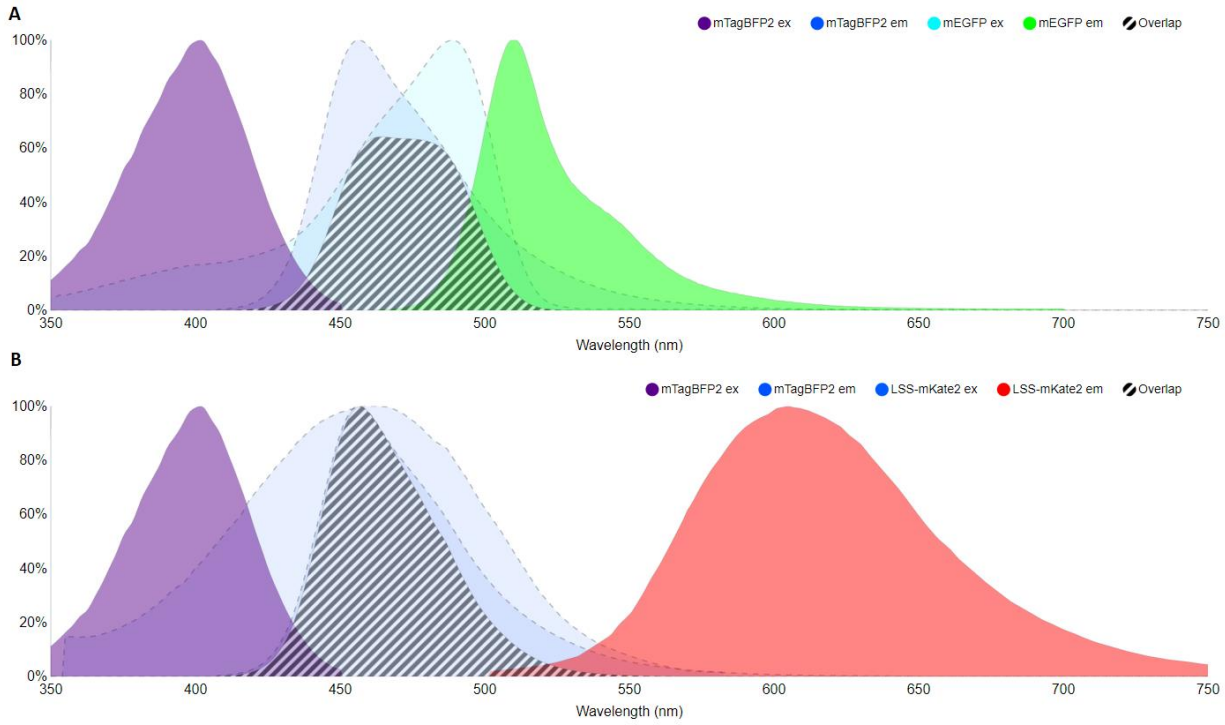


Figure 6.4: Overlap of mTAGBFP2 as donor and (A) mEGFP as acceptor and (B) LSS-mKate as acceptor. Plots were taken from fpbase.org/fret.

LSS-mKate is a new generation of fluorescence proteins that have a long-stokes shift emission allowing for the excitation of two acceptors with similar excitation profiles but well separated emission that can be separated into channels.

APPENDIX A

SUPPLEMENTARY INFORMATION FOR CHAPTER 2: DODINE AS A TRANSPARENT PROTEIN DENATURANT FOR CIRCULAR DICHROISM AND INFRARED STUDIES

A.1 ADDITIONAL DESCRIPTION OF PROTEIN PHOSPHOGLYCERATE KINASE (PGK)

Phosphoglycerate kinase (PGK) is an enzyme protein that assists in the conversion of ADP to ATP during the glycolytic cycle. PGK is highly conserved in all living organisms highlighting its importance in the glycolytic cycle. It's a 415 residue protein with two nearly equal sized domains.[18] The substrate of PGK 3-phosphoglycerate binds to the N-terminal while nucleotides, MgADP or MgATP, bind to the C-terminal. The two domains are connected by a cleft consisting of two α-helices and are made of 6-stranded parallel β-sheets surrounded by α-helices (Figure A.1).[400]

Figure A.1: Ribbon structure of PGK with N-terminal domain in red and C-terminal in black (PDB ID – 3PGK).

We used the yeast PGK sequence for our studies. Wild type yeast PGK melts at a T_m of 52 °C. The folding rate of wild type PGK follows a slow single exponential with t1≈ 8 μs, t2 ≈ 2 ms.[401] The nucleotide sequence for yeast PGK is given below.

NUCLEOTIDE SEQUENCE:

ATGTCTTTATCTTCAAAGTTGTCTGTCCAAGATTTGGACTTGAAGGACAAGCGTGTCTTCATC
AGAGTTGACTTCAACGTCCCATTGGACGGTAAGAAGATCACTTCTAACCAAAGAATTGTTGC
TGCTTTGCCAACCATCAAGTACGTTTTGGAACACCACCCAAGATACGTTGTCTTGGCTTCTCA
CTTGGGTAGACCAAACGGTGAAAGAAACGAAAAATACTCTTTGGCTCCAGTTGCTAAGGAA
TTGCAATCATTGTTGGGTAAGGATGTCACCTTCTTGAACGACTGTGTCGGTCCAGAAGTTGA
AGCCGCTGTCAAGGCTTCTGCCCCAGGTTCCGTTATTTTGTTGGAAAACTTGCGTTATCACAT
CGAAGAAGAAGGTTCCAGAAAGGTCGATGGTCAAAAGGTCAAGGCTTCCAAGGAAGATGTT
CAAAAGTTCAGACACGAATTGAGCTCTTTGGCTGATGTTTACATCAACGATGCCTTCGGTAC
CGCTCACAGAGCTCACTCTTCTATGGTCGGTTTCGACTTGCCACAACGTGCTGCCGGTTTCTT
GTTGGAAAAGGAATTGAAGTACTTCGGTAAGGCTTTGGAGAACCCAACCAGACCATTCTTGG
CCATCTTAGGTGGTGCCAAGGTTGCTGACAAGATTCAATTGATTGACAACTTGTTGGACAAG
GTCGACTCTATCATCATTGGTGGTGGTATGGCTTTCACCTTCAAGAAGGTTTTGGAAAACACT
GAAATCGGTGACTCCATCTTCGACAAGGCTGGTGCTGAAATCGTTCCAAAGTTGATGGAAAA
GGCCAAGGCCAAGGGTGTCGAAGTCGTCTTGCCAGTCGACTTCATCATTGCTGATGCTTTCT
CTGCTGATGCCAACACCAAGACTGTCACTGACAAGGAAGGTATTCCAGCCGGCTGGCAAGG
GTTGGACAATGGTCCAGAATCTAGAAAGTTGTTTGCTGCTACTGTTGCAAAGGCTAAGACAA
TTGTCTGGAACGGTCCACCAGGTGTTTTCGAATTCGAAAAGTTCGCTGCTGGTACTAAGGCT
TTGTTAGACGAAGTTGTCAAGAGCTCTGCTGCTGGTAACACCGTCATCATTGGTGGTGGTGA
CACTGCCACTGTCGCTAAGAAGTACGGTGTCACTGACAAGATCTCCCATGTCTCTACTGGTG
GTGGTGCTTCTTTGGAATTATTGGAAGGTAAGGAATTGCCAGGTGTTGCTTTCTTATCCGAAA
AGAAA

A.2 DETAILED METHOD TO PERFORM ISOTHERMAL DENATURANT TITRATION

In order to conduct an isothermal denaturant titration, the following procedure is followed:

1. Prepare three mixtures, a zero denaturant protein mixture at a concentration appropriate for fluorescence measurements (300 μL, 5 μM), a concentrated protein stock for recompensating protein removed and a concentrated denaturant solution.
2. The total volumes needed of the concentrated protein and denaturant mixtures can be calculated using the MATLAB code in section A.3 as well the volume of the solution in the cuvette to be removed and the amounts of stock protein and stock denaturant to be added.
3. Pipette the first 300 μL mixture into a cuvette and measure the spectrum with the required settings.
4. Remove the volume of protein as indicated by the MATLAB code output column 2.
5. The removed protein is compensated by pipetting in a small amount of concentrated protein stock, MATLAB code output column 3.

6. The amount of denaturant in the cuvette is increased by pipetting in a small volume of denaturant stock, MATLAB code output column 4.
7. The solution is mixed by pipetted up and down vigorously 20-30 times and allowed to equilibrate for 1 minute
8. A spectrum is then collected at the new denaturant concentration and steps 4-8 are repeated until the full titration is completed.

A.3 MATLAB CODE FOR CALCULATING TITRATION VOLUMES TO BE ADDED/REMOVED

```
function [ output ] = titration( Pin, Pstock, Dstock, Dstart, Dend, Ddiff, Vol )

%titration takes in Pin - protein solution concentration, Pstock - a
%concentrated protein stock for recompensating protein removed,Dstock -
%stock denaturant concentration
%   Pin, Pstock, Dstock are scalar numbers, default 2 uM 20 uM and 10mM.
%   Dstart is the starting denaturant conc Dend the ending and Ddiff is
%   step size. Vol is total cuvette volume, assume 700 uL for bigger
%   cuvettes.

    var = [];
    var(:,1) = (Dstart:Ddiff:Dend)';
    num_conc = size(var,1);
    var(:,2) = 0;
    var(:,3) = 0;
    var(:,4) = 0;

    for i = 1:num_conc-1
        %Moles of denaturant to add = Vol*(titration(i+1,1)-titration(i,1)) +
        %moles of denaturant lost
        %If x is the volume removed then moles of protein removed = Pin.x and
        %moles of denaturant removed = Din.x
        %Volume of protein required to compensate for lost moles = Pin.x/Pstock
        %Dstock.(x - Pin.x/Pstock) = Vol*(titration(i+1,1)-titration(i,1)) +
        %Din.x
        var(i+1,2) = (Vol*(var(i+1,1)-var(i,1)))/(Dstock - (Pin*Dstock/Pstock) - var(i,1));%volume removed
at each titration step
        var(i+1,3) = var(i+1,2)*Pin/Pstock; %volume of Pstock to add
        var(i+1,4) = var(i+1,2) - var(i+1,3);%volume of Dstock to add
    end

    output = var;
end
```

A.4 MATLAB CODE FOR ANAYZING FLUOROMETER SPECTRA

```
function FluorAnalyzeTRP_NEWFLUOR(filename, numsamples, txtname, numfitpoints, offset, fitcutoff)

%Function imports melt fluorimeter data and saves variables with txtname base.

%cd = datapath;
%set up for analysis
wlexcutoff = 305;  %wl to start analysis at (to avoid scatter)
usegauss2 = 0;

all = csvread(strcat(filename, '.csv'),offset,0);
intensity(:,:) = all;

%extract data and temperatures
numtemps = (size(intensity, 2)-1)/numsamples;
temps=intensity(1,2:end);
wavelength=intensity(2:end,1);

%wlround = round(wavelength);
[~, wloffset] = min(abs(wavelength-wlexcutoff));

for j = 1:numsamples
   celldata(:,:,j) = intensity(2:end, j+1:numsamples:end);
end
clearvars a i j;
[~, numtemps, numsamples] = size(celldata);

%automatically calculate integrated intensity, spectral center of mass

%set wloffset to 305 nm
%wloffset = find(round(wavelength) == 305);
inti = trapz(celldata(wloffset:end,:,:),1);

%do SVD
[SVD.U,SVD.S,SVD.V] = svd(celldata(wloffset:end,:));
SVD.signal = SVD.V(:,2)./SVD.V(:,1);

for i = 1:numtemps
   for j = 1:numsamples
      cofm(i,j) = sum(wavelength(wloffset:end).*celldata(wloffset:end,i,j))/sum(celldata(wloffset:end,i,j));
      if size(fitcutoff,1) == 1
         cutoffi = fitcutoff;
      else
```

```
            cutoffi = fitcutoff(find(i<=fitcutoff(:,2),1),1);
        end
        [~, fitoffset] = min(abs(wavelength-cutoffi));
        [~, peakpos] = max(celldata(fitoffset:end,i,j));
        peak = wavelength(fitoffset+peakpos);
        if peak<325 && usegauss2==1
            gaussian(i,1) = 2;
        else
            gaussian(i,1) = 1;
        end
        if gaussian(i,1) == 1
            gaussianfit = fit(wavelength(fitoffset+peakpos-numfitpoints:fitoffset+peakpos+numfitpoints),
celldata(fitoffset+peakpos-numfitpoints:fitoffset+peakpos+numfitpoints,i,j), 'gauss1');
            gausspeaks(i,4,j) = gaussianfit.a1;
            gausspeaks(i,5,j) = gaussianfit.b1;
            gausspeaks(i,6,j) = gaussianfit.c1;
        else
            gaussianfit = fit(wavelength(fitoffset+peakpos-numfitpoints:fitoffset+peakpos+numfitpoints),
celldata(fitoffset+peakpos-numfitpoints:fitoffset+peakpos+numfitpoints,i,j), 'gauss2');
            gausspeaks(i,1,j) = gaussianfit.a1;
            gausspeaks(i,2,j) = gaussianfit.b1;
            gausspeaks(i,3,j) = gaussianfit.c1;
            gausspeaks(i,4,j) = gaussianfit.a2;
            gausspeaks(i,5,j) = gaussianfit.b2;
            gausspeaks(i,6,j) = gaussianfit.c2;
        end
    end
end

for j = 1:numsamples
        peaks(:,:,j) = polyfitdg(wavelength(wloffset:end), celldata(wloffset:end,:,j), 2);
end

%set up basic data structure
data.celldata = celldata;
data.wavelength = wavelength;
data.temps = temps;
data.inti = inti;
data.cofm = cofm;
data.offset = wloffset;
data.peaks = peaks;
data.gausspeaks = gausspeaks;
data.SVD = SVD;
%save data struct in variable txtname
```

```
eval(sprintf('%s = data', txtname))

figure
set(gca, 'colororder', jet(numtemps));
hold all
plot(wavelength, celldata);
set(gca, 'colororder', jet(numtemps));
for i = 1:numtemps
   plot(wavelength, plotGaussian(gausspeaks(i,:), wavelength, gaussian(i,1)),'--')
end

save(strcat(txtname, '_compile.mat'), txtname);

end
```

A.5 MATLAB CODE FOR ANALYZING CIRCULAR DICHROISM DATA

```
function [wavelength, cdMRE, ht] = CDAnalyze(datapath, datastring, analyze, numRes, pl, conc, T, dT)


%given a data file and filename root, detect the files, extract
%temperatures or concentrations, and load into matrix [wavelengths, traces]
%for both CD and HT

%find files and set up
cd(datapath);

allfiles = dir(strcat(datastring,'*.txt')); %lists all files in directory by wildcard datastring*.txt
numfiles = numel(allfiles); %numel - number of elements

if analyze == 1
   offset = 23;
else
   offset = 37;
end

params = {allfiles.name};
params = params';
params = regexprep(params, datastring, ''); %replace string 'datastring' in params using regular expression
params = regexprep(params, '.txt', '');
params = cell2mat(params);
params = str2num(params);
%current = importdata(allfiles(1).name, '\t', 23); %imports data with delimiter \t as column seperator,
starting from line 24.
```

```
current = importdata(allfiles(1).name, '\t', offset); %use for single text files, for example spectrum
measurement
wavelength = current.data(:,1);

cdt = zeros(length(wavelength), numfiles);
ht = zeros(length(wavelength), numfiles);

cdt(:,1) = current.data(:,2);
ht(:,1) = current.data(:,3);

for i = 2:numfiles
   %current = importdata(allfiles(i).name, '\t', 23);
   current = importdata(allfiles(i).name, '\t', offset); %use for single text files, for example spectrum
measurement
   cdt(:,i) = current.data(:,2);
   ht(:,i) = current.data(:,3);
end

%plot standard data
cb = jet(numfiles);

%do basic analysis
if analyze == 1
   %plot MRE
   cdMRE = zeros(size(cdt));
   % units for MRE is deg cm^2/dmol, http://ctrstbio.org.uic.edu/manuals/kellybba.pdf
   % pl is input in m and conc in M then MRE = degcm^2/dmol

   cdMRE = (1*cdt)/(1000*pl*conc*numRes);

   %plot and calculate 222 and 208
   peak222 = cdMRE(wavelength==222, :);
   peak208 = cdMRE(wavelength==208,:);
   temps = (T:dT:((params(end,1)-1)*dT)+T)';

   for i = 1:numfiles
      smoothed_MRE(:,i) = smooth(cdMRE(:,i),10);
   end
   peak222_smoothed = smoothed_MRE(wavelength==222, :);

%     calculate and plot basic svd
   [U, S, V] = svd(cdMRE);
   S(3:end,:) = 0;
   cdMRE_reconstructed = U*S*V';
```

```
    peak222_reconstructed = cdMRE_reconstructed(wavelength==222, :);
end
save(strcat(datastring, '.mat'), 'wavelength', 'cdt', 'ht', 'cdMRE', 'peak222', 'peak208', 'temps',
'peak222_reconstructed', 'peak222_smoothed','smoothed_MRE','cdMRE_reconstructed');

end
```

APPENDIX B

SUPPLEMENTARY INFORMATION FOR CHAPTER 3: DODINE AS A KOSMO-CHAOTROPIC AGENT

B.1 PRB21 HELIX CONFIGURATION

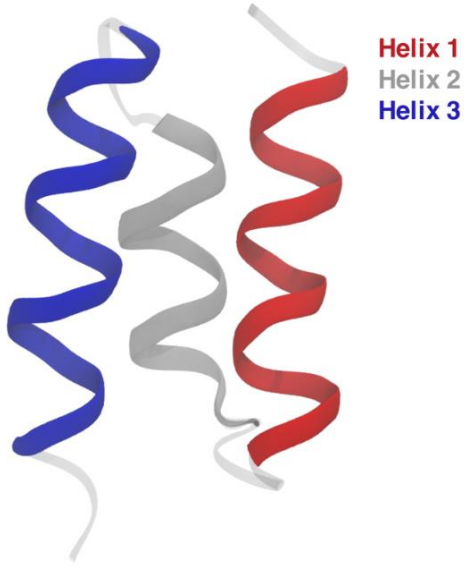


Figure B.1: The individual helices for the analysis in this work are defined as residues 4-15 (helix 1, red), residues 19-27 (helix 2, grey), and residues 31-45 (helix 3, blue).

B.2 MD SIMULATIONS WITH PRB21 WITHOUT THE V39K MUTATION

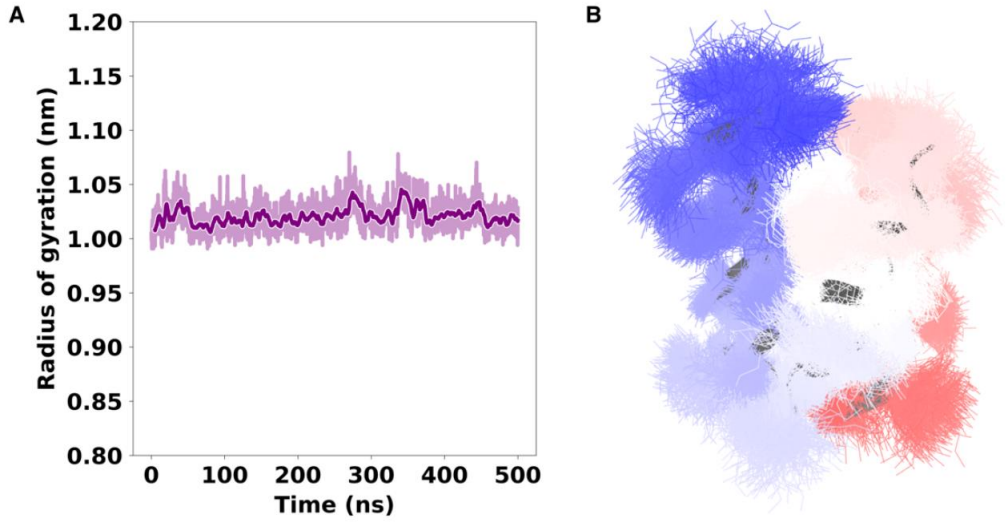


Figure B.2: Radius of gyration for PrB21. (A) Radius of gyration of the protein is plotted as a function of time for 500 ns simulation of apo protein at 300 K. This simulation is started from folded structure of the PrB protein with sequence as used in experiments. This is consistent with radius of gyration data reported for PrB21 V39K mutant. (B) An overlay of ~500 structures obtained from a 500 ns simulation of PrB21 without dodine at 300 K. The first frame is shown in gray as cartoon representation and frames sampled at every 1 ns are shown in line representation where residues 1 to 47 are colored red to blue.

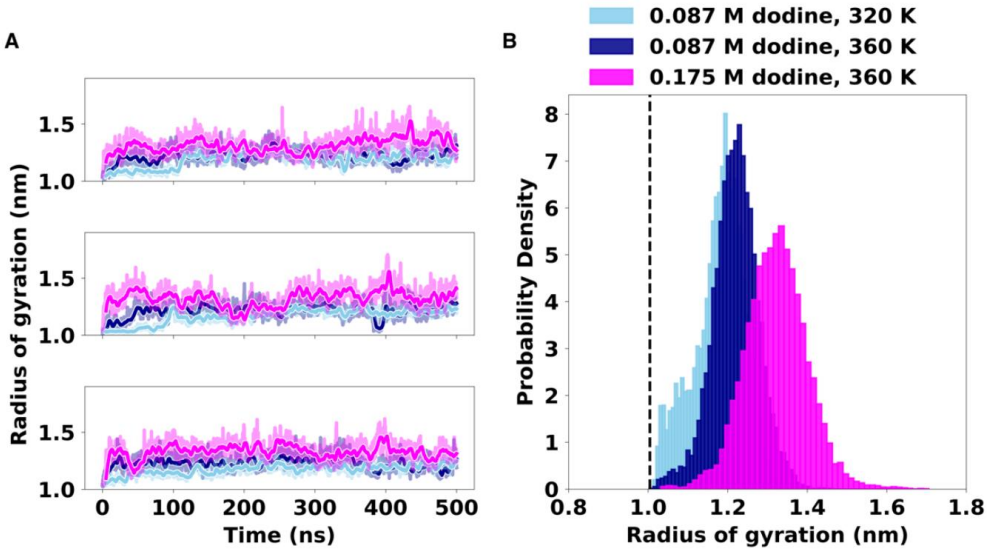


Figure B.3: (A) Radius of gyration of the protein as a function of time, and (B) Combined probability distribution (from three replicas) for radius of gyration of the protein with 0.087 M dodine at 320 K (sky blue), 360 K (dark blue), and 0.175 M dodine at 360 K (magenta). Time series and distributions shown here are from simulations of PrB21 sequence same as experiments. The black dashed line indicates the crystal structure (native state) values. We observe dynamics consistent with conclusions from our simulations on PrB21 V39K mutant.

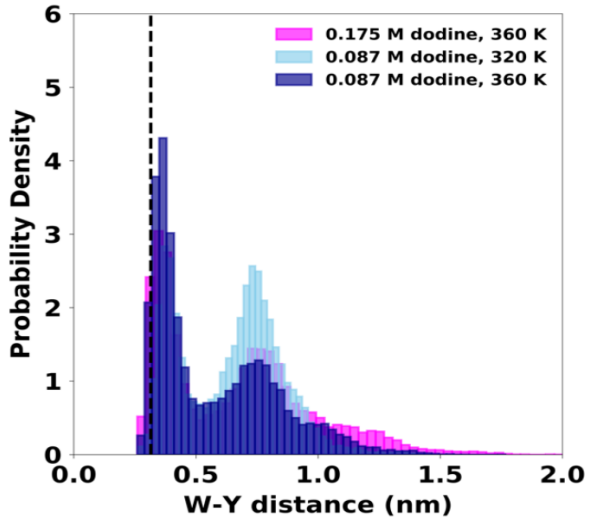


Figure B.4: Combined probability distribution from three replicas for Trp-Tyr (W-Y) distance with 0.087 M dodine at 320 K (sky blue), 360 K (dark blue), and 0.175 M dodine at 360 K (magenta). Distributions shown here are from simulations of PrB21 sequence same as experiments. The black dashed line indicates the crystal structure (native state) values. Similar to our results for PrB21 V39K mutant, we observe that the Trp-Tyr distance increases in dodine. This state when Tyr and Trp residues are farther apart (~0.75 nm) is less populated at the higher temperatures with 0.087 M dodine.

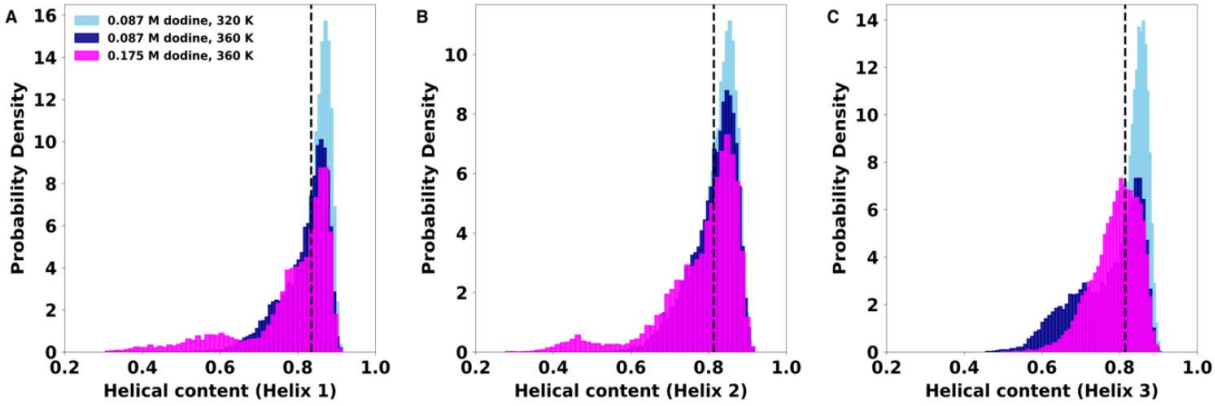


Figure B.5: Combined probability distributions for helical content of (A) helix 1, (B) helix 2 and (C) helix 3 in simulations with 0.087 M dodine at 320 K (sky blue), 360 K (dark blue), and 0.175 M dodine at 360 K (magenta) for three replicas. Helical content increased for helix 3 in 0.175 M dodine. We observed similar behavior for helices 1 and 2 for the PrB21 V39K mutant. We also observe this for helix 1 in one out of three simulated replicas. Distributions shown here are from simulations of PrB21 sequence same as experiments. The black dashed line indicates the crystal structure (native state) values.

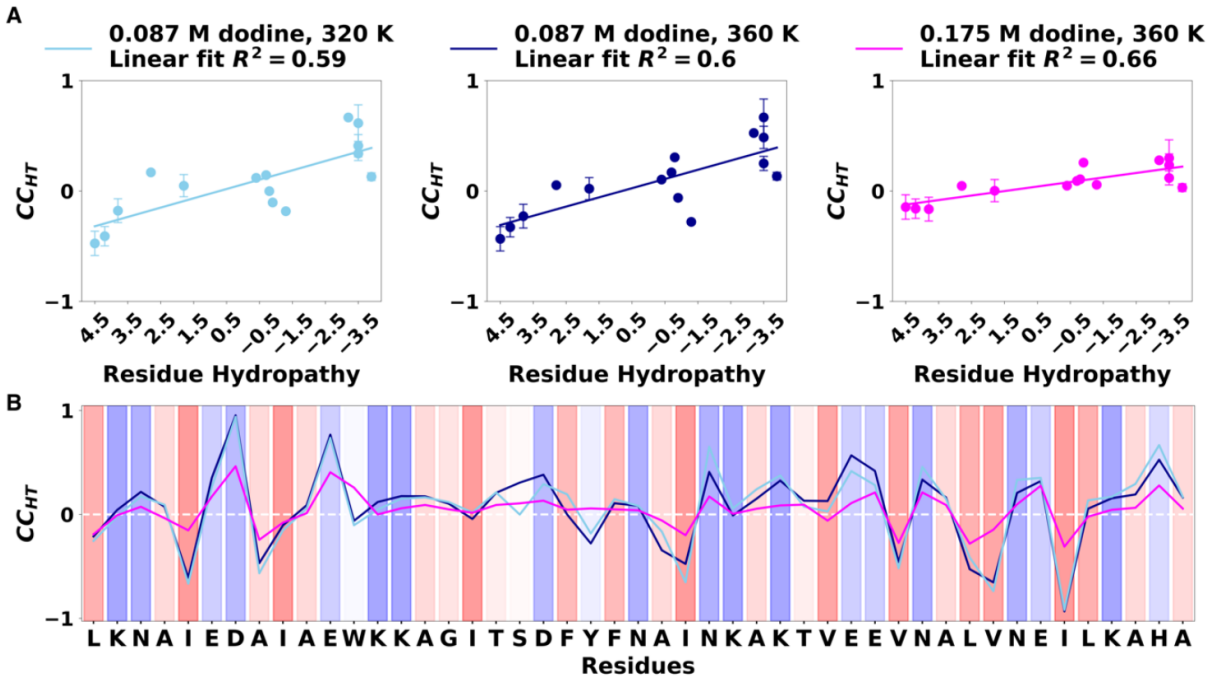


Figure B.6: Interaction of PrB21 with dodine. (A) Contact-coefficient$_{\text{head-tail}}$ (CC_{HT}) as a function of residue hydropathy for all residues (backbone+sidechain) of PrB21 in 0.087 M dodine at 320 K (cyan), 360 K (dark blue) and 0.175 M dodine at 360 K (magenta). Positive values indicate preference for guanidinium head while negative values indicate preference for the hydrophobic tail. (B) Contact-coefficient$_{\text{head-tail}}$ (CC_{HT}) as a function of sequence. The background colors represent residue hydrophobicity with blue colors being more polar and redder more hydrophobic. Values shown here are from simulations of PrB21 sequence same as experiments.

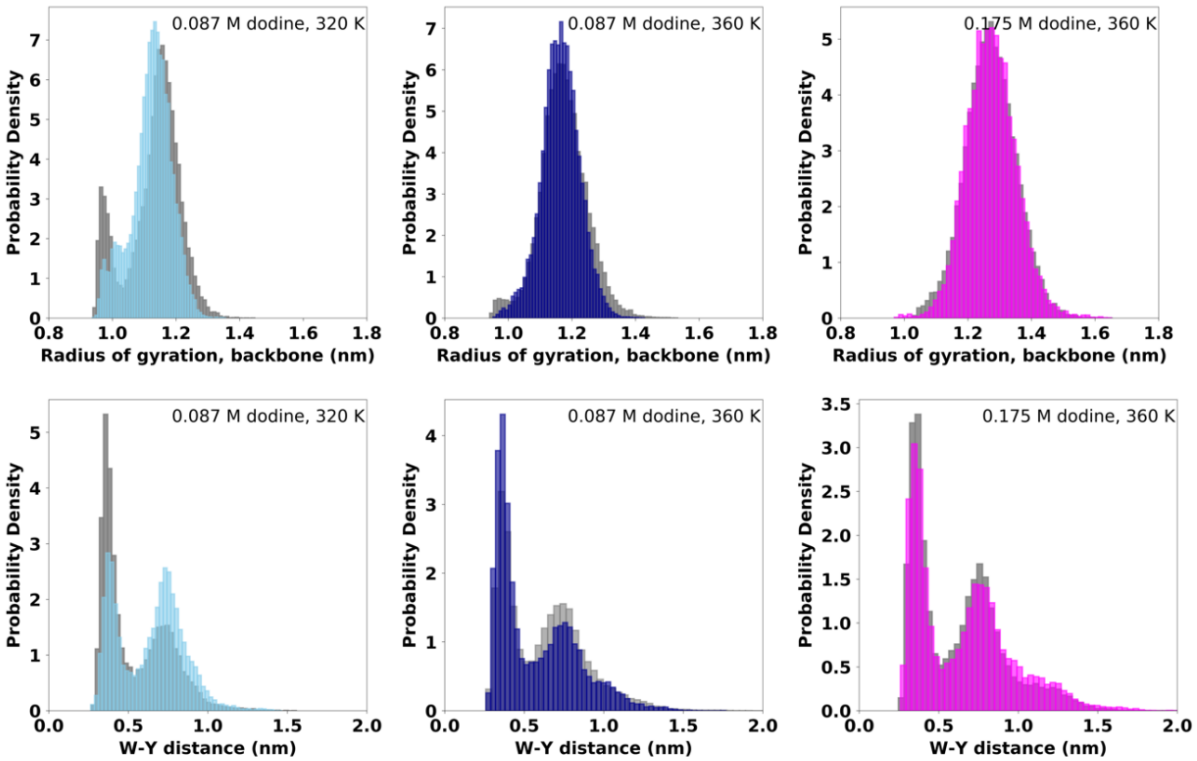


Figure B.7: Comparative combined probability distributions for radius of gyration and Trp-Tyr (W-Y) distance from MD simulations of the two sequences of PrB21. Distributions in gray are for PrB21 V39K mutant and the colored distributions are from simulations of PrB21 sequence same as experiments.

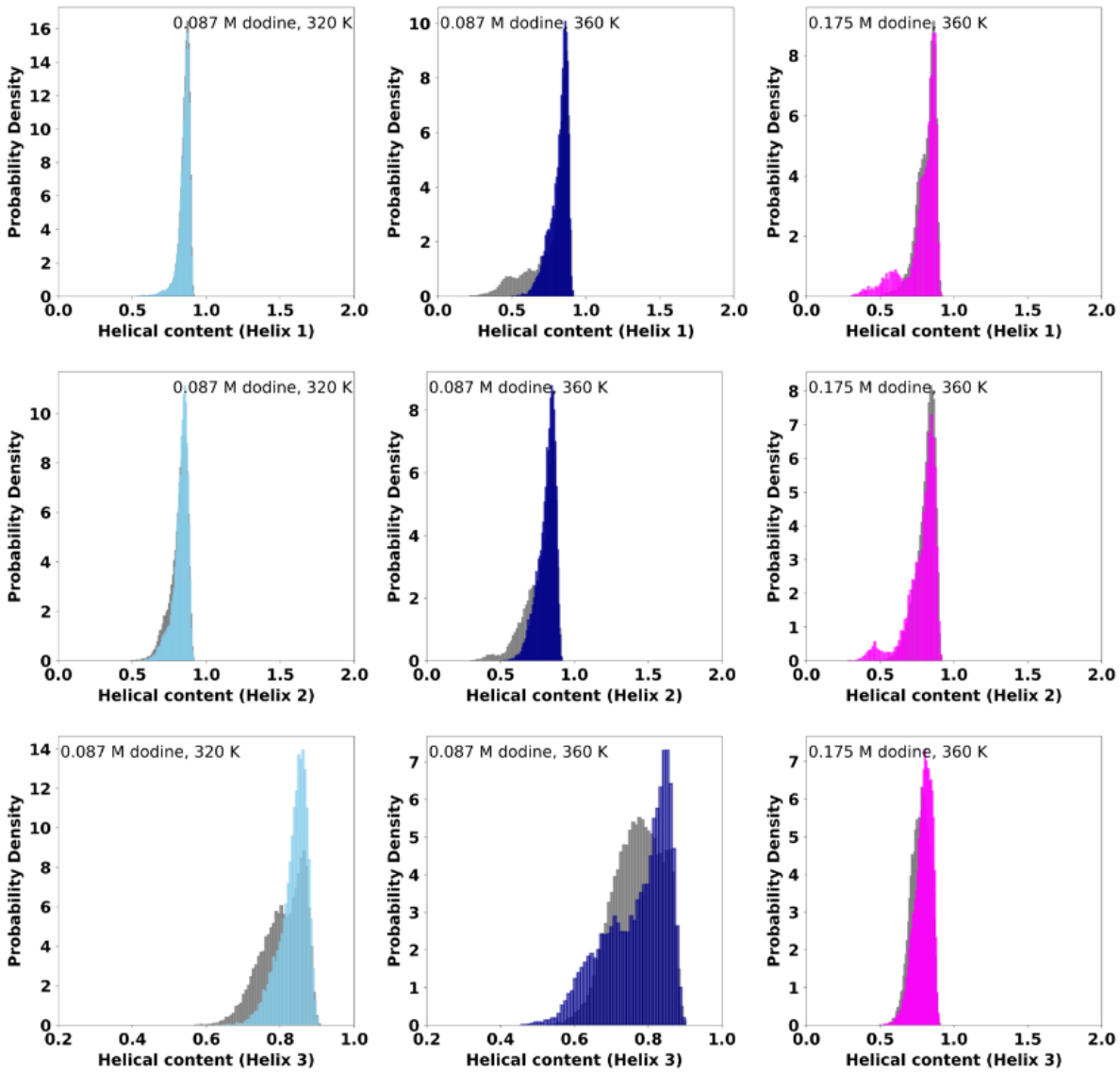


Figure B.8: Comparative combined probability distributions for helical content from MD simulations of the two sequences of PrB21. Distributions in gray are for PrB21 V39K mutant and the colored distributions are from simulations of PrB21 sequence same as experiments.

APPENDIX C

SUPPLEMENTARY INFORMATION FOR CHAPTER 4: THE DESIGN AND CONSTRUCTION OF A THREE-COLOR FAST RELAXATION IMAGING (FReI) SETUP

C.1 PARTS LIST FOR THE CUSTOM COMPUTER FOR INSTRUMENT CONTROL AND DATA ANALYSIS

Component	Product name
CPU	Intel Core i7-4770K Haswell 3.5GHz LGA 1150 84W Quad-Core Desktop Processor Intel HD Graphics BX80646I74770K
SSD	Corsair Force Series GS CSSD-F240GBGS-BK 2.5" 240GB SATA III Internal Solid State Drive (SSD)
Motherboard	ASUS SABERTOOTH Z87 LGA 1150 Intel Z87 HDMI SATA 6Gb/s USB 3.0 ATX Intel Motherboard
Power supply	Rosewill Stallion Series RD450-2-DB 450W ATX V2.2 Power Supply
Memory	CORSAIR Vengeance 8GB 240-Pin DDR3 SDRAM DDR3 1866 Desktop Memory Model CMZ8GX3M1A1866C10
GPU	ASUS GTX650TI-O-1GD5 GeForce GTX 650 Ti 1GB 128-Bit GDDR5 PCI Express 3.0 x16 HDCP Ready Video Card
CD drive	ASUS Black SATA DVD-ROM Drive Model DVD-E818AAT (DVD-E818AAT/BLK/B/GE) - OEM
Case	Rosewill BLACKHAWK Gaming ATX Mid Tower Computer Case, come with Five Fans, window side panel, top HDD dock

C.2 PARTS LIST FOR THE FREI INSTRUMENTATION

Component	Product name/part number	Company name
LED	UHP-T-LED white	Prizmatix Inc.
Laser	AP-CW-1	AdValue Photonics
Power Supply	BK9130	BK Precision Corp.
DAQ card	PCI-6221	National Instruments
BNC terminal blocks	BNC-2110	National Instruments
Camera	Lt225	Lumenera Corp.
Lens	Azure 1060ZM	Azure Photonics

C.3 SPECTRAL OVERLAP BETWEEN MEGFP AND MCHERRY

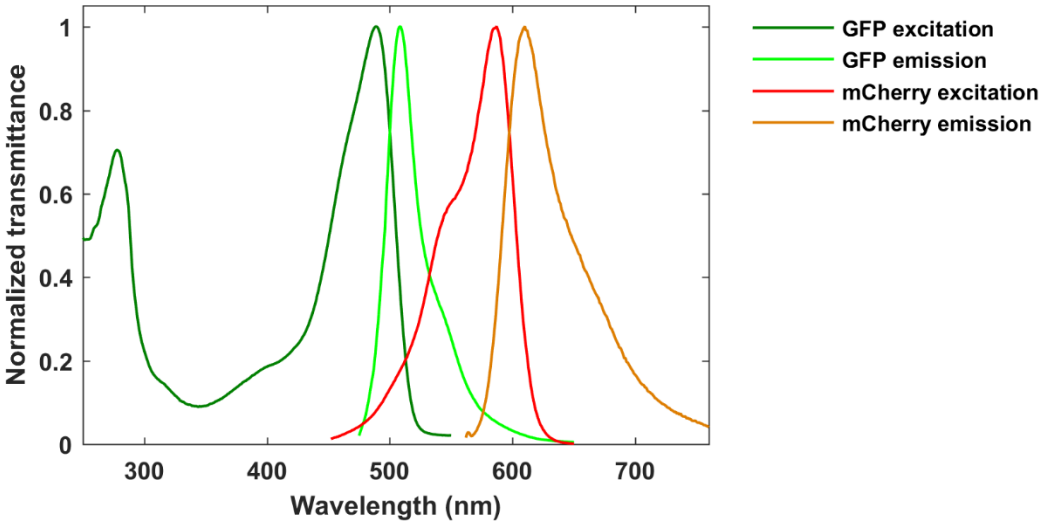


Figure C.1: GFP-mCherry spectral overlap leading for FRET.

C.4 FILTER SET CURVES FOR MEGFP AND MCHERRY

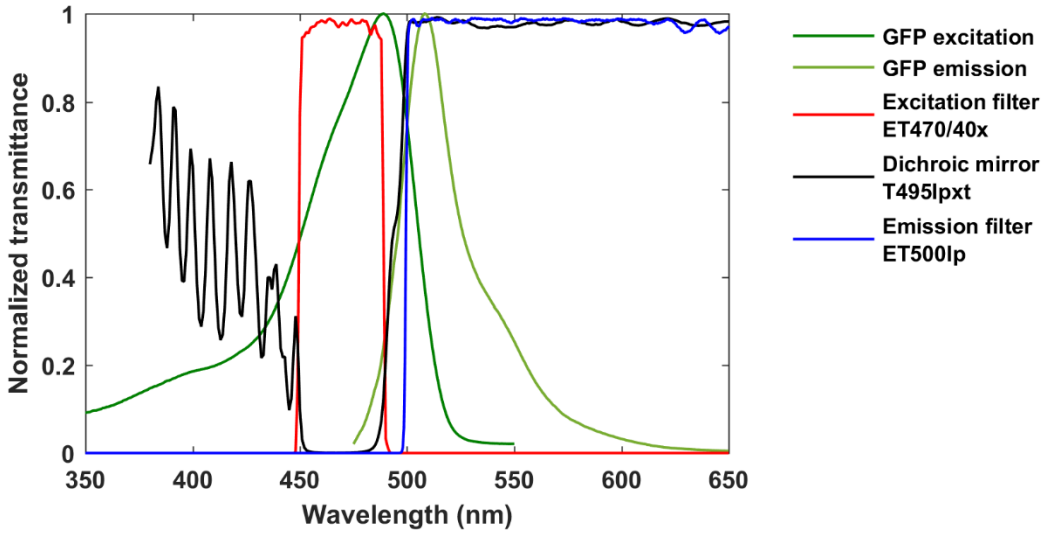


Figure C.2: Filter set for EGFP.

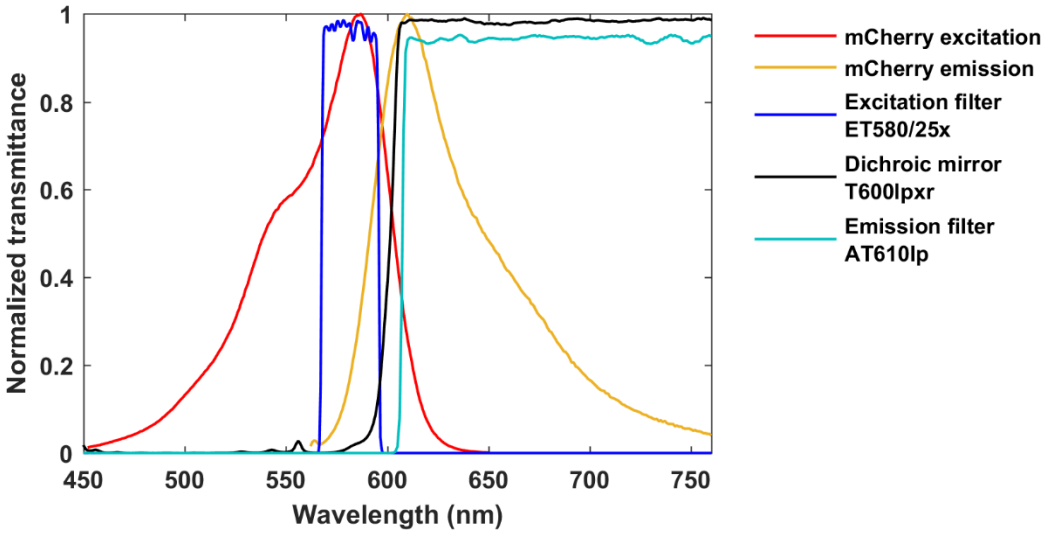


Figure C.3: Filter set for mCherry.

APPENDIX D

SUPPLEMENTARY INFORMATION FOR CHAPTER 5: HEAT SHOCK-INDUCED CHAPERONING BY HSP70 IS ENABLED IN-CELL

D.1 ADDITIONAL DESCRIPTION OF HEAT SHOCK PROTEIN 70 (HSP70), HEAT SHOCK PROTEIN 40 (HDJ1), PERCEVAL HR AND FRET-PGK1

D.1.1 HSP70

Newly synthesized proteins must fold in to their characteristic 3-D structure for the cell to be viable. Unfolded or misfolded proteins are prone to aggregation thereby exposing the cell to toxic fibrils or aggregates. The heat shock proteins (HSPs) form a family of the highly important molecular chaperones that occur naturally (2% of the proteome) and are upregulated (20-25% of the proteome) during stress.[402] These proteins protect small hydrophobic stretches in unfolded/misfolded proteins that are prone to accelerated aggregation.[403,404]

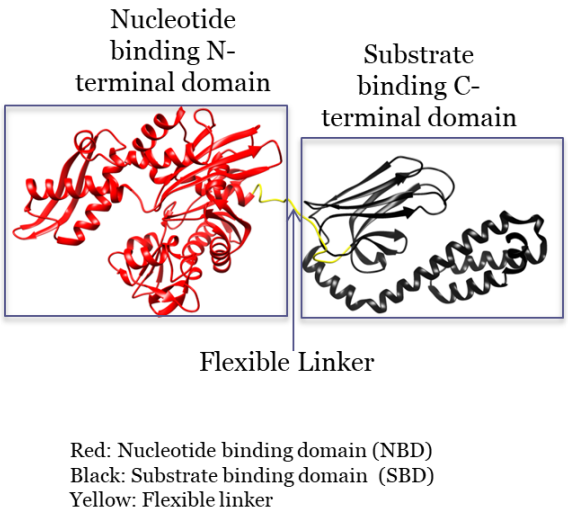


Figure D.1: Domain structure of human heat inducible Hsp70 (PDB ID – 2KHO).

HSP70s or the 70 kDa heat shock proteins participate in essential cellular function such as folding of nascent chain polypeptide, protein degradation and protein translocation across membranes.[405,406] We utilize the human cytoplasmic heat inducible Hsp70 also known as HSPA1 or Hsp72. HSPs function in-cell with

the help of co-chaperones and nucleotide exchange factors. The co-chaperone of the human HSP70 is HSP40. All proteins in the HSP70 family share a conserved architecture that is important to the HSP70 function. Shown below (Figure D.1) is the structure of the bacterial HSP70, DnaK (PDB ID: 2KHO). The structure consists of two domains connected by a flexible linker. Shown in red is the nucleotide binding ~44kDa N-terminal domain (NBD) and in black is the ~27kDa substrate binding Cterminal domain (SBD). The residues that connect and transduce communication between both domains form the flexible linker, shown in yellow.

The SBD further consists of a β-strand structure that binds short hydrophobic stretches of amino acids (~7 amino acids long) and an α-helical lid. Such short stretches are found universally in many proteins in the cell and this potentially points to the fact that Hsp70 can bind many cellular proteins. The NBD on the other hand binds either ATP or ADP forming the HSP allosteric cycle that consist of interconversion between the ATP and ADP bound states and is important for HSP function. In the normal cell HSP in found in the ATP bound state where the SBD α-helical lid is docked stably onto the NBD (Figure D.2). Substrate binds ATP-HSP with a high dissociation and association rate resulting in a low overall affinity, resulting in a weakly bound substrate. Substrate binding increases ATPase activity of the NBD and the ATP is hydrolyzed to ADP. The ATP hydrolysis rates are generally low with one molecule of ATP hydrolyzed per 20-30 minutes at 30 °C.[403]

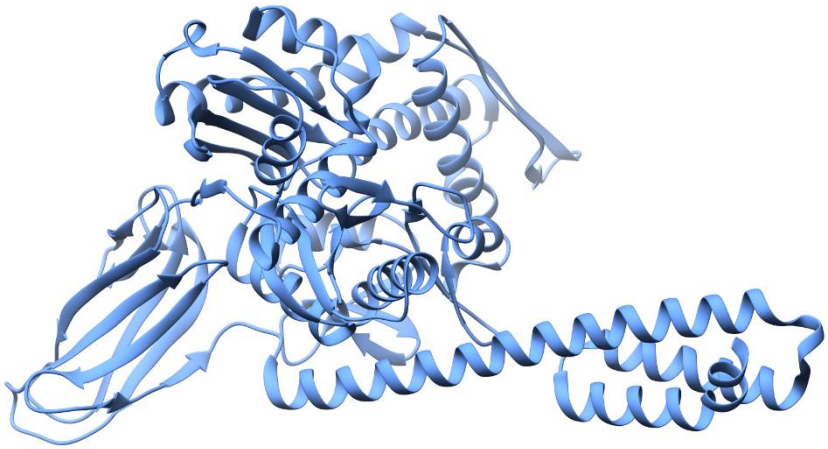


Figure D.2: Bacterial Hsp70 conformation in the ATP-bound state (PDB ID: 4B9Q).

In the ADP bound state, the α-helical lid forms a cap over the bound substrate reducing the dissociation rate of the substrate (Figure D.1). The ADP bound state has a low dissociation and association rate resulting in a high affinity. HSP70 hence binds substrate with a high association rate in the ATP bound state; the substrate then promotes hydrolysis of the ATP and gets trapped in the ADP bound state. The proteins are

allowed to fold or are translocated across membranes to the final destination. Nucleotide exchange factors then exchange ADP for ATP and the folded proteins are released. During heat stress the amount of ATP is depleted in the cell and is replaced by ADP. Sustained binding by the HSP70-ADP suppresses aggregation by stabilizing damaged proteins in partially folded states as exposed hydrophobic residues are temporarily trapped.

D.1.2 HDJ1

Hdj1 is the 40 kDa Hsp70 co-chaperone that assists in substrate recruitment as well chaperoning of bound substrates. Binding of Hsp40 to Hsp70 increases Hsp70 ATP hydrolysis and substrate binding. All proteins in the Hsp40 family contain a J-domain that can stimulate Hsp70 ATPase activity.

D.1.3 PERCEVAL-HR

Perceval-HR was created by Gary Yellen's lab as an ATP/ADP ratio sensor. A circularly permuted variant of GFP was fused to a bacterial regulatory protein GlnK1 from *Methanococcus jannaschii*. A circularly permutated variant of a fluorophore generally means that the fluorophore termini are connected to two termini of the protein of interest, in this case GlnK1. The fluorophore is cleaved at the active site. As the parent protein changes conformation, the chromophore active site amino acids come into proximity and this can be visualized as a change in the level of fluorescence emission. The protein GlnK1 is a trimeric intracellular protein that regulates ammonia transport associated with the synthesis of glutamine. GlnK1 binds Mg-ATP before it can act as an ammonia transporter, however GlnK1 undergoes a dramatic conformational change on binding ATP and hence, was chosen as an ATP sensor backbone.[407]

D.1.4 FRET-PGK

A PGK mutant was labeled with AcGFP1 on the N-terminal and mCherry at the C-terminal. The FRET pair AcGFP1-mCherry report on the folding state of PGK by monitoring donor/acceptor fluorescence. The R_0 for mEGFP-mCherry is 5.4 nm.[356]

D.2 NUCLEOTIDE SEQUENCES OF HSP70, HSP40, PERCEVAL-HR, MEGFP AND MCHERRY

The full DNA sequences of all the proteins used in the chapter are listed below.

HSP70 NUCLEOTIDE SEQUENCE:

ATGGCCAAAGCCGCGGCGATCGGCATCGACCTGGGCACCACCTACTCCTGCGTGGGGGTGTT
CCAACACGGCAAGGTGGAGATCATCGCCAACGACCAGGGCAACCGCACCACCCCCAGCTAC
GTGGCCTTCACGGACACCGAGCGGCTCATCGGGGATGCGGCCAAGAACCAGGTGGCGCTGA

ACCCGCAGAACACCGTGTTTGACGCGAAGCGGCTGATTGGCCGCAAGTTCGGCGACCCGGT
GGTGCAGTCGGACATGAAGCACTGGCCTTTCCAGGTGATCAACGACGGAGACAAGCCCAAG
GTGCAGGTGAGCTACAAGGGGGAGACCAAGGCATTCTACCCCGAGGAGATCTCGTCCATGG
TGCTGACCAAGATGAAGGAGATCGCCGAGGCGTACCTGGGCTACCCGGTGACCAACGCGGT
GATCACCGTGCCGGCCTACTTCAACGACTCGCAGCGCCAGGCCACCAAGGATGCGGGTGTG
ATCGCGGGGCTCAACGTGCTGCGGATCATCAACGAGCCCACGGCCGCCGCCATCGCCTACG
GCCTGGACAGAACGGGCAAGGGGGAGCGCAACGTGCTCATCTTTGACCTGGGCGGGGGCAC
CTTCGACGTGTCCATCCTGACGATCGACGACGGCATCTTCGAGGTGAAGGCCACGGCCGGGG
ACACCCACCTGGGTGGGGAGGACTTTGACAACAGGCTGGTGAACCACTTCGTGGAGGAGTT
CAAGAGAAAACACAAGAAGGACATCAGCCAGAACAAGCGAGCCGTGAGGCGGCTGCGCAC
CGCCTGCGAGAGGGCCAAGAGGACCCTGTCGTCCAGCACCCAGGCCAGCCTGGAGATCGAC
TCCCTGTTTGAGGGCATCGACTTCTACACGTCCATCACCAGGGCGAGGTTCGAGGAGCTGTG
CTCCGACCTGTTCCGAAGCACCCTGGAGCCCGTGGAGAAGGCTCTGCGCGACGCCAAGCTG
GACAAGGCCCAGATTCACGACCTGGTCCTGGTCGGGGGCTCCACCCGCATCCCCAAGGTGCA
GAAGCTGCTGCAGGACTTCTTCAACGGGCGCGACCTGAACAAGAGCATCAACCCCGACGAG
GCTGTGGCCTACGGGGCGGCGGTGCAGGCGGCCATCCTGATGGGGGACAAGTCCGAGAACG
TGCAGGACCTGCTGCTGCTGGACGTGGCTCCCCTGTCGCTGGGGCTGGAGACGGCCGGAGGC
GTGATGACTGCCCTGATCAAGCGCAACTCCACCATCCCCACCAAGCAGACGCAGATCTTCAC
CACCTACTCCGACAACCAACCCGGGGTGCTGATCCAGGTGTACGAGGGCGAGAGGGCCATG
ACGAAAGACAACAATCTGTTGGGGCGCTTCGAGCTGAGCGGCATCCCTCCGGCCCCCAGGG
GCGTGCCCCAGATCGAGGTGACCTTCGACATCGATGCCAACGGCATCCTGAACGTCACGGCC
ACGGACAAGAGCACCGGCAAGGCCAACAAGATCACCATCACCAACGACAAGGGCCGCCTG
AGCAAGGAGGAGATCGAGCGCATGGTGCAGGAGGCGGAGAAGTACAAAGCGGAGGACGAG
GTGCAGCGCGAGAGGGTGTCAGCCAAGAACGCCCTGGAGTCCTACGCCTTCAACATGAAGA
GCGCCGTGGAGGATGAGGGGCTCAAGGGCAAGATCAGCGAGGCGGACAAGAAGAAGGTGC
TGGACAAGTGTCAAGAGGTCATCTCGTGGCTGGACGCCAACACCTTGGCCGAGAAGGACGA
GTTTGAGCACAAGAGGAAGGAGCTGGAGCAGGTGTGTAACCCCATCATCAGCGGACTGTAC
CAGGGTGCCGGTGGTCCCGGGCCTGGGGGCTTCGGGGCTCAGGGTCCCAAGGGAGGGTCTG
GGTCAGGCCCCACCATTGAGGAGGTAGAT

MEGFP NUCLEOTIDE SEQUENCE:

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACG
GCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG
CAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCG
TGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAC
GACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA
CGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
ATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGT
ACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGT
GAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAG
CAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCA
GTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTG
ACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG

ACGFP1 NUCLEOTIDE SEQUENCE:

ATGGTGAGCAAGGGCGCCGAGCTGTTCACCGGCATCGTGCCCATCCTGATCGAGCTGAATGG
CGATGTGAATGGCCACAAGTTCAGCGTGAGCGGCGAGGGCGAGGGCGATGCCACCTACGGC
AAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCTGTGCCCTGGCCCACCCTGGT
GACCACCCTGAGCTACGGCGTGCAGTGCTTCTCACGCTACCCCGATCACATGAAGCAGCACG
ACTTCTTCAAGAGCGCCATGCCTGAGGGCTACATCCAGGAGCGCACCATCTTCTTCGAGGAT
GACGGCAACTACAAGTCGCGCGCCGAGGTGAAGTTCGAGGGCGATACCCTGGTGAATCGCA
TCGAGCTGACCGGCACCGATTTCAAGGAGGATGGCAACATCCTGGGCAATAAGATGGAGTA
CAACTACAACGCCCACAATGTGTACATCATGACCGACAAGGCCAAGAATGGCATCAAGGTG
AACTTCAAGATCCGCCACAACATCGAGGATGGCAGCGTGCAGCTGGCCGACCACTACCAGC
AGAATACCCCCATCGGCGATGGCCCTGTGCTGCTGCCCGATAACCACTACCTGTCCACCCAG
AGCGCCCTGTCCAAGGACCCCAACGAGAAGCGCGATCACATGATCTACTTCGGCTTCGTGAC
CGCCGCCGCCATCACCCACGGCATGGATGAGCTGTACAAG

MCHERRY NUCLEOTIDE SEQUENCE:

ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGG
TGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCG
CCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTC
GCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGC
CGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGA
ACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTT
CATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAG
AAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGG
GCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGA
CCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTT
GGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGC
CGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAG

D.3 AMINO ACID SEQENCES FOR HSP70, HSP40, PGK, MEGFP, ACGFP1 AND MCHERRY

Table D.1: Full protein sequences of proteins used. All sequences listed are wild type proteins. For Hsp70 red residues show N-terminal nucleotide binding domain and blue residues show C-terminal substrate binding domain.

Protein (Organism)	**Sequence**
Phosphoglycerate kinase *(Saccharomyces cerevisiae)*	MSLSSKLSVQDLDLKDKRVFIRVDFNVPLDGKKITSNQRIVAALPTIKYVL EHHPRYVVLASHLGRPNGERNEKYSLAPVAKELQSLLGKDVTFLNDCVGP EVEAAVKASAPGSVILLENLRYHIEEEGSRKVDGQKVKASKEDVQKFRHE LSSLADVYINDAFGTAHRAHSSMVGFDLPQRAAGFLLEKELKYFGKALEN PTRPFLAILGGAKVADKIQLIDNLLDKVDSIIIGGGMAFTFKKVLENTEIGDS IFDKAGAEIVPKLMEKAKAKGVEVVLPVDFIIADAFSADANTKTVTDKEGI PAGWQGLDNGPESRKLFAATVAKAKTIVWNGPPGVFEFEKFAAGTKALL DEVVKSSAAGNTVIIGGGDTATVAKKYGVTDKISHVSTGGGASLELLEGK ELPGVAFLSEKK
Hsp70 *(Homo sapiens)*	MAKAAAIGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERLI GDAAKNQVALNPQNTVFDAKRLIGRKFGDPVVQSDMKHWPFQVINDGD KPKVQVSYKGETKAFYPEEISSMVLTKMKEIAEAYLGYPVTNAVITVPAYF NDSQRQATKDAGVIAGLNVLRIINEPTAAAIAYGLDRTGKGERNVLIFDLG GGTFDVSILTIDDGIFEVKATAGDTHLGGEDFDNRLVNHFVEEFKRKHKKD ISQNKRAVRRLRTACERAKRTLSSSTQASLEIDSLFEGIDFYTSITRARFEEL CSDLFRSTLEPVEKALRDAKLDKAQIHDLVLVGGSTRIPKVQKLLQDFFNG RDLNKSINPDEAVAYGAAVQAAILMGDKSENVQDLLLLDVAPLSLGLETA GGVMTALIKRNSTIPTKQTQIFTTYSDNQPGVLIQVYEGERAMTKDNNLLG RFELSGIPPAPRGVPQIEVTFDIDANGILNVTATDKSTGKANKITITNDKGRL SKEEIERMVQEAEKYKAEDEVQRERVSAKNALESYAFNMKSAVEDEGLK GKISEADKKKVLDKCQEVISWLDANTLAEKDEFEHKRKELEQVCNPIISGL YQGAGGPGPGGFGAQGPKGGSGSGPTIEEVD
Hdj1 (Hsp40) *(Homo sapiens)*	MGKDYYQTLGLARGASDEEIKRAYRRQALRYHPDKNKEPGAEEKFKEIAE AYDVLSDPRKREIFDRYGEEGLKGSGPSGGSGGGANGTSFSYTFHGDPHA MFAEFFGGRNPFDTFFGQRNGEEGMDIDDPFSGFPMGMGGFTNVNFGRSR SAQEPARKKQDPPVTHDLRVSLEEIYSGCTKKMKISHKRLNPDGKSIRNED KILTIEVKKGWKEGTKITFPKEGDQTSNNIPADIVFVLKDKPHNIFKRDGSD VIYPARISLREALCGCTVNVPTLDGRTIPVVFKDVIRPGMRRKVPGEGLPLP KTPEKRGDLIIEFEVIFPERIPQTSRTVLEQVLPI
mEGFP	MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTT GKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFF KDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNV YIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHY LSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYK
AcGFP1	MVSKGAELFTGIVPILIELNGDVNGHKFSVSGEGEGDATYGKLTLKFICTT GKLPVPWPTLVTTLSYGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFE DDGNYKSRAEVKFEGDTLVNRIELTGTDFKEDGNILGNKMEYNYNAHNV YIMTDKAKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHY LSTQSALSKDPNEKRDHMIYFGFVTAAAITHGMDELYK

Table D.1 (cont.)

mCherry	MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAK LKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWER VMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWE ASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAY NVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK

D.4 CD SPECTRA OF PURIFIED PROTEINS AND SUPPLEMENTARY THERMAL MELTS

MCHERRY-HSP70 CD SPECTRA:

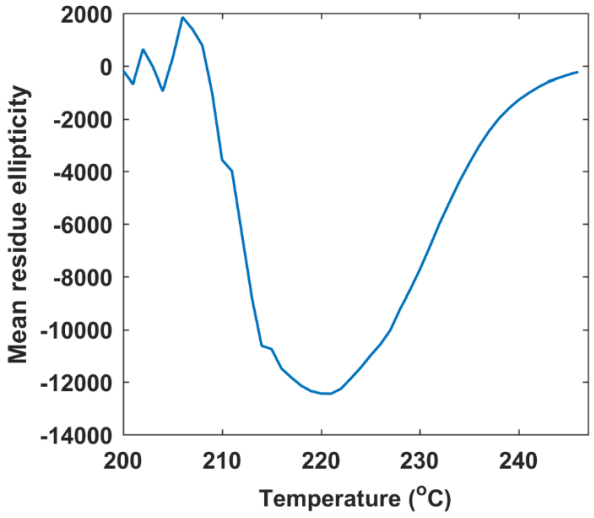


Figure D.3: CD Spectra of mCherry labeled Hsp70.

HDJ1 CD SPECTRA:

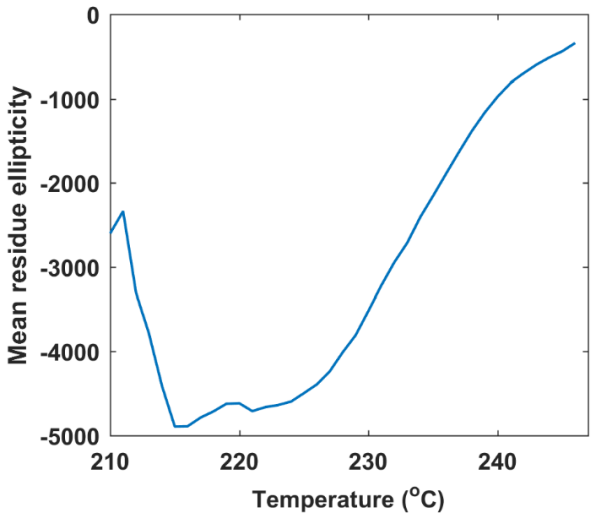


Figure D.4: CD Spectra of Hdj1.

ADDITONAL THERMAL MELTS PROBED BY TRYPTOPHAN FLUORESCENCE AND CD

Tryptophan is an intrinsically fluorescent amino acid used to monitor folding/unfolding of proteins. Its fluorescence is sensitive to the local environment where the peak emission red-shifts as the environment changes from hydrophobic to polar. Being a hydrophobic amino acid, it is generally buried in the core of the protein in the native state. On unfolding the tryptophan becomes exposed to the more polar aqueous environment. In addition to the peak shift, the polar environment also quenches fluorescence intensity and can also be used as a spectroscopic probe for protein unfolding (Figure D.5).

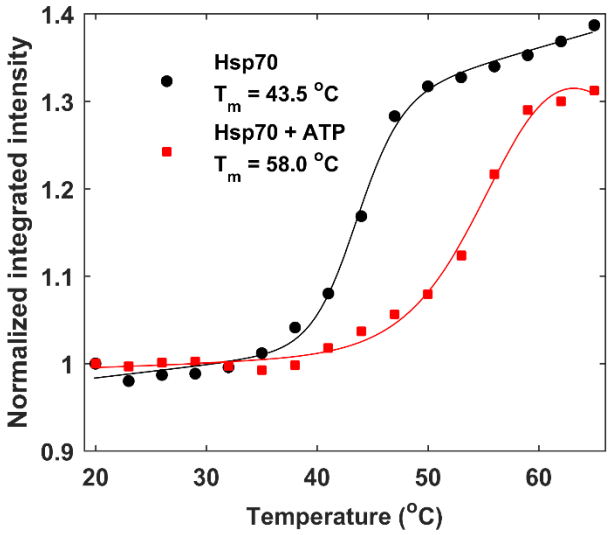


Figure D.5: Thermal unfolding of Hsp70 probed by tryptophan fluorescence integrated intensity change.

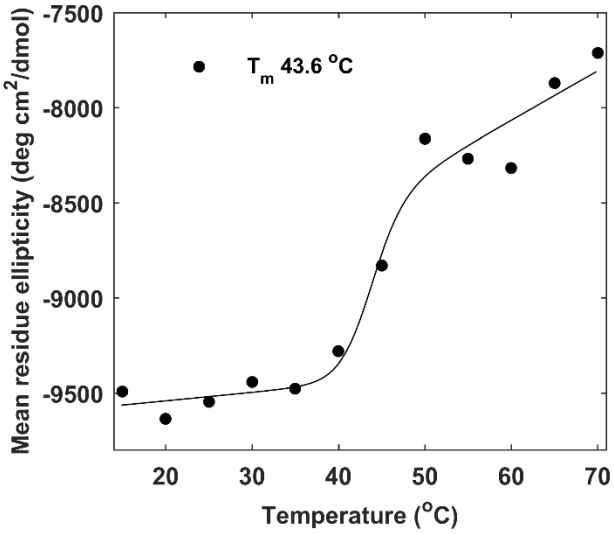


Figure D.6: Thermal unfolding of Hsp70 probed by mean residue ellipticity.

Since mHsp70 contains mCherry that is stable up to 65 °C the mCherry tryptophan fluorescence does not significantly change in the temperature range where Hsp70 unfolds. This background fluorescence overwhelms the signal from the single tryptophan in Hsp70. In order to tease out the subtle change in fluorescence from Hsp70 singular value decomposition (SVD) is performed. For a detailed description of SVD see chapter 3. In figure D.7, component 2 of matrix V after SVD is plotted.

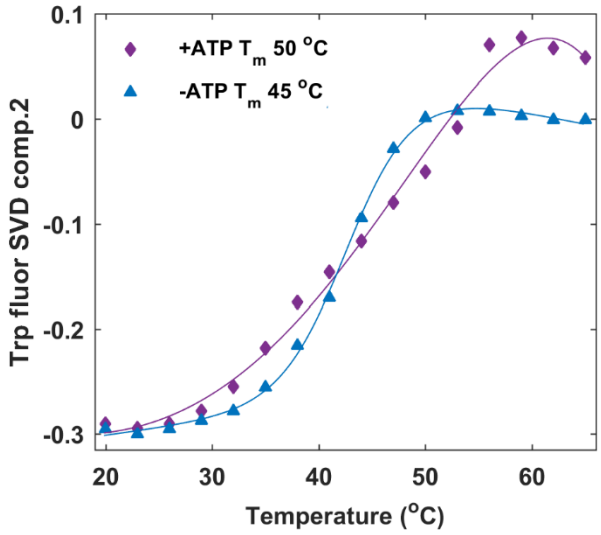


Figure D.7: Thermal unfolding of mHsp70 probed by tryptophan fluorescence.

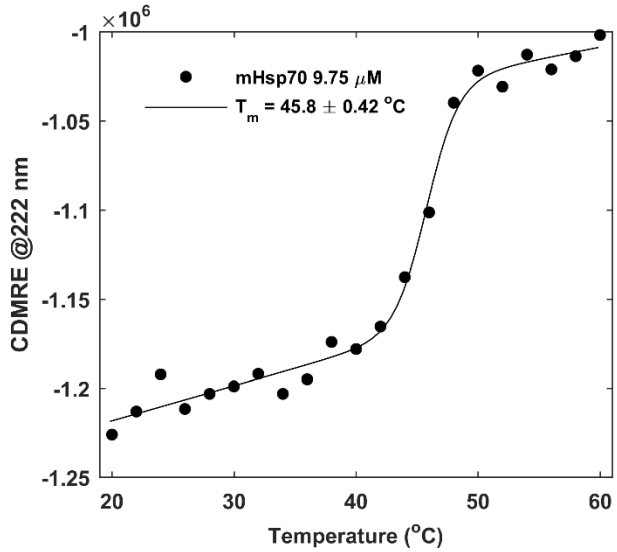


Figure D.8: Thermal unfolding of mHsp70 probed by mean residue ellipticity.

D.5 DETAILED PROTOCOL FOR B-GALACTOSIDASE ASSAY

- Prepare in advance:
    1. Need stocks of mHSP, HDJ1 and HSP70.
    2. 1 M glycylglycine pH 7.4.
    3. BSA is sold by NEB (20 mg/mL, MW ~69000Da) at 290 µM BSA.
    4. Dissolve 250 mg ONPG in 15 mL MQ (16.67 mg/mL final concentration), warm to dissolve. Aliquot and freeze 10 µL stocks.
    5. Prepare 0.5 M sodium carbonate.

6. 8 mL 10 M GuHCl.
7. 1 mL 20 mM ATP.
8. 1 mL 1 M DTT.

- Prepare 1.25 mL of 2Xrefolding buffer: 250 µL 10X K1, 25 µL 1 M DTT, 250 µL 20 mM ATP and 725.0 µL MQ.
- Prepare .625 mL of 2Xdenaturation buffer without GuHCl: 125 µL 10X K1, 0.44 µL 14.3 M BME and 500 µL MQ.
- For every reaction prepare ONPG supplemented refolding buffer – 300 µL 2Xrefolding buffer, 28.8 µL ONPG stock (final concentration 0.8 mg/mL) and 271 µL MQ.
- Add 10 µL of ONPG supplemented refolding buffer to separate reaction tubes corresponding to time points 0, 30, 60, 90, 120, 180 and 240 minutes for 7 reactions, total 49 tubes.
- Prepare 21.5 µM β-Gal **stock solution** in 1 M glycylglycine pH 7.4 by dissolving 200 µg in 20 µL 1M glycylglycine. Freeze the stock solution. Check concentration using UV A280.
- E1% (280 nm) = 20.9, ε = (E1%*molecular weight)/10. According to expasy molecular weight of β-Gal is 465877.38 so ε = 973683.
- Unfold β-Gal by first a 20-fold dilution (1 µL β-Gal + 19 µL 1 M glycylglycine) into 1 M glycylglycine to 1.07 µM. Freeze excess and use within a week.
- Denature 20-fold diluted β-Gal by 10-fold dilution in denaturation buffer, add 2 µL 1.07 µM β-Gal to 10 µL denaturation buffer with 6 µL 10 M GuHCl and 2 µL MQ. For control add 1 µL β-Gal to 5 µL refolding buffer and 4 µL MQ. Incubate 30 minutes at 30 C.
- Initiate the refolding reaction by dilution of denatured β-galactosidase 125-fold into refolding buffer at 4°C supplemented with 1.6 µM Hsp70 and 3.2 µM Hdj-1:
    1. β-Gal in Refolding buffer with 3.2 µM BSA only
       62.5 µL refolding buffer + 1.4 µL BSA + 60.1 µL MQ
    2. β-Gal in Denaturing buffer with 3.2 µM BSA only
       62.5 µL refolding buffer + 1.4 µL BSA + 60.1 µL MQ
    3. β-Gal in Denaturing buffer with 1.6 µM HSP70 only
       62.5 µL refolding buffer + 7 µL HSP70 + 54.5 µL MQ
    4. β-Gal in Denaturing buffer with 3.2 µM HDJ1 only
       62.5 µL refolding buffer + 25 µL HDJ1 + 36.5 µL MQ
    5. β-Gal in Denaturing buffer with 1.6 µM mHSP70 only
       62.5 µL refolding buffer + 16 µL mHSP70 + 45.2 µL MQ

6. β-Gal in Denaturing buffer with 1.6 µM HSP70 + 3.2 µM HDJ1 only 62.5 µL refolding buffer + 7 µL HSP70 + 25 µL HDJ1 + 29.6 µL MQ
7. β-Gal in Denaturing buffer with 1.6 µM mHSP70 + 3.2 µM HDJ1 only 62.5 µL refolding buffer + 16 µL mHSP70 + 25 µL HDJ1 + 20.2 µL MQ

- Add 1 µL control β-Gal in reaction tube 1, and 1 µL denatured β-Gal in all other tubes and mix by pipetting up and down and incubate at 37 C. Immediately after mixing remove 10 µL for time point 0 and incubate at 37 C for 15 minutes.
- Add 80 µL 0.5 M sodium carbonate to stop chromogenic reaction and measure absorbance at 412 nm.

D.6 DETAILED PROTOCOL FOR PREPARING TRANSFECTED U2-OS CELLS

- Get things together. During transfection you will need:

Day 1

1. A pre-split flask at around 60% confluence
2. PBS 1X
3. Trypsin (1 – 2 mL per flask) *
4. DMEM with PS and FBS *
5. Cover slips in petri dishes (as many as your experiments plus some extras)

Day 2

6. DMEM with FBS and without PS *
7. DMEM with FBS and PS *
8. Plasmids (As many as you are transfecting)
9. Opti-MEM 1X *
10. Lipofectamine

Day 3 (Day of imaging)

11. Opti-MEM 1X with 40% FBS *

- On day 1 around 12 hours before transfection warm up the starred items in the 37 C water bath for about 30 minutes
- Spray and sterilize (make sure you are aware of the appropriate hood protocol) the hood with 70% ethanol, attach a sterile pipette for aspiration and start vacuum.
- Aspirate the media out of the flask. Wash with PBS and aspirate.
- Add 1-2 mL of Trypsin and incubate at 37 °C for 5-10 minutes depending on the brand of trypsin.

- In the meantime, get the petri dishes ready in the hood and fill each with 2mL of DMEM with PS and FBS.
- Check the flask to see if the cells have popped off. Tap it gently if cells are not freely swimming.
- When the cells have popped of, add in 10 mL of DMEM with PS and FBS to the flask in the hood. Gently rock it.
- Add 1 mL of the cell media onto each cover slip. Add less if you want slower growth, the total volume should be 3 mL.
- Incubate at 37 C for 12 hours for 1 mL cover slips.
- On day 2, your cover slips should be around 70% confluent to be ready for transfection; warm the required reagents to 37 °C.
- Aspirate out the media with PS and wash with PBS and aspirate out the PBS.
- Add in 3 mL of DMEM with FBS without PS on to each of the cover slips and incubate for 30 minutes at 37 °C.
- For a total of 2 ug calculate the appropriate volume(For two plasmids use 1 ug of each plasmid) and mix it in 250 uL Opti-MEM 1X in an Eppendorf tube. This is tube A.
- Prepare appropriate amount of Lipofectamine in 250 uL of Opti-MEM 1X. (For every ug of plasmid use 2.5X volume in uL of Lipofectamine). This is tube B.
- Incubate tube A and B for 10 minutes at room temperature.
- Add in contents of tube B to tube A to avoid plasmid loss. Incubate for 20 minutes at room temperature. This is the more important incubation of the two.
- Divide the total volume in equal parts and pipette them on to the cover slips mixing them in the no PS media.
- Incubate for 3-4 hours.
- After incubation change the media back to DMEM with PS and FBS. Incubate for another 8-10 hours.
- Warm up an aliquot of Opti-MEM with FBS in the 37 °C water bath.
- Clean a slide and stick your spacer in the hood. Pipette 120 uL of Opti-MEM with FBS on the slide. This is the media for the cells on the cover slip.
- Bring out your cover slips from the incubator. Check to make sure your cells look okay.
- Aspirate the media from a petri dish and wash with PBS. Gently pick up the cover slip with a pair of forceps and flip it upside down to stick it on to the slide.
- Incubate for 5 minutes at 37 °C to equilibrate the cells.

D.7 DETAILED PROTOCOL FOR CELL DATA ANALYSIS

1. First run extract_cells_withsegmentation

   Note that this only runs on mat files generated from cell cropping and alignment for Crop_Align_Cell. This step lists all mat files in current folder and runs cellfind_multijumps_withsegmentation_redchannel_hist_thresh_sfr. The code thresholds green channel and picks pixels that stay above the threshold throughout the video. No user input required other than going to said folder at start of code or making a list of all folders to run this code on.

2. Then run findjumps_withsegmentedcells

   This will further analyze *cells.mat generated by step 1. This code runs findjumps_multijumps that picks jump positions and assigns the full trace to specific jumps for further analysis. User will need to pick peaks/enter value for auto peak recognition. User then approves jump picking. The code generates split trace variables jump_traces_ (green, red, da and time) and appends them to the said *cells.mat workspace.

3. Start = 1; append_cellsize_cellmat to add cellsize to each *cells.mat workspace

4. Initialize cell "folders" with folder names of folders containing desired folders with cells to average. For example, to average all cells with mHSP and mEGFP-PGK3 distributed over three days; folders may contain all such folders with these cells to be averaged.

5. Finally run Compile_all_cells(folders) which will then allow the user to look at traces and select them to either be averaged or excluded from averaging. A line is drawn through the raw traces using interpolation between 20 to 46 °C with a 2 °C temperature step. The traces are normalized and corrected for quantum yield slope by fitting the four starting points to a line and subtracting the slope. Finally, these traces are averaged.

D.8 MATLAB CODE FOR CELL FINDING

The cell finding code cellfind_multijumps_withsegmentation_redchannel_hist_thresh_sfr is highlighted below.

```
function cellfind_multijumps_withsegmentation_redchannel_hist_thresh_sfr( datapath, dataname, fps, thresh )
    %cellfind_multijumps_withsegmentation identifies cells using the
    %segmentation protocol in
    %https://blogs.mathworks.com/steve/2006/06/02/cell-segmentation/
    %   1. Threshold cells and align channels
    %   2. Save partially processed data
```

```
%   3. Convert labels to .avi movie for easier visualization of area
%   changes

%load data
strpath = pwd;
cd(datapath);
load(dataname, '-mat');

red_filename = strcat(dataname, 'red.mat');
red_exist = exist(red_filename);

mincellarea = 1000;
constant_thresh = thresh;

%Get video info
[height, width, numFrames] = size(dataright);
time = (1:numFrames)*1/fps - 1/fps;

%first threshold green channel(dataright) using frame 1 and pick out cell
%red channel image are low contrast so use adapthisteq to increase
%contrast slightly. adapthisteq implements a technique called
%contrast-limited adaptive histogram equalization, or CLAHE.
%im2bw converts the image to logical threshholded image
%thresholded image is then fill, dilated and cleared of amll objects
%<100 pixels to avoid getting small cells or areas of noise that make
%it throught the thresholding algorithm.
alignfr = 1;
I_eq = adapthisteq(uint8(dataleft(:,:,alignfr)), 'Distribution', 'rayleigh');

%Before proceeding align both channels
%Estimate the transformation needed to align the images using
%imregtform. Imregtform estimates geometric transformation that aligns
%two 2-D or 3-D images
[optimizer, metric] = imregconfig('multimodal');
tformEstimate = imregtform(uint8(dataright(:,:,1)), I_eq, 'rigid', optimizer, metric);

%Apply estimated geometric transform to the moving image. This example
%uses the 'OutputView' parameter to obtain a registered image the same
%size and with the same world limits as the reference image.
dataright_realigned = imwarp(uint8(dataright),tformEstimate,'OutputView',imref2d(size(I_eq)));

f1=figure;
%images overlayed
Im1 = imfuse(I_eq, dataright_realigned(:,:,1), 'falsecolor', 'Scaling', 'joint', 'ColorChannels', [1 2 0]);
```

```
    %images side-by-side
    Im2 = imfuse(I_eq, dataright_realigned(:,:,1), 'montage');
    %title('Panel 1 images overlayed');

    %initial threshold frame
    threshfr = 1;

    %panel 2 graythresh with non adaptive processing
    subplot(3,3,2);
    for i=1:100
       setThresh(i)=(i-1)*(1/100);
       getThresh_green(i)=sum(sum(im2bw(uint8(dataright_realigned(:,:,threshfr)),setThresh(i))));
       getThresh_red(i)=sum(sum(im2bw(uint8(dataleft(:,:,threshfr)),setThresh(i))));
    end
    plotyy((((1:100)-1)*(1/100)), getThresh_green, (((1:100)-1)*(1/100)), getThresh_red);
    hold on
    plot(graythresh(dataright_realigned(:,:,threshfr))*ones(100,1),getThresh_green,'b');
    plot(graythresh(uint8(dataleft(:,:,threshfr)))*ones(100,1),getThresh_red,'r');
    xlabel('Thresholds');
    ylabel('Pixels included');
    %title('Panel 2 graythresh with non adaptive processing');

    [labels, cellStats, thresh, numObjects, NP] = thresh_frames(uint8(dataleft(:,:,threshfr)),
constant_thresh, mincellarea, numFrames);

    %panel 3 graythresh mask with non adaptive processing
    subplot(3,3,3);
    BW = (labels(:,:,1).*dataright_realigned(:,:,threshfr))*255;
    %[~,temp_image] = image_binarize(BW, mincellarea);
    imshowpair(BW, dataright_realigned(:,:,threshfr), 'montage');
    %title('Panel 3 graythresh cell mask with non adaptive processing');
    clear temp_image BW

    for i = 1:numObjects
       %Convert label matrix into binary for calculating cell size
       %convert labels to uint8 and write values in labels_cellsize
       labels_cellsize(:,:,i) = uint8(labels);
       %convert all non zero values in labels_movie to 1 to calculate cell
       %size
       labels_cellsize(labels~=0 & labels~=i) = 1;
    end

    clear i
```

```
    %panel 5 threshold values vs. frame
    subplot(3,3,5)
    bar(1:1:numObjects, thresh, 'k');
    hold on
    bar(numObjects+1, NP, 'r');
    xlim([0 numObjects+2])
    xlabel('Cell Number');
    ylabel('Threshold value');

    %panel 6 bar distribution of a red channel image
    subplot(3,3,6)
    [counts,edges] =
histcounts(reshape(dataleft(:,:,1),1,size(dataleft,1)*size(dataleft,2)),256,'binmethod','integers');
    centers = (edges(1:end-1) + edges(2:end))/2;
    plot(fit(centers', counts','gauss2'), centers, counts, 'b')
    legend('off');
    xlabel('Pixel Intensity');
    ylabel('Number of pixels');

    %initialize variables
    leftTrace = zeros(numObjects, numFrames);
    rightTrace = zeros(numObjects, numFrames);
    daTrace = zeros(numObjects, numFrames);
    cell_size = zeros(numFrames, numObjects);

    for j = 1:numObjects
       %find row/column index assigned to cell 'j'
       [currentRow, currentColumn] = find(labels==j);
       if isempty(currentRow)
          continue
       end
       for i = 1:numFrames
          %pick out intensity values for pixels in currentRow and
          %currentColumn
          for k = 1:length(currentRow)
             currentLeft(k,1) = dataleft(currentRow(k), currentColumn(k), i);
             currentRight(k,1) = dataright_realigned(currentRow(k), currentColumn(k), i);
          end
          if isempty(currentLeft)
             continue
          end
          [low_left] = find(abs(currentLeft)<NP);
          currentLeft(low_left) = [];
          currentRight(low_left) = [];
```

```
        cell_size(i,j)=size(currentLeft,1);
        %Average cell pixels leftChannel and rightChannel
        leftTrace(j,i) = mean(double(currentLeft));
        rightTrace(j,i) = mean(double(currentRight));
        daTrace(j,i) = rightTrace(j,i)./leftTrace(j,i);
        %save currentLeft and currentRight in cells Left and Right
        %Left{i,j} = currentLeft;
        %Right{i,j} = currentRight;
        clear currentLeft currentRight
    end
end

if ~isempty(find(leftTrace==0,1))
    leftTrace(leftTrace == 0) = [];
    rightTrace(rightTrace == 0) = [];
    daTrace(daTrace == 0) = [];

    leftTrace = reshape(leftTrace, size(leftTrace,2)/5760, 5760);
    rightTrace = reshape(rightTrace, size(rightTrace,2)/5760, 5760);
    daTrace = reshape(daTrace, size(daTrace,2)/5760, 5760);
end

%panel 4 cell size vs. frame
subplot(3,3,4)
plot(1:1:numFrames, cell_size);
xlabel('Cell Number');
ylabel('Pixels included');
xlim([0 6000])

clear currentRow currentColumn i j k

%panel 6 leftTrace values vs. time
subplot(3,3,7)
plot(time, leftTrace, 'r');
xlabel('time (secs)');
ylabel('Red Trace');

%panel 7 leftTrace values vs. time
subplot(3,3,8)
plot(time, rightTrace, 'g');
xlabel('time (secs)');
ylabel('Green Trace');

%panel 8 daTrace values vs. time
```

```
    subplot(3,3,9)
    plot(time, daTrace, 'b');
    xlabel('time (secs)');
    ylabel('D/A Trace');

%     %VideoName will consist of slide and cell number and is saved in parent
%     %datapath
%     VideoName = strsplit(dataname, '_');
%     %example Video datapath ---> datapath\s1c1b.avi
%     LabelVideo = VideoWriter(fullfile(datapath, strcat(char(VideoName{1,end}),'.avi')));
%     LabelVideo.FrameRate = 60;
%     open(LabelVideo)
%     disp(strcat('Writing label file to video ->', strcat(char(VideoName{1,end}),'.avi')));
%     for i = 1:numFrames
%         %Convert label matrix into movie for manual checking
%         %convert labels to uint8 and write values in labels_movie
%         labels_movie_green = labels.*dataright_realigned(:,:,i);
%         labels_movie_red = labels.*uint8(dataleft(:,:,i));
%         %Increase contrast of images
%         labels_contrast_inc_green = (labels_movie_green);
%         labels_contrast_inc_red = (labels_movie_red);
%         %Fuse both constrast adjusted images together
%         labels_movie = imfuse(labels_contrast_inc_green, labels_contrast_inc_red, 'montage');
%         %write labels_movie into video
%         writeVideo(LabelVideo,labels_movie);
%         clear labels_movie_green labels_movie_red labels_contrast_inc_green labels_contrast_inc_red
%     end
%     close(LabelVideo)

    disp(strcat('save all variables to file:  ',dataname,'_cells.mat'));
    save(strcat(dataname, '_cells'),'labels', 'cellStats', 'time', 'leftTrace', 'rightTrace', 'daTrace', 'height',
'width', 'thresh', '-v7.3');

    %Handle amber only red channel
    if red_exist==2
        load(red_filename);
        %Align amber images
        tformEstimate_red = imregtform(uint8(red(:,:,1)), I_eq, 'rigid', optimizer, metric);
        red_realigned = imwarp(uint8(red),tformEstimate_red,'OutputView',imref2d(size(I_eq)));
        %panel 3 Amber red overlayed with FRET red
        Imred = imfuse(I_eq, red_realigned(:,:,1), 'falsecolor', 'Scaling', 'joint', 'ColorChannels', [1 2 0]);
        %display overlayed FRET red with green and overlayed with amber red
        %side by side
        Im3 = imfuse(Im1, Imred, 'montage');
```

```
        %display all images side by side
        Im4 = imfuse(Im3, Im2, 'montage');
        subplot(3,3,1);
        imshow(Im4);
        %save aligned images
        saveas(f1, strcat(dataname, '_aligned.jpg'));
        saveas(f1, strcat(dataname, '_aligned.fig'));
        close(f1);
        for j = 1:numObjects
            [currentRow_red, currentColumn_red] = find(labels(:,:,1)==j);
            if isempty(currentRow_red)
                continue
            end
            for k = 1:length(currentRow_red)
                currentRed(k,:,j) = red_realigned(currentRow_red(k), currentColumn_red(k),:);
            end
            red_mean(j,:) = mean(double(currentRed(:,:,j)));
        end
        if ~isempty(find(red_mean==0,2))
            red_mean(red_mean == 0) = [];
            red_mean = reshape(red_mean, size(red_mean,2)/60, 60);
        end
        disp(strcat('save amber red to file: ',dataname,'_cells.mat'));
        save(strcat(dataname, '_cells'), 'red_mean', '-append');
    else
        %display overlayed FRET red with green and side by side
        Im3 = imfuse(Im1, Im2, 'montage');
        subplot(3,3,1);
        imshow(Im3);
        %save aligned images
        disp(strcat('No red video for ->', dataname));
        saveas(f1, strcat(dataname, '_aligned.jpg'));
        saveas(f1, strcat(dataname, '_aligned.fig'));
        close(f1);
    end
    clear I_eq
end

function [labels, cellStats, thresh, numObjects, NP] = thresh_frames(data, constant_thresh, mincellarea,
~)
    %temporarily convert images to binary using graythresh
    BW0 = imadjust(data(:,:,1));
    BW = im2bw(BW0, graythresh(BW0));
    clear BW0
```

```
    [cells,~] = image_binarize(BW, mincellarea);
    cell_IDs = regionprops(cells, 'BoundingBox');
    cell_Labels = labelmatrix(cells);
    numObjects = size(cell_IDs, 1);
%     numObjects = 1;
    frames_to_thresh = 1;
    thresh = zeros(numObjects,frames_to_thresh);
    labels = uint8(zeros(size(data,1),size(data,2),frames_to_thresh));

    for j = 1:numObjects
        I_cell = data;
        I_rect = [1 1 size(data,2)-1 size(data,1)-1];
        width_cropim = round(I_rect(1,1)+I_rect(1,3));
        height_cropim = round(I_rect(1,2)+I_rect(1,4));
        for i = 1:frames_to_thresh
            %First fit the image using single gaussian
            %The histogram of the values from 0 to 255 is stored in counts,
            %Edges gives the left and right edge of the bin
            %For instance, count(1) contains the number of occurrence of the
            %value zero in the image.
            [counts,edges] =
histcounts(reshape(I_cell(:,:,i),1,size(I_cell,1)*size(I_cell,2)),256,'binmethod','integers');
            %centers calculates the mid point of each bin; should be te integer
            %pixel intensity value between 0-255
            centers = (edges(1:end-1) + edges(2:end))/2;
            %fit the histogram using a Gauss1 fit. This should mainly fit the
            %large peak due to background pixels.
            Gauss_noise = fit(centers', counts', 'gauss1');
            %Find out the noise pixel value mu+sigma
            NP = round(Gauss_noise.b1 + (Gauss_noise.c1/sqrt(2)));

            if numObjects>1 && i==1
                clear I_cell I_rect
                %Choose smallest box containing foreground pixels
                I_rect = cell_IDs(j).BoundingBox;
                I_rect(1,1:2) = ceil(I_rect(1,1:2)-10);
                I_rect(1,3:4) = I_rect(1,3:4)+10;
                width_cropim = round(I_rect(1,1)+I_rect(1,3));
                height_cropim = round(I_rect(1,2)+I_rect(1,4));
                if ~isempty(find(I_rect<1))
                    I_rect(find(I_rect<1)) = 1;
                end
                if I_rect(1,2)+height_cropim>size(data,1)
                    height_cropim = size(data,1) - 1;
```

```
            end
            if I_rect(1,1)+width_cropim>size(data,2)
               width_cropim=size(data,2) - 1;
            end
            %Make sure number of cells is one in I_cell
            cell_IDs_afterautocrop = cell_Labels(I_rect(1,2):height_cropim, I_rect(1,1):width_cropim);
            [~,~,numObjects_afterautocrop] = find(cell_IDs_afterautocrop~=0 &
cell_IDs_afterautocrop~=j);
            if ~isempty(numObjects_afterautocrop)
               %if still more than one cell in I_cell do manual
               %cropping
               clear row_cellj col_cellj cell_IDs_afterautocrop numObjects_afterautocrop
               f2 = figure;
               image_to_show = imadjust(data(:,:,1));
               imshow(image_to_show);
               I_rect_new = imrect(gca, I_rect);
               I_rect = wait(I_rect_new);
               close(f2);
            end
            I_cell = data(I_rect(1,2):height_cropim, I_rect(1,1):width_cropim, :);
         end
         %Fit the the data again using calculated fit parameters in
         %Gauss_noise
         fity = Gauss_noise(centers);
         %Find the first index where fity is less than 1 after maximum
         %of Gauss_noise has occured. Assume this where the second gaussian
         %with foreground signal distribution starts
         [~,max_fity_idx] = max(fity);
         row_ind = find(fity(max_fity_idx:end)<(0.01*Gauss_noise.a1),1) + max_fity_idx - 1;
         counts_interp = interp1(centers(row_ind:end), counts(row_ind:end),
centers(row_ind):0.2:centers(end));
         %Fit the second smaller gaussian with brighter pixels starting at
         %row_ind
         if constant_thresh == 0
            GaussB = fit((centers(row_ind):0.2:centers(end))', counts_interp', 'gauss1');
            %Find threshold at left side FWHM x value for GaussB
            %specifically at b1-(c1/sqrt(2)) or mu-sigma. Choose only top
            %25% values
            thresh(j,i) = floor(GaussB.b1 - (0.25*(GaussB.c1/sqrt(2))));
         else
            thresh(j,i) = constant_thresh;
         end
         if thresh(j,i) < centers(row_ind)
            Gauss_noise_interp = Gauss_noise(centers(1):0.2:centers(end));
```

```
            [~,max_Gauss_noise_interp_idx] = max(Gauss_noise_interp);
            row_ind_thresh_cutoff =
find(Gauss_noise_interp(max_Gauss_noise_interp_idx:end)<(0.01*Gauss_noise.a1),1) +
max_Gauss_noise_interp_idx - 1;
            disp('Threshold too low, using gaussian noise cutoff')
            thresh(j,i) = round(centers(1) + (row_ind_thresh_cutoff-1)*0.2);
        end
        disp(thresh(j,i));
        %start image binarization based on thresh(i)
        BW = I_cell(:,:,i);
        BW(BW<thresh(j,i))=0;
        BW = logical(BW);
        [cells,~] = image_binarize(BW, mincellarea);
        CellNum_ID(:,:,i) = labelmatrix(cells);
        if max(max(CellNum_ID(:,:,i)))>1
            temp_cellstats = regionprops(cells, 'Area');
            temp_cellSizes = cat(1, temp_cellstats.Area);
            [~,idx_bigcell] = max(temp_cellSizes);
            temp_CellNum_ID = CellNum_ID(:,:,i);
            temp_CellNum_ID(temp_CellNum_ID~=idx_bigcell) = 0;
            temp_CellNum_ID(temp_CellNum_ID==idx_bigcell) = 1;
            CellNum_ID(:,:,i) = temp_CellNum_ID;
        end
        clear counts edges centers Gauss_noise fity GaussB row_ind BW cells temp_cellSizes
temp_CellNum_ID temp_cellstats
    end
    CellNum_ID(CellNum_ID~=0) = CellNum_ID(CellNum_ID~=0) + j - 1;
    labels(I_rect(1,2):height_cropim, I_rect(1,1):width_cropim, :) = CellNum_ID(:,:,:);
    clear I_cell I_rect CellNum_ID height_cropim width_cropim
  end
  cellStats{:,1} = regionprops(labels(:,:,1), data, 'BoundingBox', 'Centroid', 'PixelList', 'MeanIntensity',
'Area');
  clear i j
end

function [cells,BW6] = image_binarize(BW, mincellarea)
  %BW = im2bw(data, thresh_im);
  BW1 = bwareaopen(BW, mincellarea);
  BW2 = bwperim(BW1);
  BW3 = imfill(BW2,'holes');
  BW4 = bwmorph(BW3, 'clean');
  seD = ones(3, 'uint8');
  BW5 = imclose(BW4, seD);
  BW6 = bwareaopen(BW5, mincellarea);
```

```
    %identify the cell using bwconncomp, bwconncomp recognizes connected
    %objects. Use bwconncomp component structure to create a matrix of
    %cells where each cell is identified by a array of pixels labeled
    %1,2,... Cell 1 pixels for example are labeled 1, cell 2 2 and so on
    cells =  bwconncomp(BW6);
end
```

D.9 MATLAB CODE FOR JUMP FINDING

```
function findjumps_multijumps( datapath, dataname )

%Opens partially analysed matlab workspaces and implements jump separation
%   MATLAB workspaces with cells first analyzed by
%   cellfind_multijumps_withsegmentation. leftTrace and rightTrace are then
%   separated into jumps. Jumps are separated by manual or automatic peak
%   picking.

%Open matlab workspace
    cd(datapath);
    load(dataname, 'rightTrace', 'leftTrace', 'daTrace', 'time');
    split_name = strsplit(dataname,'.');
    split_name = char(split_name(1));
    %intitalize jump_traces_green to check sizes
    jump_traces_green = {};
    %Intialize variables
    numObjects = size(leftTrace,1);
    numJumps = 10;
    included_cells = zeros(numObjects, 1);
    pks = zeros(numObjects, numJumps+2);
    pks(:,1) = 16;
    pks(:,end) = 4900;
    for i = 1:numObjects
        fig_name = strsplit(dataname, '_cells');
        f1=openfig(strcat(char(fig_name(1)),'_aligned'));
        f = figure;
        %plot differential of mean green channel signal. Points at which the
        %temperature is jumped should appear as spikes.
        plotyy(1:1:5759,diff(rightTrace(i,1:end)),1:1:5760,rightTrace(i,1:end));
        axis([0 inf -inf 1]);
        %skip cell is jumps don't look clean or aren't apparent as spikes.
        skip_loop = input('Enter 0 to continue, 1 to skip current cell ');
        close(f1);
        close(f);
        if skip_loop == 1
```

```
    continue
else
    included_cells(i,1) = i;
end
%decide if peaks are picked manually or automatically
pickhow = input('Pick peak manually? Y/N ', 's');
%if no user input choose manual by default
if isempty(pickhow)
    pickhow = 'N';
end

%for manual picking open figure of 1st derivative of green channel and
%then let user pick out 10 points
if (pickhow == 'Y' || pickhow == 'y')
    disp('Pick peaks now.');
    f = figure;
    plotyy(1:1:5759,diff(rightTrace(i,1:end)),1:1:5760,rightTrace(i,1:end));
    axis([0 inf -inf 1]);
    [pks_picked, ~] = getpts(f);
    %If X is a vector of length m, then Y = diff(X) returns a vector of
    %length m-1. Add 1 to the point indices to offset the index by 1 to
    %account for the loss.
    pks_picked = pks_picked + 1;
    %Make sure that the index picked is actually at the peak of the
    %spike.
    for k = 1:size(pks_picked,1)
        %convert the index to whole number and expand search for maxima
        %to 10 points centered around the peak picked
        peak_index = floor(pks_picked(k));
        [~,pks_picked(k)] = max(-diff(rightTrace(i,peak_index-30:peak_index+30)));
        pks_picked(k) = pks_picked(k) + peak_index - 30;
    end
    pks(i,2:size(pks_picked)+1) = pks_picked;
    close(f);

    faketrace = -diff(rightTrace(i,:));
    x=1:length(faketrace);
    x(pks_picked-1)=[];
    faketrace(x)=0;
    f = figure;
    subplot(1,2,1)
    findpeaks(faketrace,'Npeaks',numJumps,'MinPeakDistance', 300);
    hold on
    plot(-diff(rightTrace(i,:)),'k--');
```

```
        subplot(1,2,2)
        plot(rightTrace(i,:), 'k')
        hold on
        plot(pks_picked, rightTrace(i,pks_picked), 'ro');
        disp('If jumps selection looks good click on figure to proceed');
        waitforbuttonpress;
        close(f);
        clear faketrace x
    else
        f = figure;
        plot(diff(rightTrace(i,:)));
        axis([0 inf -inf 1]);
        grid on;
        %automatic recognition picks peaks according to a minimum height
        %cutoff. Only use if all spikes are over a minimum height cutoff
        %and are separated by atleast 450 points.
        cutoff = input('Enter cutoff height for automatic jump recognition: ');
        [~,pks_picked] = findpeaks(-diff(rightTrace(i,:)),'Npeaks',numJumps,'MinPeakDistance', 400,
'MinPeakHeight', cutoff);
        pks_picked = pks_picked+1;
        close(f);

        f = figure;
        subplot(1,2,1)
        findpeaks(-diff(rightTrace(i,:)),'Npeaks',numJumps,'MinPeakDistance', 400, 'MinPeakHeight',
cutoff);
        hold on
        plot(-diff(rightTrace(i,:)), 'k--');
        subplot(1,2,2)
        plot(rightTrace(i,:), 'k')
        hold on
        plot(pks_picked, rightTrace(i,pks_picked), 'ro');
        disp('If jumps selection looks good click on figure to proceed');
        waitforbuttonpress;
        close(f);
        pks(i,2:size(pks_picked,2)+1) = pks_picked;
    end
    pks(i,:) = sort(pks(i,:),2);
    for j = 1:size(pks,2)-1 %separate video into jumps
        jump_traces_green{j,i} = rightTrace(i, pks(i,j) - 4: pks(i,j+1) - 5);
        jump_traces_red{j,i} = leftTrace(i, pks(i,j) - 4: pks(i,j+1) - 5);
        jump_traces_da{j,i} = daTrace(i, pks(i,j) - 4: pks(i,j+1) - 5);
        jump_time{j} = time(1, pks(i,j) - 4: pks(i,j+1) - 5);
    end
```

```
    clear pks_picked
    end

    if size(jump_traces_green,1)~=0
        disp(strcat('save and append to file: ',split_name));
        save(split_name, 'jump_traces_green', 'jump_traces_red', 'jump_traces_da', 'jump_time', 'pks',
'included_cells', '-append');
    else
        save(split_name, 'included_cells', '-append');
    end
end
```

D.10 RAW CELL TRACES

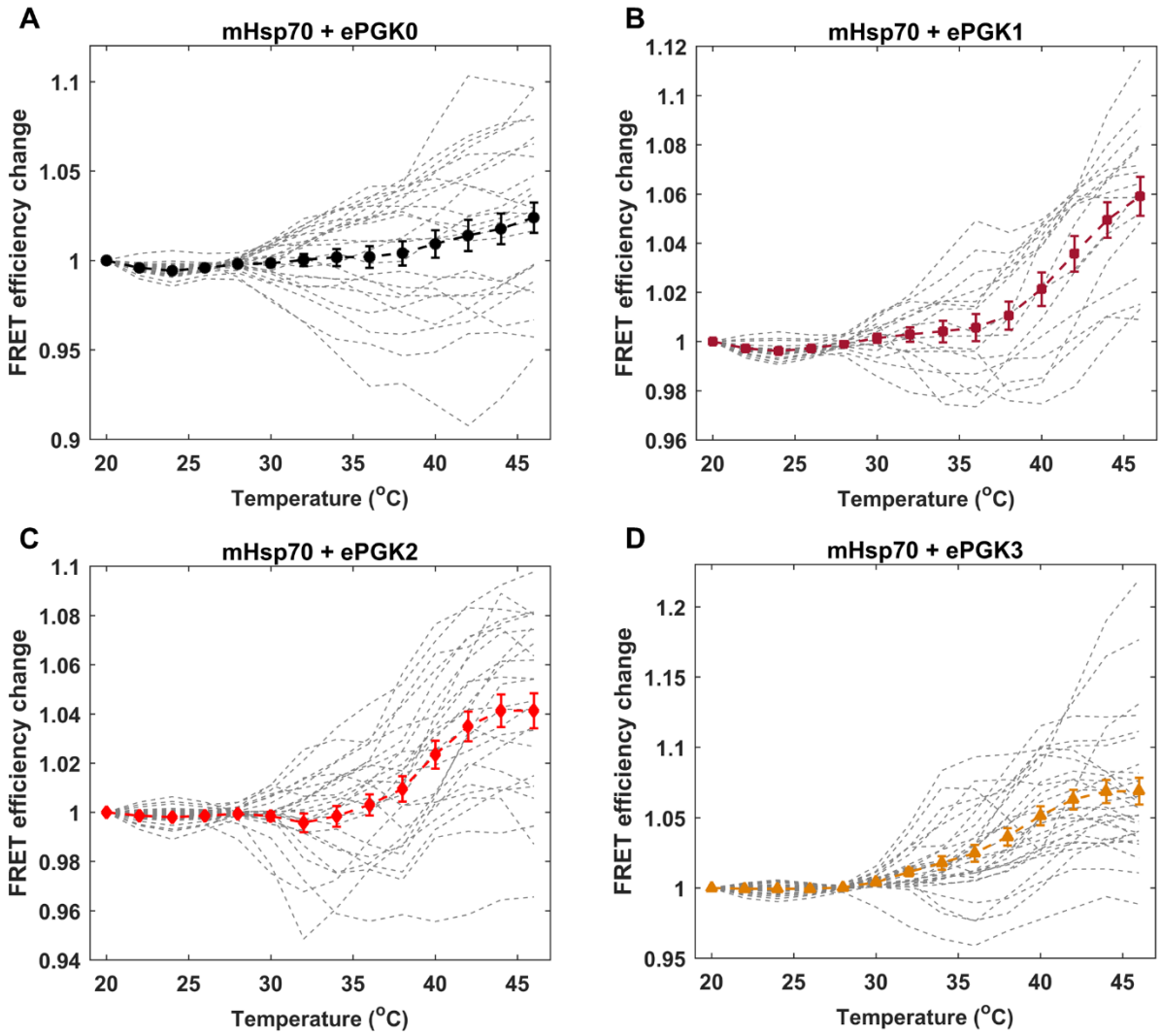


Figure D.9: mHsp70-ePGKs in-cell binding raw cell traces.

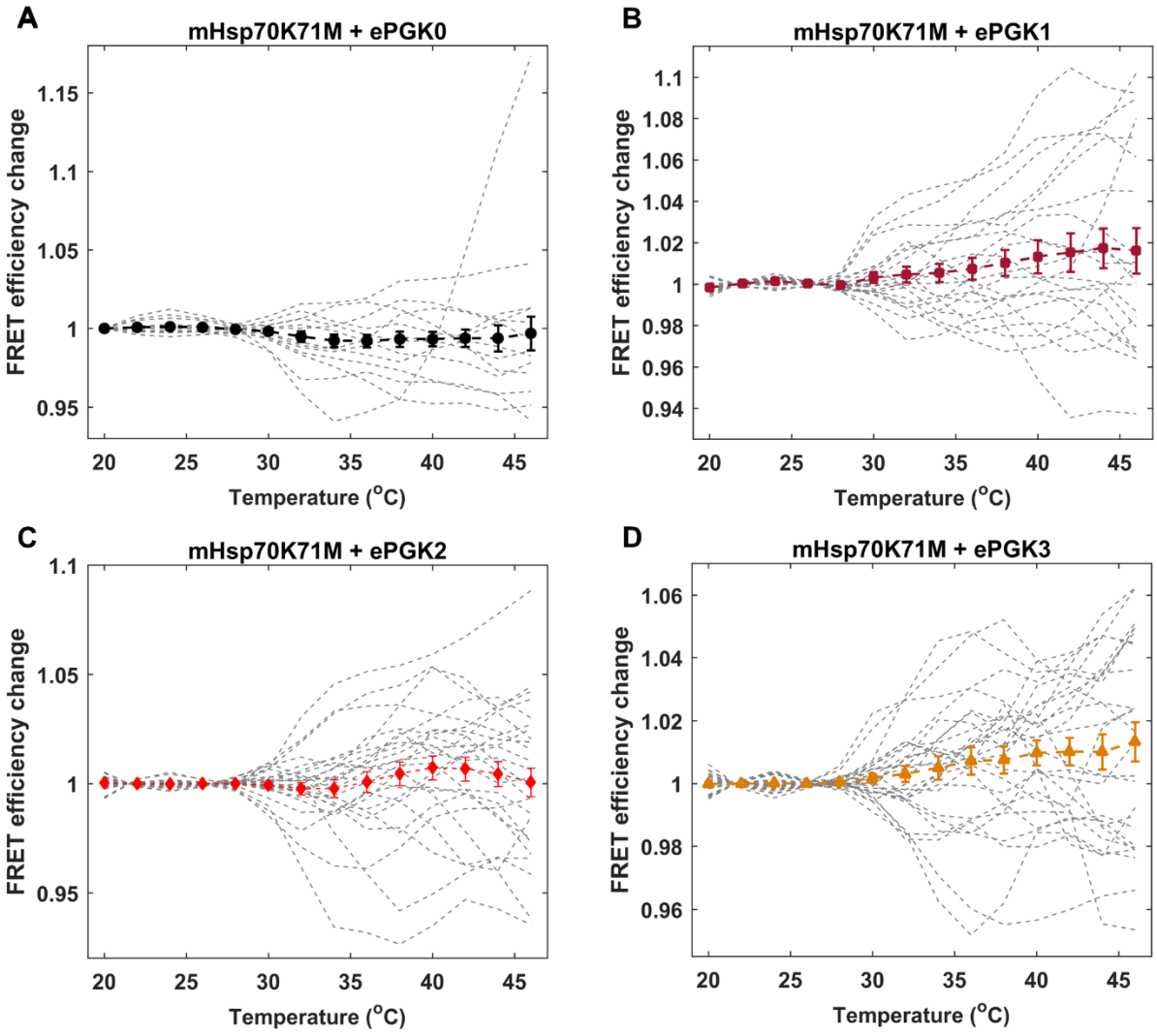


Figure D.10: mHsp70K71M-ePGKs in-cell binding raw cell traces.

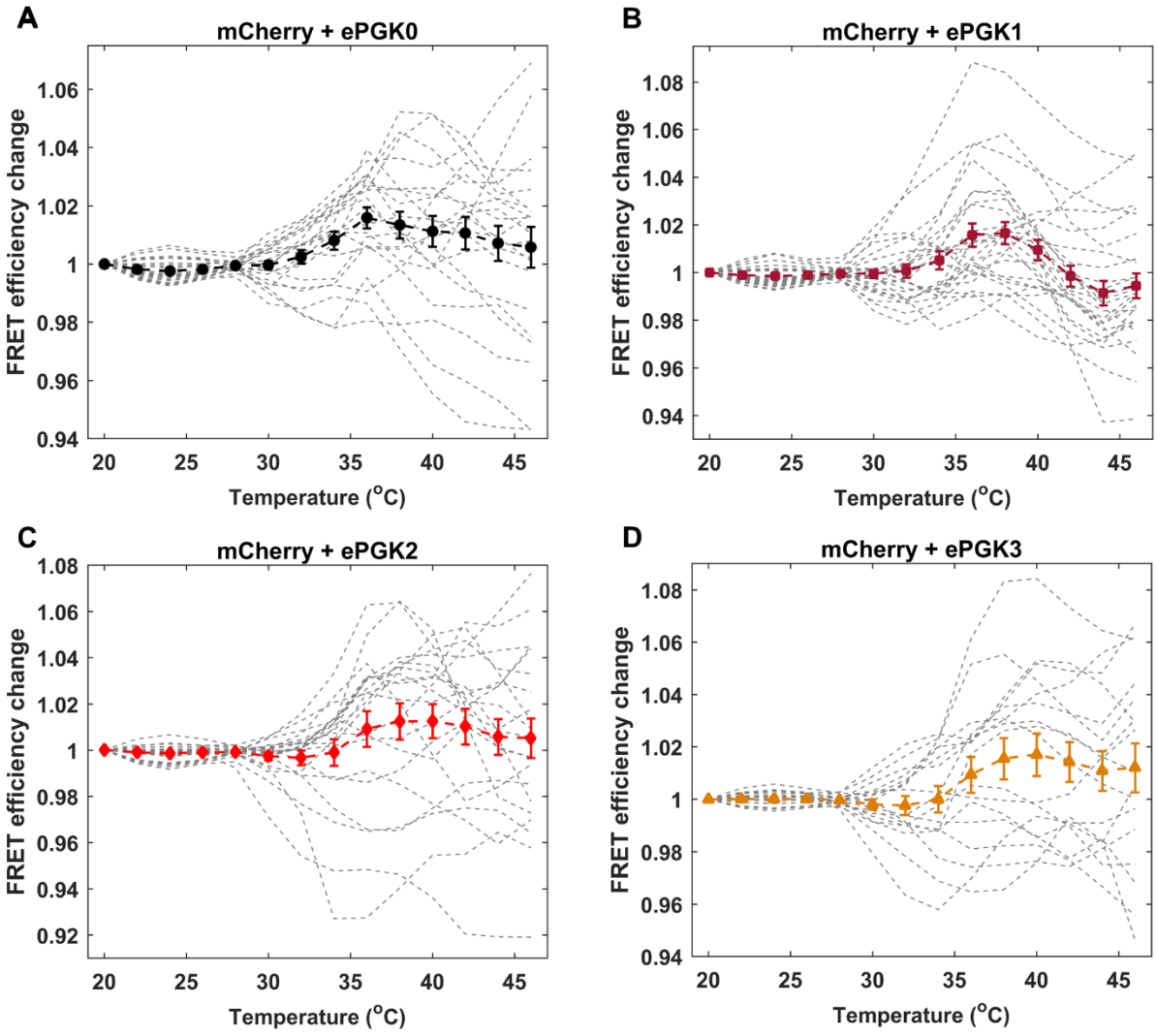


Figure D.11: mCherry-ePGKs in-cell binding raw cell traces

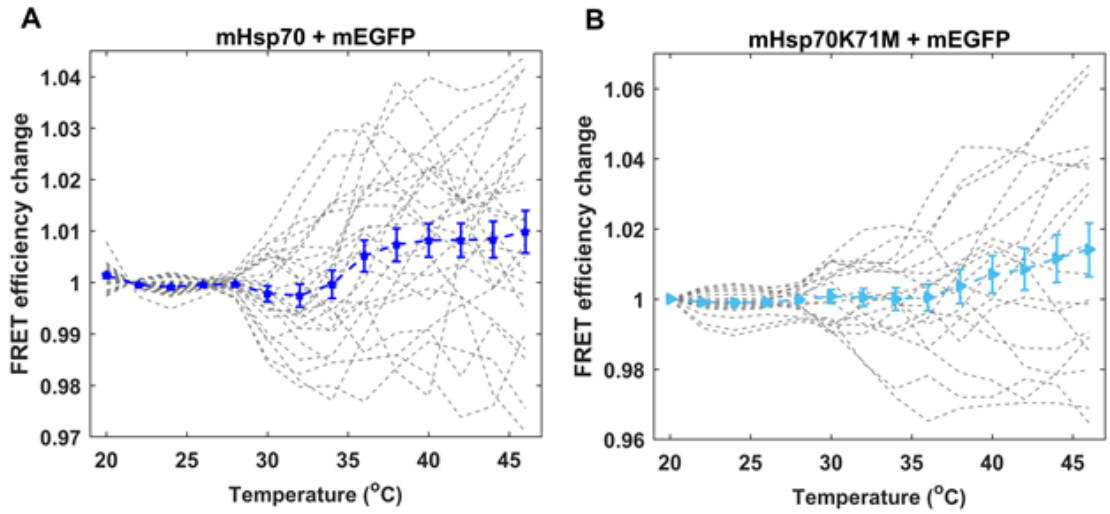


Figure D.12: mHsp70-mEGFP and mHsp70K71M-mEGFP in-cell binding raw cell traces.

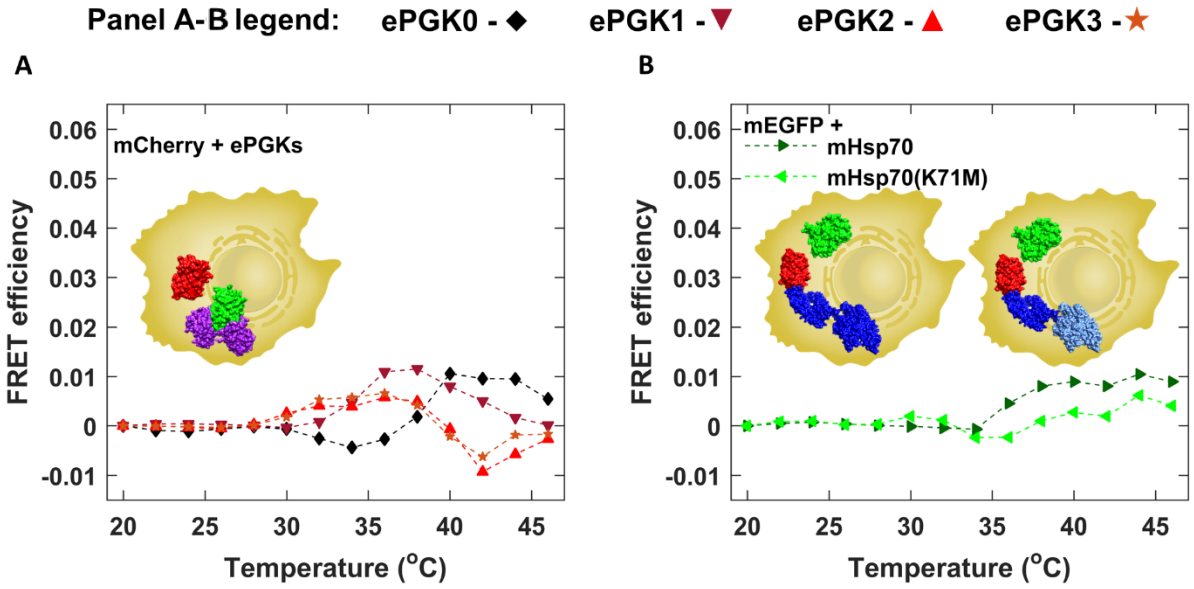


Figure D.13: mCherry-PGKs and mEGFP-mHsp70/mHsp70K71M control experiments average from figures D.11-D.12.

APPENDIX E

SUPPLEMENTARY INFORMATION FOR CHAPTER 6: CHARACTERIZING IN-CELL AND *IN VITRO* BINDING FOR THE CONSTITUTIVELY EXPRESSED CYTOPLASMIC HSP70 – HSC70

E.1 NUCLEOTIDE SEQUENCE FOR HSC70

CATATGAGCAAAGGCCCGGCGGTTGGCATTGACCTGGGCACCACCTACTCCTGCGTTGGC
GTGTTTCAGCATGGCAAAGTTGAGATTATCGCGAACGACCAGGGCAACCGTACCACCCCG
AGCTACGTGGCTTTCACCGACACCGAGCGCCTGATTGGCGACGCTGCGAAAAACCAGGTG
GCTATGAACCCGACCAACACCGTGTTCGATGCCAAGCGTCTGATTGGCCGTCGCTTTGAC
GATGCTGTGGTGCAGAGCGACATGAAACACTGGCCGTTCATGGTGGTGAACGATGCCGGT
CGCCCGAAGGTGCAGGTGGAGTACAAGGGCGAAACCAAGAGCTTCTACCCGGAGGAAGTG
AGCAGCATGGTGCTGACCAAGATGAAAGAGATCGCCGAAGCGTACCTGGGTAAAACCGTG
ACCAATGCGGTGGTGACCGTGCCAGCGTACTTCAACGACAGCCAGCGTCAGGCGACCAAG
GATGCTGGTACCATTGCCGGCCTGAACGTGCTGCGCATCATTAACGAGCCAACCGCTGCC
GCGATTGCTTATGGTCTGGACAAGAAAGTGGGCGCGGAACGTAACGTGCTGATTTTCGAC
CTGGGCGGTGGCACCTTTGATGTGAGCATCCTGACCATTGAGGACGGTATCTTCGAAGTG
AAGAGCACCGCGGGCGATACCCACCTGGGTGGCGAGGACTTTGATAACCGTATGGTGAAC
CACTTCATCGCCGAGTTTAAGCGCAAACACAAGAAAGACATTAGCGAAAACAAACGTGCG
GTGCGTCGCCTGCGTACCGCCTGCGAACGTGCTAAACGCACCCTGAGCAGCAGCACCCAG
GCGAGCATCGAGATTGACAGCCTGTACGAAGGCATCGATTTCTATACCAGCATTACCCGT
GCTCGCTTCGAGGAACTGAACGCCGACCTGTTTCGTGGCACCCTGGACCCGGTGGAGAAA
GCCCTGCGTGACGCGAAGCTGGATAAAAGCCAGATCCACGACATCGTTCTGGTGGGCGGC
AGCACCCGTATCCCGAAGATTCAGAAACTGCTGCAGGACTTCTTTAACGGCAAGGAGCTG
AACAAAAGCATCAACCCGGATGAAGCGGTGGCTTATGGTGCTGCGGTGCAGGCCGCTATT
CTGAGCGGCGATAAGAGCGAGAACGTGCAGGACCTGCTGCTGCTGGATGTGACCCCGCTG
AGCCTGGGTATCGAAACCGCCGGTGGCGTGATGACCGTGCTGATCAAACGCAACACCACC
ATTCCGACCAAGCAGACCCAGACCTTCACCACCTACAGCGACAACCAGCCGGGTGTGCTG
ATCCAGGTGTATGAGGGCGAACGTGCGATGACCAAAGATAACAACCTGCTGGGCAAGTTC
GAGCTGACCGGCATCCCGCCGGCTCCGCGCGGCGTGCCGCAGATTGAAGTGACCTTTGAC
ATCGATGCGAACGGCATCCTGAACGTGAGCGCCGTGGACAAAAGCACCGGCAAGGAGAAC
AAAATCACCATTACCAACGACAAAGGTCGTCTGAGCAAAGAGGACATCGAACGCATGGTG
CAGGAAGCCGAAAAGTACAAAGCCGAGGACGAAAAGCAGCGTGATAAAGTGAGCAGCAAG
AACAGCCTGGAGAGCTATGCGTTTAACATGAAAGCTACCGTGGAGGATGAAAAACTGCAG
GGCAAGATCAACGACGAGGATAAGCAGAAAATTCTGGACAAGTGCAACGAAATCATTAAC
TGGCTGGATAAAAACCAGACCGCCGAGAAAGAGGAGTTCGAACACCAGCAGAAAGAGCTG
GAAAAGGTGTGCAACCCGATCATTACCAAGCTGTATCAGAGCGCGGGTGGCATGCCGGGT
GGCATGCCGGGTGGCTTTCCGGGCGGCGGTGCCCCACCAAGCGGCGGCGCAAGCAGCGGC
CCGACCATTGAGGAAGTTGACTAAGGATCC

E.2 AMINO ACID SEQUENCE FOR HSC70

Red residues show N-terminal nucleotide binding domain and blue residues show C-terminal substrate binding domain.

MSKGPAVGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERLIGDAAKNQVAMNPT
NTVFDAKRLIGRRFDDAVVQSDMKHWPFMVVNDAGRPKVQVEYKGETKSFYPEEVSSMVLTK
MKEIAEAYLGKTVTNAVVTVPAYFNDSQRQATKDAGTIAGLNVLRIINEPTAAAIAYGLDKKVG
AERNVLIFDLGGGTFDVSILTIEDGIFEVKSTAGDTHLGGEDFDNRMVNHFIAEFKRKHKKDISEN
KRAVRRLRTACERAKRTLSSSTQASIEIDSLYEGIDFYTSITRARFEELNADLFRGTLDPVEKALRD
AKLDKSQIHDIVLVGGSTRIPKIQKLLQDFFNGKELNKSINPDEAVAYGAAVQAAILSGDKSENV
QDLLLLDVTPLSLGIETAGGVMTVLIKRNTTIPTKQTQTFTTYSDNQPGVLIQVYEGERAMTKDN
NLLGKFELTGIPPAPRGVPQIEVTFDIDANGILNVSAVDKSTGKENKITITNDKGRLSKEDIERMVQ
EAEKYKAEDEKQRDKVSSKNSLESYAFNMKATVEDEKLQGKINDEDKQKILDKCNEIINWLDK
NQTAEKEEFEHQQKELEKVCNPIITKLYQSAGGMPGGMPGGFPGGGAPPSGGASSGPTIEEVD

E.3 SUPPLEMENTARY THERMAL MELTS

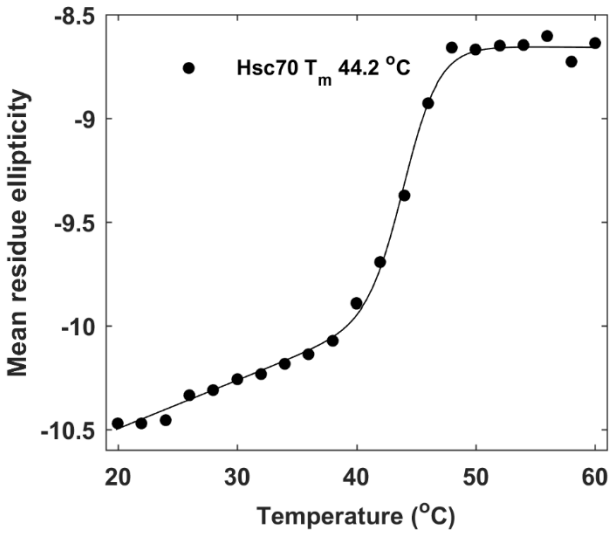


Figure E.1: Thermal unfolding of Hsc70 probed by mean residue ellipticity.

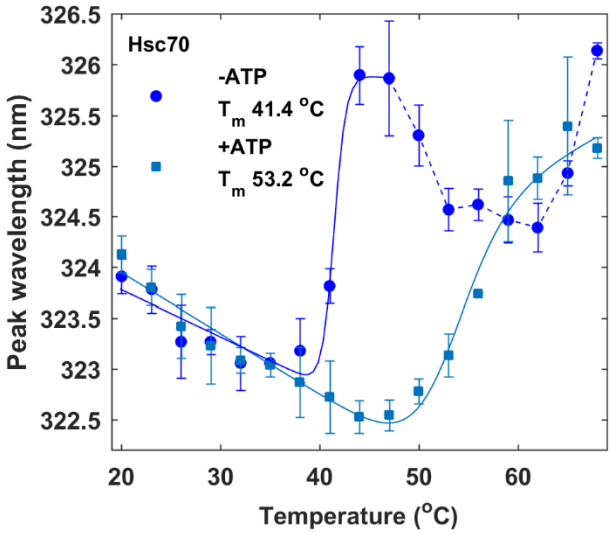


Figure E.2: Thermal unfolding of Hsc70 probed by tryptophan fluorescence peak wavelength shift.

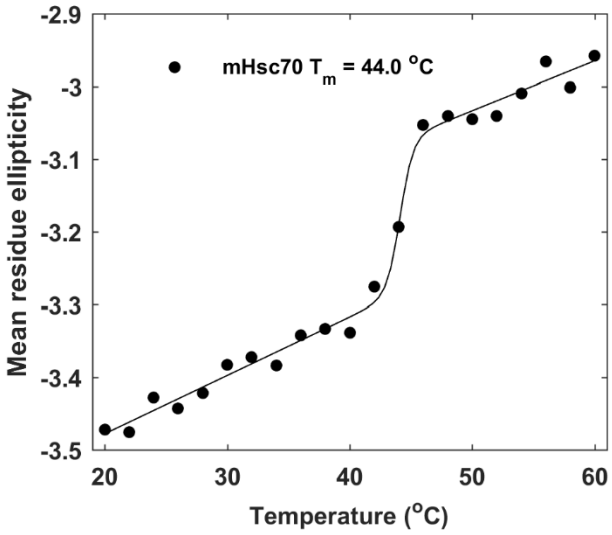


Figure E.3: Thermal unfolding of mHsc70 probed by mean residue ellipticity.

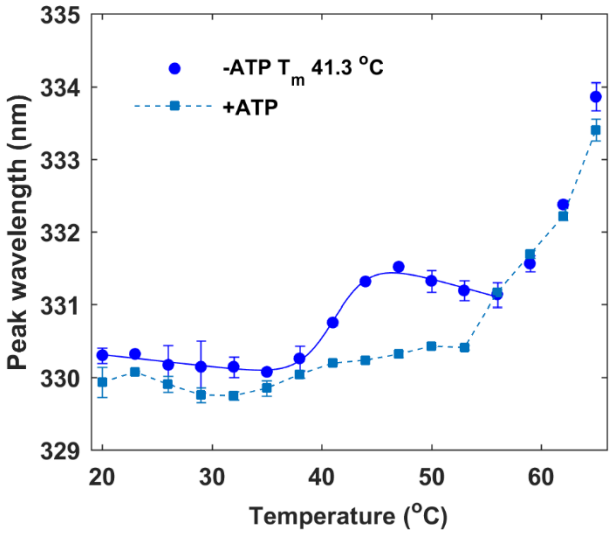


Figure E.4: Thermal unfolding of mHsc70 probed by tryptophan fluorescence peak wavelength shift.

REFERENCES

1. Caplow M, Fee L. Dissociation of the Tubulin Dimer Is Extremely Slow, Thermodynamically Very Unfavorable, and Reversible in the Absence of an Energy Source. *Mol Biol Cell*. **2002**;*13*: 2120–2131.

2. Lynch M, Hagner K. Evolutionary meandering of intermolecular interactions along the drift barrier. *Proc Natl Acad Sci*. **2015**;*112*: E30–E38.

3. Zhou H-X, Rivas G, Minton AP. Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences. Annual Review of Biophysics. Annual Reviews; 2008. pp. 375–397.

4. Sapir L, Harries D. Is the depletion force entropic? Molecular crowding beyond steric interactions. *Curr Opin Colloid Interface Sci*. **2015**;*20*: 3–10.

5. Cohen RD, Pielak GJ. A cell is more than the sum of its (dilute) parts: A brief history of quinary structure. *Protein Sci*. **2017**;*26*: 403–413.

6. Wirth AJ, Gruebele M. Quinary protein structure and the consequences of crowding in living cells: leaving the test-tube behind. *BioEssays*. **2013**;*35*: 984–993.

7. Chien P, Gierasch LM. Challenges and dreams: physics of weak interactions essential to life. *Mol Biol Cell*. American Society for Cell Biology; **2014**;*25*: 3474–7.

8. Gierasch LM, Gershenson A. Post-reductionist protein science, or putting Humpty Dumpty back together again. *Nat Chem Biol*. Nature Publishing Group; **2009**;*5*: 774–777.

9. ALLISON AC. Properties of sickle-cell haemoglobin. *Biochem J*. Portland Press Ltd; **1957**;*65*: 212–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/13403895

10. Mu X, Choi S, Lang L, Mowray D, Dokholyan N V., Danielsson J, et al. Physicochemical code for quinary protein interactions in Escherichia coli. *Proc Natl Acad Sci*. **2017**;*114*: E4556–E4563.

11. Zimmerman SB, Trach SO. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *J Mol Biol*. **1991**;*222*: 599–620.

12. Minton AP. Confinement as a determinant of macromolecular structure and reactivity. *Biophys J*. **1992**;*63*: 1090–1100.

13. Minton AP. Excluded volume as a determinant of macromolecular structure and reactivity. *Biopolymers*. Wiley-Blackwell; **1981**;*20*: 2093–2120.

14. Condon EU, Mack JE. An Interpretation of Pauli's Exclusion Principle. *Phys Rev*. American Physical Society; **1930**;*35*: 579–582.

15. Medalia O, Weber I, Frangakis AS, Nicastro D, Gerisch G, Baumeister W. Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. *Science (80-)*. **2002**;*298*: 1209–1213.

16. Saxton MJ. Anomalous diffusion due to binding: A Monte Carlo study. *Biophys J*. **1996**;*70*: 1250–1262.

17. Vöpel T, Makhatadze GI. Enzyme activity in the crowded Milieu. *PLoS One*. **2012**;*7*: 3–8.

18. Dhar A, Samiotakis A, Ebbinghaus S, Nienhaus L, Homouz D, Gruebele M, et al. Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding. *Proc Natl Acad Sci*. **2010**;*107*: 17586–17591.

19. Jiang M, Guo Z. Effect macromolecular crowding on the intrinsic cat eff of Enterobactin isochorismate synthase. *J Am Chem Soc*. **2007**;*129*: 730–731. Available: http://pubs3.acs.org/acs/journals/doilookup?in_doi=10.1021/ja065064+

20. Munishkina L, Cooper E, Uversky V, Fink A. The effect of macromolecular crowding on protein aggregation and amyloid fibril formation. *J Mol Recognit*. **2004**;*17*: 456–464. Available: http://doi.wiley.com/10.1002/jmr.699%5Cnpapers2://publication/uuid/0997C9C3-6AEE-4686-9FD1-955247ADF98A

21. del Alamo M, Rivas G, Mateu MG. Effect of macromolecular crowding agents on human immunodeficiency virus type 1 capsid protein assembly in vitro. *J Virol*. **2005**;*79*: 14271–81.

22. Zorrilla S, Rivas G, Acuña AU, Lillo MP. Protein self-association in crowded protein solutions: A time-resolved fluorescence polarization study. *Protein Sci*. **2008**;*13*: 2960–2969.

23. Díaz-López T, Dávila-Fajardo C, Blaesing F, Lillo MP, Giraldo R. Early Events in the Binding of the pPS10 Replication Protein RepA to Single Iteron and Operator DNA Sequences. *J Mol Biol*. **2006**;*364*: 909–920.

24. Davis CM, Gruebele M. Non-Steric Interactions Predict the Trend and Steric Interactions the Offset of Protein Stability in Cells. *ChemPhysChem*. **2018**;*19*: 2290–2294.

25. Smith AE, Zhou LZ, Gorensek AH, Senske M, Pielak GJ. In-cell thermodynamics and a new role for protein surfaces. *Proc Natl Acad Sci U S A*. National Academy of Sciences; **2016**;*113*: 1725–30.

26. Benton L a, Smith AE, Young GB, Pielak GJ. Unexpected Effects of Macromolecular Crowding on Protein Stability Unexpected Effects of Macromolecular Crowding on Protein Stability. *Biochemistry*. **2012**; 9773–9775.

27. Guzman I, Gelman H, Tai J, Gruebele M. The extracellular protein VlsE is destabilized inside cells. *J Mol Biol*. **2014**;*426*: 11–20.

28. Derham BK, Harding JJ. The effect of the presence of globular proteins and elongated polymers on enzyme activity. *Biochim Biophys Acta - Proteins Proteomics*. **2006**;*1764*: 1000–1006.

29. Guo M, Gelman H, Gruebele M. Coupled protein diffusion and folding in the cell. Levy YK, editor. *PLoS One*. Public Library of Science; **2014**;*9*: e113040.

30. Vaĭnshteĭn BK. Three-dimensional electron microscopy of biological macromolecules. *Sov Phys Uspekhi*. IOP Publishing; **1973**;*16*: 185–206.

31. Edelstein SJ. Patterns in the quinary structures of proteins. Plasticity and inequivalence of individual molecules in helical arrays of sickle cell hemoglobin and tubulin. *Biophys J*. **1980**;*32*: 347–60.

32. McConkey EH. Molecular evolution, intracellular organization, and the quinary structure of proteins. *Proc Natl Acad Sci U S A*. **1982**;*79*: 3236–3240.

33. Rivas G, Minton AP. Macromolecular Crowding In Vitro, In Vivo, and In Between [Internet]. Trends in Biochemical Sciences Elsevier; Nov 1, 2016 pp. 970–981.

34. Elcock AH. Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Curr Opin Struct Biol*. **2010**;*20*: 196–206.

35. Rivas G, Ferrone F, Herzfeld J. Life in a crowded world. EMBO Reports. 2004. pp. 23–27.

36. Hall D, Minton AP. Macromolecular crowding: Qualitative and semiquantitative successes, quantitative challenges. Biochimica et Biophysica Acta - Proteins and Proteomics. 2003. pp. 127–139.

37. Kuznetsova IM, Turoverov KK, Uversky VN. What macromolecular crowding can do to a protein. *Int J Mol Sci*. Multidisciplinary Digital Publishing Institute (MDPI); **2014**;*15*: 23090–140.

38. Christiansen A, Wang Q, Samiotakis A, Cheung MS, Wittung-Stafshede P. Factors Defining Effects of Macromolecular Crowding on Protein Stability: An in Vitro/in Silico Case Study Using Cytochrome *c*. *Biochemistry*. **2010**;*49*: 6519–6530.

39. Gnutt D, Gao M, Brylski O, Heyden M, Ebbinghaus S. Excluded-volume effects in living cells. *Angew Chemie - Int Ed*. **2015**;*54*: 2548–2551.

40. Lebowitz JL, Helfand E, Psaestgaakd E. Scaled particle theory of fluid mixtures. *J Chem Phys*. **1965**;*43*: 774–779.

41. Giddings JC, Kucera E, Russell CP, Myers MN. Statistical theory for the equilibrium distribution of rigid molecules in inert porous networks. Exclusion chromatography. *J Phys Chem*. **1968**;*72*: 4397–4408.

42. Zaccone A, Terentjev EM. Theory of molecular crowding in Brownian hard-sphere liquids. *Phys Rev E - Stat Nonlinear, Soft Matter Phys*. **2012**;*85*.

43. Widom B. Some Topics in the Theory of Fluids. *J Chem Phys*. American Institute of Physics; **1963**;*39*: 2808–2812.

44. Cheung MS, Klimov D, Thirumalai D. Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc Natl Acad Sci*. **2005**;*102*: 4753–8.

45. Minton AP. Effect of a concentrated “inert” macromolecular cosolute on the stability of a globular protein with respect to denaturation by heat and by chaotropes: A statistical-thermodynamic model. *Biophys J*. **2000**;*78*: 101–109.

46. Minton AP. Models for Excluded Volume Interaction between an Unfolded Protein and Rigid Macromolecular Cosolutes: Macromolecular Crowding and Protein Stability Revisited. *Biophys J*.

2005;*88*: 971–985.

47. Hayer-Hartl M, Minton AP. A simple semiempirical model for the effect of molecular confinement upon the rate of protein folding. *Biochemistry*. **2006**;*45*: 13356–13360.

48. Tubio G, Nerli B, Picó G. Relationship between the protein surface hydrophobicity and its partitioning behaviour in aqueous two-phase systems of polyethyleneglycol–dextran. *J Chromatogr B*. **2004**;*799*: 293–301.

49. Winzor DJ, Wills PR. Molecular crowding effects of linear polymers in protein solutions. *Biophys Chem*. **2006**;*119*: 186–195.

50. Bloustine J, Virmani T, Thurston GM, Fraden S. Light scattering and phase behavior of lysozyme-poly(ethylene glycol) mixtures. *Phys Rev Lett*. **2006**;*96*.

51. Chao SH, Matthews SS, Paxman R, Aksimentiev A, Gruebele M, Price JL. Two structural scenarios for protein stabilization by PEG. *J Phys Chem B*. American Chemical Society; **2014**;*118*: 8388–8395.

52. Delgado C, Francis GE, Fisher D. The uses and properties of PEG-linked proteins. *Crit Rev Ther Drug Carr Syst*. **1992**;*9*: 249–304.

53. Chao SH, Schäfer J, Gruebele M. The Surface of Protein λ6-85Can Act as a Template for Recurring Poly(ethylene glycol) Structure. *Biochemistry*. **2017**;*56*: 5671–5678.

54. Balbo J, Mereghetti P, Herten DP, Wade RC. The shape of protein crowders is a major determinant of protein diffusion. *Biophys J*. **2013**;*104*: 1576–1584.

55. Olsen SN. Applications of isothermal titration calorimetry to measure enzyme kinetics and activity in complex solutions. *Thermochim Acta*. **2006**;*448*: 12–18.

56. Denos S, Dhar A, Gruebele M. Crowding effects on the small, fast-folding protein λ6–85. *Faraday Discuss*. **2012**;*157*: 451–500.

57. Minton a P, Wilf J. Effect of macromolecular crowding upon the structure and function of an enzyme: glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry*. **1981**;*20*: 4821–4826.

58. Zhou BR, Liang Y, Du F, Zhou Z, Chen J. Mixed macromolecular crowding accelerates the oxidative refolding of reduced, denatured lysozyme: Implications for protein folding in intracellular environments. *J Biol Chem*. **2004**;*279*: 55109–55116.

59. Bai J, Liu M, Pielak GJ, Li C. Macromolecular and Small Molecular Crowding Have Similar Effects on α-Synuclein Structure. *ChemPhysChem*. **2017**;*18*: 55–58.

60. van den Berg B, Wain R, Dobson CM, Ellis RJ. Macromolecular crowding perturbs protein refolding kinetics: implications for folding inside the cell. *Embo J*. **2000**;*19*: 3870–3875.

61. Biswas S, Kundu J, Mukherjee SK, Chowdhury PK. Mixed Macromolecular Crowding: A Protein and Solvent Perspective. *ACS Omega*. **2018**;*3*: 4316–4330.

62. Goins AB, Sanabria H, Waxham MN. Macromolecular crowding and size effects on probe

microviscosity. *Biophys J*. **2008**;*95*: 5362–5373.

63. Banks DS, Fradin C. Anomalous diffusion of proteins due to molecular crowding. *Biophys J*. The Biophysical Society; **2005**;*89*: 2960–2971.

64. Schermann JP. Hole-Burning Spectroscopy. *Encycl Spectrosc Spectrom*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA ; **2016**;*40*: 126–130.

65. Volker S. Hole-Burning Spectroscopy. *Annu Rev Phys Chem*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA ; **1989**;*40*: 499–530.

66. Zhou HX. Protein folding in confined and crowded environments. Archives of Biochemistry and Biophysics. 2008. pp. 76–82.

67. Qin S, Zhou HX. Atomistic modeling of macromolecular crowding predicts modest increases in protein folding and binding stability. *Biophys J*. **2009**;*97*: 12–19.

68. Zhou HX. Effect of mixed macromolecular crowding agents on protein folding. *Proteins Struct Funct Genet*. **2008**;*72*: 1109–1113.

69. Zhou HX. Protein folding and binding in confined spaces and in crowded solutions. *J Mol Recognit*. **2004**;*17*: 368–375.

70. Batra J, Xu K, Zhou HX. Nonadditive effects of mixed crowding on protein stability. *Proteins Struct Funct Bioinforma*. **2009**;*77*: 133–138.

71. Batra J, Xu K, Qin S, Zhou HX. Effect of macromolecular crowding on protein binding stability: Modest stabilization and significant biological consequences. *Biophys J*. **2009**;*97*: 906–911.

72. Zhou H-X. Crowding Effects of Membrane Proteins. *J Phys Chem B*. **2009**;*113*: 7995–8005.

73. Guseman AJ, Perez Goncalves GM, Speer SL, Young GB, Pielak GJ. Protein shape modulates crowding effects. *Proc Natl Acad Sci*. National Academy of Sciences; **2018**;*115*: 10965–10970.

74. Chen E, Christiansen A, Wang Q, Cheung MS, Kliger DS, Wittung-Stafshede P. Effects of macromolecular crowding on burst phase kinetics of cytochrome c folding. *Biochemistry*. **2012**;*51*: 9836–9845.

75. Homouz D, Stagg L, Wittung-Stafshede P, Cheung MS. Macromolecular crowding modulates folding mechanism of α/β protein apoflavodoxin. *Biophys J*. **2009**;*96*: 671–680.

76. Tripathi S, Wang Q, Zhang P, Hoffman L, Waxham MN, Cheung MS. Conformational frustration in calmodulin-target recognition. *J Mol Recognit*. **2015**;*28*: 74–86.

77. Wang Q, Zhang P, Hoffman L, Tripathi S, Homouz D, Liu Y, et al. Protein recognition and selection through conformational and mutually induced fit. *Proc Natl Acad Sci*. **2013**;*110*: 20545–20550.

78. Wang Q, Cheung MS. A physics-based approach of coarse-graining the cytoplasm of Escherichia Coli (CGCYTO). *Biophys J*. **2012**;*102*: 2353–2361.

79. Specht EA, Braselmann E, Palmer AE. A Critical and Comparative Review of Fluorescent Tools for Live-Cell Imaging. *Annu Rev Physiol*. **2017**;*79*: 93–117.

80. Guigas G, Kalla C, Weiss M. The degree of macromolecular crowding in the cytoplasm and nucleoplasm of mammalian cells is conserved. *FEBS Lett*. **2007**;*581*: 5094–5098.

81. Boersma AJ, Zuhorn IS, Poolman B. A sensor for quantification of macromolecular crowding in living cells. *Nat Methods*. **2015**;*12*: 227–9, 1 p following 229.

82. Machiyama H, J. Morikawa T, Okamoto K, M. Watanabe T, Fujita H. The use of a genetically encoded molecular crowding sensor in various biological phenomena. *Biophys Physicobiology*. **2017**;*14*: 119–125.

83. Liu B, Åberg C, van Eerden FJ, Marrink SJ, Poolman B, Boersma AJ. Design and Properties of Genetically Encoded Probes for Sensing Macromolecular Crowding. *Biophys J*. **2017**;*112*: 1929–1939.

84. Morikawa TJ, Fujita H, Kitamura A, Horio T, Yamamoto J, Kinjo M, et al. Dependence of fluorescent protein brightness on protein concentration in solution and enhancement of it. *Sci Rep*. **2016**;*6*.

85. Sukenik S, Ren P, Gruebele M. Weak protein–protein interactions in live cells are quantified by cell-volume modulation. *Proc Natl Acad Sci*. **2017**;*114*: 6776–6781.

86. Davis FF, Abuchowski A, van Es T, Palczuk NC, Chen R, Savoca K, et al. Enzyme-Polyethylene Glycol Adducts: Modified Enzymes with Unique Properties. Enzyme Engineering. Boston, MA: Springer US; 1978. pp. 169–173.

87. Milton Harris J, Chess RB. Effect of pegylation on pharmaceuticals. Nature Reviews Drug Discovery. 2003. pp. 214–221.

88. Zhang Z, Chen X, Chen L, Yu S, Cao Y, He C, et al. Intracellular pH-sensitive PEG- Block - Acetalated-dextrans as efficient drug delivery platforms. *ACS Appl Mater Interfaces*. **2013**;*5*: 10760–10766.

89. Huang J, Wang L, Xiong C, Yuan F. Elastic hydrogel as a sensor for detection of mechanical stress generated by single cells grown in three-dimensional environment. *Biomaterials*. **2016**;*98*: 103–112.

90. Jin W, Brennan JD. Properties and applications of proteins encapsulated within sol-gel derived materials. Analytica Chimica Acta. 2002. pp. 1–36.

91. Kisley L, Landes CF. Molecular approaches to chromatography using single molecule spectroscopy. Analytical Chemistry. 2015. pp. 83–98.

92. Dong R, Jensen TW, Engberg K, Nuzzo RG, Leckband DE. Variably elastic hydrogel patterned via capillary action in microchannels. *Langmuir*. **2007**;*23*: 1483–1488.

93. Eggers DK, Valentine JS. Molecular confinement influences protein structure and enhances thermal protein stability. *Protein Sci*. Wiley-Blackwell; **2001**;*10*: 250–61.

94. Peterson ES, Leonard EF, Foulke JA, Oliff MC, Salisbury RD, Kim DY. Folding myoglobin within a sol-gel glass: Protein folding constrained to a small volume. *Biophys J*. **2008**;*95*: 322–332.

95. Kisley L, Serrano KA, Guin D, Kong X, Gruebele M, Leckband DE. Direct Imaging of Protein Stability and Folding Kinetics in Hydrogels. *ACS Appl Mater Interfaces*. **2017**;*9*: 21606–21617.

96. Kisley L, Serrano KA, Davis CM, Guin D, Murphy EA, Gruebele M, et al. Soluble Zwitterionic Poly(sulfobetaine) Destabilizes Proteins. *Biomacromolecules*. American Chemical Society; **2018**;*19*: 3894–3901.

97. Sarkar M, Smith AE, Pielak GJ. Impact of reconstituted cytosol on protein stability. *Proc Natl Acad Sci*. **2013**;*110*: 19342–19347.

98. Monteith WB, Cohen RD, Smith AE, Guzman-Cisneros E, Pielak GJ. Quinary structure modulates protein stability in cells. *Proc Natl Acad Sci U S A*. National Academy of Sciences; **2015**;*112*: 1739–42.

99. Wang Q, Zhuravleva A, Gierasch LM. Exploring weak, transient protein-protein interactions in crowded in vivo environments by in-cell nuclear magnetic resonance spectroscopy. *Biochemistry*. **2011**;*50*: 9225–9236.

100. Barbieri L, Luchinat E, Banci L. Protein interaction patterns in different cellular environments are revealed by in-cell NMR. *Sci Rep*. **2015**;*5*.

101. Monteith WB, Pielak GJ. Residue level quantification of protein stability in living cells. *Proc Natl Acad Sci*. **2014**;*111*: 11335–11340.

102. Danielsson J, Mu X, Lang L, Wang H, Binolfi A, Theillet F-X, et al. Thermodynamics of protein destabilization in live cells. *Proc Natl Acad Sci*. **2015**;*112*: 12402–12407.

103. Miklos AC, Sarkar M, Wang Y, Pielak GJ. Protein crowding tunes protein stability. *J Am Chem Soc*. **2011**;*133*: 7116–7120.

104. Sarkar M, Lu J, Pielak GJ. Protein Crowder Charge and Protein Stability. *Biochemistry*. American Chemical Society; **2014**;*53*: 1601–1606.

105. Wang Y, Sarkar M, Smith AE, Krois AS, Pielak GJ. Macromolecular crowding and protein stability. *J Am Chem Soc*. **2012**;*134*: 16614–16618.

106. Cleland JL, Powell MF, Shire SJ. The development of stable protein formulations: a close look at protein aggregation, deamidation, and oxidation. *Crit Rev Ther Drug Carrier Syst*. **1993**;*10*: 307–77.

107. Carpenter JF, Kendrick BS, Chang BS, Manning MC, Randolph TW. Inhibition of stress-induced aggregation of protein therapeutics. *Methods Enzymol*. **1999**;*309*: 236–255.

108. Macdougall IC. Antibody-mediated pure red cell aplasia (PRCA): Epidemiology, immunogenicity and risks. *Nephrol Dial Transplant*. **2005**;*20*: iv9-15.

109. Jiskoot W, Randolph TW, Volkin DB, Middaugh CR, Schöneich C, Winter G, et al. Protein

instability and immunogenicity: Roadblocks to clinical application of injectable protein delivery systems for sustained release. *J Pharm Sci*. **2012**;*101*: 946–954.

110. Moore W V., Leppert P. Role of aggregated human growth hormone (hGH) in development of antibodies to hGH. *J Clin Endocrinol Metab*. **1980**;*51*: 691–697.

111. Ranke MB, Wit JM. Growth hormone-past, present and future. *Nat Rev Endocrinol*. **2018**;*14*: 285–300.

112. St. John RJ, Carpenter JF, Balny C, Randolph TW. High pressure refolding of recombinant human growth hormone from insoluble aggregates: Structural transformations, kinetic barriers, and energetics. *J Biol Chem*. **2001**;*276*: 46856–46863.

113. Eckhardt BM, Oeswein JQ, Bewley TA. Effect of Freezing on Aggregation of Human Growth Hormone. Pharmaceutical Research: An Official Journal of the American Association of Pharmaceutical Scientists. 1991. pp. 1360–1364.

114. Gomez-Orellana I, Variano B, Miura-Fraboni J, Milstein S, Paton DR. Thermodynamic characterization of an intermediate state of human growth hormone. *Protein Sci*. **1998**;*7*: 1352–1358.

115. Johansson H, Jensen MR, Gesmar H, Meier S, Vinther JM, Keeler C, et al. Specific and nonspecific interactions in ultraweak protein-protein associations revealed by solvent paramagnetic relaxation enhancements. *J Am Chem Soc*. **2014**;*136*: 10277–10286.

116. Jensen MR, Kristensen SM, Keeler C, Christensen HEM, Hodsdon ME, Led JJ. Weak self-association of human growth hormone investigated by nitrogen-15 NMR relaxation. *Proteins Struct Funct Genet*. **2008**;*73*: 161–172.

117. Tang C, Ghirlando R, Clore GM. Visualization of transient ultra-weak protein self-association in solution using paramagnetic relaxation enhancement. *J Am Chem Soc*. **2008**;*130*: 4048–4056.

118. Dhar A, Ebbinghaus S, Shen Z, Mishra T, Gruebele M. The Diffusion Coefficient for PGK Folding in Eukaryotic Cells. *Biophys J*. **2010**;*99*: L69–L71.

119. Phillip Y, Kiss V, Schreiber G. Protein-binding dynamics imaged in a living cell. *Proc Natl Acad Sci*. **2012**;*109*: 1461–1466.

120. Margineanu A, Chan JJ, Kelly DJ, Warren SC, Flatters D, Kumar S, et al. Screening for protein-protein interactions using Förster resonance energy transfer (FRET) and fluorescence lifetime imaging microscopy (FLIM). *Sci Rep*. **2016**;*6*.

121. Dhar A, Girdhar K, Singh D, Gelman H, Ebbinghaus S, Gruebele M. Protein Stability and Folding Kinetics in the Nucleus and Endoplasmic Reticulum of Eucaryotic Cells. *Biophys J*. **2011**;*101*: 421–430.

122. Ebbinghaus S, Dhar A, McDonald JD, Gruebele M. Protein folding stability and dynamics imaged in a living cell. *Nat Methods*. **2010**;*7*: 319–323.

123. Willem Borst J, Visser AJWG. Fluorescence lifetime imaging microscopy in life sciences. Measurement Science and Technology. 2010.

124. Krieger JW, Singh AP, Bag N, Garbe CS, Saunders TE, Langowski J, et al. Imaging fluorescence (cross-) correlation spectroscopy in live cells and organisms. *Nat Protoc*. **2015**;*10*: 1948–74.

125. Seksek O, Biwersi J, Verkman AS. Translational diffusion of macromolecule-sized solutes in cytoplasm and nucleus. *J Cell Biol*. **1997**;*138*: 131–42.

126. Etoc F, Balloul E, Vicario C, Normanno D, Liße D, Sittner A, et al. Non-specific interactions govern cytosolic diffusion of nanosized objects in mammalian cells. *Nat Mater*. **2018**;*17*: 740–746.

127. Dave K, Gelman H, Thu CTH, Guin D, Gruebele M. The Effect of Fluorescent Protein Tags on Phosphoglycerate Kinase Stability Is Nonadditive. *J Phys Chem B*. American Chemical Society; **2016**;*120*: 2878–85.

128. Serber Z, Selenko P, Hänsel R, Reckel S, Löhr F, Ferrell JE, et al. Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. *Nat Protoc*. **2007**;*1*: 2701–2709.

129. Barbieri L, Luchinat E, Banci L. Characterization of proteins by in-cell NMR spectroscopy in cultured mammalian cells. *Nat Protoc*. **2016**;*11*: 1101–1111.

130. Luchinat E, Banci L. A unique tool for cellular structural biology: In-cell NMR. Journal of Biological Chemistry. 2016. pp. 3776–3784.

131. Schlesinger AP, Wang Y, Tadeo X, Millet O, Pielak GJ. Macromolecular crowding fails to fold a globular protein in cells. *J Am Chem Soc*. **2011**;*133*: 8082–8085.

132. Inomata K, Ohno A, Tochio H, Isogai S, Tenno T, Nakase I, et al. High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature*. **2009**;*458*: 106–109.

133. Bicout DJ, Field MJ. Stochastic Dynamics Simulations of Macromolecular Diffusion in a Model of the Cytoplasm of Escherichia coli . *J Phys Chem*. **2002**;*100*: 2489–2497.

134. Ridgway D, Broderick G, Lopez-Campistrous A, Ru'aini M, Winter P, Hamilton M, et al. Coarse-grained molecular simulation of diffusion and reaction kinetics in a crowded virtual cytoplasm. *Biophys J*. **2008**;*94*: 3748–59.

135. Konopka MC, Shkel IA, Cayley S, Record MT, Weisshaar JC. Crowding and confinement effects on protein diffusion in vivo. *J Bacteriol*. **2006**;*188*: 6115–6123.

136. Feig M, Harada R, Mori T, Yu I, Takahashi K, Sugita Y. Complete atomistic model of a bacterial cytoplasm for integrating physics, biochemistry, and systems biology. *J Mol Graph Model*. NIH Public Access; **2015**;*58*: 1–9.

137. Spitzer J, Poolman B. The Role of Biomacromolecular Crowding, Ionic Strength, and Physicochemical Gradients in the Complexities of Life's Emergence. *Microbiol Mol Biol Rev*. **2009**;*73*: 371–388.

138. Schwartz R, Ting CS, King J. Whole proteome pI values correlate with subcellular localizations of proteins for organisms within the three domains of life. *Genome Res*. **2001**;*11*: 703–709.

139. Kiraga J, Mackiewicz P, Mackiewicz D, Kowalczuk M, Biecek P, Polak N, et al. The relationships

between the isoelectric point and: Length of proteins, taxonomy and ecology of organisms. *BMC Genomics*. **2007**;*8*: 163.

140. Vaynberg J, Qin J. Weak protein-protein interactions as probed by NMR spectroscopy. Trends in Biotechnology. 2006. pp. 22–27.

141. Srere PA. The metabolon. Trends in Biochemical Sciences. 1985. pp. 109–110.

142. Srere PA. Complexes of Sequential Metabolic Enzymes. *Annu Rev Biochem*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA ; **1987**;*56*: 89–124.

143. Wu F, Minteer S. Krebs cycle metabolon: Structural evidence of substrate channeling revealed by cross-linking and mass spectrometry. *Angew Chemie - Int Ed*. **2015**;*54*: 1851–1854.

144. Bulutoglu B, Garcia KE, Wu F, Minteer SD, Banta S. Direct Evidence for Metabolon Formation and Substrate Channeling in Recombinant TCA Cycle Enzymes. *ACS Chem Biol*. **2016**;*11*: 2847–2853.

145. Guo M, Ignatov M, Musier-Forsyth K, Schimmel P, Yang X-L. Crystal structure of tetrameric form of human lysyl-tRNA synthetase: Implications for multisynthetase complex formation. *Proc Natl Acad Sci*. **2008**;*105*: 2331–2336.

146. Jang SR, Nelson JC, Bend EG, Rodríguez-Laureano L, Tueros FG, Cartagenova L, et al. Glycolytic Enzymes Localize to Synapses under Energy Stress to Support Synaptic Function. *Neuron*. **2016**;*90*: 278–291.

147. Zhang Y, Beard KFM, Swart C, Bergmann S, Krahnert I, Nikoloski Z, et al. Protein-protein interactions and metabolite channelling in the plant tricarboxylic acid cycle. *Nat Commun*. **2017**;*8*.

148. Fernie AR, Carrari F, Sweetlove LJ. Respiratory metabolism: Glycolysis, the TCA cycle and mitochondrial electron transport. Current Opinion in Plant Biology. 2004. pp. 254–261.

149. Vélot C, Mixon MB, Teige M, Srere PA. Model of a quinary structure between krebs TCA cycle enzymes: A model for the metabolon. *Biochemistry*. **1997**;*36*: 14271–14276.

150. Easterby JS. A generalized theory of the transition time for sequential enzyme reactions. *Biochem J*. **2015**;*199*: 155–161.

151. Ovádi J, Tompa P, Vértessy B, Orosz F, Keleti T, Welch GR. Transient-time analysis of substrate-channelling in interacting enzyme systems. *Biochem J*. **2015**;*257*: 187–190.

152. Elcock AH, Huber GA, Andrew McCammon J. Electrostatic channeling of substrates between enzyme active sites: Comparison of simulation and experiment. *Biochemistry*. **1997**;

153. Robinson JB, Inman L, Sumegi B, Srere PA. Further characterization of the Krebs tricarboxylic acid cycle metabolon. *J Biol Chem*. **1987**;*262*: 1786–1790.

154. BEECKMANS S, KANAREK L. Demonstration of Physical Interactions between Consecutive Enzymes of the Citric Acid Cycle and of the Aspartate-Malate Shuttle: A Study Involving Fumarase, Malate Dehydrogenase, Citrate Synthase and Aspartate Aminotransferase. *Eur J Biochem*. **1981**;*117*: 527–535.

155. Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, et al. The quantitative proteome of a human cell line. *Mol Syst Biol*. **2011**;*7*: 549.

156. Liu Y, Hickey DP, Guo JY, Earl E, Abdellaoui S, Milton RD, et al. Substrate Channeling in an Artificial Metabolon: A Molecular Dynamics Blueprint for an Experimental Peptide Bridge. *ACS Catal*. **2017**;*7*: 2486–2493.

157. Negrutskii BS, Deutscher MP. Channeling of aminoacyl-tRNA for protein synthesis in vivo. *Proc Natl Acad Sci*. **2006**;*88*: 4991–4995.

158. Hausmann CD, Ibba M. Aminoacyl-tRNA synthetase complexes: Molecular multitasking revealed. FEMS Microbiology Reviews. 2008. pp. 705–721.

159. Bhattacharyya S, Bershtein S, Yan J, Argun T, Gilson AI, Trauger SA, et al. Transient protein-protein interactions perturb E. coli metabolome and cause gene dosage toxicity. *Elife*. **2016**;*5*: 1–22.

160. Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: Organizers of cellular biochemistry. Nature Reviews Molecular Cell Biology. 2017. pp. 285–298.

161. Grousl T, Ivanov P, Frydlova I, Vasicova P, Janda F, Vojtova J, et al. Robust heat shock induces eIF2 -phosphorylation-independent assembly of stress granules containing eIF3 and 40S ribosomal subunits in budding yeast, Saccharomyces cerevisiae. *J Cell Sci*. **2009**;*122*: 2078–2088.

162. Buchan JR, Parker R. Eukaryotic Stress Granules: The Ins and Outs of Translation. Molecular Cell. 2009. pp. 932–941.

163. Hoyle NP, Castelli LM, Campbell SG, Holmes LEA, Ashe MP. Stress-dependent relocalization of translationally primed mRNPs to cytoplasmic granules that are kinetically and spatially distinct from P-bodies. *J Cell Biol*. **2007**;*179*: 65–74.

164. Li P, Banjade S, Cheng HC, Kim S, Chen B, Guo L, et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature*. **2012**;*483*: 336–340.

165. Protter DSWW, Rao BS, Van Treeck B, Lin Y, Mizoue L, Rosen MK, et al. Intrinsically Disordered Regions Can Contribute Promiscuous Interactions to RNP Granule Assembly. *Cell Rep*. NIH Public Access; **2018**;*22*: 1401–1412.

166. Riback JA, Katanski CD, Kear-Scott JL, Pilipenko E V., Rojek AE, Sosnick TR, et al. Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell*. **2017**;*168*: 1028-1040.e19.

167. Cherkasov V, Grousl T, Theer P, Vainshtein Y, Gläßer C, Mongis C, et al. Systemic control of protein synthesis through sequestration of translation and ribosome biogenesis factors during severe heat stress. *FEBS Lett*. **2015**;*589*: 3654–3664.

168. Wallace EWJ, Kear-Scott JL, Pilipenko E V., Schwartz MH, Laskowski PR, Rojek AE, et al. Reversible, Specific, Active Aggregates of Endogenous Proteins Assemble upon Heat Stress. *Cell*. **2015**;*162*: 1286–1298.

169. Tatomer DC, Terzo E, Curry KP, Salzler H, Sabath I, Zapotoczny G, et al. Concentrating pre-

mRNA processing factors in the histone locus body facilitates efficient histone mRNA biogenesis. *J Cell Biol*. **2016**;*213*: 557–570.

170. Strzelecka M, Trowitzsch S, Weber G, Lührmann R, Oates AC, Neugebauer KM. Coilin-dependent snRNP assembly is essential for zebrafish embryogenesis. *Nat Struct Mol Biol*. **2010**;*17*: 403–409.

171. Novotny I, Blazikova M, Stanek D, Herman P, Malinsky J. In vivo kinetics of U4/U6{middle dot}U5 tri-snRNP formation in Cajal bodies. *Mol Biol Cell*. **2011**;*22*: 513–523.

172. Read EB, Okamura HH, Drubin DG. Actin- and tubulin-dependent functions during Saccharomyces cerevisiae mating projection formation. *Mol Biol Cell*. **1992**;*3*: 429–444.

173. Goode BL, Wong JJ, Butty AC, Peter M, McCormack AL, Yates JR, et al. Coronin promotes the rapid assembly and cross-linking of actin filaments and may link the actin and microtubule cytoskeletons in yeast. *J Cell Biol*. **1999**;*144*: 83–98.

174. Gill SR, Schroer TA, Szilak I, Steuer ER, Sheetz MP, Cleveland DW. Dynactin, a conserved, ubiquitously expressed component of an activator of vesicle motility mediated by cytoplasmic dynein. *J Cell Biol*. **1991**;*115*: 1639–1650.

175. Waterman-Storer CM, Salmon ED. Actomyosin-based retrograde flow of microtubules in the lamella of migrating epithelial cells influences microtubule dynamic instability and turnover and is associated with microtubule breakage and treadmilling. *J Cell Biol*. **1997**;*139*: 417–434.

176. Gavin A-C, Bösche M, Krause R, Grandi P, Marzioch M, Bauer A, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature. **2002**;*415*: 141–147.

177. Blundell TL, Fernández-Recio J. Cell biology: Brief encounters bolster contacts. Nature. 2006. pp. 279–280.

178. Tang C, Iwahara J, Clore GM. Visualization of transient encounter complexes in protein-protein association. *Nature*. **2006**;*444*: 383–386.

179. Schlecht R, Erbse AH, Bukau B, Mayer MP. Mechanics of Hsp70 chaperones enables differential interaction with client proteins. *Nat Struct Mol Biol*. **2011**;*18*: 345–351.

180. Mashaghi A, Bezrukavnikov S, Minde DP, Wentink AS, Kityk R, Zachmann-Brand B, et al. Alternative modes of client binding enable functional plasticity of Hsp70. *Nature*. Nature Research; **2016**;*539*: 448–451.

181. Slepenkov S V., Witt SN. Peptide-induced conformational changes in the molecular chaperone DnaK. *Biochemistry*. American Chemical Society ; **1998**;*37*: 16749–16756.

182. Iwahara J, Clore GM. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature*. **2006**;*440*: 1227–1230.

183. Ubbink M. The courtship of proteins: Understanding the encounter complex. FEBS Letters. 2009. pp. 1060–1066.

184. Schreiber G, Haran G, Zhou H-X. Fundamental Aspects of Protein−Protein Association Kinetics.

Chem Rev. **2009**;*109*: 839–860.

185. Fawzi NL, Doucleff M, Suh J-Y, Clore GM. Mechanistic details of a protein-protein association pathway revealed by paramagnetic relaxation enhancement titration measurements. *Proc Natl Acad Sci*. **2010**;*107*: 1379–1384.

186. Misteli T. Protein dynamics: For nuclear architecture and gene expression. *Science (80-)*. **2001**;*291*: 843–847.

187. Iwahara J, Clore GM. Direct observation of enhanced translocation of a homeodomain between DNA cognate sites by NMR exchange spectroscopy. *J Am Chem Soc*. **2006**;*128*: 404–405.

188. Hansen DF, Hass MAS, Christensen HM, Ulstrup J, Led JJ. Detection of short-lived transient protein-protein interactions by intermolecular nuclear paramagnetic relaxation: Plastocyanin from Anabaena variabilis. *J Am Chem Soc*. **2003**;*125*: 6858–6859.

189. Volkov AN, Worrall JAR, Holtzmann E, Ubbink M. Solution structure and dynamics of the complex between cytochrome c and cytochrome c peroxidase determined by paramagnetic NMR. *Proc Natl Acad Sci*. **2006**;*103*: 18945–18950.

190. Bashir Q, Scanu S, Ubbink M. Dynamics in electron transfer protein complexes. *FEBS J*. **2011**;*278*: 1391–1400.

191. Demott CM, Majumder S, Burz DS, Reverdatto S, Shekhtman A. Ribosome Mediated Quinary Interactions Modulate In-Cell Protein Activities. *Biochemistry*. **2017**;*56*: 4117–4126.

192. Davidi D, Noor E, Liebermeister W, Bar-Even A, Flamholz A, Tummler K, et al. Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k cat measurements . *Proc Natl Acad Sci*. **2016**;*113*: 3401–3406.

193. Breindel L, DeMott C, Burz DS, Shekhtman A. Real-Time In-Cell Nuclear Magnetic Resonance: Ribosome-Targeted Antibiotics Modulate Quinary Protein Interactions. *Biochemistry*. **2018**;*57*: 540–546.

194. Selenko P, Serber Z, Gadea B, Ruderman J, Wagner G. Quantitative NMR analysis of the protein G B1 domain in Xenopus laevis egg extracts and intact oocytes. *Proc Natl Acad Sci*. **2006**;*103*: 11904–11909.

195. Bertrand K, Reverdatto S, Burz DS, Zitomer R, Shekhtman A. Structure of proteins in eukaryotic compartments. *J Am Chem Soc*. **2012**;*134*: 12798–12806.

196. Sakai T, Tochio H, Inomata K, Sasaki Y, Tenno T, Tanaka T, et al. Fluoroscopic assessment of protein leakage during Xenopus oocytes in-cell NMR experiment by co-injected EGFP. *Anal Biochem*. **2007**;*371*: 247–249.

197. Simsek D, Tiu GC, Flynn RA, Byeon GW, Leppek K, Xu AF, et al. The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. *Cell*. **2017**;*169*: 1051-1065.e18.

198. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, et al. Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell*. **2012**;*149*: 1393–1406.

199. Majumder S, DeMott CM, Reverdatto S, Burz DS, Shekhtman A. Total Cellular RNA Modulates Protein Activity. *Biochemistry*. **2016**;*55*: 4568–4573.

200. Narayanaswamy R, Levy M, Tsechansky M, Stovall GM, O'Connell JD, Mirrielees J, et al. Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. *Proc Natl Acad Sci*. **2009**;*106*: 10147–10152.

201. Gancedo JM. Yeast carbon catabolite repression. *Microbiol Mol Biol Rev*. **1998**;*62*: 334–61. Available: http://www.ncbi.nlm.nih.gov/pubmed/9618445%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC98918

202. Thongwichian R, Selenko P. In-Cell NMR in Xenopus laevis Oocytes. Humana Press, Totowa, NJ; 2012. pp. 33–41.

203. Sakai T, Tochio H, Tenno T, Ito Y, Kokubo T, Hiroaki H, et al. In-cell NMR spectroscopy of proteins inside Xenopus laevis oocytes. *J Biomol NMR*. **2006**;*36*: 179–188.

204. Hancock R. Internal organisation of the nucleus: Assembly of compartments by macromolecular crowding and the nuclear matrix model. Biology of the Cell. 2004. pp. 595–601.

205. Spitzer J, Pielak GJ, Poolman B. Emergence of life: Physical chemistry changes the paradigm. Biology Direct. 2015.

206. Adkar B V., Manhart M, Bhattacharyya S, Tian J, Musharbash M, Shakhnovich EI. Optimization of lag phase shapes the evolution of a bacterial enzyme. *Nat Ecol Evol*. **2017**;*1*.

207. Bershtein S, Mu W, Serohijos AWR, Zhou J, Shakhnovich EI. Protein Quality Control Acts on Folding Intermediates to Shape the Effects of Mutations on Organismal Fitness. *Mol Cell*. **2013**;*49*: 133–144.

208. Dickerson RE, Takano T, Eisenberg D, Kallai OB, Samson L, Cooper A, et al. Ferricytochrome c A General Features of the Horse and Bonito Proteins at 2.8 Å Resolution. *J Biol Chem*. **1971**;*246*: 1511–1535.

209. Dickerson RE. The structure of cytochrome c and the rates of molecular evolution. *J Mol Evol*. **1971**;*1*: 26–45.

210. Çetinbaş M, Shakhnovich EI. Catalysis of Protein Folding by Chaperones Accelerates Evolutionary Dynamics in Adapting Cell Populations. *PLoS Comput Biol*. **2013**;*9*: e1003269.

211. Verma AK, Diwan D, Raut S, Dobriyal N, Brown RE, Gowda V, et al. Evolutionary Conservation and Emerging Functional Diversity of the Cytosolic Hsp70:J Protein Chaperone Network of Arabidopsis thaliana. *G3-Genes Genomes Genet*. **2017**;*7*: 1941–1954.

212. Sørensen JG, Kristensen TN, Loeschcke V. The evolutionary and ecological role of heat shock proteins. Ecology Letters. 2003. pp. 1025–1037.

213. Dixit PD, Maslov S. Evolutionary Capacitance and Control of Protein Stability in Protein-Protein Interaction Networks. *PLoS Comput Biol*. **2013**;*9*: e1003023.

214. Schlesinger MJ. Heat-Shock Proteins. *J Biol Chem*. **1990**;*265*: 12111–12114.

215. Craig EA, Schlesinger MJ. The heat shock response. *Crit Rev Biochem Mol Biol*. **1985**;*18*: 239–280.

216. Queitsch C, Sangstert TA, Lindquist S. Hsp90 as a capacitor of phenotypic variation. *Nature*. **2002**;*417*: 618–624.

217. Rutherford SL, Lindquist S. Hsp90 as a capacitor for morphological evolution. *Nature*. **1998**;*396*: 336–342.

218. Rutherford S, Hirate Y, Swalla BJ. The Hsp90 capacitor, developmental remodeling, and evolution: The robustness of gene networks and the curious evolvability of metamorphosis. Critical Reviews in Biochemistry and Molecular Biology. 2007. pp. 355–372.

219. Wade MJ, Johnson NA, Jones R, Siguel V, McNaughton M. Genetic variation segregating in natural populations of Tribolium castaneum affecting traits observed in hybrids with T. freemani. *Genetics*. **1997**;*147*: 1235–1247.

220. Lebesi DM, Tzia C. Effect of the Addition of Different Dietary Fiber and Edible Cereal Bran Sources on the Baking and Sensory Characteristics of Cupcakes. *Food Bioprocess Technol*. **2011**;*4*: 710–722.

221. Chow K-C. Hsp70 (DnaK) – an evolution facilitator? Trends in Genetics. 2000. pp. 484–485.

222. Gaucher EA, Govindarajan S, Ganesh OK. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature*. **2008**;*451*: 704–707.

223. Gouy M, Chaussidon M. Evolutionary biology: Ancient bacteria liked it hot. Nature. 2008. pp. 635–636.

224. Sleep NH. The Hadean-Archaean environment. Cold Spring Harbor perspectives in biology. 2010.

225. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci*. **2005**;*102*: 14338–14343.

226. Zeldovich KB, Chen P, Shakhnovich EI. Protein stability imposes limits on organism complexity and speed of molecular evolution. *Proc Natl Acad Sci*. **2007**;*104*: 16152–16157.

227. Monsellier E, Chiti F. Prevention of amyloid-like aggregation as a driving force of protein evolution. EMBO Reports. 2007. pp. 737–742.

228. Patki AU, Hausrath AC, Cordes MHJ. High Polar Content of Long Buried Blocks of Sequence in Protein Domains Suggests Selection Against Amyloidogenic Non-polar Sequences. *J Mol Biol*. **2006**;*362*: 800–809.

229. Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M. Life on the edge: a link between gene expression levels and aggregation rates of human proteins. Trends in Biochemical Sciences. 2007. pp. 204–206.

230. Troadec JP, Gervois A, Oger L. Statistics of Voronoi cells of slightly perturbed face-centered

cubic and hexagonal close-packed lattices. *Europhys Lett*. **1998**;*42*: 167–172.

231. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. Evolutionary rate in the protein interaction network. *Science*. American Association for the Advancement of Science; **2002**;*296*: 750–2.

232. Givens GD, Seidemann MF. Middle ear measurements in a difficult to test mentally retarded population. [Internet]. Mental retardation. 1977.

233. Johnson ME, Hummer G. Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks. *Proc Natl Acad Sci*. National Academy of Sciences; **2011**;*108*: 603–608.

234. Zhang J, Maslov S, Shakhnovich EI. Constraints imposed by non-functional protein-protein interactions on gene expression and proteome size. *Mol Syst Biol*. **2008**;*4*.

235. Claverie JM. Gene number: What if there are only 30,000 human genes? Science. 2001. pp. 1255–1257.

236. Maslov S, Sneppen K. Specificity and stability in topology of protein networks. *Science (80-)*. **2002**;*296*: 910–913.

237. Heo M, Maslov S, Shakhnovich E. Topology of protein interaction network shapes protein abundances and strengths of their functional and nonspecific interactions. *Proc Natl Acad Sci*. **2011**;*108*: 4258–4263.

238. Levy ED, De S, Teichmann SA. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proc Natl Acad Sci*. **2012**;*109*: 20461–20466.

239. Keskin O, Gursoy A, Ma B, Nussinov R. Principles of Protein−Protein Interactions: What are the Preferred Ways For Proteins To Interact? Chemical Reviews. 2008. pp. 1225–1244.

240. Janin J, Bahadur RP, Chakrabarti P. Protein-protein interaction and quaternary structure. Quarterly Reviews of Biophysics. 2008. pp. 133–180.

241. Ishihama Y, Schmidt T, Rappsilber J, Mann M, Harlt FU, Kerner MJ, et al. Protein abundance profiling of the Escherichia coli cytosol. *BMC Genomics*. **2008**;*9*.

242. Theillet F-X, Binolfi A, Frembgen-Kesner T, Hingorani K, Sarkar M, Kyne C, et al. Physicochemical Properties of Cells and Their Effects on Intrinsically Disordered Proteins (IDPs). Chemical Reviews. 2014. pp. 6661–6714.

243. Yang J-R, Liao B-Y, Zhuang S-M, Zhang J. Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. *Proc Natl Acad Sci*. **2012**;*109*: E831–E840.

244. Zhang J, Yang JR. Determinants of the rate of protein sequence evolution. *Nat Rev Genet*. **2015**;*16*: 409–420.

245. Forsdyke DR. Functional Constraint and Molecular Evolution. *eLS*. **2005**;

246. Harms MJ, Thornton JW. Evolutionary biochemistry: Revealing the historical and physical causes

of protein properties. Nature Reviews Genetics. 2013. pp. 559–571.

247. Mirny L, Shakhnovich E. Evolutionary conservation of the folding nucleus. *J Mol Biol*. **2001**;*308*: 123–129.

248. Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS. The Stability Effects of Protein Mutations Appear to be Universally Distributed. *J Mol Biol*. **2007**;*369*: 1318–1332.

249. Pazos F, Helmer-Citterich M, Ausiello G, Valencia A. Correlated mutations contain information about protein - protein interaction. *J Mol Biol*. **1997**;*271*: 511–523.

250. Ferńndez A, Lynch M. Non-adaptive origins of interactome complexity. *Nature*. **2011**;*474*: 502–506.

251. Lynch M. The evolution of multimeric protein assemblages. *Mol Biol Evol*. **2012**;*29*: 1353–1366.

252. Khersonsky O, Roodveldt C, Tawfik DS. Enzyme promiscuity: evolutionary and mechanistic aspects. Current Opinion in Chemical Biology. 2006. pp. 498–508.

253. Glasner ME, Gerlt J a, Babbitt PC. Mechanisms of protein evolution and their application to protein engineering. *Adv Enzym Relat Areas Mol Biol*. **2007**;*360*: 228–250.

254. Serohijos AWR, Rimas Z, Shakhnovich EI. Protein Biophysics Explains Why Highly Abundant Proteins Evolve Slowly. *Cell Rep*. Cell Press; **2012**;*2*: 249–256.

255. Wylie CS, Shakhnovich EI. A biophysical protein folding model accounts for most mutational fitness effects in viruses. *Proc Natl Acad Sci*. **2011**;*108*: 9916–9921.

256. Serohijos AWR, Shakhnovich EI. Merging molecular mechanism and evolution: Theory and computation at the interface of biophysics and evolutionary population genetics. Current Opinion in Structural Biology. 2014. pp. 84–91.

257. Bershtein S, Serohijos AW, Shakhnovich EI. Bridging the physical scales in evolutionary biology: from protein sequence space to fitness of organisms and populations. Current Opinion in Structural Biology. 2017. pp. 31–40.

258. Mackay TFC. Epistasis and quantitative traits: Using model organisms to study gene-gene interactions. *Nat Rev Genet*. **2014**;*15*: 22–33.

259. Joy JB, Liang RH, McCloskey RM, Nguyen T, Poon AFY. Ancestral Reconstruction. *PLOS Comput Biol*. **2016**;*12*: e1004763.

260. Bateson W, Mendel G. Mendel's principles of heredity, by W. Bateson. [Internet]. Mendel's principles of heredity, by W. Bateson. Cambridge [Eng.]: University Press,; 2011.

261. Sailer ZR, Harms MJ. High-order epistasis shapes evolutionary trajectories. *PLoS Comput Biol*. **2017**;*13*: e1005541.

262. Shah P, McCandlish DM, Plotkin JB. Historical contingency and entrenchment in protein evolution under purifying selection. *Proc Natl Acad Sci U S A*. National Academy of Sciences; **2014**;*112*: E3226-35.

263. Blount ZD, Borland CZ, Lenski RE. Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. *Proc Natl Acad Sci*. **2008**;*105*: 7899–7906.

264. Weinreich DM, Lan Y, Wylie CS, Heckendorn RB. Should evolutionary geneticists worry about higher-order epistasis? Current Opinion in Genetics and Development. 2013. pp. 700–707.

265. Sailer ZR, Harms MJ. Molecular ensembles make evolution unpredictable. *Proc Natl Acad Sci*. **2017**;*114*: 11938–11943.

266. Sailer ZR, Harms MJ. Detecting high-order epistasis in nonlinear genotype-phenotype maps. *Genetics*. **2017**;*205*: 1079–1088.

267. Pollock DD, Thiltgen G, Goldstein RA. Amino acid coevolution induces an evolutionary Stokes shift. *Proc Natl Acad Sci*. **2012**;*109*: E1352–E1359.

268. Harms MJ, Thornton JW. Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature*. **2014**;*512*: 203–207.

269. Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science (80-)*. **2014**;*344*: 1519–1522.

270. Sanger F, Thompson EOP, Kitai R. The amide groups of insulin. *Biochem J*. Portland Press Limited; **2015**;*59*: 509–518.

271. Merkl R, Sterner R. Ancestral protein reconstruction: Techniques and applications. Biological Chemistry. 2016. pp. 1–21.

272. Hall BG. Simple and accurate estimation of ancestral protein sequences. *Proc Natl Acad Sci*. **2006**;*103*: 5431–5436.

273. Zhang J, Rosenberg HF. Complementary advantageous substitutions in the evolution of an antiviral RNase of higher primates. *Proc Natl Acad Sci U S A*. **2002**;*99*: 5486–5491.

274. Malcolm BA, Wilson KP, Matthews BW, Kirsch JF, Wilson AC. Ancestral lysozymes reconstructed, neutrality tested, and thermostability linked to hydrocarbon packing. *Nature*. **1990**;*345*: 86–89.

275. Chandrasekharan UM, Sanker S, Glynias MJ, Karnik SS, Husain A. Angiotensin II-forming activity in a reconstructed ancestral chymase. *Science (80-)*. **1996**;*271*: 502–505.

276. Thomson JM, Gaucher EA, Burgan MF, De Kee DW, Li T, Aris JP, et al. Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat Genet*. **2005**;*37*: 630–635.

277. Pretorius IS. Tailoring wine yeast for the new millennium: Novel approaches to the ancient art of winemaking. Yeast. 2000. pp. 675–729.

278. Merkl R, Sterner R. Reconstruction of ancestral enzymes. *Perspect Sci*. **2016**;*9*: 17–23.

279. Lim SA, Hart KM, Harms MJ, Marqusee S. Evolutionary trend toward kinetic stability in the folding trajectory of RNases H. *Proc Natl Acad Sci*. **2016**;*113*: 13045–13050.

280. Lim SA, Bolin ER, Marqusee S. Tracing a protein's folding pathway over evolutionary time using ancestral sequence reconstruction and hydrogen exchange. *Elife*. **2018**;*7*: e38369.

281. Lim SA, Marqusee S. The burst-phase folding intermediate of ribonuclease H changes conformation over evolutionary history. *Biopolymers*. **2018**;*109*: e23086.

282. Copley SD. An evolutionary biochemist's perspective on promiscuity. Trends in Biochemical Sciences. 2015. pp. 72–78.

283. Nam H, Lewis NE, Lerman JA, Lee DH, Chang RL, Kim D, et al. Network context and selection in the evolution to enzyme specificity. *Science (80-)*. **2012**;*337*: 1101–1104.

284. Almaas E, Kovács B, Vicsek T, Oltvai ZN, Barabási AL. Global organization of metabolic fluxes in the bacterium Escherichia coli. *Nature*. **2004**;*427*: 839–843.

285. Cecchini DJ, Guan KL, Giese RW. Staphylococcal nuclease high-performance liquid chromatographic characterization of diaminooctane-modified DNA and its biotin and fluorescein derivatives. *J Chromatogr A*. **1988**;*444*: 97–106.

286. Makhatadze GI. Thermodynamics of Protein Interactions with Urea and Guanidinium Hydrochloride. *J Phys Chem B*. American Chemical Society; **1999**;*103*: 4781–4785.

287. Courtenay ES, Capp MW, Saecker RM, Record MT. Thermodynamic analysis of interactions between denaturants and protein surface exposed on unfolding: interpretation of urea and guanidinium chloride m-values and their correlation with changes in accessible surface area (ASA) using preferential interactio. *Proteins*. **2000**;*Suppl 4*: 72–85.

288. Johnson Jr. WC. Protein secondary structure and circular dichroism: a practical guide. *Proteins Struct Funct Genet*. **1990**;*7*: 205–214.

289. Surewicz WK, Mantsch HH, Chapman D. Determination of protein secondary structure by Fourier transform infrared spectroscopy: A critical assessment. *Biochemistry*. **1993**;*32*: 389–394.

290. Otzen D. Protein–surfactant interactions: A tale of many states. *Biochim Biophys Acta - Proteins Proteomics*. **2011**;*1814*: 562–591.

291. Otzen DE. Proteins in a brave new surfactant world. *Curr Opin Colloid Interface Sci*. **2015**;*20*: 161–169.

292. Voelz VA, Singh VR, Wedemeyer WJ, Lapidus LJ, Pande VS. Unfolded-state dynamics and structure of protein L characterized by simulation and experiment. *J Am Chem Soc*. **2010**;*132*: 4702–9.

293. Religa TL, Markson JS, Mayor U, Freund SM V, Fersht a R. Solution structure of a protein denatured state and folding intermediate. *Nature*. **2005**;*437*: 1053–1056.

294. Mitchinson C, Pain RH, Vinson JR, Walker T. The relative effectiveness of guanidinium and some biguanide salts as denaturants. Assessment against penicillinase. *Biochim Biophys Acta*. **1983**;*743*: 31–36.

295. Gelman H, Perlova T, Gruebele M. Dodine as a Protein Denaturant: The Best of Two Worlds? *J*

Phys Chem B. **2013**;*117*: 13090–13097.

296. Missiakas D, Betton JM, Chaffotte A, Minard P, Yon JM. Kinetic studies of the refolding of yeast phosphoglycerate kinase: comparison with the isolated engineered domains. *Protein Sci*. **1992**;*1*: 1485–1493.

297. Sherman MA, Beechem JM, Mas MT. Probing Intradomain and Interdomain Conformational Changes during Equilibrium Unfolding of Phosphoglycerate Kinase: Fluorescence and Circular Dichroism Study of Tryptophan Mutants. *Biochemistry*. **1995**;*34*: 13934–13942.

298. Missiakas D, Betton JM, Minard P, Yon JM. Unfolding-refolding of the domains in yeast phosphoglycerate kinase: comparison with the isolated engineered domains. *Biochemistry*. **1990**;*29*: 8683–9.

299. Damaschun G, Damaschun H, Fabian H, Gast K, Kröber R, Wieske M, et al. Conversion of yeast phosphoglycerate kinase into amyloid-like structure. *Proteins Struct Funct Genet*. **2000**;*39*: 204–211.

300. Fairbrother WJ, Walker PA, Minard P, Littlechild JA, Watson HC, William RJP. NMR analysis of site-specific mutants of yeast phosphoglycerate kinase. An investigation of the triose-binding site. *Eur J Biochem*. **1989**;*183*: 57–67.

301. Rosenkranz T, Schlesinger R, Gabba M, Fitter J. Native and Unfolded States of Phosphoglycerate Kinase Studied by Single-Molecule FRET. *ChemPhysChem*. **2011**;*12*: 704–710.

302. Lillo MP, Szpikowska BK, Mas MT, Sutin JD, Beechem JM. Real-time measurement of multiple intramolecular distances during protein folding reactions: a multisite stopped-flow fluorescence energy-transfer study of yeast phosphoglycerate kinase. *Biochemistry*. **1997**;*36*: 11273–11281.

303. Szpikowska BK, Mas MT. Urea-Induced Equilibrium Unfolding of Single Tryptophan Mutants of Yeast Phosphoglycerate Kinase: Evidence for a Stable Intermediate. *Arch Biochem Biophys*. **1996**;*335*: 173–182.

304. Osváth S, Gruebele M. Proline can have opposite effects on fast and slow protein folding phases. *Biophys J*. **2003**;*85*: 1215–1222.

305. Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc*. **2006**;*1*: 2876–90.

306. Nishii I, Kataoka M, Tokunaga F, Goto Y. Cold denaturation of the molten globule states of apomyoglobin and a profile for protein folding. *Biochemistry*. **1994**;*33*: 4903–9.

307. Tanford C. Protein Denaturation. Part C. Theoretical Models for The Mechanism of Denaturation. *Adv Protein Chem*. **1970**;*24*: 1–95.

308. Hu CQ, Sturtevant JM. Thermodynamic study of yeast phosphoglycerate kinase. *Biochemistry*. **1987**;*26*: 178–182.

309. Szpikowska BK, Beechem JM, Sherman M a, Mas MT. Equilibrium unfolding of yeast phosphoglycerate kinase and its mutants lacking one or both native tryptophans: a circular dichroism and steady-state and time-resolved fluorescence study. *Biochemistry*. **1994**;*33*: 2217–

25. Available: http://www.ncbi.nlm.nih.gov/pubmed/8117679

310. Chardot T, Mitraki A, Amigues Y, Desmadril M, Betton JM, Yon JM. The effect of phosphate on the unfolding-refolding of phosphoglycerate kinase induced by guanidine hydrochloride. *FEBS Lett*. **1988**;*228*: 65–68.

311. Osváth S, Sabelko JJ, Gruebele M. Tuning the Heterogeneous Early Folding Dynamics of Phosphoglycerate Kinase. *J Mol Biol*. **2003**;*333*: 187–199.

312. Garcia P, Desmadril M, Minard P, Yon JM. Evidence for residual structures in an unfolded form of yeast phosphoglycerate kinase. *Biochemistry*. **1995**;*34*: 397–404. Available: http://www.ncbi.nlm.nih.gov/pubmed/7819231

313. Dempsey CE, Piggot TJ, Mason PE. Dissecting contributions to the denaturant sensitivities of proteins. *Biochemistry*. **2005**;*44*: 775–781.

314. Lim WK, Rösgen J, Englander SW. Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. *Proc Natl Acad Sci U S A*. **2009**;*106*: 2595–2600.

315. Lad MD, Ledger VM, Briggs B, Green RJ, Frazier RA. Analysis of the SDS-lysozyme binding isotherm. *Langmuir*. **2003**;*19*: 5098–5103.

316. Malmendal A, Underhaug J, Otzen DE, Nielsen NC. Fast mapping of global protein folding states by multivariate NMR: A GPS for proteins. *PLoS One*. **2010**;*5*.

317. England JL, Haran G. Role of solvation effects in protein denaturation: from thermodynamics to single molecules and back. *Annu Rev Phys Chem*. **2011**;*62*: 257–77.

318. O'Brien EP, Dima RI, Brooks B, Thirumalai D. Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: Lessons for protein denaturation mechanism. *J Am Chem Soc*. **2007**;*129*: 7346–7353.

319. Otzen DE. Protein unfolding in detergents: effect of micelle structure, ionic strength, pH, and temperature. *Biophys J*. **2002**;*83*: 2219–2230.

320. Ervin J, Larios E, Osváth S, Schulten K, Gruebele M. What causes hyperfluorescence: folding intermediates or conformationally flexible native states? *Biophys J*. Cell Press; **2002**;*83*: 473–83.

321. Beechem JM, Sherman M a, Mas MT. Sequential domain unfolding in phosphoglycerate kinase: fluorescence intensity and anisotropy stopped-flow kinetics of several tryptophan mutants. *Biochemistry*. **1995**;*34*: 13943–13948.

322. Dodine. *World Heal Organ - Pestic Residues Ser*. **1975**;

323. Dodine - A (Fairly) New Fungicide. *Int J Pest Manag Part C*. Taylor & Francis Group ; **1961**;*7*: 128–134.

324. Guin D, Sye K, Dave K, Gruebele M. Dodine as a transparent protein denaturant for circular dichroism and infrared studies. *Protein Sci*. Wiley-Blackwell; **2016**;*25*: 1061–1068.

325. Monera OD, Kay CM, Hodges RS. Protein denaturation with guanidine hydrochloride or urea

provides a different estimate of stability depending on the contributions of electrostatic interactions. *Protein Sci*. **1994**;*3*: 1984–1991.

326. Carvalho FAO, Alves FR, Carvalho JWP, Tabak M. Guanidine hydrochloride and urea effects upon thermal stability of Glossoscolex paulistus hemoglobin (HbGp). *Int J Biol Macromol*. **2015**;*74*: 18–28.

327. Huerta-Viga A, Woutersen S. Protein denaturation with guanidinium: A 2D-IR study. *J Phys Chem Lett*. **2013**;*4*: 3397–3401.

328. Johansson MU, De Château M, Wikström M, Forsén S, Drakenberg T, Björck L. Solution structure of the albumin-binding GA module: A versatile bacterial protein domain. *J Mol Biol*. **1997**;*266*: 859–865.

329. Royer CA. Probing Protein Folding and Conformational Transitions with Fluorescence. Chemical Reviews. 2006. pp. 1769–1784.

330. Kelly S, Price N. The Use of Circular Dichroism in the Investigation of Protein Structure and Function. *Curr Protein Pept Sci*. **2005**;*1*: 349–384.

331. Chen RF. Fluorescence quantum yields of tryptophan and tyrosine. *Anal Lett*. **1967**;*1*: 35–42.

332. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. *Science (80-)*. **2011**;*334*: 517–520.

333. Schrödinger. The PyMOL Molecular Graphics System. Schrödinger LLC wwwpymolorg. 2015.

334. Case DA, Darden T, Iii TEC, Simmerling C, Brook S, Roitberg A, et al. Amber 14. *Univ California, San Fr*. **2015**;

335. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. **1983**;*79*: 926.

336. PubChem Identifier: CID 17110 [Internet]. [cited 4 Nov 2018]. Available: https://pubchem.ncbi.nlm.nih.gov/compound/17110

337. Sousa Da Silva AW, Vranken WF. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res Notes*. **2012**;

338. Borgohain G, Mandal B, Paul S. Molecular dynamics approach to understand the denaturing effect of a millimolar concentration of dodine on a λ-repressor and counteraction by trehalose. *Phys Chem Chem Phys*. **2017**;

339. Martinez L, Andrade R, Birgin EG, Martínez JM. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J Comput Chem*. **2009**;

340. Berendsen HJCC, Postma JPMM, van Gunsteren WF, Dinola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys*. **1984**;*81*: 3684–3690.

341. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys*. **1993**;*98*: 10089–10092.

342. Kräutler V, Van Gunsteren WF, Hünenberger PH. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J Comput Chem*. **2001**;*22*: 501–508.

343. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput*. **2015**;*11*: 3696–3713.

344. McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, et al. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J*. **2015**;*109*: 1528–32.

345. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput*. **2013**;*9*: 3084–95.

346. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph*. **1996**;*14*: 33–38.

347. Stumpe MC, Grubmüller H. Interaction of urea with amino acids: Implications for urea-induced protein denaturation. *J Am Chem Soc*. **2007**;*129*: 16126–16131.

348. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*. **1982**;*157*: 105–132.

349. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *J Comput Chem*. **2005**;*26*: 1781–1802.

350. Lakowicz JR, Masters BR. Principles of Fluorescence Spectroscopy, Third Edition. *J Biomed Opt*. **2008**;*13*: 029901.

351. Rees DC, Robertson AD. Some thermodynamic implications for the thermostability of proteins. *Protein Sci*. **2002**;*10*: 1187–1194.

352. Frauenfelder H, Hartmann H, Parak F, Karplus M, Kuriyan J, Kuntz ID, et al. Thermal Expansion of a Protein. *Biochemistry*. **1987**;*26*: 254–261.

353. Thirumalai D, Samanta H, Maity H, Reddy G. Universal nature of collapsibility in the context of protein folding and evolution. *bioRxiv*. Cold Spring Harbor Laboratory; **2018**; 461046.

354. Facchiano AM, Colonna G, Ragone R. Helix stabilizing factors and stabilization of thermophilic proteins: an X-ray based study. *Protein Eng Des Sel*. **2002**;*11*: 753–760.

355. Dhar A, Gruebele M. Fast relaxation imaging in living cells. *Curr Protoc Protein Sci*. **2011**;*1*.

356. Bajar BT, Wang ES, Zhang S, Lin MZ, Chu J. A guide to fluorescent protein FRET pairs. *Sensors (Switzerland)*. Multidisciplinary Digital Publishing Institute (MDPI); **2016**;*16*.

357. Gelman HM. Cellular influence on protein folding. **2015**; Available: https://www.ideals.illinois.edu/handle/2142/88268

358. Guo M, Xu Y, Gruebele M. Temperature dependence of protein folding kinetics in living cells.

Proc Natl Acad Sci U S A. **2012**;*109*: 17863–7.

359. Palleros DR, Welch WJ, Fink AL. Interaction of hsp70 with unfolded proteins: effects of temperature and nucleotides on the kinetics of binding. *Proc Natl Acad Sci*. **1991**;*88*: 5719–5723.

360. Clerico EM, Tilitsky JM, Meng W, Gierasch LM. How hsp70 molecular machines interact with their substrates to mediate diverse physiological functions. *J Mol Biol*. **2015**;*427*: 1575–88.

361. Diamant S, Goloubinoff P. Temperature-Controlled Activity of DnaK−DnaJ−GrpE Chaperones: Protein-Folding Arrest and Recovery during and after Heat Shock Depends on the Substrate Protein and the GrpE Concentration. *Biochemistry*. **1998**;*37*: 9688–9694.

362. Rosenzweig R, Sekhar A, Nagesh J, Kay LE. Promiscuous binding by Hsp70 results in conformational heterogeneity and fuzzy chaperone-substrate ensembles. *Elife*. eLife Sciences Publications, Ltd; **2017**;*6*: e28030.

363. Behnke J, Mann MJJ, Scruggs F-LL, Feige MJJ, Hendershot LMM. Members of the Hsp70 Family Recognize Distinct Types of Sequences to Execute ER Quality Control. *Mol Cell*. **2016**;*63*: 739–752.

364. Zhuravleva A, Gierasch LM. Substrate-binding domain conformational dynamics mediate Hsp70 allostery. *Proc Natl Acad Sci*. **2015**;*112*: E2865–E2873.

365. Carbonell P, Nussinov R, Del Sol A. Energetic determinants of protein binding specificity: Insights into protein interaction networks. *Proteomics*. **2009**;*9*: 1744–1753.

366. McCarty JS, Buchberger A, Reinstein J, Bukau B. The Role of ATP in the Functional Cycle of the DnaK Chaperone System. *J Mol Biol*. **1995**;*249*: 126–137.

367. Hurley JH. The Sugar Kinase/Heat Shock Protein 70/Actin Superfamily: Implications of Conserved Structure for Mechanism. *Annu Rev Biophys Biomol Struct*. **1996**;*25*: 137–162.

368. Bork P, Sander C, Valencia A. An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. *Proc Natl Acad Sci U S A*. National Academy of Sciences; **1992**;*89*: 7290–4.

369. Mayer MP, Schröder H, Rüdiger S, Paal K, Laufen T, Bukau B. Multistep mechanism of substrate binding determines chaperone activity of Hsp70. *Nat Struct Biol*. **2000**;*7*: 586–593.

370. Fourie AM, Sambrook JF, Gething MJH. Common and divergent peptide binding specificities of hsp70 molecular chaperones. *J Biol Chem*. **1994**;*269*: 30470–30478.

371. Bösl B, Grimminger V, Walter S. Substrate binding to the molecular chaperone Hsp104 and its regulation by nucleotides. *J Biol Chem*. **2005**;*280*: 38170–38176.

372. Fernández-Higuero JA, Aguado A, Perales-Calvo J, Moro F, Muga A. Activation of the DnaK-ClpB Complex is Regulated by the Properties of the Bound Substrate. *Sci Rep*. **2018**;*8*.

373. Freeman BC, Michels A, Song J, Kampinga HH, Morimoto RI. Analysis of molecular chaperone activities using in vitro and in vivo approaches. *Methods Mol Biol*. New Jersey: Humana Press; **2000**;*99*: 393–419.

374. Palleros DR, Shi L, Reid KL, Fink a L. hsp70-protein complexes. Complex stability and conformation of bound substrate protein. *J Biol Chem*. **1994**;*269*: 13107–14. Available: http://www.ncbi.nlm.nih.gov/pubmed/8175736

375. Sharma SK, De Los Rios P, Christen P, Lustig A, Goloubinoff P. The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase. *Nat Chem Biol*. Nature Research; **2010**;*6*: 914–920.

376. Zhuravleva A, Clerico EM, Gierasch LM. An interdomain energetic tug-of-war creates the allosterically active state in Hsp70 molecular chaperones. *Cell*. **2012**;*151*: 1296–1307.

377. Kityk R, Kopp J, Sinning I, Mayer MP. Structure and Dynamics of the ATP-Bound Open Conformation of Hsp70 Chaperones. *Mol Cell*. **2012**;*48*: 863–874.

378. Mogk a, Tomoyasu T, Goloubinoff P, Rüdiger S, Röder D, Langen H, et al. Identification of thermolabile Escherichia coli proteins: prevention and reversion of aggregation by DnaK and ClpB. *EMBO J*. **1999**;*18*: 6934–6949.

379. Tantama M, Martínez-François JR, Mongeon R, Yellen G. Imaging energy status in live cells with a fluorescent biosensor of the intracellular ATP-to-ADP ratio. *Nat Commun*. **2013**;*4*.

380. Schumacher RJ, Hurst R, Sullivan WP, McMahon NJ, Toft DO, Matts RL. ATP-dependent chaperoning activity of reticulocyte lysate. *J Biol Chem*. **1994**;*269*: 9493–9499.

381. Wang S, Xie W, Rylander MN, Tucker PW, Aggarwal S, Diller KR. HSP70 kinetics study by continuous observation of HSP-GFP fusion protein expression on a perfusion heating stage. *Biotechnol Bioeng*. **2008**;*99*: 146–154.

382. O'Brien MC, Flaherty KM, McKay DB. Lysine 71 of the chaperone protein Hsc70 Is essential for ATP hydrolysis. *J Biol Chem*. **1996**;*271*: 15874–8. Available: http://www.ncbi.nlm.nih.gov/pubmed/8663302

383. Palleros DR, Reid KL, McCarty JS, Walker GC, Fink AL. DnaK, hsp73, and their molten globules. Two different ways heat shock proteins respond to heat. *J Biol Chem*. **1992**;*267*: 5279–5285.

384. Mattoo RUH, Goloubinoff P. Molecular chaperones are nanomachines that catalytically unfold misfolded and alternatively folded proteins. *Cell Mol Life Sci*. Springer; **2014**;*71*: 3311–25.

385. Theodorakis NG, Morimoto RI. Posttranscriptional regulation of hsp70 expression in human cells: effects of heat shock, inhibition of protein synthesis, and adenovirus infection on translation and mRNA stability. *Mol Cell Biol*. **1987**;*7*: 4357–4368.

386. Flynn GC, Pohl J, Flocco MT, Rothman JE. Peptide-binding specificity of the molecular chaperone BiP. *Nature*. **1991**;*353*: 726–730.

387. Platkov M, Gruebele M. Periodic and stochastic thermal modulation of protein folding kinetics. *J Chem Phys*. American Institute of Physics; **2014**;*141*: 035103.

388. Inoue R, Biehl R, Rosenkranz T, Fitter J, Monkenbusch M, Radulescu A, et al. Large Domain Fluctuations on 50-ns Timescale Enable Catalytic Activity in Phosphoglycerate Kinase. *Biophys J*.

2010;*99*: 2309–2317.

389. Sekhar A, Velyvis A, Zoltsman G, Rosenzweig R, Bouvignies G, Kay LE. Conserved conformational selection mechanism of Hsp70 chaperone-substrate interactions. *Elife*. eLife Sciences Publications Limited; **2018**;*7*: e32764.

390. Van Durme J, Maurer-Stroh S, Gallardo R, Wilkinson H, Rousseau F, Schymkowitz J. Accurate prediction of DnaK-peptide binding via homology modelling and experimental data. *PLoS Comput Biol*. **2009**;*5*.

391. Kampinga HH, Hageman J, Vos MJ, Kubota H, Tanguay RM, Bruford EA, et al. Guidelines for the nomenclature of the human heat shock proteins. Cell Stress and Chaperones. 2009. pp. 105–111.

392. Radons J. The human HSP70 family of chaperones: where do we stand? Cell Stress and Chaperones. 2016. pp. 379–404.

393. Rédei GP. Needleman-Wunsch Algorithm. Encyclopedia of Genetics, Genomics, Proteomics and Informatics. 2008. pp. 1327–1327.

394. Gething MJ, Sambrook J. Protein folding in the cell. Nature. 1992. pp. 33–45.

395. Liu T, Daniels CK, Cao S. Comprehensive review on the HSC70 functions, interactions with related molecules and involvement in clinical diseases and therapeutic potential. Pharmacology and Therapeutics. 2012. pp. 354–374.

396. Stricher F, Macri C, Ruff M, Muller S. HSPA8/HSC70 chaperone protein. *Autophagy*. Taylor & Francis; **2013**;*9*: 1937–1954.

397. Goldfarb SB, Kashlan OB, Watkins JN, Suaud L, Yan W, Kleyman TR, et al. Differential effects of Hsc70 and Hsp70 on the intracellular trafficking and functional expression of epithelial sodium channels. *Proc Natl Acad Sci*. **2006**;*103*: 5817–5822.

398. Čvoro A, Korać A, Matić G. Intracellular localization of constitutive and inducible heat shock protein 70 in rat liver after in vivo heat stress. *Mol Cell Biochem*. **2004**;*265*: 27–35.

399. Ellis S, Killender M, Anderson RL. Heat-induced alterations in the localization of HSP72 and HSP73 as measured by indirect immunohistochemistry and immunogold electron microscopy. *J Histochem Cytochem*. **2000**;*48*: 321–32.

400. Chiarelli LR, Morera SM, Bianchi P, Fermo E, Zanella A, Galizzi A, et al. Molecular insights on pathogenic effects of mutations causing phosphoglycerate kinase deficiency. PLoS ONE. 2012.

401. Sabelko J, Ervin J, Gruebele M. Observation of strange kinetics in protein folding. *Proc Natl Acad Sci U S A*. **1999**;*96*: 6031–6036.

402. El-Samad H, Kurata H, Doyle JC, Gross CA, Khammash M. Surviving heat shock: control strategies for robustness and performance. *Proc Natl Acad Sci U S A*. **2005**;*102*: 2736–41.

403. Mayer MP. Hsp70 chaperone dynamics and molecular mechanism. *Trends Biochem Sci*. **2013**;*38*: 507–514.

404. Mayer MP, Bukau B. Hsp70 chaperones: Cellular functions and molecular mechanism. Cellular and Molecular Life Sciences. 2005. pp. 670–684.

405. De Los Rios P, Ben-Zvi A, Slutsky O, Azem A, Goloubinoff P. Hsp70 chaperones accelerate protein translocation and the unfolding of stable protein aggregates by entropic pulling. *Proc Natl Acad Sci U S A*. National Academy of Sciences; **2006**;*103*: 6166–6171.

406. Hartl FU, Hayer-Hartl M. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science*. **2002**;*295*: 1852–8.

407. Yildiz Ö, Kalthoff C, Raunser S, Kühlbrandt W. Structure of GlnK1 with bound effectors indicates regulatory mechanism for ammonia uptake. *EMBO J*. **2007**;*26*: 589–599.