DISORDERED PROTEINS *IN VITRO* AND IN CELLS

Joseph F. Thole

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Chemistry Department in the College of Arts and Sciences.

Chapel Hill
2022

Approved by:

Gary J. Pielak

Dorothy Erie

Elizabeth Brunk

Amy Gladfelter

Rebecca Berlow

ABSTRACT

Joseph F. Thole: Disordered proteins *in vitro* and in cells
(Under the direction of Gary J. Pielak)

One requirement for the survival and reproduction of living organisms is the ability to receive external stimuli and then respond appropriately. It is increasingly understood that intrinsically disordered proteins are key to many of these biological processes, providing a flexible foundation that allow cells to coordinate and tune signal transduction cascades. To understand how these proteins behave and function, researchers from diverse intellectual backgrounds have examined their biological functions, physical properties, and interactions with other species. My dissertation is an attempt to build on prior research and integrate knowledge of intrinsically disordered proteins within the field of macromolecular crowding—i.e., how the packing of many biological molecules inside of a single cell impact protein and protein-complex stability. I first discuss intrinsically disordered proteins and their properties and predict what types of disordered protein properties might be affected by macromolecular crowding. I then examine mitogen-activated protein kinase pathway signaling through the lens of Epidermal Growth Factor signaling to highlight some of the biological roles that disordered proteins perform. I quantify the binding between the C-terminal tail of Son of sevenless and its native SH3-binding domain partner and demonstrate that precise quantification is possible even when the disordered domain is extended. I then demonstrate an improved method for eukaryotic in-cell NMR for quantifying the effects

of the cytoplasm on protein and protein-complexes in living cells. Finally, I provide comments regarding questions and experiments that I attempted thought my PhD training that will hopefully provide assistance for researchers who come after me. Overall, the perspectives and results in this document provide a roadmap for those who want to study disordered proteins in complex and crowded environments. Our understanding of disordered protein properties will continue to evolve, and I attempt to highlight relevant questions and areas researchers interested in developing macromolecular crowding theories must address and provide a foundation for the actual biological processes that underlie disordered protein function.

To Hannah
For encouraging me when I lacked the confidence
Cheering me when I achieved
You gave me the strength to reach for things I thought impossible

ACKNOWLEDGEMENTS

Graduate school has been an incredible experience that I have worked to get to but also have had numerous occasions of luck and serendipity that pushed me to where I am today. First, I must thank Radhika Plakkot who taught me AP Biology and Elizabeth Megonigal who taught me AP Chemistry. Their passion for education, science, and especially their interest in mechanics and details left an indelible mark on me, and kept me on a life-long journey of wanting to understand the way the world works. Their efforts led me to study biochemistry in college, where I met Taarika Babu. It was only because of Taarika's insistence that I applied for my first laboratory role as an Intern at the FDA. Throughout my roles in scientific research, Timmothy Blake, Eleanor Robinson, Sarah McClymont, and Bill Law showed me how to work hard, perform research ethically, think critically, and be supportive to colleagues.

Gary, thank you for accepting me as a member of the Pielak mafia, teaching me about enzymes, biophysics, how to treat people well, and that you have to have a life and intellectual interests outside of your immediate research. You gave me to space and resources to try things, and made graduate school a truly engaging and fun experience. Your support has been incredibly meaningful and have helped me see that the only limits in life are those we set for ourselves, and laws of thermodynamics.

To my former lab members who taught me much of what I know—Pixie Piszkiewicz, Sam Stadmiller, Shannon Speer, Candice Crilly, Harrison Esterley. To current lab members who make every day entertaining—Claire Stewart, Jack Eicher,

Julia Brom, I-Te Chu, Oskar Hutcheson. I especially want to shout out Sam, who answered every question I had, and always made time to help me when I needed it. Also, Candice, whose friendship, sharp questions, and dragging me to Friends kept me sane during the stressful times.

My friendships from before, and new ones formed here that kept me grounded—Shaheer Hasan, Taarika Babu, Dan Schuldenfrei, Bill and Ashley Law, Jeff Bonin, Rhese and Tika Thompson, Jared Baisden, Michelle Currie, Adam Waterbury, Parth Jariwala, Rachel Johnson, and Dylan DiGioa. All of you have brought me so much joy and kept me refreshed during the grueling times.

My mom Theresa and my dad Joe, my brother Jake and sister Annalise, my family close and extended, who have all listened and nodded very generously after asking me what I’ve been up to. Your love and support has been irreplaceable.

UNC is a wonderful place to do research and live, primarily because of the people here. Other researchers, and especially the staff make UNC an incredibly warm place to learn and grow. I have had the good fortune to be able to utilize many of the core research facilities at UNC, and get to know the amazing people who staff them. Thank you, especially Stu Parnham. Whole days of titrations at the NMR core were tough, but our chats, me bugging you about minutia, and your tales made the long days fun and unforgettable. Thank you, Ash Tripathy, you always went out of your way to help me get the data I needed, and fit me in when I was under pressure. You taught me so much, and I am forever grateful for your trust and encouragement. Thank you, Lauren Altemara, you were an incredible resource, and helped me get my first project off the ground. Your knowledge, generosity, and support truly made all of my success possible.

Finally, Hannah, you have given me a lifetime of support and encouragement. You have pushed me to pursue my dreams, and even apply to graduate school, despite the physical distance that separated us. You have always helped me when I needed it, and cheered me on when times were challenging, and you never hesitated or made me feel guilty when you made sacrifices so that I could pursue my career. The phrase “wind beneath my wings” doesn’t begin to cover how fundamental you are to my life and success.

TABLE OF CONTENTS

LIST OF FIGURES

TABLE OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

1D	one-dimensional
2D	two-dimensional
ADP	adenosine diphosphate
AMP	adenosine monophosphate
ATP	adenosine triphosphate
BSA	bovine serum albumin
CD	circular dichroism spectropolarimetry
CDC25	cell division control protein 25
CPMG	Carr-Purcell-Meiboom-Gill
Crk-L	CT10 Regulator of Kinase-like
CSP	chemical shift perturbation
DLS	dynamic light scattering
DMSO	dimethyl sulfoxide
Drk	Downstream of Receptor Kinase
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
Elk	ETS domain-containing protein Elk-1
ERK	Extracellular signal-related kinase
g	grams
GAB1/2	GRB2 Associated Binding protein
GAREM	GRB2-Associated and Regulator of MAPK/ERK
GB1	B1 domain of streptococcal protein G
GDP	guanosine diphosphate
GRB2	Growth factor Receptor Bound protein 2

GTP	guanosine triphosphate
h	Planck's constant
HMQC	Heteronuclear multiple quantum coherence
HSQC	heteronuclear single quantum coherence
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
Jnk	Jun amino-terminal kinases
K	Kelvin
k_B	Boltzmann constant
K_A	association constant
K_D	dissociation constant
kDa	kilodalton
k_{off}	dissociation rate constant
k_{on}	association rate constant
KSR	Kinase Suppressor of Ras
L	liters
LLPS	liquid-liquid phase separation
LSCM	Laser-scanning confocal microscopy
M	molar
MAPK	mitogen activated protein kinase
MBD	Maltose Binding Domain
NOE	nucelar Overhauser effect
NMR	nuclear magnetic resonance

OD_{600}	optical density at 600 nm
PAGE4	Prostate-Associated Gene 4
PDB	Protein Data Bank
PEG	polyethylene glycol
pH_i	intracellular pH
PI3K	Phosphatidylinositol 3-Kinase Regulatory
PI3KR1	Phosphatidylinositol 3-Kinase Regulatory subunit alpha
PID	phosphotyrosine interaction domain
PIP_2	phosphatidylinositol (4,5)-trisphosphate
PIP_3	phosphatidylinositol (3,4,5)-trisphosphate
pSer	phsophoserine
pThr	phosphothreonine
pTyr	phosphotyrosine
R	gas constant
R_1	longitudinal relaxation rate
R_2	transverse relaxation rate
Raf	rapidly accelerated fibrosarcoma
RTK	receptor tyrosine kinase
S^2	Order parameter
SAXS	small angle X-ray scattering
SH2	src-homology 2 domain
SH3	src-homology 3 domain
SHC1	SHC-transforming protein 1
SOS	Son of Sevenless
SPR	surface plasmon resonance
SPRY	sprouty

T	absolute temperature
T_1	longitudinal relaxation time
T_2	transverse relaxation time
TEV	tobacco etch virus
TITAN	titration analysis
T_m	melting temperature
T_{ref}	reference temperature
TROSY	Transverse relaxation optimized spectroscopy
T_s	temperature of maximum stability
α-syn	α-synuclein
δ	chemical shift
$\Delta C_P^{\circ\prime}$	equilibrium change in heat capacity
$\Delta C_{P,D}^{\circ\prime}$	equilibrium change in heat capacity of dissociation
$\Delta G_D^{\circ\prime\ddagger}$	activation free energy of dissociation
$\Delta G_D^{\circ\prime}$	equilibrium free energy of dissociation
$\Delta G_A^{\circ\prime\ddagger}$	activation free energy of association
$\Delta G_U^{\circ\prime}$	equilibrium free energy of unfolding
$\Delta H_A^{\circ\prime\ddagger}$	activation enthalpy of association
$\Delta H_D^{\circ\prime}$	equilibrium enthalpy of dissociation
$\Delta H_D^{\circ\prime\ddagger}$	activation enthalpy of dissociation
$\Delta H_U^{\circ\prime}$	equilibrium enthalpy of unfolding
$\Delta S_A^{\circ\prime\ddagger}$	activation entropy of association
$\Delta S_D^{\circ\prime}$	equilibrium entropy of dissociation
$\Delta S_D^{\circ\prime\ddagger}$	activation entropy of dissociation

$\Delta S_U^{\circ\prime}$	equilibrium entropy of unfolding
τ_c	rotational correlation time
τ_S	effective correlation time
‡	transition state

CHAPTER 1: IMPACT OF MACROMOLECULAR CROWDING ON DISORDERED PROTEINS

PROTEIN DISORDER

Protein function is classically tied to protein structure. Until recently, this structure-function paradigm was the sole framework tying the relatedness, activity and role of proteins directly to their structure. This chapter discusses how the paradigm is changed and enriched by the inclusion of disordered proteins.

The organization, folding, structure, stability and function of globular and structural proteins arise from their primary structure and the linear order of amino acids, numbered starting at the N terminus.[1,2] The linear sequence of amino acids organizes itself into local structures via networks of backbone hydrogen bonds, referred to as the secondary structure.[3,4] Common structures include α-helices, β-sheets, turns, loops and short stretches of left-handed helices. These structures are determined by the local composition of amino acids, which determines the torsion angles that the backbone can adopt.[5] Each protein has a unique primary structure, which comprises a mix of secondary structural elements. In contrast, structural proteins, such as keratin and collagen, have a repeating primary structure pattern that gives rise to a repeating higher order structure.

In globular proteins, the non-repeating sequences cause local structures to organize into larger unique, ~100 residue, domains called tertiary structures, which is the highest level of structure for monomeric proteins in dilute solutions. Many proteins have multiple domains, separated by flexible, unstructured linkers. These domains can be independent or have interactions with other intra-protein domains. Monomeric proteins can also form higher-order units via homo- or hetero-monomer associations to generate a quaternary structure.

Some researchers look beyond the quaternary level to understand how the cellular milieu impacts or changes the structures that are available to proteins, the so-called quinary level.[6] Although it is unlikely that non-specific interactions will drastically change the energy-landscape of a protein-fold, these interactions alter protein stability, and less-stable proteins have a broader ensemble of structures, which can increase their functional repertoire or lead to aggregation.

The earliest methods for seeing protein structure at high resolution, particularly x-ray crystallography,[7,8] drove the sequence-structure paradigm for decades. Crystallography requires highly-concentrated, well-behaved proteins that can withstand extreme conditions yet retain a narrow ensemble of structures. Crystallography enabled much of our understanding and classification of protein domains and led to the observation that many proteins are composites of domains (super secondary structures), with homologous domains having differences in their amino acid sequences that provide the specificity required to carry out their evolved function. While this description seems concise, the modern accumulation of many high-resolution structures has revealed that protein structure is not simple. Several proteins have two distinct folds,[9] showing that our underlying understanding of folding and structure remains incomplete.

It was only later, when techniques like fluorescence spectroscopy, high-field nuclear magnetic resonance spectroscopy (NMR), and small angle X-ray scattering (SAXS) showed that there is another class of proteins. This class lacks persistent secondary or higher levels of structure.[10] Although such proteins had been known to exist for decades, they were viewed as artifacts of evolution or isolation.[10-13] Intrinsically disordered regions (IDRs) not only serve as extended linkers between structured domains but also exist as monomers (intrinsically disordered proteins, or IDPs), as well as domains within larger proteins, and carry out specific functions.[12,14] The observation that disordered proteins perform essential biological function was surprising because it was believed that function required structure. Protein disorder exists across all domains of life but is enriched in eukaryotes compared to prokaryotes and archaea.[15]

MACROMOLECULAR CROWDING

Efforts to understand biological macromolecules most often utilize a reductionist approach, studying these molecules in well controlled, dilute, buffered solutions. Such efforts are important, especially as a starting point, so that the activity of the macromolecules of interest can be directly measured. This approach, however, ignores an important fact—most biological macromolecules have evolved to exist and function in the crowded and complex environment found inside cells. The prokaryote *Escherichia coli* contains upwards of 300 g/L of protein,[16] and eukaryotes such as human cell cultures and chordate oocytes have concentrations of 100 g/L[17] and 30-60 g/L,[18-22] respectively. Although, eukaryotes are less crowded than prokaryotes, investigations in buffer typically employ macromolecule concentrations of less than 10 g/L.

Macromolecular crowding comprises two general types of forces. The first is purely entropic. Hardcore repulsion is a purely steric effect—that is, two molecules do not affect each other until they come into contact, at which point repulsion becomes infinite—and arises due to the system finding its most entropically favorable arrangement. Furthermore, the center of mass of these two molecules cannot enter an excluded volume that surrounds the two species.[23] Because macromolecules are not rigid spheres and can take on multiple arrangements, hard-core repulsions favor states where macromolecules occupy the most compact state available. In terms of protein stability or protein-protein complexes, this compaction can only stabilize the protein or complex.[24,25] The degree of stabilization is proportional to the fraction of volume occupied by crowders, with a higher fraction having a stronger effect. As crowder size increases at constant volume occupancy, stabilization initially increases, reaching a maximum when the crowder size matches the macromolecule, and then decreases as the crowder becomes much larger than the protein of interest. This is in contrast to the predictions using scaled particle theory,[23] which predict that small crowders should have the largest crowding effect and decrease at a constant occupancy.[26,27] Another way to view the hard-core entropic force is via what are called depletion forces, which favor a compact state. Under this model, it is more favorable for the excluded volumes of two species to overlap because it allows more total volume for everything else in the solution, minimizing volume occupancy.[28,29] Although hard-core repulsions are entirely entropic, depletion forces can also have an enthalpic component.[30]

As discussed above, proteins and other macromolecules are neither rigid spheres nor chemically inert. All molecules possess the ability to form non-covalent interactions, from London dispersion forces to hydrogen bonds, hydrophobic interactions, dipole-dipole interactions and others.[31] For biological macromolecules, these interactions are complex and numerous and can lead to a variety of outcomes depending on the molecule and its surroundings. Proteins and their complexes can be stabilized or destabilized depending on the sum of total interactions. Importantly, their strength can overcome the stabilizing effect of hard-core repulsions, as shown in numerous systems.[32-38] While the general principles of crowding are understood, determining how each type of interaction impacts a given protein fold or protein complex remains challenging. Our group, in collaboration with the Harries group[30] in Jerusalem, is developing models to account for the important enthalpic and entropic effects that make up macromolecular crowding.

THEORETICAL IMPACT OF CROWDING ON DISORDERED-PROTEIN BINDING

This section addresses recent efforts to understand the effects of macromolecular crowding on protein-protein complexes when one or both partners are disordered.

Using the framework of Fonin *et al*.[39] for IDPs involved in binding, we can divide them into “unfoldable”, “partially foldable” and “readily foldable”. In most studies of “partially-foldable” and “readily-foldable” IDPs, the disordered protein undergoes a large disorder-to-order transition to enter the bound state, while for “unfoldable” IDPs, a small, local disorder-to-order transition is required. An important exception is the strong interaction between histone H1 and its nuclear chaperone prothymosin.[40] These peptides are highly- and oppositely-charged and associate in a dynamic complex with no traditional order in the bound state. This complex is an extreme example of what are deemed “fuzzy interactions.”

Folding and binding

Through the lens of traditional protein-domain interactions,[2] proteins typically reach a low-energy state to properly generate a surface recognized by their binding partners. Only after this recognition do the partners enter the energy transition state and then fall to the lower free energy state bound state.[41] Two paths are generally considered: 1) conformational selection, where the conformations of the unbound states are similar to those of the bound states, or 2) induced fit, where binding occurs first, with a subsequent change in the conformation of one or both partners.[42] Crowding can change the free energy difference between the folded and unfolded monomeric states and the free and bound states and impact the transition states to alter the kinetics of binding. To make matters even more complex, IDPs can undergo conformational selection and induced fit pathways,[42,43] with the propensity of structure driving one path or another.[44]

Generally, we can view the crowding forces that affect protein binding as: 1) enthalpic forces corresponding to changes in protein surfaces between states, and 2) entropic forces corresponding to changes in volume occupied between states. IDPs that undergo large binding-induced disorder-to-order transitions to form a relatively static final structures are likely to be perturbed by crowding (Figure 1.1). Folding would create a large change in surface between the unbound state and bound state, and the reaction must pay a large entropic penalty for de-mixing and forming a more ordered bound state, while potentially gaining favorable enthalpic interactions. Following this line of

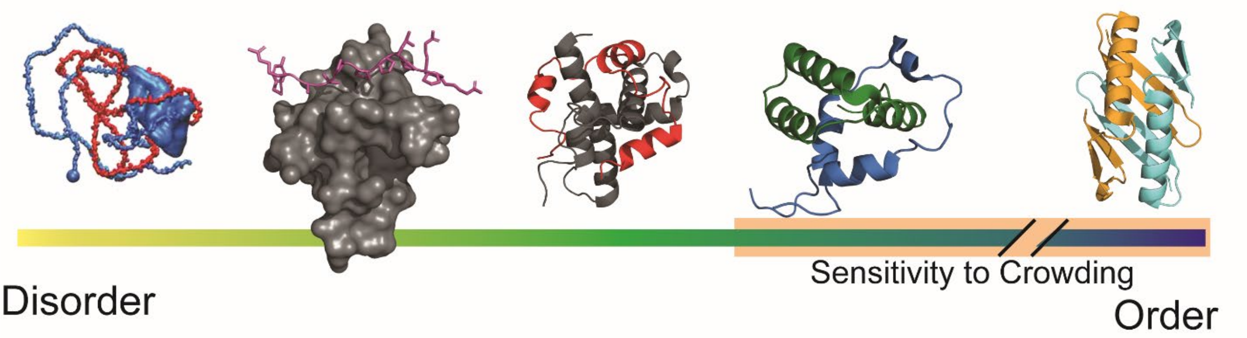


Figure 1.1. Model of IDP disorder-to-order transition upon binding. From left to right: Histone H1-prothymosin (blue/red; adapted with permission[40]) interaction; Drk N-terminal SH3-Sos site 2 peptide (grey/purple) binding;[45] HIF-1α-CBP Taz domain (red/grey; PDB ID: 1L8C) binding; CBP nuclear receptor coactivator binding domain-p53 TAD binding (green/blue; PDB ID: 2L14); GB1 domain-swapping heterodimer (PDB ID: 1Q10). Folded domains are colored in grey. Pink box represents the types of interactions sensitive to crowding.

argument, "marginally foldable" IDPs should be more sensitive to crowding perturbations—if the stability of the folded state is lower, larger population shifts due to small energy changes would be expected.

Beyond equilibrium, transition state analysis of globular proteins is a well-known method to distinguish the important residues in folding and binding transitions, as well as to determine if a residue is structured in the transition state.[41,46] This method also works for IDP binding.[47] However, its application may be challenging for IDPs whose interactions are weak, and therefore more sensitive to crowding. A study of the kinetics and equilibrium thermodynamics of the interaction between the N-terminal disordered tail of measles virus nucleoprotein and the X domain, using a systematic series of alanine substitutions, showed that changes in equilibrium-free energy of the bound and transition states are too small to distinguish between changes in the folding and binding steps.[48] Only techniques with high sensitivity and precision are likely to permit analysis.

IDPs that do not undergo large conformational changes, i.e., "unfoldable" IDPs, are likely to be less sensitive to crowding. With a smaller change in surface area or volume upon binding, there will likely be small or no change in the equilibrium thermodynamics. This situation takes the view of a single disordered site binding to another molecule. IDPs are not always simple in this way, using avidity[49,50] or allovalency [51] to enhance the net binding. Avidity describes binding between an IDP and its partner where there are multiple, unique binding sites simultaneously available on the IDP. This situation creates additional surfaces on which crowding can act: the binding sites plus the disordered linker(s) separating them. Allovalency is the presence of multiple binding sites on a disordered protein that compete for the same target. This type of interaction is predicted to increase the local concentration of the target and allow the strongest sites to carry out binding. Crowding, especially the increased volume occupancy that comes with having many other macromolecules nearby, should enhance these effects and further increase the effective concentration.

IDPs are not simple random coils. Even those that are “unfoldable” will have regions that can take on transient secondary structure or have residues that are “stickers” and capable of enhanced non-specific interactions with their surroundings.[52] The understanding of crowding effects may focus on specific sequences or regions that are distributed throughout IDPs.[53]

Other forces and caveats

IDPs tend to become more compact as temperature increases. This effect has been noted in several systems: secondary structure content increases,[54-58] smaller hydrodynamic radii are measured,[59-63] and more order is observed.[64] Compaction is predicted to arise because of unfavorable solvation at higher temperatures, leading to less hydrogen bonding with water and improved intramolecular hydrophobic interactions.[61,62,65,66] This scenario suggests that solvent depletion may strongly influence IDP folding, favoring a more compact ensemble, especially if crowders have more favorable interactions with water than with the IDP. Observations of phase-separation, a common binding property of IDPs, show that as temperature is increased, the more turbid and less solvated protein-rich phase, is enhanced. This phase is also enhanced when hydroscopic chemicals, such as PEG, are added.[66] This observation strongly suggests that solvent-IDP interactions are key to many properties of IDPs.

A *caveat* arises from the presence of strong interactions. Binding events with large free energies of dissociation are not perturbed by crowding.[67] This effect will likely hold for IDPs. Crowding effects are often the same order of magnitude as the available thermal energy,[A] so only processes with differences in free energy at a similar order of magnitude are likely to be affected. Many currently characterized IDP binding reactions are likely to remain unperturbed for this reason. Weaker interactions,[68] including the interaction between GRB2 and SOS1,[69,70] as part of the mitogen activated protein kinase pathway described in the next chapter, have µM to mM affinities, and more potential for crowding-induced effects.

THERMODYNAMICS OF IDP CROWDING *IN VITRO*

As of October 2022, five thorough studies have reported the effects of crowding on IDP binding. All focus on small IDPs binding larger, folded domains. The first, a study by Stadmiller *et al.*,[71] measured the effects of crowding on the well-characterized binding of a stable SH3 domain and a 12-residue proline-rich peptide. The authors found that the changes in free energy are generally < 1 kcal/mol near room temperature. The analysis involved testing sugars and protein crowders, and found that

---

[A] *RT*, where *R* is the gas constant and *T* is the absolute temperature sugars are relatively inert, have small effects on the energetics, and decrease the on- and off-rates. Protein crowders, such as lysozyme, bovine serum albumin and the small model protein *Streptococcal* B1 domain of protein G (GB1), showed varied effects. The analysis was limited to favorable/unfavorable electrostatics based on pI values, showing no consistent trend. The changes are partially explained by slowed translational and rotational diffusion in sugars, and favorable protein-protein sugar interactions with SH3. However, there was no analysis of the changes in translational or rotational diffusion of the peptide; further experimentation is required to develop a full understanding of the results.

The second study, from Schnatwinkel and Herrmann,[72] measured the effects of crowding on the binding of a 25-residue disordered c-Myb peptide to the folded KIX domain, another well characterized, moderate-affinity system. The authors used a series of osmolytes, sugars, and both synthetic and protein crowders and reported the same result described by Stadmiller *et al.*—small equilibrium thermodynamic changes. Only osmolytes and ethylene glycol/polyethylene glycol perturbed the equilibrium, albeit minimally.

The third study, by Gruber *et al.*,[73] compared the binding of an 80-residue disordered region of the mostly disordered Gab1 protein to the folded SH2 domain of SHP2, which requires a phosphorylated tyrosine on Gab1. The authors reported that 270 g/L bovine serum albumin (BSA) stabilized a binding-competent transition state of Gab1, but the interaction was weak and unquantified. This contrasts with the complete absence of an interaction when BSA is absent. When phosphorylated, Gab1 interacted strongly with SH2, with a dissociation constant of 10 nM. BSA decreased the affinity to

18 nM, a change of 350 cal/mol at 298 K, which is in line with the results from Stadmiller *et al*. and Schnatwinkel and Herrmann. The change was small compared to the $\Delta G^{\circ\prime}_{D,298\,K}$ of the complex (11 kcal/mol), suggesting that efforts should be focused on weaker-associating complexes. This result also exemplifies the various ways that crowding can perturb systems— it both alters the conformation of one of the partners to potentiate binding and simultaneously makes the $\Delta G^{\circ\prime}_{D}$ less favorable.

Most experiment-based and computational studies of IDPs in crowded environments have focused on the hydrodynamic radii of IDPs and effects on local structure.[53,74-77] These results are in line with the theoretical outcomes described above. For example, Cino *et al*. showed that in 160 g/L Ficoll 70, the IDP TC-1 remains disordered, but residues that have the propensity for local structure show changes in their R_2 relaxation values via NMR and no changes in R_1 or ^{15}N NOEs, suggesting that the overall ensemble does not change, but the residues that have the propensity for structure have non-specific interactions.[74] One result that stands out is the measurement of membrane-bound α-synuclein (α-syn).[76] When exposed to the crowder Hsp27 (which also interacts with the membrane), α-syn takes on a conformation that is not seen in the solution or membrane-bound states. This observation suggests that when constrained by binding, non-specific interactions might favor particular conformations.

The final study by Zosel *et al.*[78] measured the binding of the molten globule nuclear coactivator domain of CBP/p300 and the disordered activation domain of steroid receptor coactivator 3. They quantified equilibrium and kinetic parameter values in the presence of PEGs at various concentrations and lengths, as well as dextran, poyvinylpyrrolidone, and polyvinyl alcohol. Overall, their findings point to solvent depletion as the most important contributor to the changes in binding. This is in line with what has been observed and predicted for IDP conformation ensembles (described above), and supports the idea that IDPs are sensitive to changes in solvation. They suggested that their results show that IDPs are likely to be less sensitive to the interactions that are traditionally considered in macromolecular crowding, compared to folded proteins of the same size. This is reasonable, but further testing is required in solutions of less "inert" crowders, such as bovine serum albumin, lysozyme and cell lysates, to determine if only depletion interactions play a role; experiments on IDP interaction with larger disorder-to-order transitions upon binding are also needed.

DISORDERED PROTEINS IN LIQUID-LIQUID PHASE SEPARATION

The IDP binding phenomenon most studied *in vivo* and *in vitro* is liquid-liquid phase separation (LLPS), the process by which two or more protein/protein or protein/RNA binding partners interact in fuzzy complexes, often with allovalency, to form condensates. That is, they sort into a different physical phase from the surrounding solution based on differential interactions with "stickers" and solvent.[52,66,79-82] Phase separation coupled to percolation has been proposed to play a role in a variety of biological functions, including the formation of membrane-less organelles (e.g., stress granules, processing bodies, the centriole, cajal bodies), ribosome biogenesis, gene regulation, and RNA processing.[83,84] The shapes and dynamics of LLPS structures are extremely sensitive to buffer conditions, and can take on unique super-structures depending on the environment.[85,86] A 2020 review characterized observations of how crowders can alter the behavior of phase separating partners *in vitro*,[87] although caution should be exercised in directly correlating the results to biological outcomes because the conditions and systems tested may not mimic those found in cells.[88-90]

Three recent studies have shown that biological processes are enhanced while in a phase-separated state: ribozyme and deoxy-ribozyme cleavage,[91] the activation of Ras by Sos1,[92] and SUMOylation enzyme E2 activity.[93]

Careful measurements of any *in vivo* system and direct comparisons to *in vitro* results are not simple.[89] Many *in vivo* studies of phase separation have focused on the contributions of specific actors—binding competent species, structure, and additional modulates such as ATP—and how they impact lifetimes, granule structures, and dynamics.[84,94] Sang *et al.* attached client proteins to an engineered system that undergoes LLPS[95]: a constitutively-active MAPK3 (mitogen-activated protein kinase), and one of its targets, the disordered C-terminus of Elk1 (ETS domain-containing protein Elk-1). The authors quantified the phosphorylation of ELK1 as a measure of association *in vitro* and *in vivo*.[96] They first showed, compared to dilute solution, recruitment into condensates increases total phosphorylation. When the same kinase and target were expressed on modified scaffolds in the yeast *Saccharomyces cerevisiae*, ELK1 was phosphorylated at even higher rates and at multiple sites, including some that were less favorable *in vitro*. The granule volumes were smaller, despite containing more protein. They tested the effects of osmotic compression in 1 M sorbitol on a strain of yeast that is incapable of responding to changes in osmotic stress (*HOG1Δ*), and observed a two-fold increase in phosphorylation when the system was not already near saturation. For many signaling pathways, the phosphorylation of IDPs is a critical step in signal propagation and is another modulation of IDP binding and biological outcomes.[73,97]

CONCLUSIONS

The ensembles, biological roles, mechanisms, and the degree of disorder-to-order transitions that IDPs undergo during complex formation are wide-ranging. IDPs are sensitive to their environments, but it remains largely unknown if IDP binding is sensitive to macromolecular crowding. Crowding forces operate on changes in volume and surfaces between states, suggesting that IDPs that undergo the largest disorder-to-order transitions will be the most affected. Crowding seems to impact LLPS, although it is unknown if the effects are a result of the same types of interactions that have been quantified in traditional macromolecular crowding studies. The development of more systems where precise measurements can be made in complex crowded environments are critical to understanding how macromolecular crowding perturbs the activity of IDPs.

CHAPTER 2: DISORDER IN THE MITOGEN-ACTIVATED PROTEIN KINASE PATHWAY

INTRODUCTION

The mitogen-activated protein kinase (MAPK) pathway is an evolutionarily conserved protein-interaction framework that regulates a host of cell processes including division, differentiation, migration, stress, and apoptosis.[98] Several pathways follow the MAPK framework: the extracellular signal-related kinases (ERK1/2), ERK3/4 and ERK 5, Jun amino-terminal kinases (JNK1/2/3), and p38-MAPK. These pathways are found across eukaryotes, from plants to humans.[99-101] All share a similar scheme—an extracellular signal mediated by a membrane-bound receptor, is passed on to a series of serine-threonine kinases: MAPK kinase kinase (MKKK), which activates a MAPK kinase (MKK), which activates a MAPK, which then activates a transcription factor that generates the cellular response. This chapter focuses on Epidermal growth factor (EGF) signaling (part of the ERK1/2 pathway) as an example for how intrinsically disordered proteins (IDPs) and disordered regions (IDRs) direct cellular action and feedback regulation to maintain homeostasis and healthy growth.

CANONICAL EGFR SIGNALING

Tissue-specific signaling peptides, such as EGF,[102,103] collectively called mitogens, are exported from neighboring cells and undergo proteolysis to bind membrane-bound receptor tyrosine kinases (RTK) such as Epidermal Growth Factor Receptor (EGFR).[104,105] Upon binding, RTKs form oligomers,[106] which results in a conformational change, allowing autophosphorylation of cytosolic tyrosines (pTyr) that propagates the signal.[107-109]

After this phosphorylation, signal transduction proteins that contain Src Homology 2 domains (SH2) or phosphotyrosine interaction domains (PID), which specifically bind phosphotyrosine are recruited. One of the canonical proteins, Growth factor Receptor Bound protein 2 (GRB2), binds EGFR and then binds and activates the protein Son of sevenless 1 (SOS1).[110-112]

Once activated, Sos binds one or two Ras motifs, exchanging a guanosine diphosphate (GDP) for a guanosine triphosphate (GTP), turning on Ras. Activated Ras recruits and binds the MKKK protein Rapidly accelerated fibrosarcoma (Raf) kinase. Upon binding, Raf (B-Raf, C-Raf) homo- or hetero-dimerize,[113,114] and then recruits the MKK MEK (Mitogen-Activated protein Kinase Kinase 1), which continues signal propagation by recruiting and phosphorylating the MAPK ERK1 (Extra-cellular Response Kinase 1). ERK1 then dimerizes and localizes to the nucleus,[115] where it activates transcription factors (e.g., MYC, JUN, FOS)[98] to begin transcription of cyclin D to pass the G1/S checkpoint.

Many of the proteins involved in EGF signaling were identified decades ago, and their binary interactions are well studied. However, with modern techniques such as high-resolution mass spectrometry, researchers have begun to dissect network dynamics to understand how these multiprotein complexes function in concert to coordinate cellular responses.[116-118]

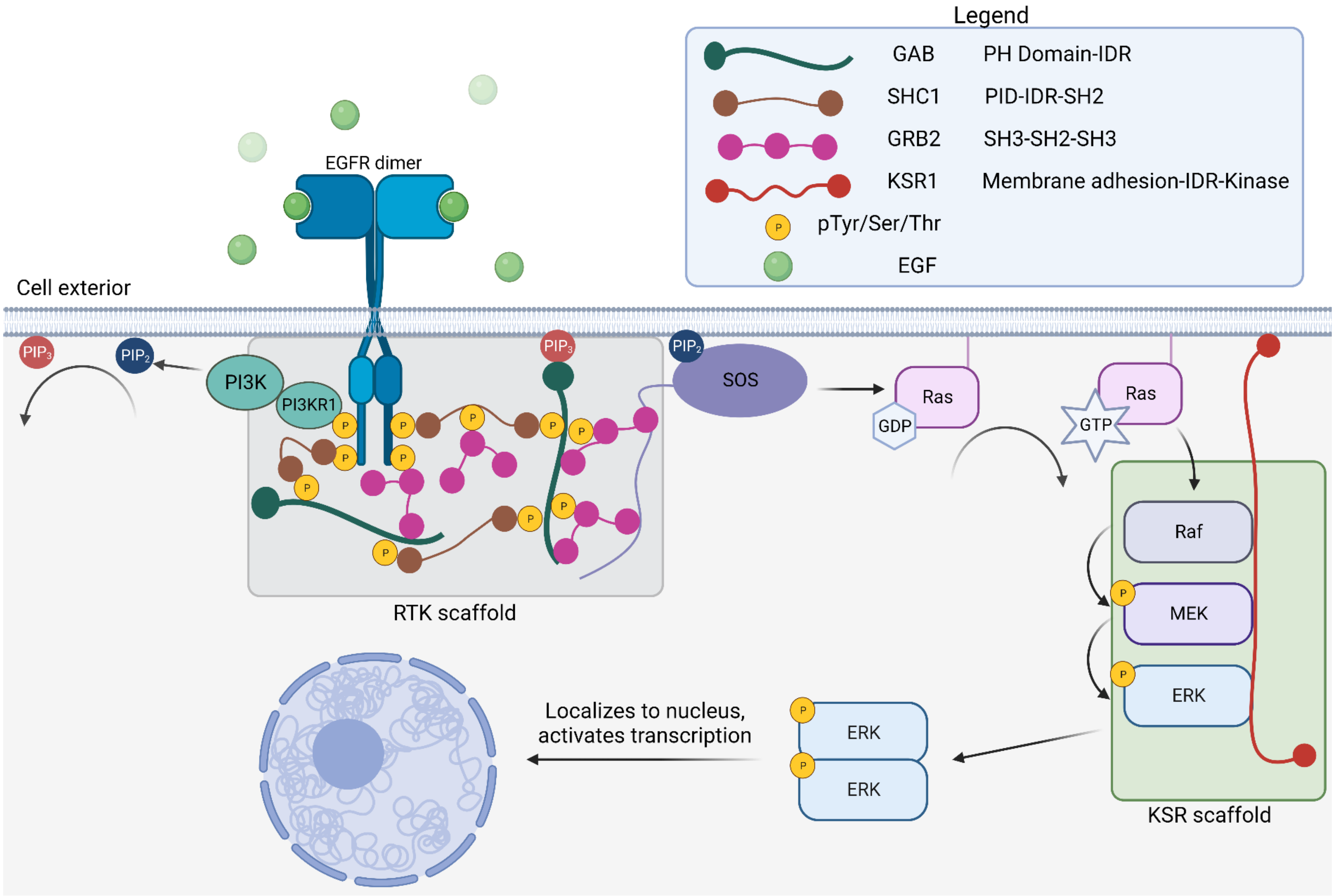


Figure 2.1. A simplified representation of EGF signaling, highlighting the roles of intrinsically disordered proteins. EGF binds to Epidermal Growth Factor Receptor (EGFR), which oligomerizes upon binding. Activated EGFR autophosphorylates a series of

tyrosines on its disordered C-terminal domain, which recruit Phosphatidylinositol 3-Kinase Regulatory subunit alpha (PIK3R1), activating Phosphatidylinositol 3-Kinase (PIK3), which catalyzes phosphatidylinositol (3,4,5)-trisphosphate (PIP_3), from phosphatidylinositol (3,4,5)-trisphosphate (PIP_2). The phosphorylated tyrosines also recruit SHC-transforming protein 1 (SHC1), Growth factor Receptor Bound protein 2 (GRB2), and GRB2 associated binding protein 1/2 (GAB), which form a scaffold and helps GRB2 bind the C-terminal disordered domain of Son of Sevenless 1 (SOS1). To become active, SOS1 must also contemporaneously bind PIP_2, which in turn, allows SOS1 to activate Ras by exchanging the bound GDP for GTP. Once active, Ras binds to Raf, which phosphorylates MEK, which in turn phosphorylates Extracellular Signal-related Kinase (ERK). Raf/MEK/ERK are coordinated by Kinase Suppressor of Ras (KSR). Phosphorylated ERKs dimerize and localize to the nucleus to activate a transcriptional response.

ROLES OF DISORDERED PROTEINS IN EGFR SIGNALING

The simplified description of EGFR activation, which leads to G1/S checkpoint passage for division, is complicated by multiple layers of feedback mechanisms and coordinated complexes (Figure 2.1) that cells evolved to regulate this process.[98,116,117,119] When this process is dysregulated, numerous diseases and cancers arise.[120-124]

EGFR localized scaffolding

When EGFR is activated, it autophosphorylates multiple tyrosines on its disordered C-terminal tail, which recruits multiple proteins. One of the first proteins recruited is Phosphatidylinositol 3-Kinase Regulatory subunit alpha (PIK3R1), which activates Phosphatidylinositol 3-Kinase (PIK3), in turn catalyzing the formation of phosphatidylinositol (3,4,5)-trisphosphate (PIP_3), a phospholipid, from phosphatidylinositol (3,4,5)-trisphosphate (PIP_2).[117] EGFR also binds SHC-transforming protein 1 (SHC1),[116] which comprises a Phosphotyrosine binding Domain (PID), an SH2 domain and extended disordered regions, as well as GRB2-Associated and Regulator of MAPK/ERK (GAREM),[125] which is largely disordered and has a proline-rich sequence capable of binding GRB2. The PID and SH2 domains of SRC1 bind EGFR, and can then be phosphorylated by the RTK itself to recruit other proteins, such as GRB2[126] and GRB2 associated binding protein 1/2 (GAB); GAB has a Pleckstrin Homology (PH) domain that is capable of binding PIP_3 and an extended disordered region that is hyper-phosphorylated so more SHC1 and GRB2 can bind.[127] Additionally, GAB has a proline-rich sequence that is specific to the C-terminal SH3 of GRB2.

SHC1/GAREM/GAB/GRB2 forms an obligate multivalent network[128] that contributes to the organization and recruitment of pathway-relevant kinases and SOS1. All of these species are phosphorylated to recruit other proteins or change biological outcomes. The specific role of GAREM is not known, but it has been shown to interact with the scaffold, as well as the 14-3-3 adaptor proteins that are associated with complex disassembly.[129,130]

The affinities of SH2/PID-pTyr interactions are 0.1-2 μM,[131,132] with most on the weaker end. PH-PIP_3 has an affinity of ~0.1 μM.[133] The data suggest that anchoring of GAB to the membrane is relatively strong, but most the individual interactions alone only have moderate affinity. The multivalent nature of the complex probably improves the efficiency of signal transduction[128] and may increase complex lifetime. No direct measurements of the complex itself are available, but the phosphorylation states required for the scaffold persist for at least 30 minutes.[118] The cell actively dissembles this scaffold by disrupting individual interactions or converting PIP_3 back to PIP_2.[130,134,135]

Dynamic regulation of SOS1

SOS1 in its non-active state has two forms of autoinhibition: one form is mediated by SOS1 and the other is mediated by GRB2. When active, SOS1 has two guanine exchange domains—a CDC25 domain and a Ras exchange motif[112,136]—that can activate Ras. One mode of activation requires SOS1 to bind a PIP_2 via a pleckstrin homology domain, which causes its neighboring Dbl homology domain to undergo a conformational change, allowing Ras binding.[137,138] PIP_2 is a modification that is made to phospholipid membranes and biases Sos activity near the membrane, where RTKs should also be located. The second mode is GRB2-mediated activation via SH3-proline rich motifs throughout the 55 kDa C-terminal tail.[136,139] The N-terminal SH3 of GRB2 binds what has been classically identified as a “type 2” site [PXΨPXR, where Ψ is a hydrophobic residue (L/I/V) and R is a basic residue (R/K)].[69,140,141] Proline-rich motifs form left-handed polyproline II helices that fit into a hydrophobic binding pocket on SH3 and can form a salt bridge with an acid patch on the SH3 RT-loop. SOS1 has two GRB2 binding sites with low µM and several binding sites with µM-mM affinities.[69,141] The allovalency[B] of this interaction likely contributes to keeping the C-terminal tail and SH3 domains in close proximity. Despite the high on-off rates of the SH3-proline rich site (described in Chapter 3), the abundance of binding competent SH3s in the GAB-mediated complex keeps the C-terminal tail out of its auto-inhibitory state long enough for Ras activation.[139] The affinity of this interaction is decreased by multiple pSer residues via negative feed-back loops, some of which are associated with 14-3-3 protein binding, which is suspected to allosterically block GRB2 binding.[142,143]

IDR functions are context-dependent

Another layer of regulation is mediated by Sprouty 1-4 (SPRY).[144-147] SPRY proteins comprise a C-terminal SPR domain for PIP_2 binding plus an extended disordered region. The most well studied homologue, SPRY2, has multiple clusters of pTyr[148] and pSer[149] residues that change regulatory outcomes.[146] SPRY2 translation increases as a result of ERK1/2 signaling and enhances EGFR signaling, while suppressing signals originating from another RTK—Fibroblast Growth Factor Receptor

---

[B] Allovalency is the presence of multiple binding sites on a disordered protein that compete for the same target.

(FGFR).[148,150-152] For FGFR signaling, a SPRY2 proline-rich motif competes with the GRB2 N-terminal SH3 interaction for SOS1. In contrast, during EFGR signaling, SPRY2 represses EGFR ubiquitination and subsequent proteasomal digestion.[148,153]

IDR scaffolds enhance kinase activity

Another critical scaffold in EGFR signaling is the organization of Raf, MEK and ERK into a multi-protein complex by Kinase suppressor of Ras 1/2 (KSR1/2).[154,155] KSR is thought to be the “backbone” of the scaffold.[156-158] It is a 102 kDa protein that comprises an N-terminal folded membrane binding domain, an extended disordered region (>440 residues)[159,160], and then a C-terminal kinase domain. Before signaling, KSR is thought to be constitutively bound to MEK and two 14-3-3 proteins through pSer residues.[156,157] When a cell receives a signal activating the ERK1/2 pathway, KSR is dephosphorylated, allowing KSR/MEK to bind the membrane and recruit Raf. ERK is also bound, and the localization of all three serine-threonine kinases allows the rapid activation of ERK. A knockdown study of KSR shows that the scaffold’s presence increases MEK activation by 10-fold.[161] Additional folded proteins can interact with the scaffold and are thought to change the localization or function based on additional signaling information.[156,157]

An important note about the scaffolds described in this chapter are that they are likely stable but dynamic. IDPs lack stable structure, and are often involved in weaker interactions compared to their folded counterparts.[162] It has been shown that a variety of scaffolds in cell are not static,[163] such as the homologous KSR/Raf/MEK/ERK complex in *Saccharomyces cerevisiae* (Ste5p/Ste11p/Ste7p/Fus3p).[164] In pheromone-treated yeast, Ste5p, Ste7p and Fus3p could be seen stably in the mating projections, while Ste11p (Raf) was too dynamic and could not be imaged.

WHY DISORDERED PROTEINS?

The information from cellular signaling is passed via the flow of pTyr/Thr/Ser to activated partners.[116,119] It seems that the need for physical organization results in selection for protein disorder. Borrowing the metaphor from Good *et al.*,[165] IDPs are highly tunable[10,14,166] and offer *literal* flexible circuitry that allows the signal to be modulated as more information comes in—by adding new binding partners, altering phosphorylation states and changing biological outcomes. This flexibility is seen with the ERK1/2 pathway when different RTKs are activated. All pathways use the same central proteins, SHC1/GAB/GRB2, but add new proteins to the scaffold by changing phosphorylation states to relay that a different response is needed.[98,108,127] SPRY2 offers the same advantage; it acts as a molecular Swiss-army knife, capable of a variety of interactions based on alterations of its phosphorylation state.[146]

It is predicted that phosphorylation can alter the ensembles of disordered regions,[167-169] and we know that differential phosphorylation allows IDP PAGE4 to respond to unique cellular conditions.[170] There may be unexpected changes when we view IDPs in the context of a bound state. When α-synuclein is over-expressed in human cells, the global ensemble of soluble species is similar but more compact than that seen in solution;[171] however, when membrane-bound and in the presences of crowding molecules that also interact with membranes, a unique ensemble is seen.[76] Researchers must now strive to connect the local dynamics and ensembles of

IDPs/IDRs with the global dynamics and function of the complexes with which they associate.

CHAPTER 3: NEIGHBORING SEQUENCES AFFECT BINDING OF SH3 TO PROLINE-RICH SITES IN THE 25KDA DISORDERED SOS C-TERMINAL TAIL

INTRODUCTION

Approximately 40% of the eukaryotic proteome comprises intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs),[172,173] which have many functions in signal transduction.[14,174] Their mechanisms of interaction and binding vary, ranging from folding upon binding[166] to “fuzzy” interactions.[175] Protein-protein complexes where at least one partner is disordered tend to be less stable than complexes formed between folded species,[162] but exceptions exist.[40] Disordered interactions are enriched in signaling pathways because they are highly tunable so outcomes can be altered in response to external stimuli and feedback.[14,42,174,176] Nevertheless, we lack a general understanding of IDP/IDR function and behavior because of their diverse mechanisms and because the proteins often have challenging physical properties (e.g., they tend to aggregate and phase separate).

The mitogen-activated protein kinase (MAPK) pathway is a well-conserved signaling regime that allows cells to differentiate, divide, respond to stress, and undergo apoptosis.[99,120] In *Drosophila melanogaster*, extracellular signals (Spitz, Trunk, Bride of sevenless, etc.) begin the signaling cascade, leading to the activation of Son of sevenless (Sos).[111,177,178]

Sos is a 178 kDa multi-domain protein with guanine nucleotide exchange factor activity that further stimulates the GTPase, Ras (Figure 3.1A). The human homolog, SOS1, undergoes two forms of auto-inhibition. The N-terminal Dbl- and Plekstrin-homology domains require interaction with phosphatidylinositol 4,5-bisphosphate (PIP_2) to recruit Ras, ensuring localization to the membrane.[137,138,179] At the C-terminus, a 55 kDa proline-rich disordered domain binds the Src homology 3 domains (SH3) of Downstream receptor kinase (Drk)/GRB2 (*D. Melanogaster*/Human), which is recruited by an activated receptor tyrosine kinase via a phosphorylated cytosolic Tyr.[69,111,136] The inhibitory mechanism of the C-terminal domain is unknown.

Proline-rich sequences form type-II left-handed helices and bind SH3 using a short P-X-Ψ-P-X-R motif, where Ψ is Leu/Ile/Val and R is Arg or Lys.[69,140,141] Similar to SOS1,[69,141] Sos has at least four binding sites.[70] Sites 2 and 4 have low μM affinities. Sites 1 and 3 have high μM or mM affinities (Figure 3.1B). Efforts to quantify these affinities, like many efforts involving a disordered protein, focus only on peptides containing the key residues and immediately flanking sequences. This reasonable choice simplifies analysis, yet residues beyond the binding site probably play a role.[180] Our goal was to determine how the context of the IDR affects SH3 binding by comparing the protein with peptides.

We used techniques known to work for IDPs and IDRs—isothermal titration calorimetry (ITC) and nuclear magnetic resonance spectroscopy (NMR)—to distinguish binding of individual sites from non-specific or weak binding,[70,181] and focused on the strongest sites (2 and 4). We found that the single-site affinities are comparable to those for the peptides, but binding to the protein is less temperature sensitive because of enthalpy-entropy compensation.[182] Our approach is applicable to other disordered proteins and will help unravel the kinetics and equilibrium thermodynamics of IDP interactions.

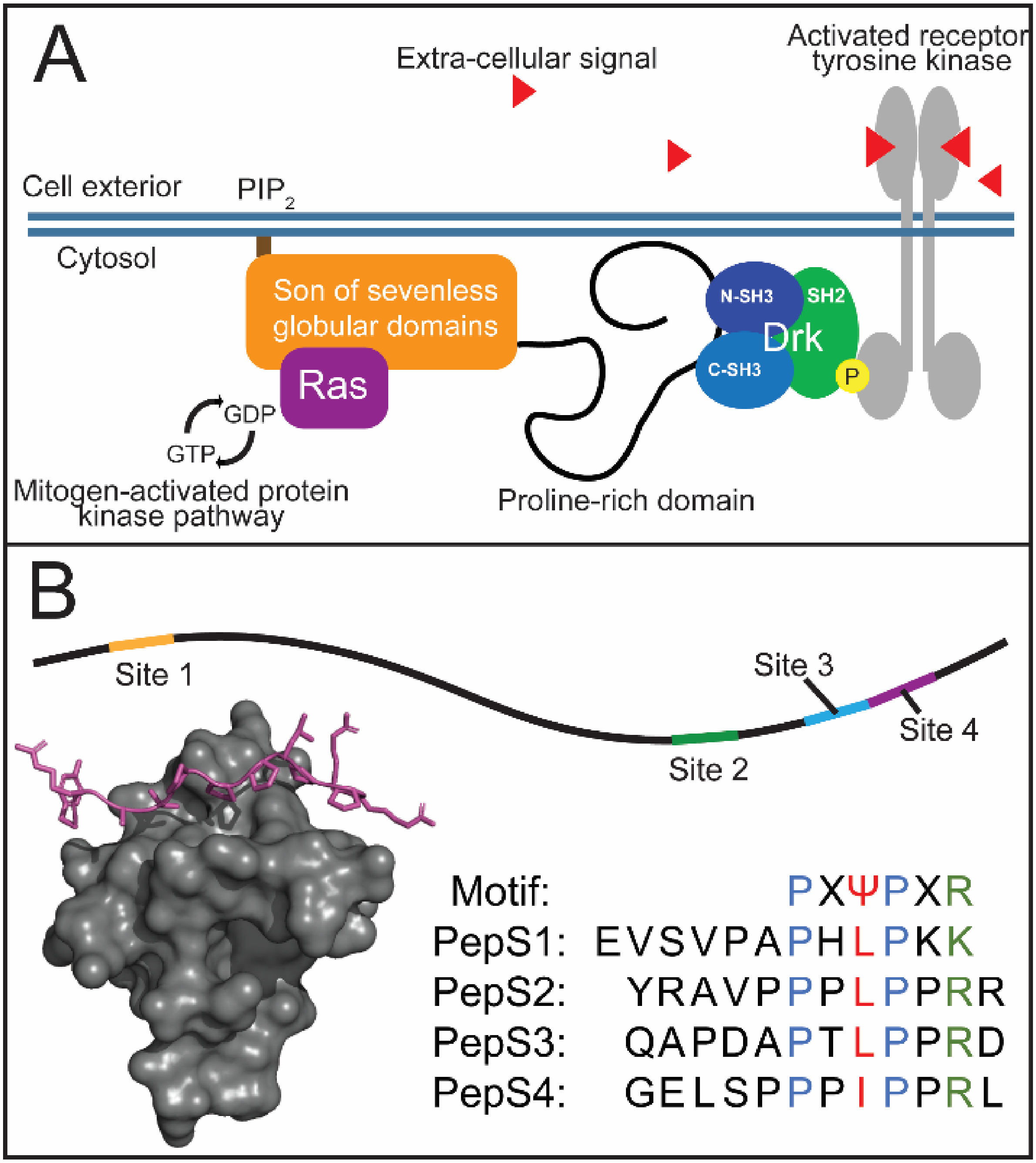


Figure 3.1. Model of Sos-SH3 binding. A) Simplified model of Ras activation by Son of sevenless. B) Representation of SH3 binding sites on Sos disordered C-terminal tail. Simulation of PepS2 binding to SH3 domain.[45] Sequence alignment of binding motif with Sos binding sites.[69,70]

METHODS

Construct design

The Sos intrinsically-disordered region was designed from *D. Melanogaster* Sos (Uniprot ID: P26675) and comprises residues 1177-1405 (Figure S3.1). Constructs were cloned into pET28b plasmids with a C-terminal cysteine self-cleaving peptide[183] followed by a deca- or hexa-histidine tag. We later observed that an N-terminal maltose binding domain (MBD), followed by a TEV cleavage site, improved the yield. Constructs contain the L1203A, as well as the L1372A and K1375A substitutions that eliminate sites 1 and 3, respectively. Also, C1306S and C1357S substitutions eliminated disulfide bond formation. The site 2 construct also contains the I1384A and K1387A substitutions that eliminate site 4. The site 4 construct has the L1346A, R1349A, and R1350A substitutions that eliminate site 2. The knockout construct contains all these alanine substitutions. Post purification (described later), constructs contain additional SMG (with MBD)- or MG (without MBD)-residues at the N-terminus and QSL residues at the C-terminus.

Expression and purification

Son of sevenless

Unenriched Sos was prepared as follows: The pET28b plasmid containing the construct of interest was transformed into BL21 (DE3) *Escherichia coli* cells (Thermo Fisher Scientific). A 5 mL liquid culture of 25 g/L Lennox broth (LB, 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) containing 50 mg/L kanamycin was inoculated from a single colony and shaken at 37 °C, 225 rpm (Innova I26, New Brunswick). After >6 h, 200 mL of LB was inoculated with 200 µL of the smaller culture. The 200 mL culture was shaken overnight at 37 °C, 225 rpm. The next day, 1 L cultures were inoculated with 10 or 20 mL of the overnight culture and grown at 37 °C, 225 rpm. Isopropyl β-d-1-thiogalactopyranoside (IPTG, 1 mL of a 1M solution) was added when the optical density at 600 nm (OD_{600}) reached 0.6 (~ 2 h). Cultures were then shaken overnight at 20 °C.

Isotopically enriched Sos was prepared as follows: A 5 mL liquid culture of 25 g/L LB and 50 mg/L kanamycin was inoculated from a single colony and shaken at 37 °C, 225 rpm. After >6 h, a 750 mL solution of 2x M9 media (100 mM Na_2HPO_4, 40 mM KH_2PO_4, 9 mM NaCl, 3 g/L glucose, 1 g/L NH_4Cl, 0.1 mM $CaCl_2$, 2 mM $MgSO_4$, 10 mg/L thiamine, 10 mg/L biotin, and 50 mg/L kanamycin, pH 7.4) was inoculated with 750 µL of the smaller culture. The 750 mL culture was shaken overnight at 37 °C, 225 rpm. The next day, 1 L cultures were inoculated with 50 mL of the overnight culture and grown at 37 °C, 225 rpm. When the OD_{600} reached 0.55 (~6 h), 60 mg/L [3,3-D_2] ^{13}C α-ketobutyric acid[184,185] was added and the temperature was lowered to 20 °C. After 30 min, 1 mL of 1 M IPTG was added and the cultures were shaken overnight at 20 °C. The cultures were transferred to 1 L bottles and centrifuged at 1000 *g* for 30 min. The pellets were resuspended in loading buffer (15.1 mM Na_2HPO_4, 4.9 mM NaH_2PO_4, 20 mM imidazole, 300 mM NaCl, 2 mM tris(2-carboxyethyl)phosphine hydrochloride, 10% v/v propylene glycol, pH 8.0) and protease inhibitors were added (Sigma Aldrich, P2714). Cells were lysed using a sonic dismembrator (Fisher, model 505) at 30% amplitude, 1/1 s power cycling in an ice-water bath, 10 min per 6 L. Lysates were centrifuged at 17,540*g* for 45 min at 4 °C followed by syringe filtration (Millex, 0.45 µm).

Lysates were loaded on Ni^{2+} columns (Cytiva HisTrap HP, 10 mL resin/6 L cell lysate) at 4 °C. The resin was then washed with four column volumes of loading buffer, and equilibrated with cleavage buffer (15.1 mM Na_2HPO_4, 4.9 mM NaH_2PO_4, 20 mM imidazole, 100 mM NaCl, 2 mM tris(2-carboxyethyl)phosphine hydrochloride, 10% propylene glycol v/v, 300 µM phytic acid, pH 8.0).[183] Cleavage was allowed to occur overnight. The next day, cleaved protein was eluted with cleavage buffer. The protein fractions were pooled and 1 ng of TEV protease was added. The sample was transferred to a 3.5 kDa molecular weight cut-off dialysis bag (Thermo Fisher Scientific) and dialyzed against 4 L of 20 mM 2-(N-morpholino)ethanesulfonic acid, 75 mM NaCl, 2.5 mM Dithiothreitol, pH 6.0 for 3 h at room temperature.

Dialyzed samples were transferred to a conical vial, propylene glycol was added to 10% v/v, and the sample was centrifuged at 4,500 *g*, 25 °C, for 5 min to remove aggregates. The supernatant was processed using cation exchange chromatography at room temperature (Cytiva, SP High Performance). A custom 25 mL column was equilibrated with 25 mM 2-(N-morpholino)ethanesulfonic acid, 4 M urea, 10% propylene glycol, and 120 mM NaCl. The column was then washed with 1 column volume of the equilibration buffer. A linear ramp to 300 mM NaCl over three column volumes was used to separate the Sos construct from closely-related degradation products. *Performing this step at 4 °C does not provide sufficient resolution; it should be performed at room temperature.*

SDS-PAGE with Coomassie visualization was used to select pure fractions. For storage, the pure fractions were combined and concentrated to about 1 mM and stored at -80 °C. The protein was exchanged into the appropriate buffer immediately before use. Determining the concentration of IPDs can be challenging;[186] we estimated the molar absorptivity at 280 nm [15,470 $(\mathrm{M\cdot cm})^{-1}$][187] and confirmed the purity of Sos using elemental analysis (Atlantic Microlab, Norcross GA) (Table S3.1). We further validated the identity and purity of the protein by mass-spectrometry (Table S3.2).

SH3

We used the *D. melanogaster* N-terminal SH3 domain of Downstream of receptor kinase (Uniprot ID: Q08012), the native binding partner to Sos. The wildtype sequence is partially unfolded under native conditions.[188] To focus on binding only, we used the stabilized T22G variant.[189] Expression and purification were performed as described,[70,190] except that after size exclusion chromatography, samples were polished via anion exchange chromatography to remove nucleic acids. A custom 25- or 50- mL Sepharose Q column (Cytiva) was equilibrated in 50 mM Tris pH 7.4, the sample was loaded, and a linear ramp to 50 mM Tris, 450 mM NaCl was performed over two column volumes. Only SH3 was eluted during using this protocol. *DNA does not stick as well to the 5 mL pre-packed columns from Cytiva and elutes close to SH3.* Pure samples were extensively dialyzed into distilled and deionized water (>17 MΩ cm), aliquoted into amounts that were appropriate for each experiment, flash frozen in ethanol/CO_2(s), and lyophilized for storage. Protein identity and purity were verified by mass spectrometry (Table S3.2).

NMR

Data were acquired using Bruker Avance III HD spectrometers equipped with QCI- (^{1}H Larmor frequencies of 500, 470 MHz for ^{19}F), or TCI- cryoprobes (^{1}H Larmor frequencies of 850 MHz, 213 MHz for ^{13}C). Data were processed using NMRpipe

(2020.171.18.39) and referenced to trimethylsilylpropanesulfonate to account for the temperature dependence of chemical shifts.

^{19}F peptide spectra comprised at least 80 and up to 400 scans, with a 7062 Hz sweep-width, 1400 complex points, an interscan delay of 2.5 s, and a center frequency of -57,646 Hz. Samples were prepared as described.[70] Titrations were performed a total of three times: once using peptide concentrations of 0, 29, 73, 145, 218, 290, 435, 580, 870, 1160, and 1450 μM with an SH3 concentration of 290 μM, then twice with peptide concentrations of 0, 6, 16, 45, 64, 91, 128, 181, 256, 363, and 1450 μM and SH3 concentrations at 290 and 145 μM.

^{1}H-^{13}C data were acquired using a TROSY-HMQC experiment[191] with at least 8 scans and up to 40 scans, depending on the signal/noise of the Ile1384 peak. The ^{1}H sweep width was 12755 Hz using 5120 complex points and a center frequency at 3996 Hz. Forty-eight complex points were collected in the indirect dimension, with a sweep width of 1069 Hz, and a center frequency of 2673 Hz. The water signal was suppressed using pre-saturation and a selective 90° pulse at the beginning of the sequence. Sos site 4 protein was exchanged into 20 mM sodium phosphate, pH 7.5 plus 5% (v/v) D_2O using a PD-10 midi desalting column (Cytiva); its concentration was then adjusted to 51.5 μM. Four hundred μL of the Sos site 4 solution was used to dissolve lyophilized SH3 for the initial titration point, then spectra were acquired at 4.2, 15, 25, 35 and 45 °C. The solution containing only Sos site 4 was used for a series of dilutions to obtain SH3 titration points of 270, 121.5, 100, 75, 50, 37.5, 25, 12.5, 6.25, 3.13, 1.56, 0.781, 0.391 and 0 μM. ^{13}C-δ1methyl Peak assignments are described in Figure S3.2.

NMR titration data analysis

Peptide titration data were fit using lineshape analysis.[70] Uncertainties in $\Delta H_D^{\circ\prime}$, T$\Delta S_{D,298.15K}^{\circ\prime}$, $\Delta H_A^{\circ\prime\ddagger}$, T$\Delta S_{A,298.15\,K}^{\circ\prime\ddagger}$, $\Delta H_D^{\circ\prime\ddagger}$, and T$\Delta S_{D,298.15K}^{\circ\prime\ddagger}$ were assessed by bootstrap analysis, fitting 1000 replicates to the van't Hoff or Eyring equations. Uncertainties represent the standard deviation of 1000 fits.

^{1}H-^{13}C measurements were analyzed using TITAN, with parameter uncertainties estimated from 100 bootstrapped replicas.[192,193] Ile1384, Ile1394, Ile1325 and one pseudo peak were fit as a single spin group for each temperature. van't Hoff and Eyring parameter uncertainties are the 95% confidence intervals of linear fits.

ITC

Each set of protein samples were exchanged using PD-10 midi desalting columns and then dialyzed overnight into the same solution of 20 mM sodium phosphate, pH 7.5 using dialysis cassettes (Thermo Fisher Scientific, 2 kDa molecular weight cut-off). Samples were filtered (Millex, 0.22 μm) prior to measurement and concentrations were validated by measuring the absorbance at 280 nm (Thermo Fisher Scientific, NanoDrop One).

ITC was performed using a MicroCal PEAQ-ITC Automated (Malvern Panalytical), with Sos in the cell and SH3 as the ligand, with one 0.4 μL injection, followed by 19, 2 μL injections. At 4.2 °C, the cell concentration was 250 μM, and the ligand concentration was 3.89 and 3.56 mM for Site 2 and Site 4, respectively. At 35 °C, the cell concentration was 350 μM (site 2 and site 4) and the ligand concentration was 4.67 mM. Knockout measurements were concentration matched and made alongside each measurement described.

ITC data analysis

Data were analyzed with PEAQ-ITC Analysis Software (Malvern Panalytical). To distinguish specific- from non-specific-binding, knockout replicates were used as controls and subtracted from the corresponding Sos replicate. Constructs were fit to a single-site model. ITC parameter uncertainties are the propagation of error from each individual fit.

RESULTS

Multi-temperature NMR datasets for site 4 peptide binding, and one-temperature datasets for site 2 peptide binding are published.[70] We began by completing the site 2 peptide dataset (Figure S3.3 and S3.4) using the same methods. Like the site 4 peptide, equilibrium binding is enthalpically favored and entropically disfavored. Kinetically, both association and dissociation are enthalpically disfavored and entropically favored. There are no important differences between the thermodynamics or kinetics of SH3 binding to the site 2 peptide and the site 4 peptide.

Initial efforts to purify the disordered Sos protein were unsuccessful due to non-specific hydrophobic interactions, causing Sos to co-elute with contaminants. Adding 10% v/v propylene glycol to the buffer solved the problem. We then tried to measure the affinities of the site 2 only- and site 4 only-constructs using ^{19}F NMR titrations in the same manner as performed for the peptide and fit the data to a 2-state model using lineshape analysis across all temperatures simultaneously. This effort was unsuccessful (see Fig. S4.5) because of non-specific SH3 binding (see Discussion), so we turned to ITC.

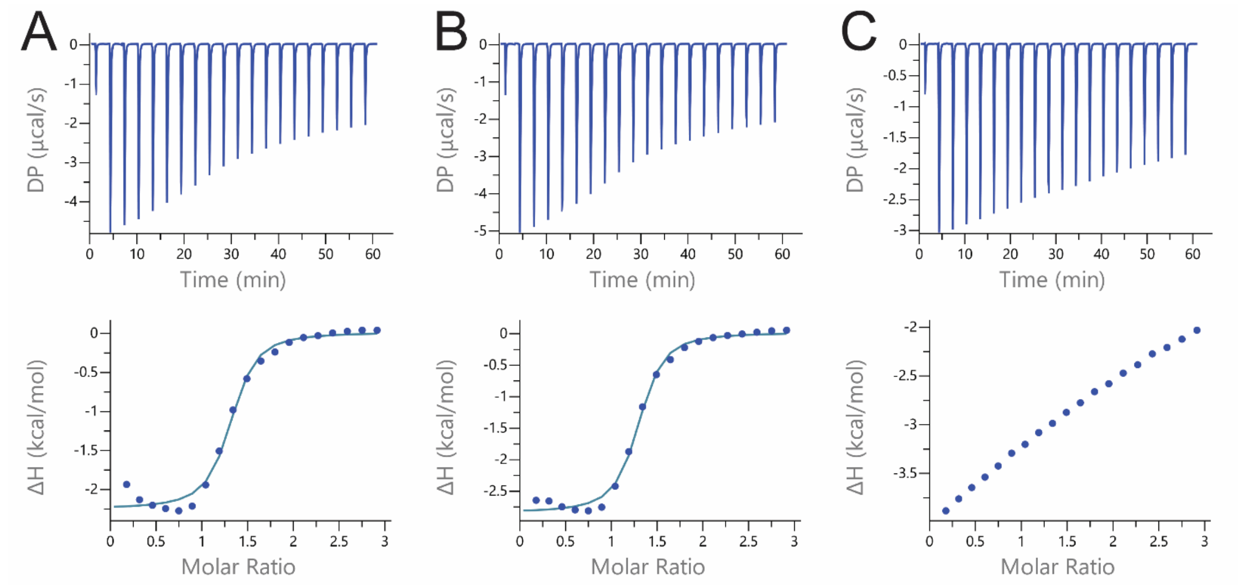


Figure 3.2. ITC thermograms of A) Sos site 2, B) Sos Site 4, and C) Sos knockout at 4.2 °C

ITC measurements of the site 2- and site 4-only constructs showed that the interaction remains unsaturated even after adding >3 molar equivalents, suggesting that the non-specific interactions observed during purification and in the ^{19}F SH3 experiments contribute to weak binding. To isolate the activity of the individual sites, we needed to overcome the heat signal of non-specific binding, leading us to produce a knockout construct where all four specific sites are abolished by amino acid changes. The knockout measurements showed injection heats of the same magnitude as the heat at the end of the titration, where we would expect heats from the single site constructs to approach baseline (Figure 3.2C). We were unable to fit the knockout to a binding isotherm, as the c-value was ≤1.[194]

We used these knockout controls, prepared side-by-side with the construct of interest, to subtract non-specific binding. This approach yielded reasonable fits (Figure 3.2, Table 3.1) and stoichiometries of one SH3 per Sos. Site 2 has an affinity of 4 ±1 µM, and an $\Delta H_D^{\circ\prime}$ of 3.2 ± 0.1 kcal/mol at 4.2 °C. Site 4 has an affinity of 3 ± 1 µM, and an $\Delta H_D^{\circ\prime}$ of 2.8 ± 0.1 kcal/mol at 4.2 °C. At 35 °C, both sites have affinities of 14 ± 3 μM and $\Delta H_D^{\circ\prime}$ values of 8.4 ± 0.3 kcal/mol, indicating a positive $\Delta C_{P,D}^{\circ\prime}$ of dissociation (Table 3.1).

The subtraction does not perfectly account for the non-random order in which SH3 binds Sos. Initially, SH3 prefers the higher-affinity site compared to the weaker, non-specific sites in the knockout. We observe this notion in the fit; an initial over-subtraction, and at the end of the titration, as the stronger binding site becomes fully occupied, more non-specific binding occurs, resulting in better alignment between the data and fit. However, when we compare this approach to direct measurements of site 4 binding via NMR (described below), the methods yield the same information. In summary, our approach is reasonable because the difference in affinities is large, allowing us to treat non-specific binding as random events. If the affinities were closer, our approach would fail.

Table 3.1. Parameter estimates for Sos site 2 and Sos site 4 ITC measurements.

Site	Temperature (°C)	N[1]	K$_D$[1] (μM)	$\Delta H_D^{\circ\prime}$[1] (kcal/mol)	$\Delta G_D^{\circ\prime}$[2] (kcal/mol)	T$\Delta S_D^{\circ\prime}$ [2] (kcal/mol)
2	4.2	1.24 ± 0.03	5 ± 2	2.5 ± 0.1	6.83 ± 0.03	-4.25 ± 0.07
	35	1.21 ± 0.01	14 ± 1	8.9 ± 0.1	6.86 ± 0.01	2.1 ± 0.1
4	4.2	1.23 ± 0.02	3 ± 1	2.9 ± 0.1	6.95 ± 0.02	-4.10 ± 0.03
	35	1.27 ± 0.02	14 ± 3	8.4 ± 0.3	6.84 ± 0.02	1.5 ± 0.1

[1]Uncertanties derived from error propagation of three fits. [2]Uncertainties are the standard deviation of three estimates.

We then reevaluated the ^{19}F IDP binding experiments and tried to fit the data to a bidentate model to represent a single known binding site that we would constrain using the enthalpies measured by ITC, and then allow the second site to represent all weak, non-specific binding observed via ITC in the knockout construct. This regimen was also unsuccessful, suggesting that the interaction from the perspective of SH3 is too complex for simple models. We then sought the simplest way to measure binding from the opposite perspective, i.e., by monitoring the interaction from via Sos. We did this by moving the NMR-active nucleus from SH3 to Sos. Specifically, we isotopically enriched (^{13}C-δ1 methyl) isoleucines in the site 4 protein. The Sos construct has three Iles: I1384 is predicted to be in site 4; I1394 is adjacent to site 4; and I1325 is likely distant the from site 4 binding site. During titration (Figure 3.4), the I1384 cross-peak is most sensitive in terms of chemical shift changes, the I1394 cross-peak undergoes some shift, and the I1325 cross-peak does not change (Figure 3.4A). Analysis yields affinities like those for

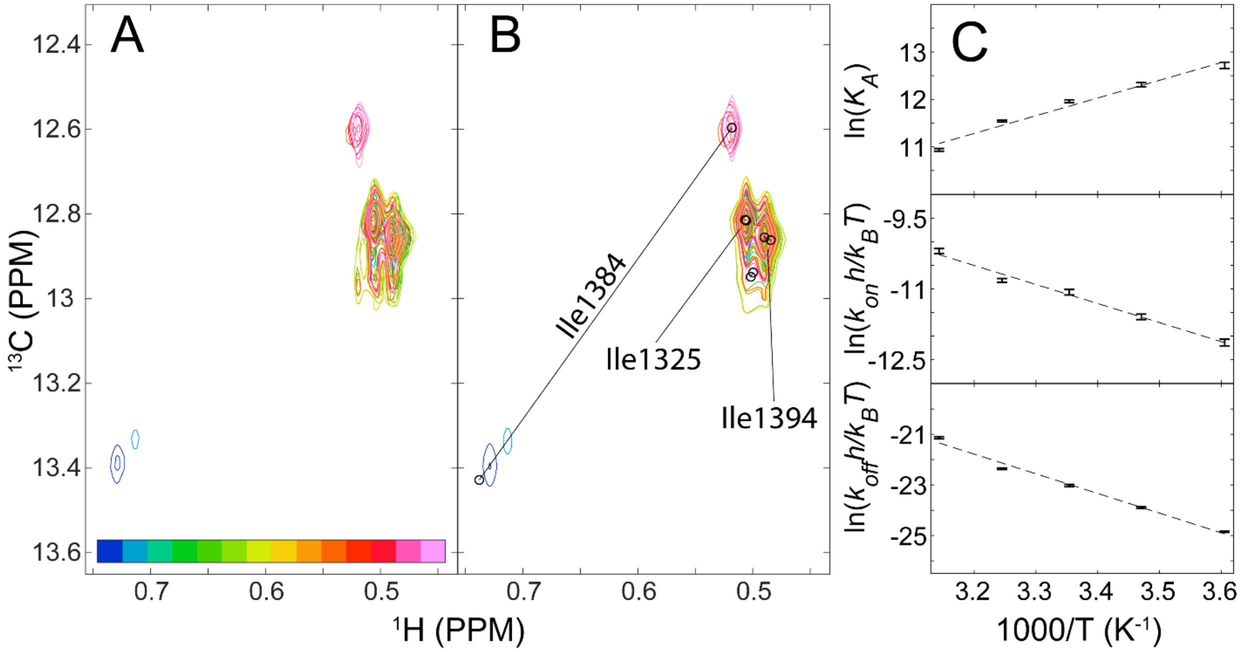


Figure 3.3 ^{1}H-^{13}C HMQC titration of Sos site 4. A) Real spectra at 35 °C. B) Simulated spectra from TITAN fits. C) van't Hoff and Eyring analyses of 1H-13C HMQC titrations.

the site 4 peptide at low temperatures and those from the ITC data (Table 3.2). However, the change in exchange rates as a function of temperature parallels that seen the ^{19}F data (Figure S3.5; S3.6-10) showing that the kinetics, as observed via SH3 and Sos site 4, are similar.

In terms of equilibrium binding, linear fitting of van't Hoff plots yields a $\Delta H_D^{\circ\prime}$ of 7 ± 2 kcal/mol, and a $\mathrm{T}\Delta S_D^{\circ\prime}$ of 0 ± 2 kcal/mol at 298 K. Although we know there is a difference in heat capacity between the free and bound forms of the proteins from the temperature dependence of $\Delta H_D^{\circ\prime}$ (Table 3.1), and there is curvature in the van't Hoff data (Figure 3.3C, S3.11), its magnitude is small, making a linear fit reasonable. In summary, binding is driven by a decrease in enthalpy; the entropic change is small.

Turning to kinetics, we used linear Eyring analysis of k_{on} and k_{off} to characterize the activation parameters. We obtain a $\Delta H_A^{\circ\prime\ddagger}$ of 8 ± 2 kcal/mol and a $T\Delta S_A^{\circ\prime\ddagger}$ of 1 ± 2 at 298 K. The activation enthalpy of dissociation, $\Delta H_D^{\circ\prime\ddagger}$, is 15 ± 3 kcal/mol with a $T\Delta S_D^{\circ\prime\ddagger}$ (at 298 K) of 2 ± 3. The reaction in both directions shows little to no entropic contribution, but must overcome an enthalpic barrier to associate, and a higher enthalpic barrier to dissociate.

In summary, the results from ITC and NMR for SH3-Sos binding are in agreement. The equilibrium enthalpic and entropic contributions to binding are indistinguishable between methods. The small differences in affinities may be due to the addition of D_2O[195] or the deuteration of the γ1-carbon that results from selective enrichment.

Table 3.2. Comparison of NMR parameters and binding Free Energies of Sos site peptides and single sites on Sos protein as a function of temperature.

	Temperature (°C)	K_D (μM)	$\Delta G_D^{o'}$ (kcal/mol)	k_{on} (10^8 $M^{-1}s^{-1}$)	$\Delta G_A^{o'\ddagger}$ (kcal/mol)	k_{off} (10^3 s^{-1})	$\Delta G_D^{o'\ddagger}$ (kcal/mol)
	4.2	9 ± 2	6.5 ± 0.3	0.6 ± 0.2	6.3 ± 0.3	0.486 ± 0.003	12.825 ± 0.007
	15	17 ± 2	6.3 ± 0.1	0.7 ± 0.1	6.5 ± 0.1	1.23 ± 0.02	12.77 ± 0.02
Peptide 2[1]	25	33 ± 3	6.12 ± 0.09	1.1 ± 0.1	6.48 ± 0.09	3.55 ± 0.05	12.61 ± 0.01
	35	59 ± 4	5.96 ± 0.07	1.5 ± 0.1	6.53 ± 0.07	8.88 ± 0.05	12.492 ± 0.006
	45	110 ± 20	6.1 ± 0.1	1.7 ± 0.3	6.7 ± 0.2	17 ± 2	12.5 ± 0.1
	5	20 ± 10	6.2 ± 0.2	0.2 ± 0.06	7.0 ± 0.1	0.2 ± 0.07	13.1 ± 0.1
	15	40 ± 10	6.0 ± 0.3	0.4 ± 0.2	7.0 ± 0.3	0.9 ± 0.1	13.0 ± 0.1
Peptide 4[1,2]	25	60 ± 10	5.8 ± 0.1	0.6 ± 0.1	7.0 ± 0.1	2.9 ± 0.2	12.7 ± 0.1
	35	110 ± 10	5.6 ± 0.1	0.57 ± 0.08	7.1 ± 0.1	5.8 ± 0.3	12.8 ± 0.1
	45	210 ± 30	5.4 ± 0.1	1.1 ± 0.1	7.0 ± 0.1	23 ± 3	12.2 ± 0.1
	4.2	3.0 ± 0.2	7.01 ± 0.07	0.31 ± 0.02	6.69 ± 0.07	0.093 ± 0.002	13.70 ± 0.02
	15	4.5 ± 0.2	7.05 ± 0.04	0.56 ± 0.03	6.64 ± 0.06	0.25 ± 0.09	13.69 ± 0.04
Sos site 4[3]	25	6.4 ± 0.2	7.09 ± 0.03	0.97 ± 0.06	6.56 ± 0.06	0.62 ± 0.03	13.64 ± 0.05
	35	9.7 ± 0.2	7.07 ± 0.02	1.29 ± 0.05	6.62 ± 0.04	1.25 ± 0.04	13.69 ± 0.03
	45	17.8 ± 0.5	6.91 ± 0.03	2.5 ± 0.1	6.45 ± 0.05	4.4 ± 0.2	13.36 ± 0.05

[1] Uncertainties from triplicate analysis. [2] Published.[70] [3] Uncertainties from bootstrap analysis of a single measurement.

DISCUSSION

The affinities of site 4 for SH3 in the peptide and in the IDP are similar at low temperatures and higher at high temperatures. Further analysis reveals two important changes.

First, the $\Delta H_D^{\circ\prime}$ of SH3-peptide binding remains favorable at all temperatures and drives binding. This observation is in line with other studies of SH3-peptide interactions[69,70,196]. Also, the peptide data are well fit by linear van't Hoff analysis (Figure S3.3), suggesting that $\Delta C_{P,D}^{\circ\prime}$ is small. Turning to the disordered protein, we observe a positive $\Delta C_{P,D}^{\circ\prime}$, although the data are too poor for quantification, because the $\Delta H_D^{\circ\prime}$ becomes more favorable with increasing temperature (Table 3.1).

Second, the entropic contribution to binding changes. At equilibrium, the change is favorable at low temperatures and becomes less favorable at higher temperatures (Table 3.1), but is always unfavorable for the peptides (Figure S3.3). If the only factor was the mixing of SH3 and Sos, we would expect unfavorable entropic contributions at all temperatures, as is consistently seen for SH3-peptide binding[69,70,111,197-199]. Analysis of entropic contributions are thorny; however, our observation is consistent with the temperature dependence of IDP hydrodynamic radii, where IDPs become more compact at higher temperatures[59-64,200]. Explanations for this observation point to the release of water, in that the strength of the hydrophobic effect increases with temperature, which is expected to cause collapse[62,65,66,78]. If water is released upon binding at low temperatures, we would expect a favorable change in entropy. At higher temperatures, less water is released, and the penalty of de-mixing is larger. We cannot, however, confirm that Sos undergoes compaction. An alternative explanation is that at lower temperatures, Sos possesses more transient structure, and that structure is disrupted when SH3 binds.

In summary, for peptides, $\Delta G_D^{\prime\circ}$ becomes less favorable as temperature increases but in the IDP, enthalpy-entropy compensation results in a $\Delta G_D^{\prime\circ}$ that is nearly invariant, at least over the temperatures examined. This compensation is potentially beneficial for *D. melanogaster* and other small ectotherms because the invariance would require fewer regulatory mechanisms to respond to temperature changes.

Turning to kinetics, the k_{on} values for the peptides and the IDP are all near the diffusion limit, but the presence of the IDP slows k_{off}. Combining this with the more favorable $\Delta G_D^{\circ\prime}$ of the IDP suggests that the bound state is simply more stable because of contacts with extra residues bordering the binding site. Looking at the activation parameters, $\Delta G_D^{\circ\prime\ddagger}$ and $\Delta G_A^{\circ\prime\ddagger}$ for Sos site 4 do not change with temperature, limiting evaluation using Leffler-like analysis of the transition complex[47]. However, a parsimonious interpretation is that the mechanism does not change over the range of thermal energy studied. It is also reasonable to assume that the changes between the site 4 peptide and the IDP will hold with the site 2 peptide compared to the IDP, given the similarities in energetics and kinetics between the site 2 and 4 peptides (Table 3.2) and the similarities in the IDP sites (Table 3.1).

In summary, we have shown that detailed quantification of extended-IDP interaction is possible. Although measuring the interaction via the ligand (^{19}F-labeled SH3) is challenging because of nonspecific binding, monitoring residues in the binding site (^{13}C enriched Sos) offers a viable alternative allowing the characterization of more IDPs. Looking forward, more kinetic and equilibrium data on IDPs are required to develop a general view of disordered proteins. To build this view, we must combine information on specific interactions as well as weaker, non-specific interactions that may contribute to biological outcomes and play a role in the allovalency of IDPs with multiple binding sites[51,201-203].

Overall, interaction of the stabilized N-terminal SH3 domain of Drk with its native binding sequences is similar mechanistically if the sites are on a short peptide or in the context of the larger disordered protein. However, the additional residues in the IDP cause enthalpy-entropy compensation such that $\Delta G_D^{\circ\prime}$ varies less with temperature compared to the peptide. We also observe that at low temperatures, binding to the IDP is entropically favorable, but becomes less favorable with increasing temperature. Our experimental framework will facilitate the study of more IDP-binding systems and a better understanding of IDPs.

SUPPORTING INFORMATION

^{13}C-δ1methyl Peak assignments

Sos of sevenless Site 4 variants I1325N and I1394L were designed to determine cross-peak identity in the ^{1}H-^{13}C TROSY-HMQC measurements. Template DNA purchased from Gene Universal (Newark, DE) was transformed into DH5α competent cells (Thermo Fisher Scientific, Waltham, MA). A 25 mL culture were grown overnight at 37 °C in Lennox broth and 50 mg/L kanamycin (LB, 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl). The next day, plasmid DNA was harvested using a QIAprep Spin Miniprep kit (Qiagen, Hilden, Germany), following the manufacturer's protocol.

To generate the I1325N variant, the forward primer GTGCGGTGCCGAACAGCC and reverse primer CGTGCGGGCTGTTCGGC were added at 500 nM to a polymerase chain reaction (PCR) using Q5 2X master mix (New England Biolabs, Ipswitch MA),1 ng/uL Site 4 purified template and 2% dimethyl sulfoxide (DMSO) v/v. I1394L was generated using the same recipe, but with forward primer GAACCACAGCACCGGCCTTAGCTAT, reverse primer GACGCAGATAGCTAAGGCCGGT, and with 6% DMSO. PCR carried out following the manufacturer's protocol using a 72°C annealing temperature.

After PCR, samples were treated with 10 U DpnI (Thermo Fisher Scientific) for 1 hour at 37 °C, followed by inactivation at 80 °C for 20 min. PCR success was confirmed by sequencing the plasmid region encoding the entire protein product to ensure the correct mutation and no other changes.

The mutant DNA was then transformed into BL21 cells and 1 L of each was expressed, purified and prepared for NMR as described in the main text. The

I1325N and I1394L variants were measured at 29 and 5 µM. Peaks were referenced to I1384 (present in both variants) and the missing peak was identified as the mutant (Figure S3.4).

^{19}F NMR

Data were acquired using Bruker Avance III HD spectrometers equipped with a QCI cryoprobe (^{1}H Larmor frequencies of 500, and 470 MHz for ^{19}F). Data was processed using NMRpipe (2020.171.18.39) and referenced to trimethylsilylpropanesulfonate to account the temperature dependence of chemical shift.

^{19}F spectra comprised at least 80 scans or until the signal/noise ratio reached 10 (with a line broadening of 2 Hz), a 7062 Hz sweep-width, 1400 complex points and an interscan delay of 2.5 s with a center frequency of -57,646 Hz. Sos site 4 (non-MBD tagged construct) was exchanged into of 50 mM HEPES, 50 mM bis-tris propane, 50 mM sodium acetate pH 7.5 plus 5% (v/v) D_2O using a PD-10 Midi column (Cytiva) selecting the fractions with the highest concentration for the titration. The Sos solution was used to dissolved one aliquot of SH3, and a second, larger volume solution of SH3 was prepared in the same buffer, at an identical concentration of SH3. Measurements were made at 4.2, 15, 25, 35 and 45 °C. The sample was then removed, diluted using the larger SH3 solution and returned to a clean tube. Site 4 titration concentrations were 0, 0.29, 2.9, 7.3, 15, 29, 73, 145, 218, 290, 435, 870 and 1160 µM, with SH3 at 270 µM.

Buffer and Construct Comparison

The Site 2 peptide was measured in 50 mM HEPES, 50 mM bis-tris propane, 50 mM sodium acetate, pH 7.5, as well as the previous work on the

Drosophila melanogaster Sos peptides.[70,71] To determine how buffer conditions might affect the equilibrium, we performed ITC at 4.2 °C with non-MBP tagged Sos as described in the main text, with Sos Sites 2, 4 and knockout at 250 μM, and SH3 at 3.75 mM (Table S3.2). The results show a similar decrease, but smaller in magnitude decrease in enthalpy relative to 20 mM sodium phosphate, with a similar affinity to that of the peptides.

To determine if an additional glycine on the N-terminus of the Sos construct (MBP-tagged Sos results in an extra SMG at the N-term, while the non-MBP tagged construct has an extra MG), we performed ITC 4.2 °C in 20 mM sodium phosphate, pH 7.5, with Sos-SMG, Sos-MG, and knockout controls. Experimental parameters are described in Table S3.2.

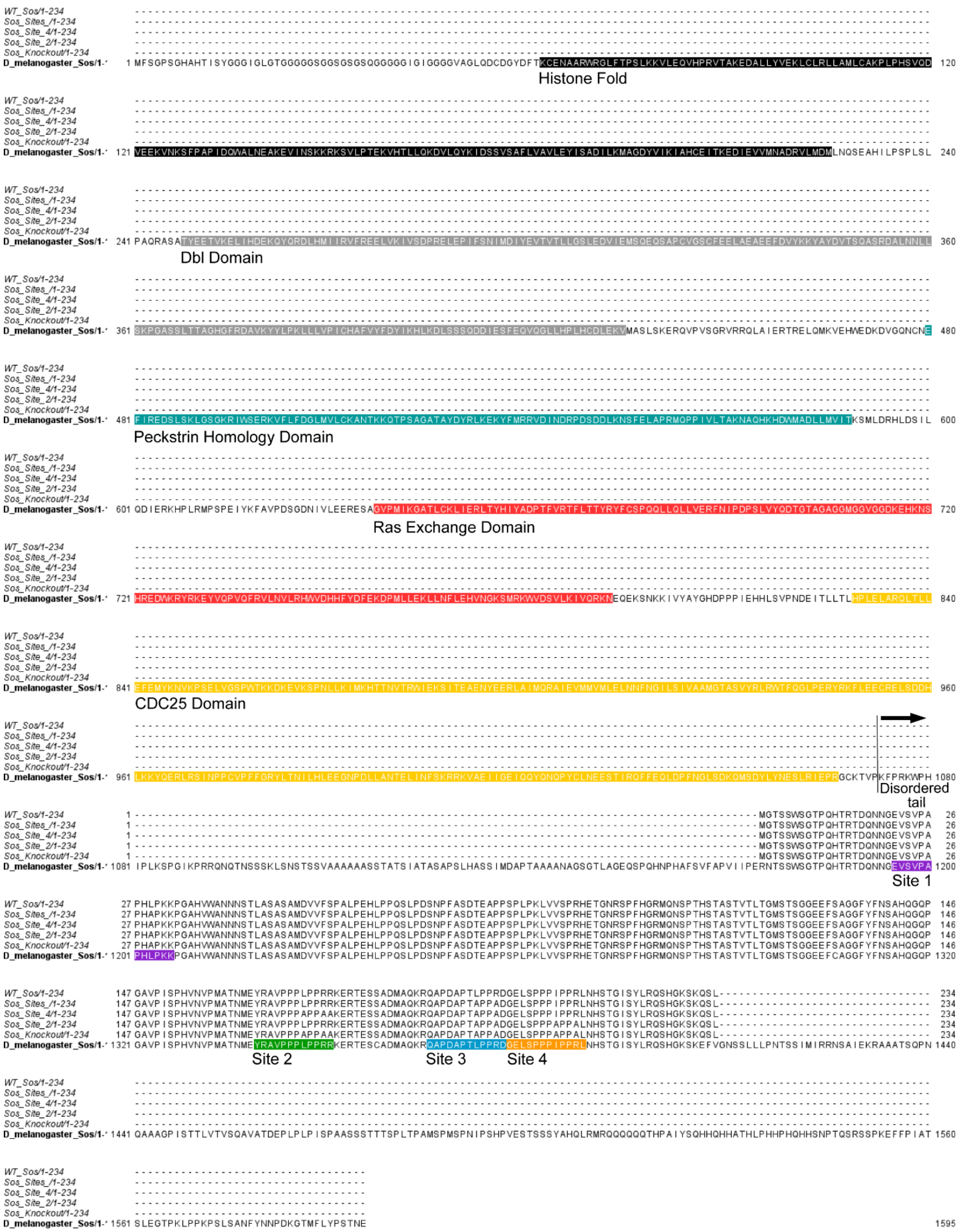


Figure S3.1. Sequence alignment WT Sos, Sos site 2, Sos site 4, Sos Sites 2 and 4, Sos knockout and the D. melanogaster Son of sevenless. Sequence alignment performed in Jalview v2.11.

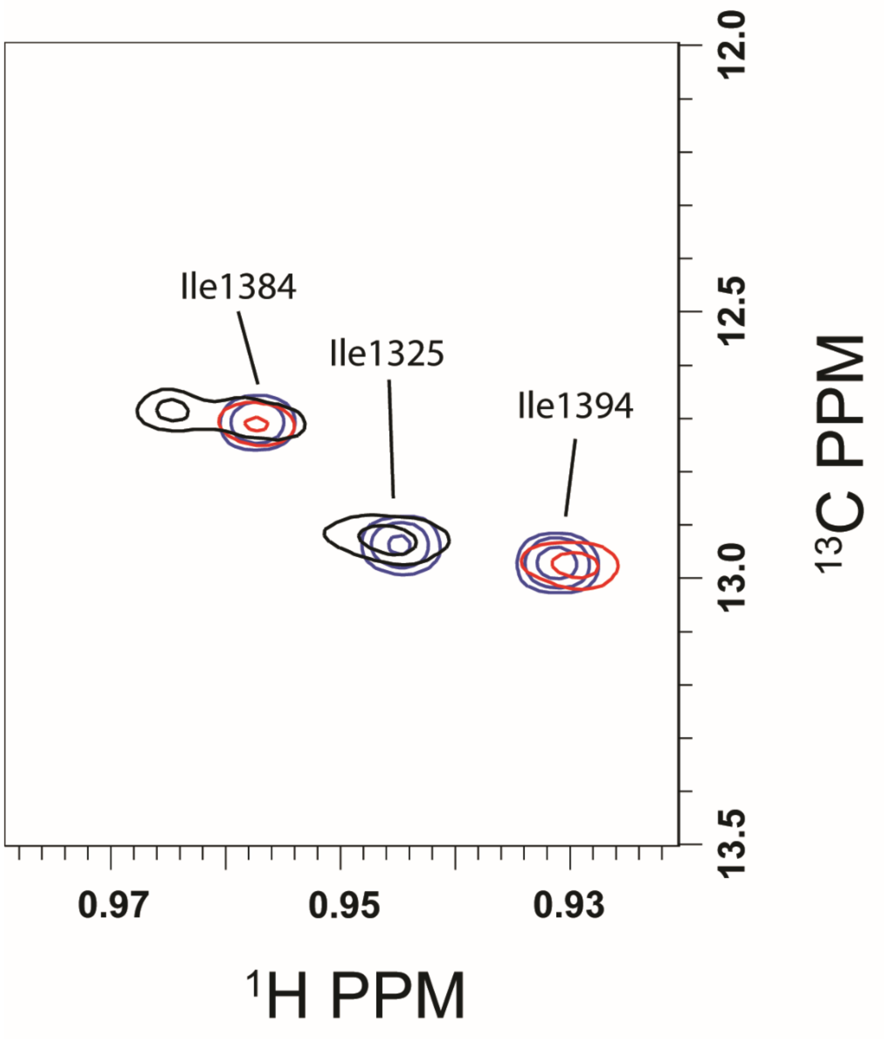
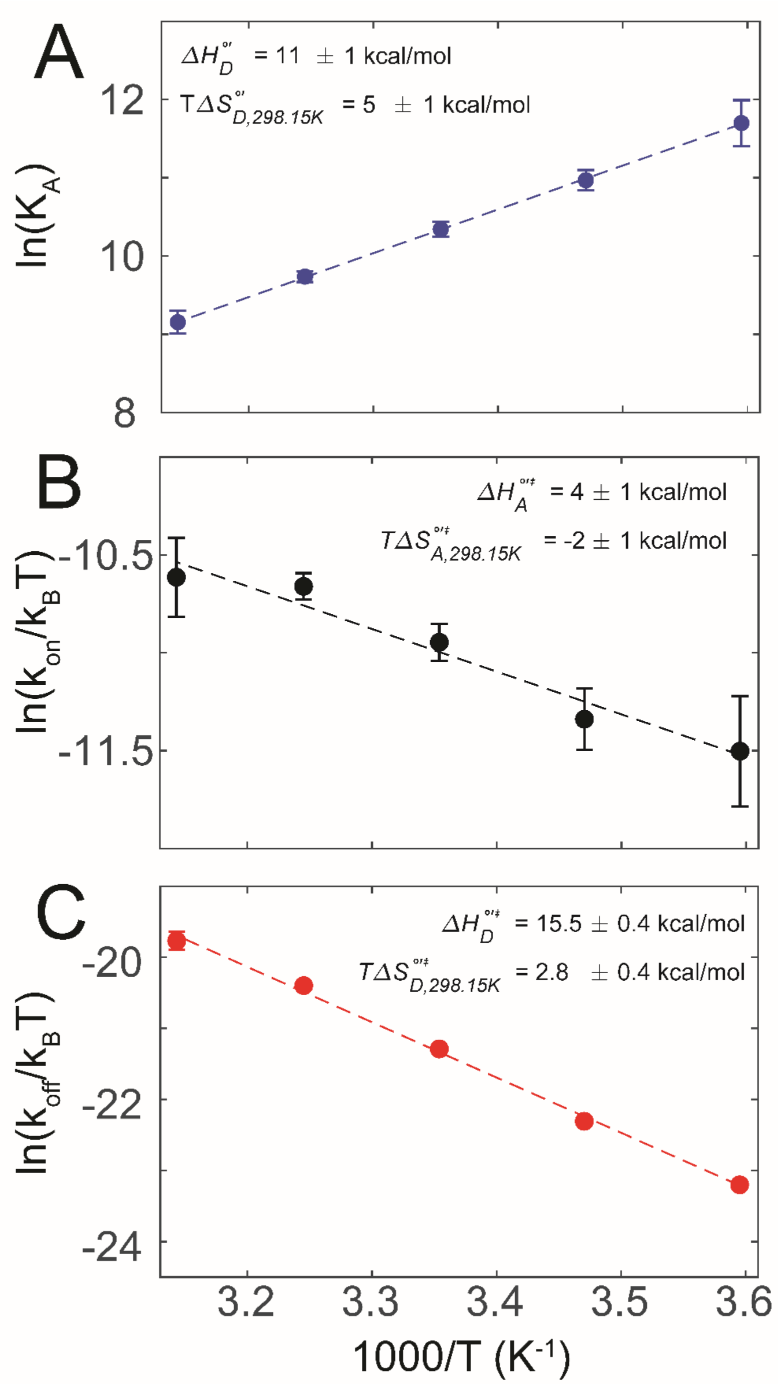


Figure S3.2. Overlay of 13C-δ1methyl isoleucine spectra of Sos site 4 (Blue), Site 4 I1325N (Red) Site 4 I1394L (Black) at 45 °C.



FigureS3.3. Analysis of SH3 - Sos site 2 peptide binding. A) van't Hoff analysis B) Eyring plots of association C) and dissociation. Error bars are the SEM of triplicate measures. Values for enthalpy and entropy are the mean and standard deviation of bootstrap analysis. Dotted lines show fits of the bootstrapped parameters.

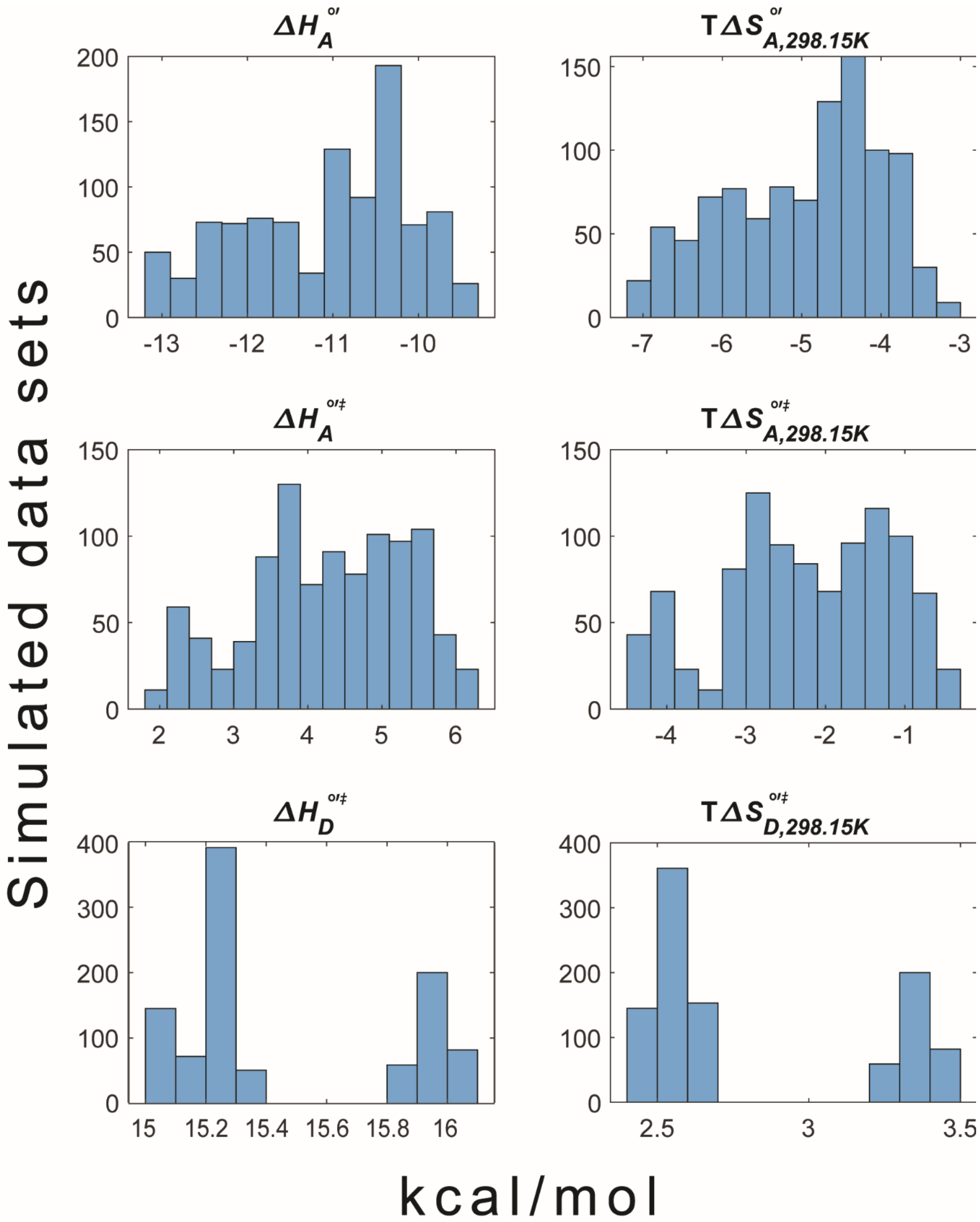


Figure S3.4. Distributions of parameters for 1000 bootstrapped fits to the A) Van't Hoff equation B) Erying equation for association and C) Eyring equation for dissociation.

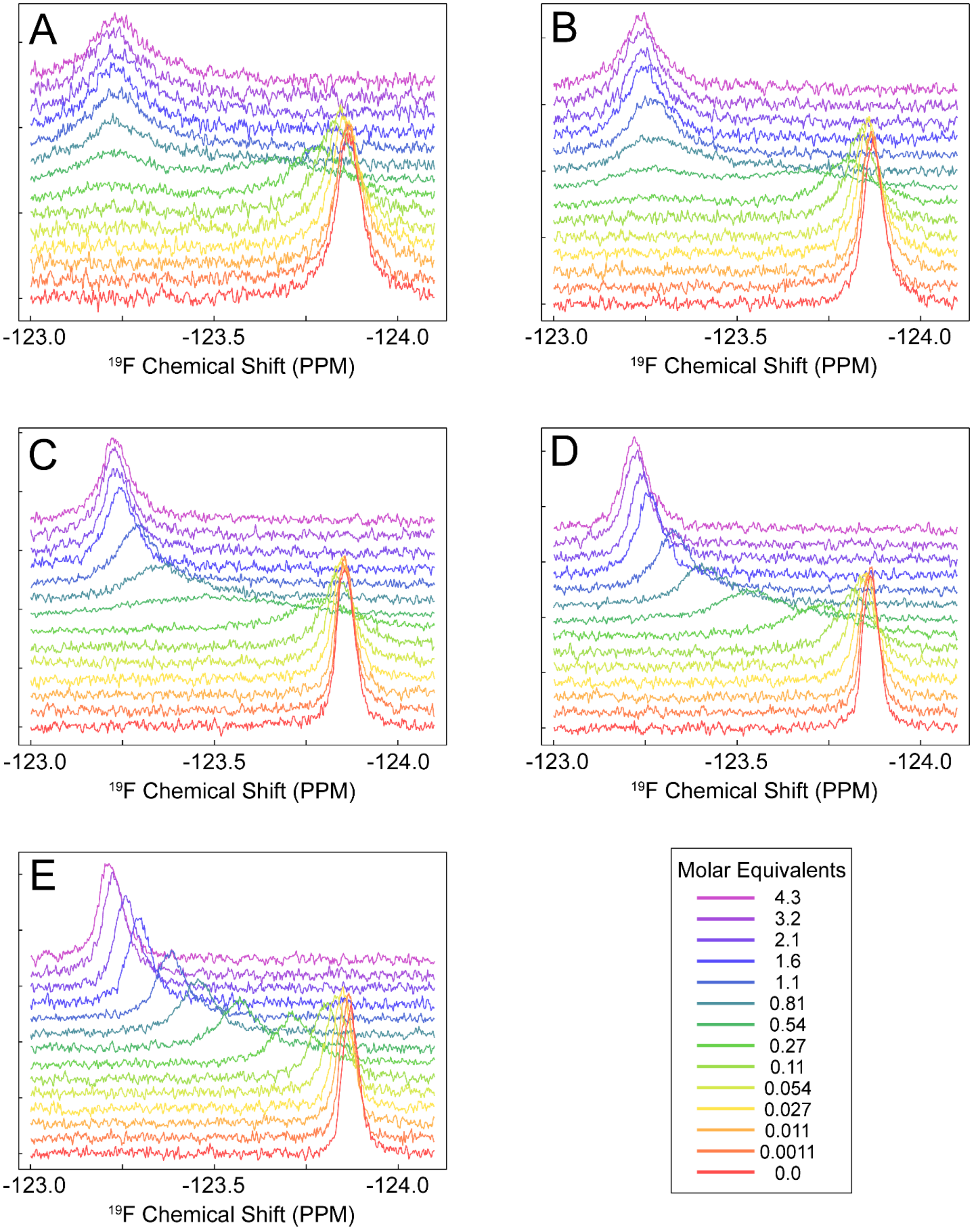


Figure S3.5. 19F NMR titration of Sos-SH3 binding. Titration was performed at A) 4.2 °C, B) 15 °C, C) 25 °C, D) 35 °C, E) 45 °C.

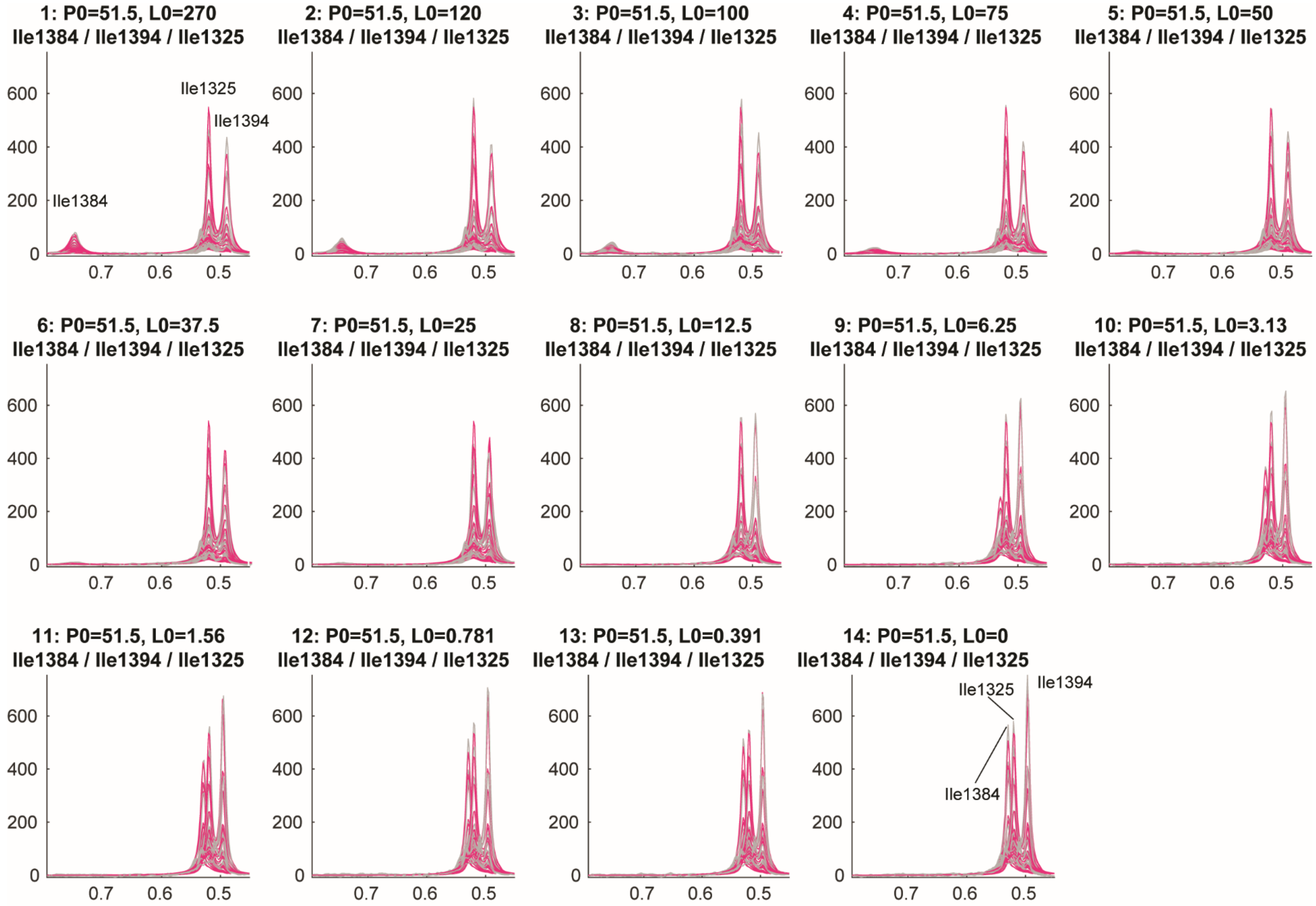


Figure S3.6. TITAN fit of Sos site 4-SH3 binding, 5 °C. Overlay of real (grey) and simulated (pink) spectra from lineshape analysis. ^{1}H-^{13}C titration of Sos site 4. X-axis is ^{1}H chemical shift in PPM.

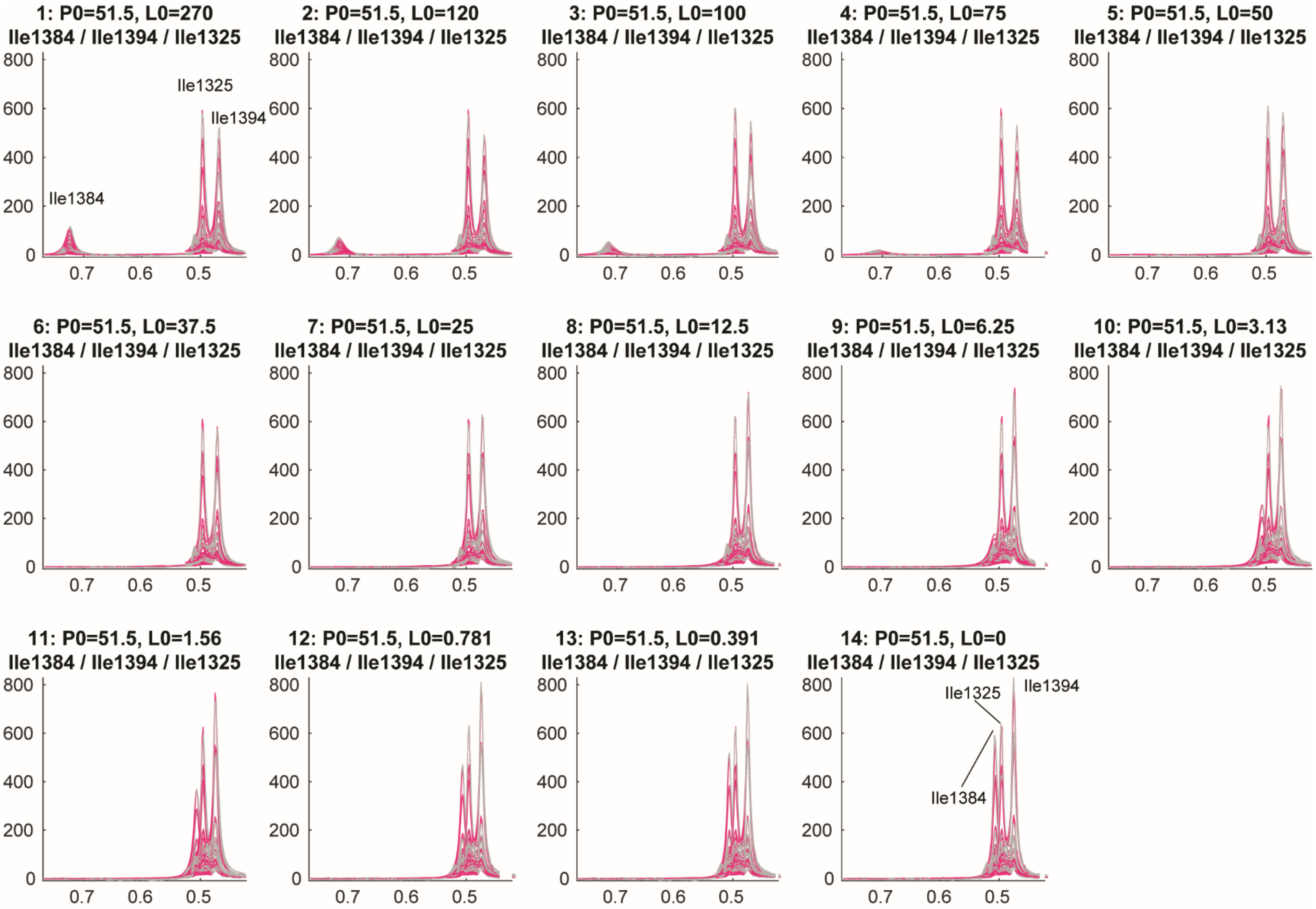


Figure S3.7. TITAN fit of Sos site 4-SH3 binding, 15 °C. Overlay of real (grey) and simulated (pink) spectra from lineshape analysis. ^{1}H-^{13}C titration of Sos site 4. X-axis is ^{1}H chemical shift in PPM.

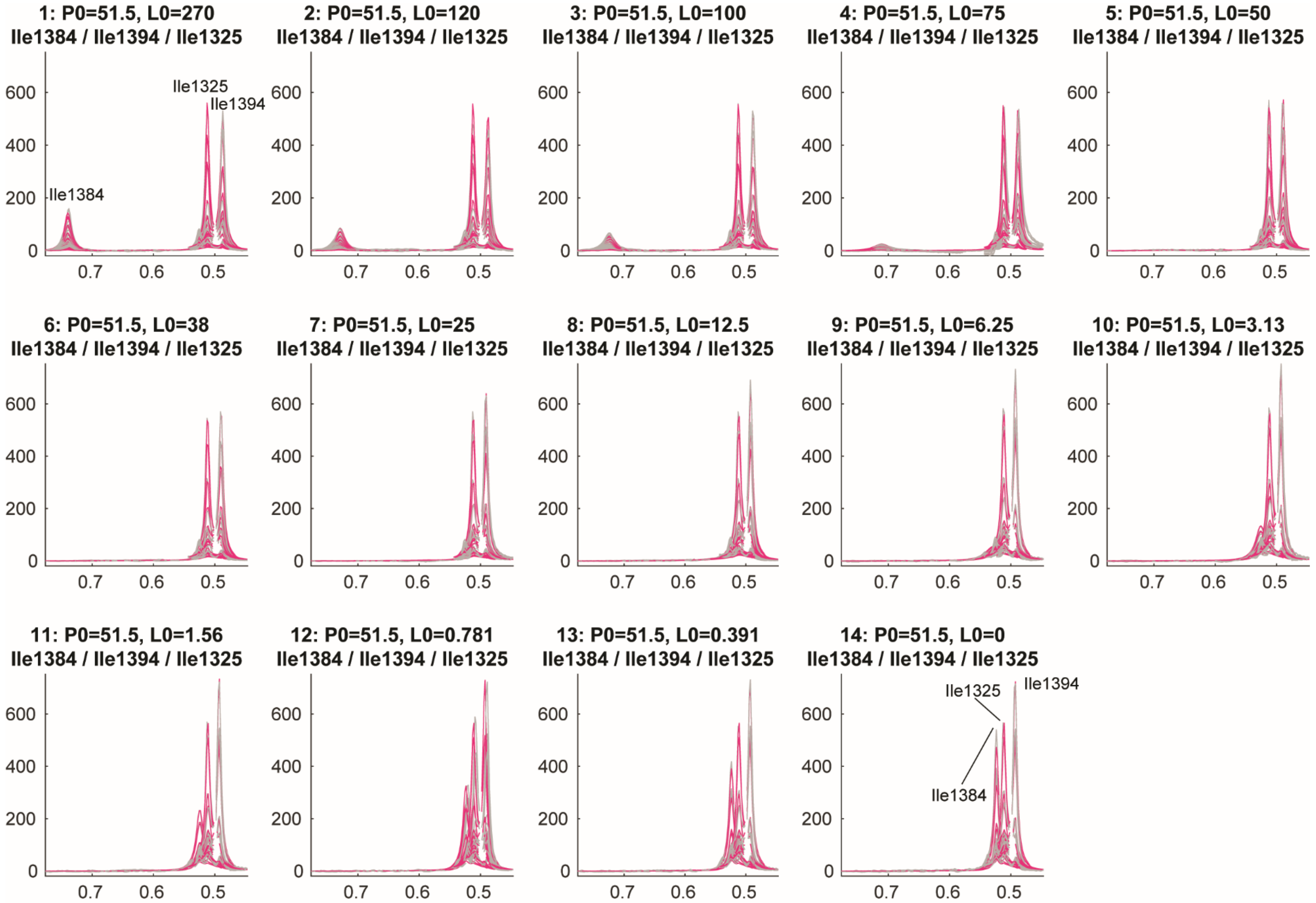


Figure S3.8. TITAN fit of Sos site 4-SH3 binding, 25 °C. Overlay of real (grey) and simulated (pink) spectra from lineshape analysis. ^{1}H-^{13}C titration of Sos site 4. X-axis is ^{1}H chemical shift in PPM.

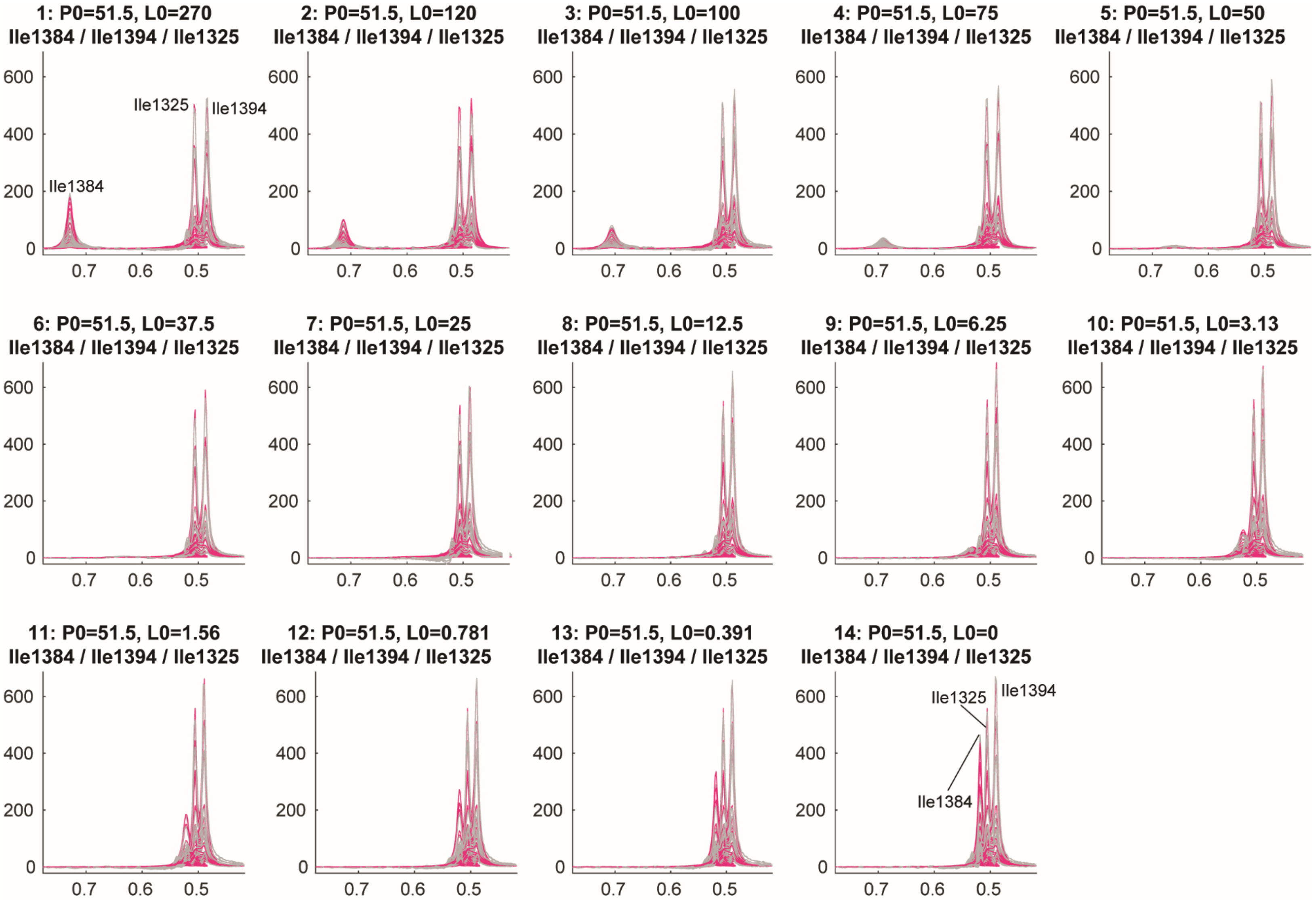


Figure S3.9. TITAN fit of Sos site 4-SH3 binding, 35 °C. Overlay of real (grey) and simulated (pink) spectra from lineshape analysis. ^{1}H-^{13}C titration of Sos site 4. X-axis is ^{1}H chemical shift in PPM.

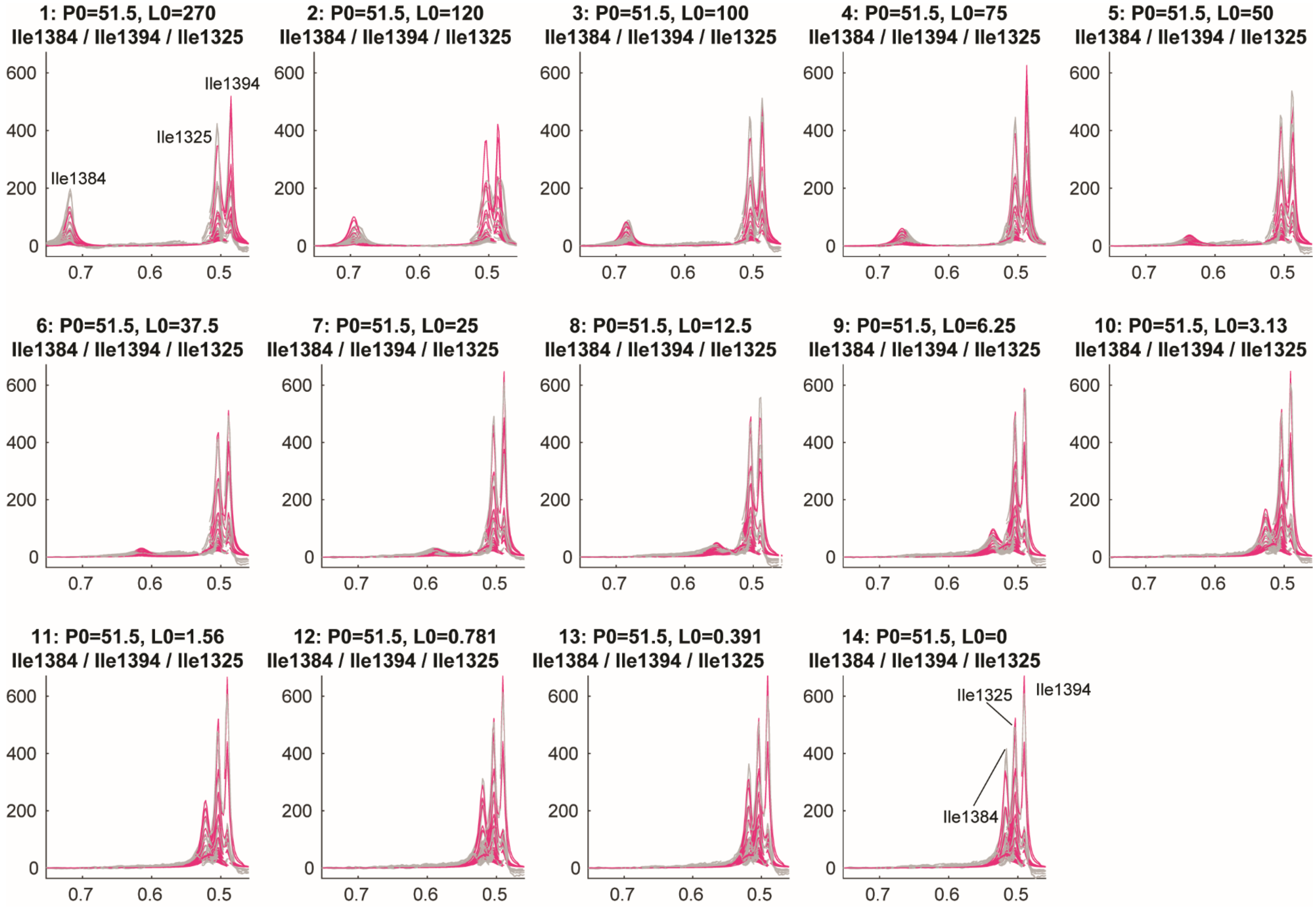


Figure S3.10. TITAN fit of Sos site 4-SH3 binding, 45 °C. Overlay of real (grey) and simulated (pink) spectra from lineshape analysis. ^{1}H-^{13}C titration of Sos site 4. X-axis is ^{1}H chemical shift in PPM.

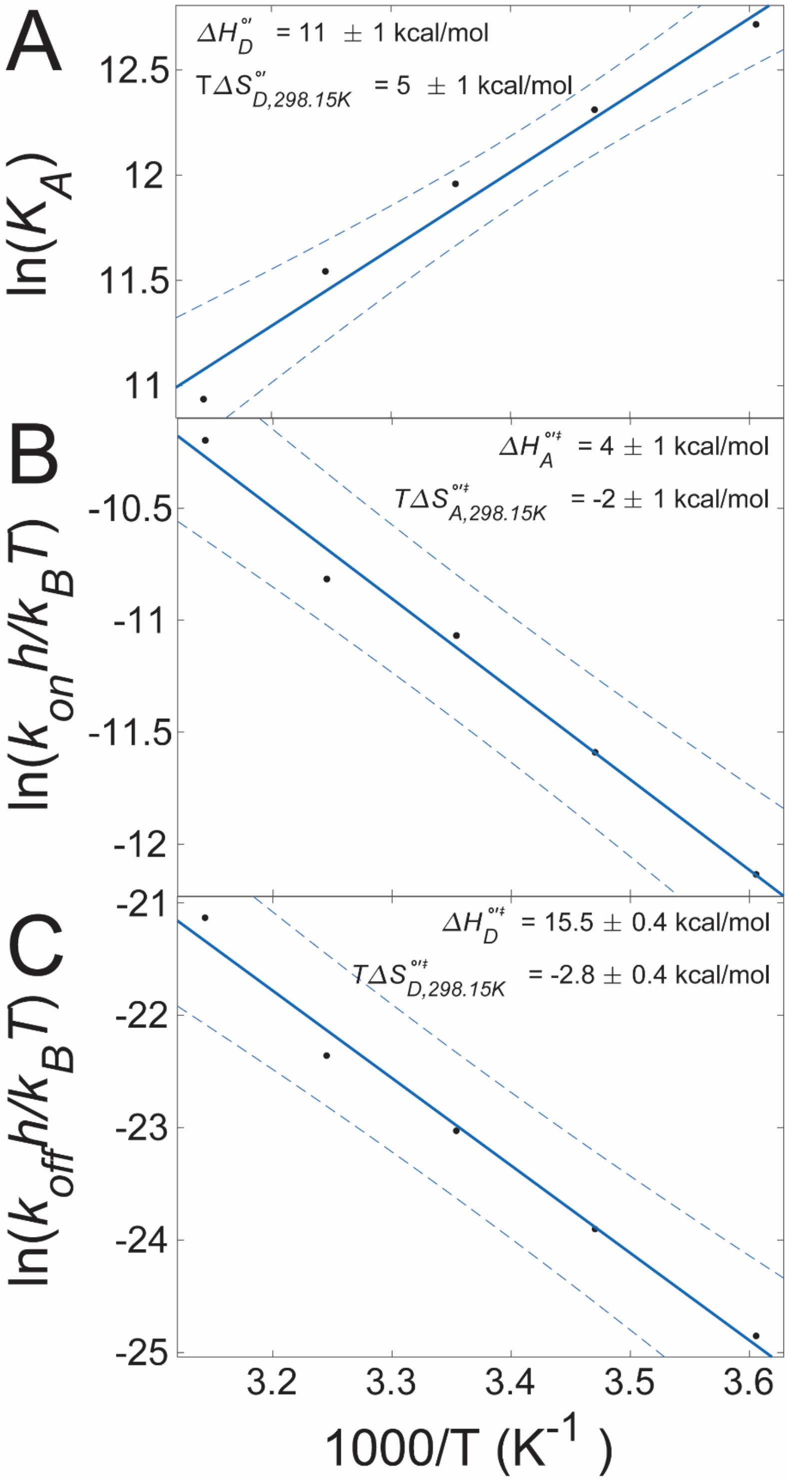


Figure S3.11. Fits of ^{1}H-^{13}C δ1methyl isoleucine lineshape analysis parameters. A) van't Hoff analysis B) Eyring plots of association C) and dissociation. Dotted lines indicated 95% Confidence intervals of the fit, with uncertainties derived from the 95% confidence intervals of each fit parameter.

Table S3.1. Results of elemental analysis.

	SOS site 2 - MW 24,616.23 g/mol					
	Number of atoms	Atomic weight (g/mol)	Total Mass (g/mol)	Theoretical %	Replicate 1	Replicate 2
H	1655	1.008	1668.24	6.78%	7.02%	7.15%
C	1068	12.011	12827.748	52.11%	46.88%	47.00%
N	317	14.007	4440.219	18.04%	15.99%	16.11%
O	341	15.999	5455.659	22.16%		
S	7	32.06	224.42	0.91%	0.87%	0.87%

Table S3.2. Theoretical and measured masses of proteins.

Protein	Theoretical Mass (Da)	Mass observed (Da)	Difference (Da)
SH3	6833.56	6832.9	-0.66
Sos site 2	24,703.24		
Sos site 4	24,618.14	24617.4	-0.74
Sos knockout	24,480.95	24,490.5	-0.45

Table S3.3. Comparison of ITC results by Sos construct and buffer.

Sos construct		Temperature (°C)	Buffer	[Sos] (mM)	[SH3] (mM)	N	K_D (μM)[1]	$\Delta H^{o'}_D$ (kcal/mol)[1]	$\Delta G^{o'}_D$ (kcal/mol)[2]	$T\Delta S^{o'}_D$ (kcal/mol)[2]	Offset[1]	Sign of $\Delta C_{P,D}$
2	MG	4.2	HBA[3]	0.250	3.75	1.08 ± 0.01	9 ± 1	4.96 ± 0.08	6.43 ± 0.05	-1.5 ± 0.2	-0.46 ± 0.03	positive
	MG		NaPi[4]	0.229	4.00	1.23 ± 0.03	7 ± 2	3.2 ± 0.1	6.56 ± 0.03	-3.41 ± 0.07	-0.53 ± 0.06	
	SMG		NaPi[4]	0.229	4.00	1.24 ± 0.03	5 ± 2	2.5 ± 0.1	6.76 ± 0.03	-4.25 ± 0.07	-0.26 ± 0.06	
	SMG	35.1	NaPi[4]	0.229	4.00	1.21 ± 0.01	14 ± 1	8.9 ± 0.1	6.86 ± 0.01	2.1 ± 0.1	0.00 ± 0.05	
4	MG	4.2	HBA[3]	0.250	3.75	1.10 ± 0.03	8 ± 4	4.7 ± 0.3	6.5 ± 0.3	-1.8 ± 0.8	-0.4 ± 0.1	positive
	SMG		NaPi[4]	0.250	3.56	1.23 ± 0.02	3 ± 1	2.9 ± 0.1	6.95 ± 0.02	-4.10 ± 0.03	-0.47 ± 0.07	
	SMG	35.1	NaPi[4]	0.350	4.67	1.29 ± 0.02	14 ± 3	8.4 ± 0.3	6.84 ± 0.02	1.5 ± 0.1	-0.5 ± 0.2	

[1] Uncertanties derived from error propagation of three fits. [2] Uncertainties are the standard deviation of three estimates. [3] 50 mM HEPES, 50 mM bis-tris propane, 50 mM sodium acetate, pH 7.5. [4] 20 mM sodium phosphate, pH 7.5

CHAPTER 4: *DANIO RERIO* OOCYTES FOR EUKARYOTIC IN-CELL NMR

Reprinted with permission from Biochemistry 2021, 60, 6, 451–459. Copyright 2021 American Chemical Society

INTRODUCTION

For over a century biochemistry focused on the properties of proteins and other macromolecules in buffer alone. Recently, investigators have begun to explore proteins[32,204,205] and RNA[206,207] in the environment where they evolved, in the cell. This crowded and complex milieu is predicted to stabilize proteins and their complexes due to hardcore repulsions,[25] although shape also plays a role.[23,35,208] Simultaneously, soft interactions— charge-charge, dipole, non-polar, and others— can be stabilizing or destabilizing, depending on whether they are repulsive or attractive, respectively.[209] While it is easy to advocate measurements in cells, realistically, these are challenging endeavors requiring a large investment of both time and supplies. Our goal is to develop simple, practical, and economical systems that are broadly applicable to a variety of questions about the effect of the cellular interior on proteins.

Quantifying in-cell biochemistry of specific target proteins requires a label. Fluorescence microscopy is commonly used for studing macromolecular localization, abundance, dynamics and interactions.[210-213] These efforts provide key insight into how the cellular environment impact tertiary- and quaternary- structure and changes in macromolecular assemblies. However, a potential drawback is that most these efforts rely on the introduction of large (0.1 kDa to >10 kDa) fluorescent labels. These labels can potentially interfere with the behavior of the protein being studied.[214]

Nuclear magnetic resonance spectroscopy (NMR) avoids large labels while providing atomic-resolution information on structure, stability and dynamics of macromolecules in living cells.[171,207,215-221] Serber et al.[215] showed the utility of *Escherichia coli* as a model organism for in-cell NMR because strains optimized for recombinant protein expression are useful for enriching proteins with NMR-active nuclei and allow observation of target proteins in the sea of other cellular molecules. However, many proteins are difficult or impossible to observe by in-cell NMR in *E. coli* because surface residues make non-specific contacts with other macromolecules, slowing their tumbling and broadening their resonances into the background.[222-224]

This disappearance of target protein signals in *E. coli* and some eukaryotic cells emphasizes the need for systems that allow a larger range of macromolecules to be assessed. Several techniques have been employed to introduce isotopically enriched- or labeled- proteins into eukaryotic cells, including, electroporation,[171,225] importation-tags,[217] expression,[217,226,227] and pore-forming toxins.[228] These methodologies, however, do not allow facile access to- and accurate control of- target protein, a property critical for the future quantification of protein-protein interactions.[229,230]

Selenko et al.[216] bypassed these limitations by microinjecting the model B1 domain of the Streptococcal immunoglobulin G-binding protein (GB1) into *Xenopus laevis* oocytes. Oocytes are useful for studying the effects of the cytoplasmic milieu on proteins because they are fixed in development, which minimizes changes due to the cell cycle and growth.[231,232] Sánchez-López et al. used transparent early-stage *Danio* *rerio* (zebrafish) embryos to assess enzyme-activity *in vivo*.[233] To exploit the transparency of zebrafish, while maintaining the stable cellular environment, we employ unfertilized zebrafish oocytes to assess their utility for in-cell NMR. Here, we validate the detectability and cytoplasmic location of a test protein (Figure 4.1), the 7-kDa N-terminal src homology 3 domain (SH3) domain of the *Drosophila melanogaster* signal transduction protein Drk protein, and then quantified its stability and dynamics.

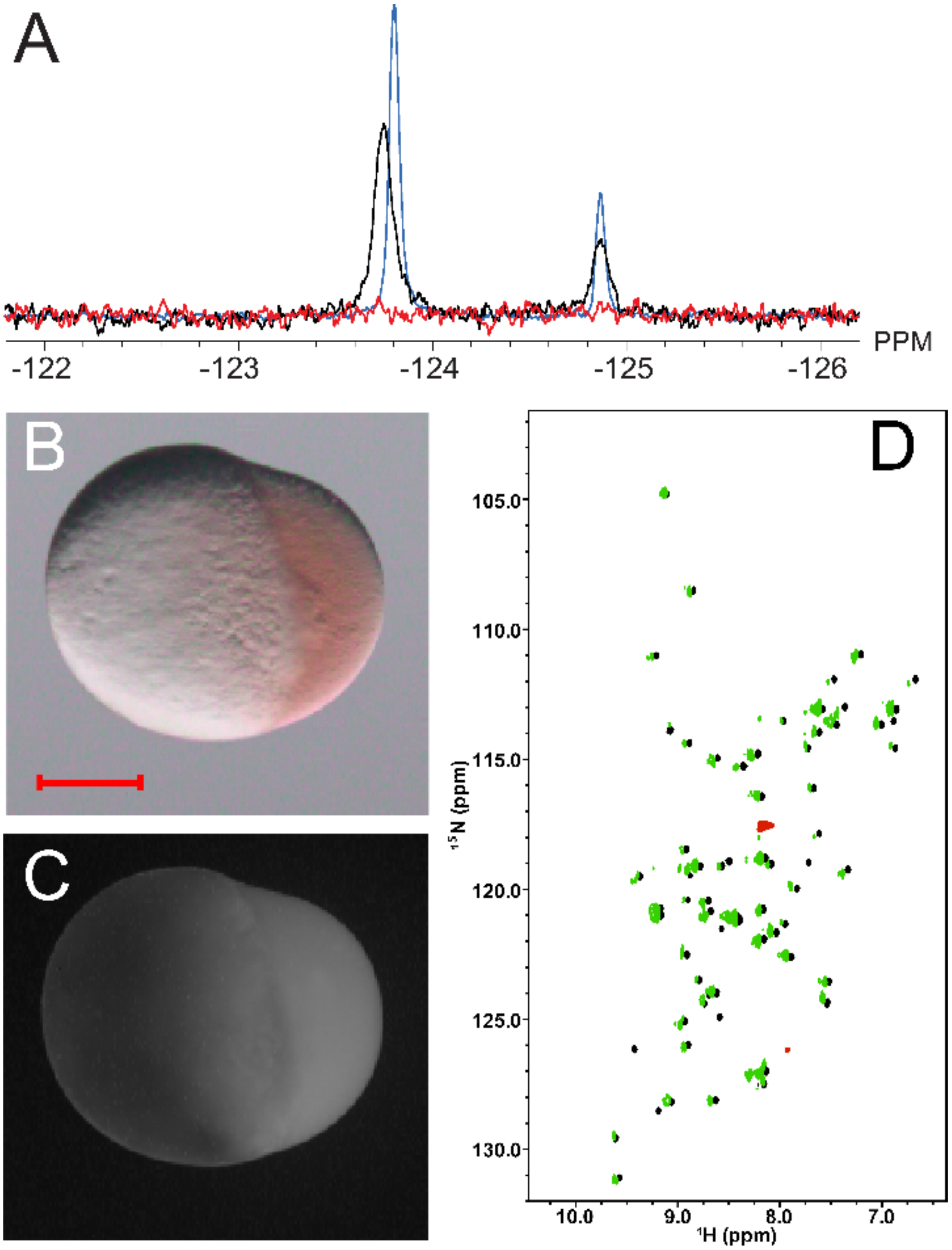


Figure 4.1. Representative NMR spectra and images of SH3 *in vitro* and in oocytes. A) Representative ^{19}F-NMR spectra at 298 K of ^{19}F labeled wildtype SH3 in buffer (blue), in oocytes (black) and the 2:1 dilution leakage control (red). The folded state and unfolded state chemical shifts are at approximately -123.7 and -124.85 PPM respectively. B) Brightfield image of a *Danio rerio* oocyte injected with 1 mM T22G;C60 SH3 covalently tagged with bromobimane. Scale bar: 200 μm. C) The same oocyte imaged with epifluorescence (excitation 436 nm - 420 nm; emission, 480 nm - 440 nm). D) ^{15}N-^{1}H HSQC spectra of ^{15}N enriched T22G SH3 in *D. rerio* oocytes (green, 62.5 μM), *E. coli* (red, ~1 mM), and in buffer at pH 7.5 (black, 490 μM). The solution surrounding the oocytes was interrogated by NMR, but no SH3 resonances were observed.

MATERIALS AND METHODS

Protein preparation

Fluorine-labeled SH3 was prepared as described,[34,195,234] except that expression was allowed to proceed for 12-14 h at 18.5 °C. Fluorine-labeled N51K SH3 was expressed in the same way, except it was cloned into a pET28b (Novagen) vector with an N-terminal hexa-histidine- and SUMO- tags, transformed into BL21 Star (DE3) One Shot cells (Invitrogen). cOmplete Protease Inhibitor Cocktail (Protease Inhibitors, Roche) was added to the cell pellets, and the samples sonicated (Fisher Scientific Sonic Dismembrator Model 500, 20% amplitude, 0.4 s on, 0.2 s off, 15 min). Protease Inhibitors were added again, the lysates centrifuged (Sorvall RC-5B, 17,540 rcf, 4°C, 45 min) and then syringe filtered (0.45 µm, Millex).

Lysates were processed by FPLC (ÄKTA pure) using a Ni^{2+} column (Cytiva HisTrap HP) pre-equilibrated with loading buffer (15.1 mM Na_2HPO_4, 4.9 mM NaH_2PO_4, 500 mM NaCl, 20 mM imidazole, pH 8). The column was washed with 50 mL loading buffer. The protein was eluted with 15.1 mM Na_2HPO_4, 4.9 mM NaH_2PO_4, 500 mM NaCl, 200 mM imidazole, pH 8. Fractions of interest were combined with Cth protease (final concentration 1-2 nM).[235] Digestion proceeded while the sample was dialyzed against 50 mM Na_2HPO_4, 20 mM NaH_2PO_4, 9 mM NaCl for 4-6 h. The digested sample was subjected to Ni^{2+} chromatography as described above. The flow-through was collected, and the protein further purified using size-exchange and hydrophobic-interaction chromatography.[234] Purity was assesses by using SDS PAGE and mass spectrometry (Thermo Q Exactive HF-X, SH3 - 6877.6 Da expected, 85% 6873.3 Da and 15% 6855.3 observed; N51K SH3—6891.7 Da expected, 90% 6891.2 Da and 10%

6873.4 Da observed) and ^{19}F NMR. The lower mass species are not ^{19}F-labeled. Wildtype SH3 yielded 3-6 mg of purified protein /L of culture. N51K SH3 yielded 6-10 mg/L.

Purified samples were buffer exchanged using PD-10 desalting columns (Cytiva) into *D. rerio* Ringer's solution (Ringer's solution, 116 mM NaCl, 2.9 mM KCl, 1.8 mM $CaCl_2$, 5 mM Hepes, pH 7.4).[236] SH3 was concentrated to >13 mM using a Millipore Amicon Ultra concentrator (3 kDa cut off). Protease inhibitors were added to 10x final concentration, and phenyl red dye was added to 0.5% weight/volume. Final protein concentrations were 11.3 mM for SH3 and 13.4 mM for N51K SH3. Samples were flash frozen and stored at -80 °C.

^{15}N-enriched K10H GB1,^{15}N-enriched/^{19}F-labeled T22G SH3 and T22G/C60 SH3 were prepared as described[190,204] and their purity validated by SDS PAGE and mass spectrometry (K10H GB1, 6300.3 Da expected, 6296.8 Da observed; T22G SH3, 6903.97 Da expected, 86.7% 6904.0 Da and 11.4% 6886.5 Da observed). The T22G/C60 SH3 variant was prepared as described above, except that 1 mM dithiothreitol was added to buffers (6918.6 Da expected, 6918.3 Da observed). Samples were flash frozen and lyophilized for storage. Prior to injection, the proteins were dissolved in *D. rerio* Ringer's solution with 10x Protease inhibitors containing 0.5% phenyl red (final concentration).

Bromobimane fluorescence

T22G;C60 SH3 was mixed with bromobimane at a 1:10 molar ratio in 15.1 mM Na_2HPO_4, 4.9 mM NaH_2PO_4, 150 mM NaCl, 5 mM EDTA, pH 7.6. The reaction was allowed to proceed for 12 h at 4°C. The protein was separated from unreacted dye by using a PD-10 column and exchanged into H_2O for lyophilization.

Zebrafish oocyte injections

AB strain adult female zebrafish were set up in breeding tanks 1 d prior to experiments. The following morning, the fish were anesthetized in 0.08% Tricaine (3-amino benzoic acid ethyl ester) in E3 (5 mM NaCl, 170 μM KCl, 330 μM $CaCl_2$, 330 μM $MgSO_4$, 0.00003% methylene blue).[236] Oocytes were gently squeezed from the fish using a plastic spoon to distribute the force. Dead or mis-shapened oocytes were removed. Oocytes were loaded onto 3% agarose plates for injection when a prominent blastodisc was observed (~25 min).

Protein injection solution was loaded into the back of a glass capillary needle. The point of the needle was clipped. The air injection system (Parker Picospritzer III) was set to 30 psi. The injection bubble diameter, 210 μm, measured in mineral oil using a stage micrometer, corresponds to a 4.5 nL injection. Injections were made through the yolk to just below, or into, the blastodisc. Assuming a cytoplasmic volume of 50 nL (see ***Microscopy***), the final cytoplasmic concentration of injected protein was 1.2 - 1.4 mM.[216]

Five- to seven- hundred oocytes were injected per trial. The oocytes were allowed to recover in E3 in a 55-mm petri dish for at least 1 h. Oocytes damaged by injection, or those where the injection solution was not taken into the cytoplasm (e.g. phenyl red localized in the yolk, rather than the blastodisc) were removed. Between 300 and 550 of the healthiest-looking oocytes were transferred to E3, 1.5% Ficoll 70 (Sigma), pH 7.5. This Ficoll concentration compressed the blastodisc within 30 min without damaging the oocytes. Oocytes were transferred to a Shigemi NMR tube, pre-filled with E3, 1.5% Ficoll 70, pH 7.5, to prevent damage to the injected oocytes. The total number of oocytes added to the tube was recorded, and additional solution was removed until the volume was 350 µL. The solution was then adjusted to E3, 1.5% Ficoll 70, 10% D_2O using E3, 1.5% Ficoll 70, 90% D_2O, pH meter reading 7.5. This study was carried out under the UNC-CH Institutional Animal Care & Use Committee under protocol 18-004.0.

Comments on injection

There are three primary determinants. The most important determinant is oocyte quality. An unhealthy fish yields few, if any, healthy oocytes. Additionally, the first two or three collections from a newly-mature fish, or a multi-week break in laying can result in poor quality oocytes. The second determinant comprises the injection parameters. An injection air-pressure between 20-30 PSI maximizes the injection volume without damaging the oocyte, and a narrow needle minimizes oocyte damage. The third factor is the operator skill. Novices learning needle placement have a success rate of about 50%, which reaches 80-90% with experience.

Identifying dead oocytes and leakage controls

NMR tubes were inspected for dead oocytes after data acquisition. Dead oocytes are bright white and therefore quickly counted. Data were rejected if 10% or more of the oocytes died. To assess leakage, the solution surrounding the oocytes was removed using a Pasteur pipette. The volume was measured and adjusted to 350 uL with E3, 1.5% Ficoll 70, 10% D_2O. The leak-check sample was subjected to the same acquisition parameters as the oocyte sample at 298 K. Signals from the protein of interest were not detected.

Microscopy

Zebrafish oocyte preparation, collection and injection were carried out as described above. Lyophilized, T22G SH3 and bromobimane-labeled T22G;C60 SH3 were resuspended in Ringer’s solution with phenyl red to give a final protein concentration of 10 mM. The final intracellular protein concentration was ~1 mM. Additional injections were made using only Ringer’s solution with phenyl red. Epifluorescence images (Figures 4.1B and 4.1C, Supporting Information 1) were acquired using an Axiozoom.V16 stereomicroscope (Zeiss) equipped with an Axiocam 503 color (Zeiss) after allowing injected oocytes to recover in E3 for 1 h (excitation 436 nm - 420 nm; emission, 480 nm - 440 nm). Oocytes were placed on a 150-mm petri dish for imaging, with chorions mechanically removed with forceps.

Cytoplasmic volume estimates were made by imaging 30 fully segregated oocytes (≥4 h post collection) using a M205FA stereomicroscope (Leica) equipped with DFC360FX camera (Leica). The methodology was adapted from Fuentes et al.[237] Ovals were drawn over blastodisc and yolk region using Adobe Illustrator. Oval radii were measured in the vertical and horizontal directions. The cytoplasm and yolk were assumed to be half-spheroids. When the yolk occupied a portion of the cytoplasmic volume, the yolk volume was subtracted from the cytoplasmic volume (Figure S4.2).

Laser-scanning confocal microscopy (LSCM) was used to observe redistribution of injected bromobimane labeled T22G;C60 SH3. Injections were performed as described above. The final intracellular protein concentration was 0.5 mM. A LU-

N4/N4S (Nikon Instruments, Inc.) laser combiner, acousto-optical tunable filter, and a solid-state, 15-mW 405-nm laser (25% intensity) was used to excite the bromobimane label. An A1R galvanometric mirror confocal scan head (Nikon Instruments, Inc.) operated in “galvo mode” was used to scan the excitation light and de-scan the emission light through a 90.67-µm hexagonal pinhole (physical size). An inverted 10X CFI Plan Fluor air-immersion 0.3 NA objective-type lens element (Nikon Instruments, Inc.) was used to deliver the scanned excitation light and collect the emission light through a #1.5 glass coverslip in a closed-top chamber to prevent media evaporation. The dimensions of the captured images were 1024 x 1024 pixels at a pixel size of 1.243 µm. A three-dimensional volume was acquired using 35 optical slices at 15-µm step sizes via an objective Ti ZDrive (Nikon Elements). Fluorescence was detected through a 450/50 filter cube (Chroma Technology Corporation) using a Multi-Alkali photomultiplier tube housed in a A1-DUG-2 Multi Detector Unit (Nikon Instruments, Inc). Images were acquired and three-dimensional reconstructions were made using NIS-Elements acquisition software (Nikon Instruments, Inc).

NMR acquisition

Data were processed using Topspin 3.6.2. Measurements were made on Bruker Avance III HD spectrometers equipped with QCI cryoprobes, (^{1}H Larmor frequencies of 500 or 600 MHz, 470 MHz or 564 MHz, respectively, for ^{19}F), or TCI cryoprobes, (^{1}H Larmor frequencies of 700 or 850 MHz). ^{19}F measurements (Figures 4.1A and S4.3) comprised 280 scans of 7000 Hz sweep width, 7000 points per acquisition with a relaxation delay of 4 s. Measurements were made from 301 K to 286 K for in-cell experiments and from 318 K to 280 K for *in vitro* experiments with equilibration times of

10 min (Figure 4.2). For in-cell experiments of the SH3 N51K variant, 80 pulses were added to each progressive temperature point. Buffer measurements were made in Ringer's solution + 10% D_2O + 0.05% phenyl red (w/v), pH 7.4 with ^{19}F-labeled SH3 and N51K SH3 concentrations between 0.2 - 0.3 mM. Reversibility was shown by returning to 298 K at the end of each buffer experiment (stability within 10%). ^{15}N-^{1}H HSQC spectra of the T22G SH3 variant in oocytes were acquired with 13600 Hz and 3000 Hz sweep widths in the ^{1}H and ^{15}N dimensions, respectively, and 128 increments of 20 acquisitions at a ^{1}H Larmor frequency of 850 MHz.

500-MHz HSQC spectra in *E. coli* were acquired with sweep widths of 8000 Hz and 2280 Hz in the ^{1}H and ^{15}N dimensions, respectively, and 130 indirect increments of 32 acquisitions. To assess leakage, cells were gently centrifuged and the measurement repeated on the supernatant. No signals from our proteins of interest were detected (Figure S4.4).

To acquire the pH titration curve, ^{15}N-enriched K10H GB1[238] was dissolved in modified *D. rerio* Ringer's solution (described above, with the addition of 5 mM bis-tris propane and 5 mM citrate) at pH values of 6.0, 6.2, 6.6, 7.0, 7.3, 7.7, 7.9, 8.1, 8.4, and 8.7 using a GB1 concentration 0.4 mM. HSQC spectra were acquired using sweep widths of 11,200 Hz and 2500 Hz in ^{1}H and ^{15}N dimensions, respectively, and 128 indirect increments of 8 acquisitions at a Larmor frequency of 700 MHz. Spectra were referenced to the invariant ε ^{15}N-^{1}H Trp43 cross peak.[238]

Intracellular pH measurements were made following the injection protocol for ^{15}N-enriched GB1 K10H as described above, with a final cellular concentration of 1 mM. Measurements were made at 301.15 K through 286.15 K in 3 K increments. Oocytes were held at each temperature for 30 min to mimic the ^{19}F-NMR wildtype SH3 measurements. ^{15}N-^{1}H HSQC spectra of the K10H GB1 variant in oocytes were acquired identically, but with a single increment of 264 acquisitions. Spectra were referenced to the invariant ε ^{15}N-^{1}H Trp43 peak (Figure 4.3).[238]

^{19}F spin lattice relaxation (T_1) times were measured at 292.15 K using a signal inversion recovery sequence with delay times of 0.00, 0.05, 0.10 (triplicated), 0.25, 0.50, 0.80, 1.00, and 1.500 s. ^{19}F transverse relaxation (T_2) times were measured at 292.15 K using a Carr-Purcell-Meiboom-Gill (CPMG) sequence with delays of 0.84, 1.69 (triplicated), 3.38, 5.06, 6.75, 9.28, 11.82, 15.19, 19.41, and 25.32 μs in oocytes, and additional delays of 33.75, 46.42, and 63.3 μs for buffer measurements. Effects of chemical exchange were limited by using an effective field of 5000 Hz. Buffer values were acquired using 520 μM solutions of wildtype or N51K SH3. No leakage was detected. Measurements were made at ^{19}F Larmor frequencies of both 470 and 564 MHz.

NMR processing, fitting and analysis of uncertainties

NMR data were analyzed in Topspin 3.6.2 (Bruker). ^{19}F free-induction decays were zero filled to 64 K points and processed with 5 Hz broadening. Integrals were measured manually. Further analysis was performed using MATLAB R2020a. For protein stability measurements, ^{19}F data were fit to the integrated Gibbs-Helmholtz equation.[33] Uncertainties were analyzed by generating 1000 data sets, resampling from measured values with replacement. Averages and standard deviations represent values from population fits.[239]

Correlation times (τ_C) values were fit using the Model Free formalism,[240,241] globally fitting T_1, T_2 data measured at ^{19}F frequencies of 470 and 564 MHz, with the folded-state order parameter (S^2) fixed to 0.82 and the effective correlation (τ_s) fixed to 20 ps.[242] To show that the relative changes observed were independent of the absolute values of S^2 and τ_s, we fixed each parameters within a range between 0.5—1.0 and 10—100 ps, respectively (Table S4.2).

Uncertainties in T_1-, and T_2- values, and correlation times (τ_C) values were estimated as described.[32] Briefly, for T_1 and T_2, the standard deviation from a triplicate measurement was applied to all measurements, then a Monte Carlo analysis (N=1000) was performed to estimate a mean and standard deviation. For τ_C, the mean and standard deviations of T_1, T_2 were used in a Monte Carlo (N=1000) analysis to estimate the mean τ_C and its standard deviation.

RESULTS AND DISCUSSION

As stated in the Introduction, test protein spectra are often undetectable in cells. To show that SH3 can be observed after microinjection, we compared the signal from ^{15}N-^{1}H HSQC spectra of the stabilized SH3 T22G variant[189] in *D. rerio* oocytes and *E. coli* (Figure 4.1C). Almost every backbone and side chain amide crosspeak is observed in oocytes, and broadening does not limit identification of nearby crosspeaks. This situation contrasts with the spectrum from inside *E. coli*, where only metabolites are observed,[243] but the protein spectrum appears upon lysis and dilution (Figure S4.4).[244]

The observation of spectra in oocytes and its absence in *E. coli* arises from the potential for chemical interactions in each cytoplasm. Prokaryotes have cytoplasmic protein concentrations of 200 to 450 g/L,[16,245] resulting in a large number of non-specific protein-protein interactions that slow protein tumbling.[224,244,246] The concentration of eukaryotic cytoplasmic proteins is about half that in *E. coli*.[17] Amphibian and fish oocytes are even more dilute, with concentration of 30-60 g/L.[18,20-22,245] This lower concentration makes oocytes an ideal system for in-cell NMR because a broader selection of macromolecules, in terms of both size and charge, will be quantifiable.

The unfolded population of wildtype SH3 in buffer is significant, even under non-denaturing conditions.[32-34,188,247] To measure SH3 stability, we introduced one fluorine atom into its sole tryptophan by adding 5-fluoroindole during expression in *E. coli.*[248,249] SH3 undergoes two-state folding.[188] The rate of exchange between the folded and unfolded states is small relative to the frequency difference of the ^{19}F resonances from each state (i.e. slow-exchange),[32,250] which means the areas of the resonances are directly proportional to their relative concentrations. The stability of the protein can therefore be quantified as the modified (neutral pH) standard-state free energy of unfolding, $\Delta G_U^{\circ\prime}$, where *R* is the gas constant and *T* is the absolute temperature.

$$\Delta G_U^{\circ\prime} = -RTln\frac{[unfolded\ population]}{[folded\ population]} \qquad (1)$$

Plotting $\Delta G_U{}^{\circ\prime}$ as a function of temperature allows us to use the integrated Gibbs-Helmholtz equation (Equation 2) to estimate the modified standard-state enthalpy ($\Delta H_U^{\circ\prime}$)-, entropy ($\Delta S_U^{\circ\prime}$)-, and heat capacity ($\Delta C_p^{\circ\prime}$)- of unfolding at a reference temperature (T_{ref}).[251] This model assumes that ΔC_P is constant over the narrow temperature range tested.[251,252]

$$\Delta G_U^{\circ\prime} = H_U^{\circ\prime} - T\Delta S_U^{\circ\prime} + \Delta C_P^{\circ\prime}\left[T - T_{ref} - Tln\frac{T}{T_{ref}}\right] \qquad (2)$$

We measured SH3 stability as a function of temperature in buffer and oocytes (Figure 4.2, Table 4.1, Figure S4.2). Due to viability limitations, measurements were confined to between 286 K and 301 K. This choice limited the ability to quantify ΔC_P, but we expected small changes to this parameter in cells and therefore used the buffer value[32] to fit the in-cell data.

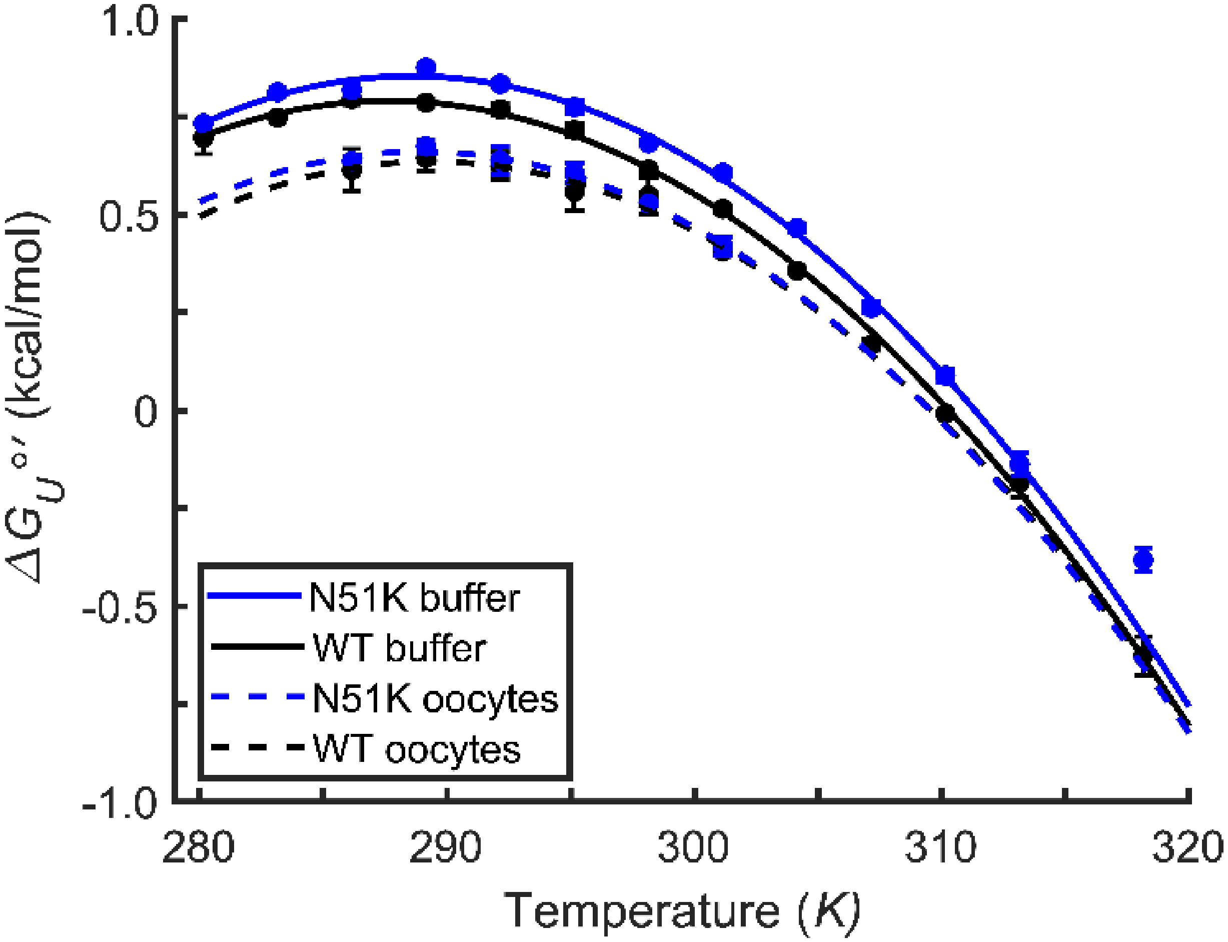


Figure 4.2. Temperature dependence of wildtype (WT) SH3 and the N51K variant stability in buffer (pH 7.5) and *D. rerio* oocytes. Averages and standard errors of the mean are shown along with fits to Equation 2.

Table 4.1. Equilibrium thermodynamic parameters for SH3 unfolding.

		T_m (K)	$\Delta H_{U,\mathrm{Tm}}$ (kcal/mol)	T_s (K)	$\Delta H_{U,\mathrm{Ts}}$ (kcal/mol)	ΔC_P (kcal/(mol·K))
WT	Buffer	310.1 ± 0.3[a]	21.6 ± 0.7	287.5 ± 0.7	0.80 ± 0.01	0.92 ± 0.06
	Oocytes	310 ± 1	19.4 ± 0.5	290 ± 2	0.64 ± 0.03	
N51K[b]	Buffer	311.3 ± 0.2	23.1 ± 0.6	288.5 ± 0.3	0.857 ± 0.007	0.97 ± 0.04
	Oocytes	309 ± 0.5	20.1 ± 0.2	289.0 ± 0.6	0.66 ± 0.02	

[a] Uncertainties are from standard deviation of parameter estimates using resampling. See Supporting Information for details.

SH3 is less stable in oocytes than it is in buffer. The destabilization is enthalpically driven (Table 4.1, Figure S4.5) because the temperature of maximum stability (T_s) is unchanged from the buffer value, but $\Delta H^{\circ\prime}_{TS}$ decreases. Additionally, the melting temperature (T_m) remains constant while $\Delta H^{\circ\prime}_{U,Tm}$ decreases. The decreased stability is caused by attractive chemical interactions between the surface of SH3 and the cellular milieu.[31,253] These attractive interactions are destabilizing because protein unfolding exposes additional attractive surface. This decreased stability in oocytes is smaller in magnitude, but similar in direction to the destabilization of the same protein in *E. coli*.[32] The observation that the more crowded *E. coli* cytoplasm and the less crowded oocyte cytoplasm are both destabilizing indicates that chemical interactions may be more important than hard core repulsions in cells.

Destabilizing attractive contacts include charge-charge interactions. The average net charge of *D. rerio* oocyte proteins is negative at physiological pH,[254,255] as it is in prokaryotes,[256] while SH3 has a net charge of -6. Simple analyses based on steric- and net charge- repulsion predict a that SH3 should be stabilized in cells compared to buffer. The destabilization shows the importance of these other attractive interactions in cells.

Nevertheless, charge still plays a key role as shown by our studies of the N51K variant, which changes the charge by +1 (Table 4.1, Figure S4.6). Given the net negative protein charge in cells, this change should further destabilize SH3 in cells compared to buffer by increasing attractive interactions. The prediction is supported; the variant is more destabilized ($\Delta\Delta G_U^{\circ\prime}$) at T_s compared to the wildtype protein. We attempted to study other charge-change variants (see Supplemental Materials). They were either aggregation-prone, clogging the needle, or did not follow two-state folding in oocytes.

Another important concern is diffusion. Several measurements have been reported in biologically relevant crowders *in vitro*,[71,257] in *E. coli*,[220,258] and in *X. laevis* oocytes.[220] We estimated the relative viscosity in zebrafish oocytes by measuring the ratio of the rotational correlation time (τ_C, Equation 3) in oocytes compared to that in buffer. In Equation 3, k_B is the Boltzmann constant, η is the viscosity ($\text{kg}\cdot\text{m}^{-1}\cdot\text{s}^{-1}$), and V is the volume of SH3 (m^3), which is known.[71]

$$\tau_C = \frac{6\eta V}{k_B T} \qquad (3)$$

Given that the cellular environment is unlikely to change the shape of a folded globular protein, the ratio of τ_C values equals the ratio of viscosities. τ_C was estimated via the Model-Free approach using spin lattice- and transverse- relaxation times (Table 4.2).[240-242]

Table 4.2. Rotational correlational times (τ_C, ns/radian) of folded SH3 and its variant in buffer (pH 7.5) and *Danio rerio* oocytes.

	Buffer τ_C[a,b]	In-cell τ_C[a,b]	Relative change[c]
WT	4.4 ± 0.1	11 ± 2	2.5 ± 0.5
N51K	5.3 ± 0.1	14 ± 3	2.6 ± 0.6

[a] Order parameter (S^2) fixed to 0.82. τ_e fixed to 20 ns.[242]
[b] See supplemental for discussion of uncertainty.
[c] Uncertainties from error propagation.

Both wildtype SH3 and the charge change variant N51K perceive a viscosity about 2.5 times greater in oocytes than in buffer. This ratio is similar to that for a small protein in *X. laevis,* but the viscosity perceived in *E. coli* is about 7-fold greater than buffer.[220] The faster diffusion in oocytes compared to *E. coli* reflects the higher macromolecule concentration in bacteria. The similar viscosities of *X. laevis* and zebrafish oocyte cytoplasms indicate that protein diffusion is comparable across oocytes from these two distantly related species. Computational studies support the idea that protein-protein chemical interactions are the primary cause of the increased viscosity compare to buffer.[259,260] The observation that changing the charge does not affect the relative viscosity of the wildtype protein and the charge-change variant shows that interactions beyond simple electrostatics are important.

We could not quantify the relative viscocity experienced by unfolded SH3 in oocytes because we had to choose a temperature that allowed acquisition of reliable data over the multi-hour measurement times. At that temperature (293 K), the population of unfolded state is too small to obtain the requisite data. Increasing the temperature increases the population of the unfolded state but it also increases the rate of oocyte death.

Another consideration is ensuring that the in-cell environment reflects typical cell function. One measure is intracellular pH (pH_i), which is tightly regulated by cells, and alterations are associated with breakdowns in homeostasis.[261,262] Our protein, homogenously occupies the cytoplasm (Figure 4.1, S4.1, Associated Content AVI) and therefore a similar measure of cytoplasmic pH_i is appropriate. The K10H variant of GB1 has been characterized as an in-cell pH probe.[238] A fit of the H10 ^{15}N-^{1}H crosspeak across a range of pH values to the Henderson-Hasselbalch equation yields a pK_a of 7.2 at 25 °C (Figure 4.3A). We accounted for the temperature-dependence of ionization by combining this information with ionization enthalpy values of surface histidines.[263] Using this correction, the pK_a was estimated for each temperature, in the same order of acquisition as was performed for SH3 and N51K SH3 stability measurements (Figure 4.3C). Initially the cytoplasm is slightly alkaline relative to a typical eukaryotic cytoplasm pH_i,[262] as reported for zebrafish embryos,[264] and then falls to 7.3-7.2 within 1 h. These data indicate that zebrafish oocytes are a reasonable model for a eukaryotic cytoplasm.

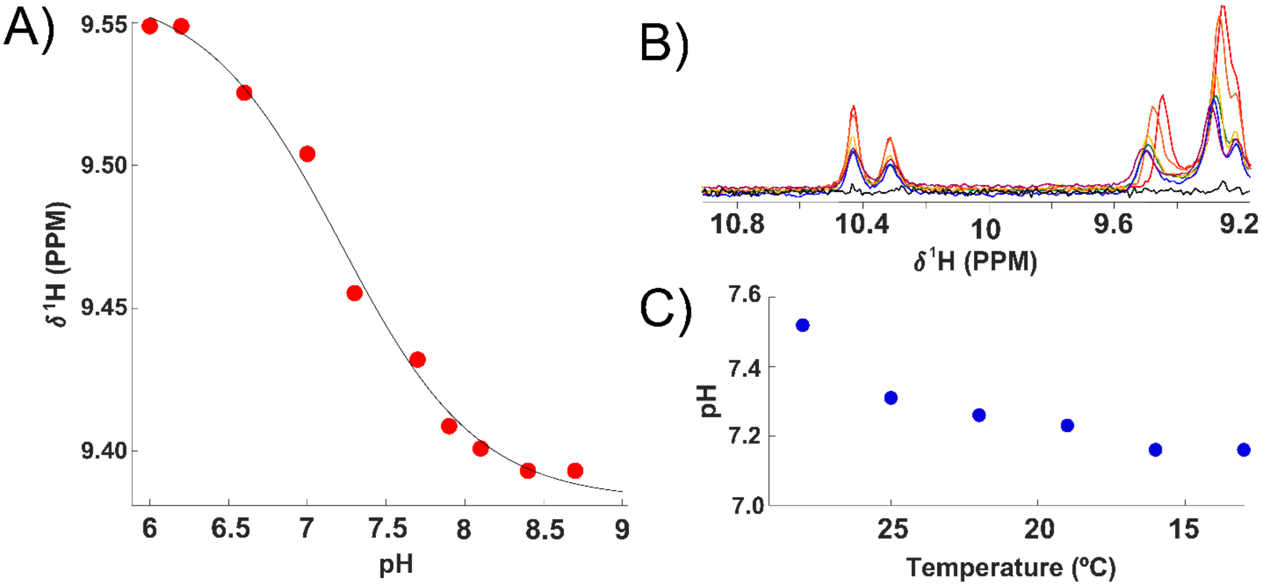


Figure 4.3. pH of the *Danio rerio* oocyte cytoplasm. A) Standard curve of histidine-10 ^{1}H chemical shift of K10H GB1 at varying pH values (298 K) in buffer. B) ^{15}N-^{1}H spectra of K10H GB1 injected into zebrafish oocytes at 301 K (red), 298 K (orange), 295 K (yellow), 292 K (green), 289 K (blue), 286 K (purple), and leakage control (black). The histidine-10 resonance is located between 9.5-9.4 ppm. The tryptophan 43 ε-^{15}N-^{1}H at 10.41 resonance was used as reference for chemical shift calculations. C) Estimated intracellular pH as a function of temperature while replicating the protocol used for SH3 stability measurements.

In summary, our results show that destabilizing attractive interactions can out-compete stabilizing repulsive interactions in cells, even when the overall intracellular protein concentration is significantly lower. However, the similar diffusion of the wildtype protein and the charge-change variant in oocytes highlights the idea that there are a broader range of intermolecular interactions, including hydrogen bonding, whose cellular effects on proteins require study.[31]

We also demonstrated that zebrafish oocytes are a viable system for in-cell NMR. *X. laevis* oocytes have been used in several publications.[216,220,265] Over the course of this study we observed that some injected oocytes fail to incorporate some or all of test protein into the cytoplasm. The use of phenyl red makes lack of incorporation easy to detect because zebrafish oocytes are transparent.

We suggest that zebrafish oocytes are a robust vehicle for in-cell NMR for several reasons. Their transparency permits validation of injection localization to ensure the NMR signal originates only from the cytoplasm. Furthermore, *D. rerio* is a widely used, lower-cost model organism because of its popularity in genetic, developmental biology and toxicology studies.[266,267] Zebrafish oocyte collection is simpler and fish have a shorter recovery time (2-3 weeks versus 1-4 months for *X. laevis*), allowing more data to be collected in a shorter time.[236,268,269] Our work demonstrates a simple and effective method to study macromolecules and their ligands in-cell. The use of zebrafish oocytes to make in-cell NMR more accessible and improve our understanding of protein chemistry in living cells.

SUPPORTING INFORMATION

Charge-change Mutant Screening

Variants were generated via site-directed mutagenesis on the pET11a vector containing wildtype SH3 gene or the pET28b vector containing an N-terminal hexa-histidine- and SUMO tagged- wildtype SH3. Primers were obtained from Integrated DNA Technologies. PCR was carried out using the Q5 high-fidelity polymerase (New England Biolabs), followed by digestion with Dpn1 (Thermo Scientific), following the manufacturer's protocol. Successful reactions were transformed into DH5α competent cells (Thermo Fisher Scientific). A single colony was chosen and grown in 5 mL Lenox broth (LB, 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) culture with 100 μg/mL ampicillin or 50 μg/mL kanamycin, and shaken at 225 rpm for 12-16 h at 37 °C. Plasmid DNA was isolated (Qiagen QIAprep Spin Miniprep Kit) and sequenced. Plasmids bearing the mutated genes were transformed into BL21 DE3 gold cells (Invitrogen). A single colony was used to inoculate 5 mL of LB with 100 μg/mL ampicillin or 50 μg/mL kanamycin. After shaking at 225 rpm for 6 to 8 h at 37 °C (New Brunswick Scientific Innova I26), 50 μL of the culture was used to inoculate a 100 mL overnight culture of M9 media (50 mM Na_2HPO_4, 20 mM KH_2PO_4, 9 mM NaCl, 4 g/L glucose, 1 g/L NH_4Cl, 0.1 mM $CaCl_2$, 2 mM $MgSO_4$, 10 mg/L thiamine, 10 mg/L biotin, and 100 mg/L ampicillin or 50 mg/L kanamycin, pH 7.4). The culture was shaken overnight as described above.

The next day, the overnight culture was diluted with 100 mL of M9 minimal media. 5-Fluoroindole (dissolved in dimethyl sulfoxide) was added to the culture to a final concentration of 100 mg/L.[248] After an additional 30 min of shaking, the cells were induced with 1 mM isopropyl-β-d-1-thiogalactopyranoside (IPTG), and shaken at 18 °C for 1.5 h.

The culture was then centrifuged at 1000 *g* for 30 min and resuspended in 200 mL of sterile M9 media, IPTG was again added (final concentration of 1 mM), and the culture was shaken for 1.5 h, 18 °C. The cells were then centrifuged at 1000 *g*, 30 min and resuspended in 250 µL 75 mM Hepes, 75 mM bis-tris propane, 75 mM citrate, pH 7.4, 10% D_2O). Protease inhibitors were added, the slurry was immediately sonicated (10% amplitude, 0.4 s on, 0.2 s off, 30 s) and then centrifuged at 4 °C, 12,000 *g* for 10 min. The supernatants were interrogated by NMR as described below.

The E54K variant exhibited a resonance (-123.8 ppm at 298 K) from the folded state, and was purified and concentrated as described for wildtype SH3. The concentrated protein aggregated within minutes of loading the glass capillary needle used for injection. To screen other variants, we repeated this experiment for wildtype SH3 and used the relative folded peak intensity as the benchmark for the minimum stability required for further testing. Only the N51D and N51K variants showed significant folded state resonances (-123.8, -124.1 ppm, respectively, at 298 K), both of which were approximately the same as that of the wildtype protein (Table S4.1). The N51D variant was expressed, purified and injected as described for N51K. During in-cell measurements in oocytes, a third peak (-119.5 ppm at 298 K) was observed whose intensity varied (reversibly) with temperature.

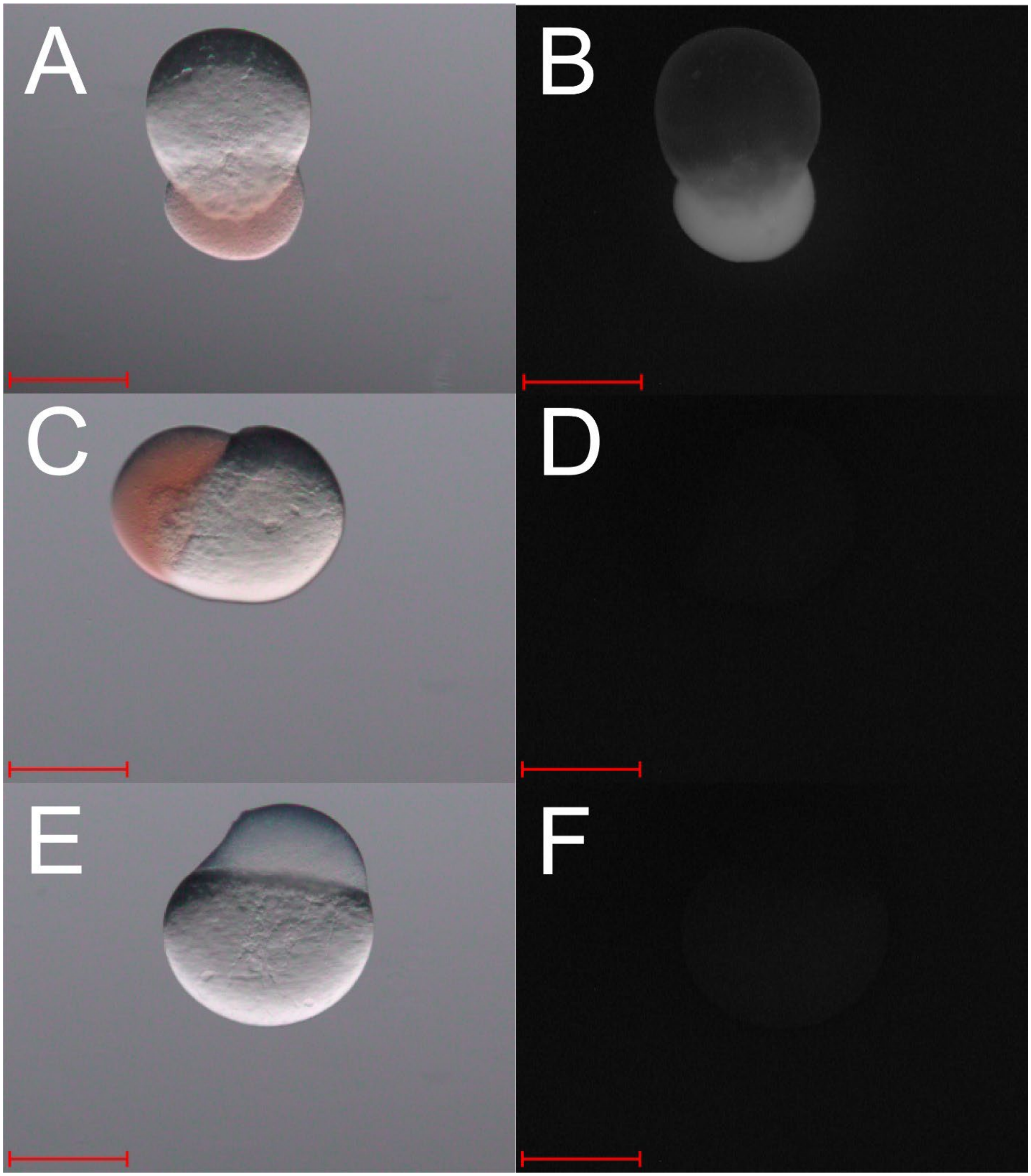


Figure S4.1. Microscopy of injected oocytes. Oocytes injected with (A and B) T22G;C60 SH3 covalently labeled with bromobimane, (C and D) T22G SH3. (E-F) uninjected oocytes. A, C, and E brightfield; B, D, and F epifluorescence (excitation, 436 nm - 420 nm; emission, 480 nm - 440 nm). Scale bar, 200 μm.

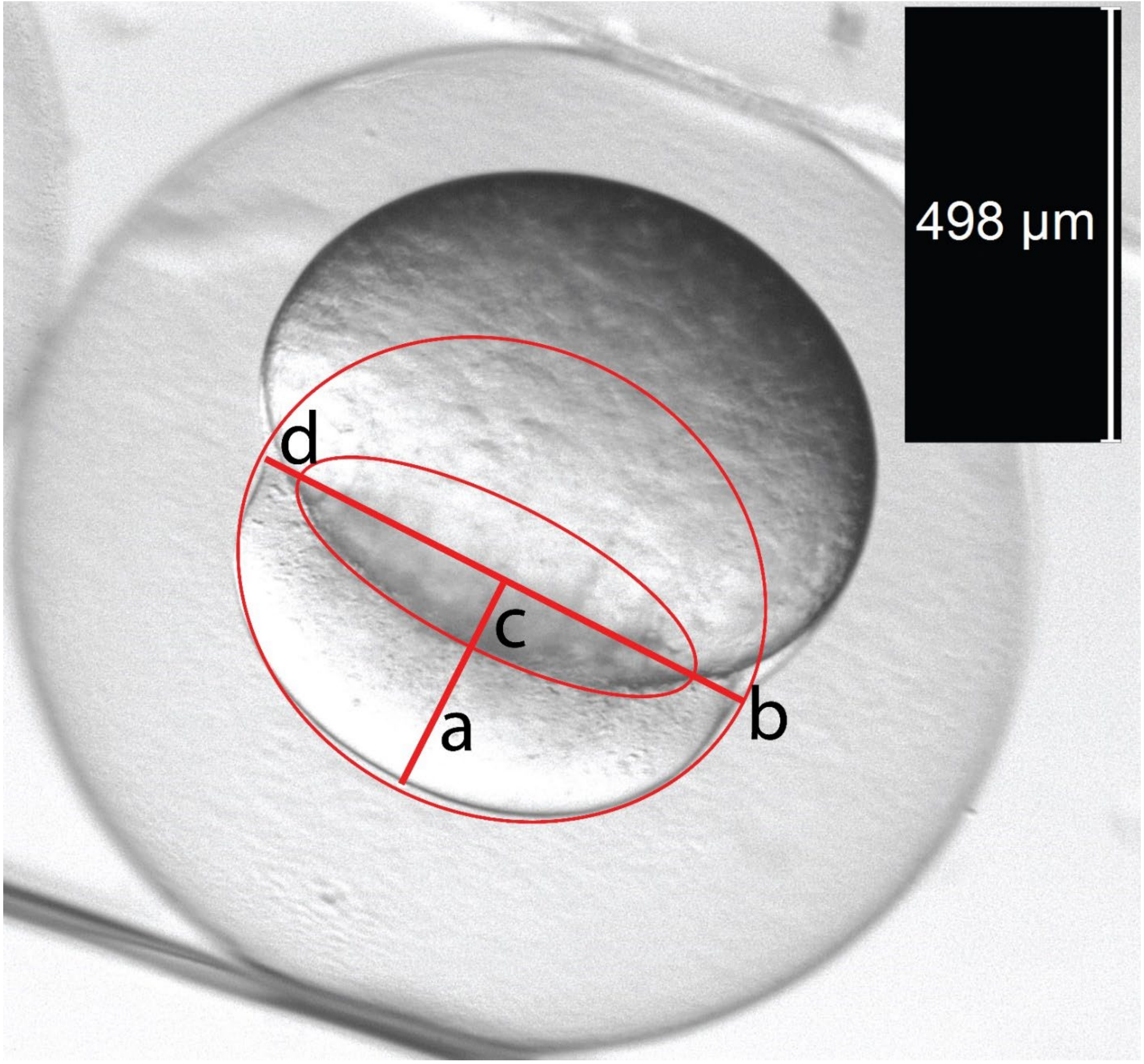


Figure S4.2. Oocyte volume estimation. Ovals were traced over the cytoplasm and the portion of the yolk that occupies the cytoplasmic volume. Vertical radii (a, c) and horizontal radii (b, d) were measured for the cytoplasm and yolk, respectively. Both the cytoplasm and yolk were assumed to be half spheroids.

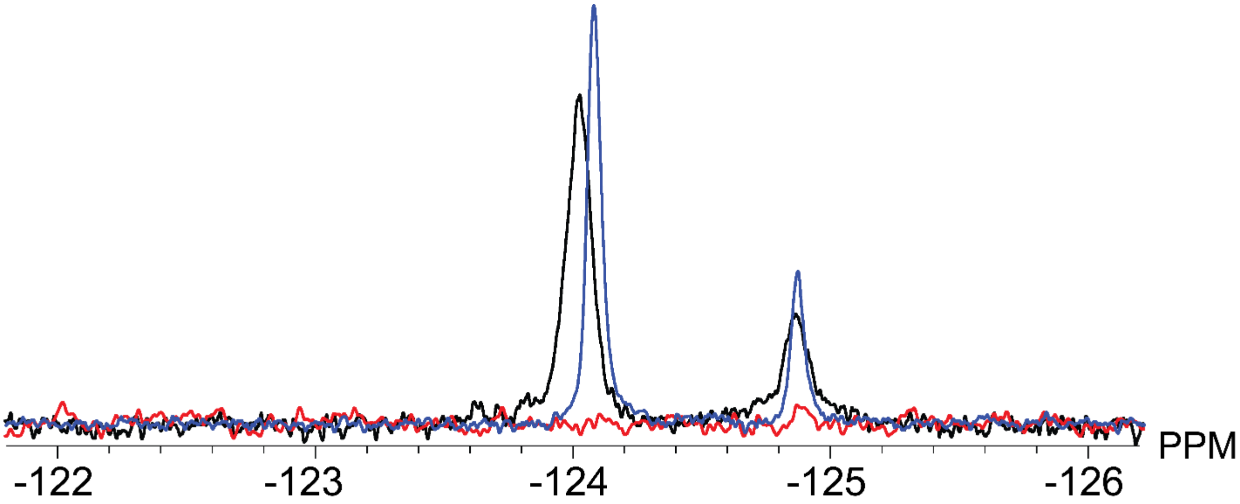


Figure S4.3. Representative ^{19}F-NMR spectra at 298 K of ^{19}F N51K SH3 in buffer (blue), in oocytes (black) and the 2:1 dilution leakage control (red). The resonances of the folded form have chemical shifts of approximately -124.0 PPM. The resonances of the unfolded form have chemical shifts of -124.85 PPM.

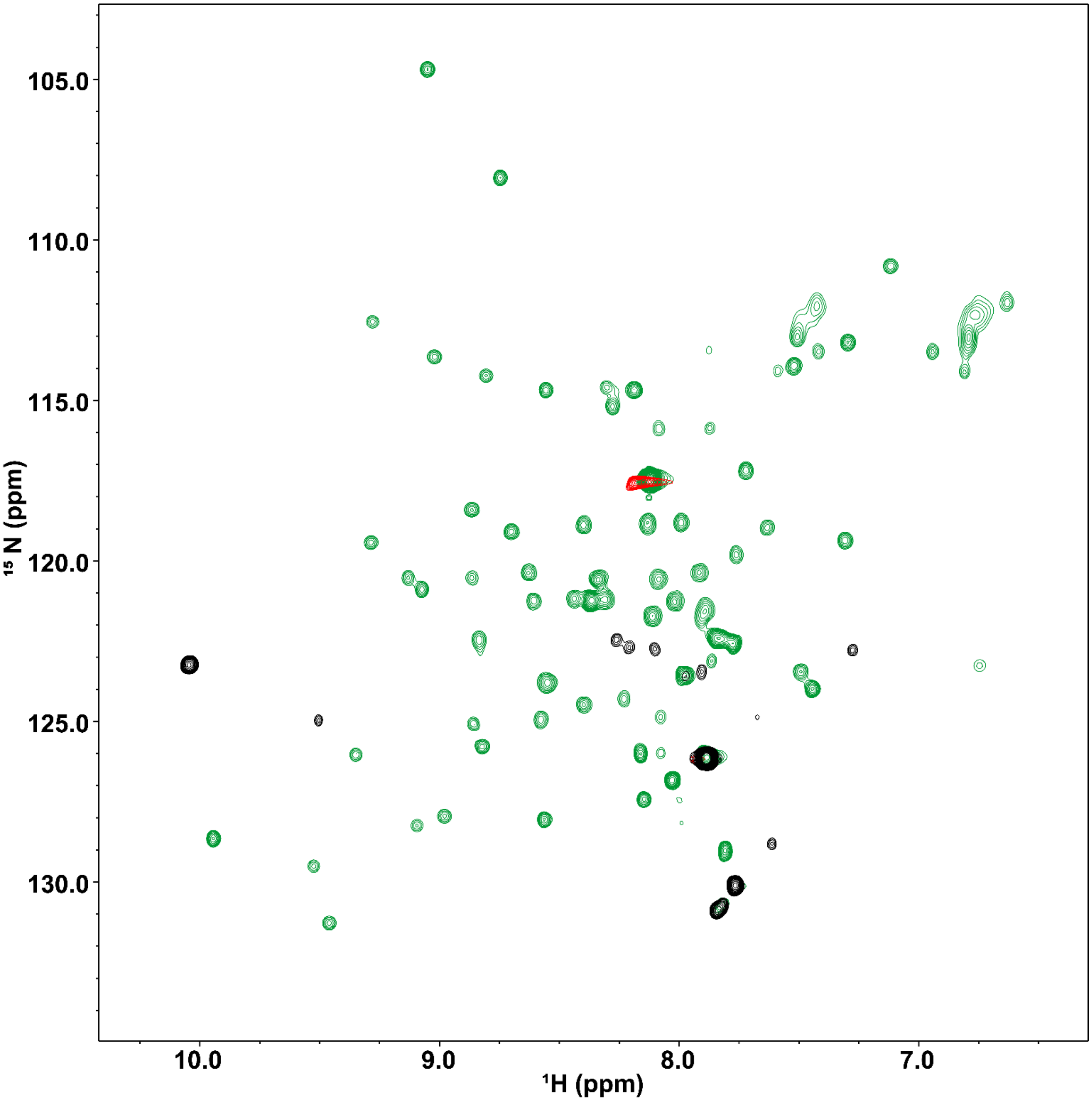


Figure S4.4. ^{15}N-^{1}H HSQC spectrum of T22G SH3 in *E. coli* cells (red), cell lysate after sonication (green), and leakage control (black, 2-fold dilution).

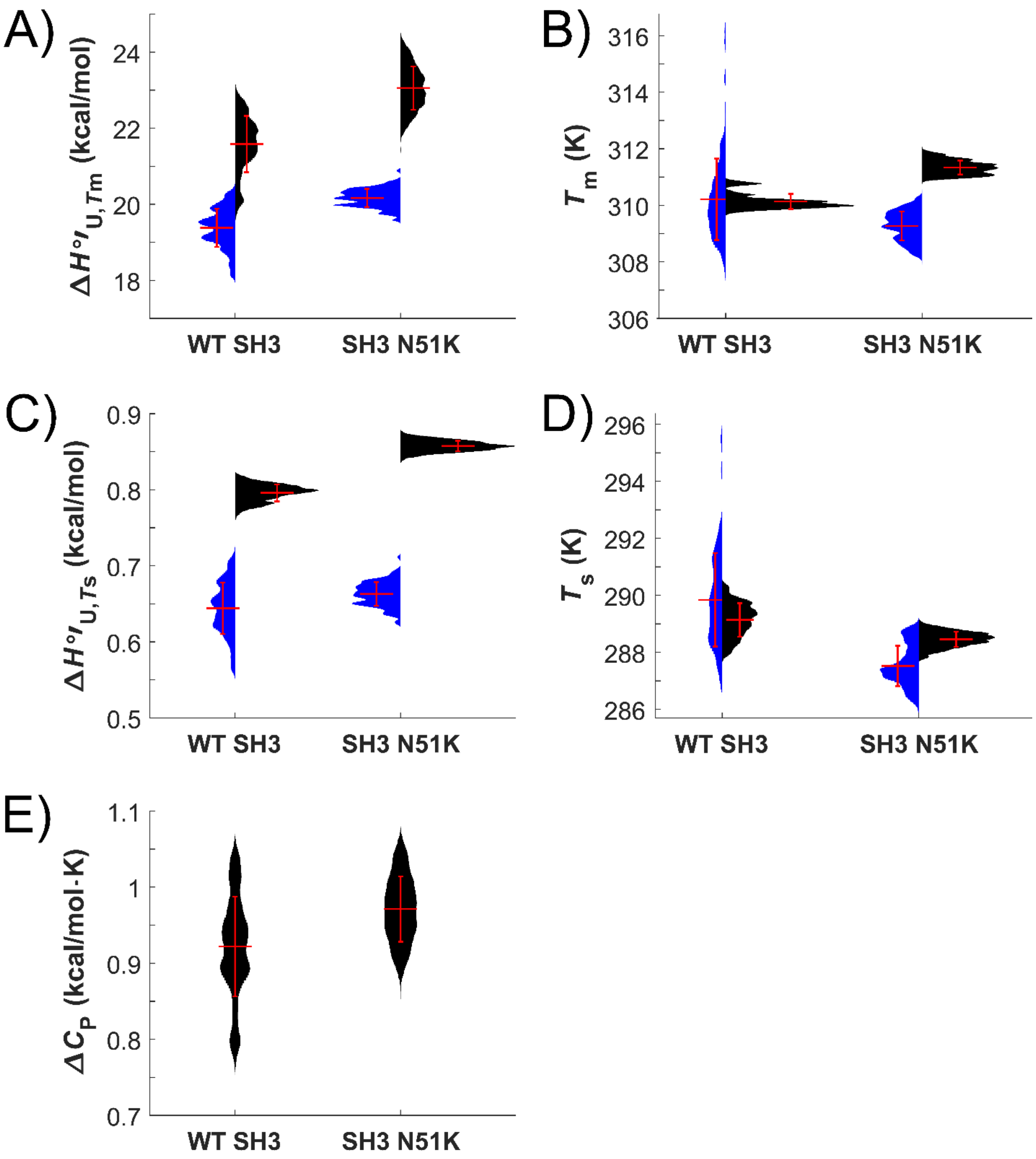


Figure S5.5. Histograms of *in-vitro* (black) and in-oocyte (blue) wildtype SH3 and charge-change variant N51K SH3 resampled stability parameters. A) enthalpy of unfolding at the melting-temperature, $\Delta H^{\circ\prime}_{U,Tm}$; B) melting temperature, T_m; C) enthalpy of unfolding at the temperature of maximum stability $\Delta H^{\circ\prime}_{U,Ts}$; D) temperature of maximum stability, T_s, E) heat capacity of unfolding (ΔC_P) from buffer measurements. Horizontal lines represent the means. Vertical lines represent the standard deviations.

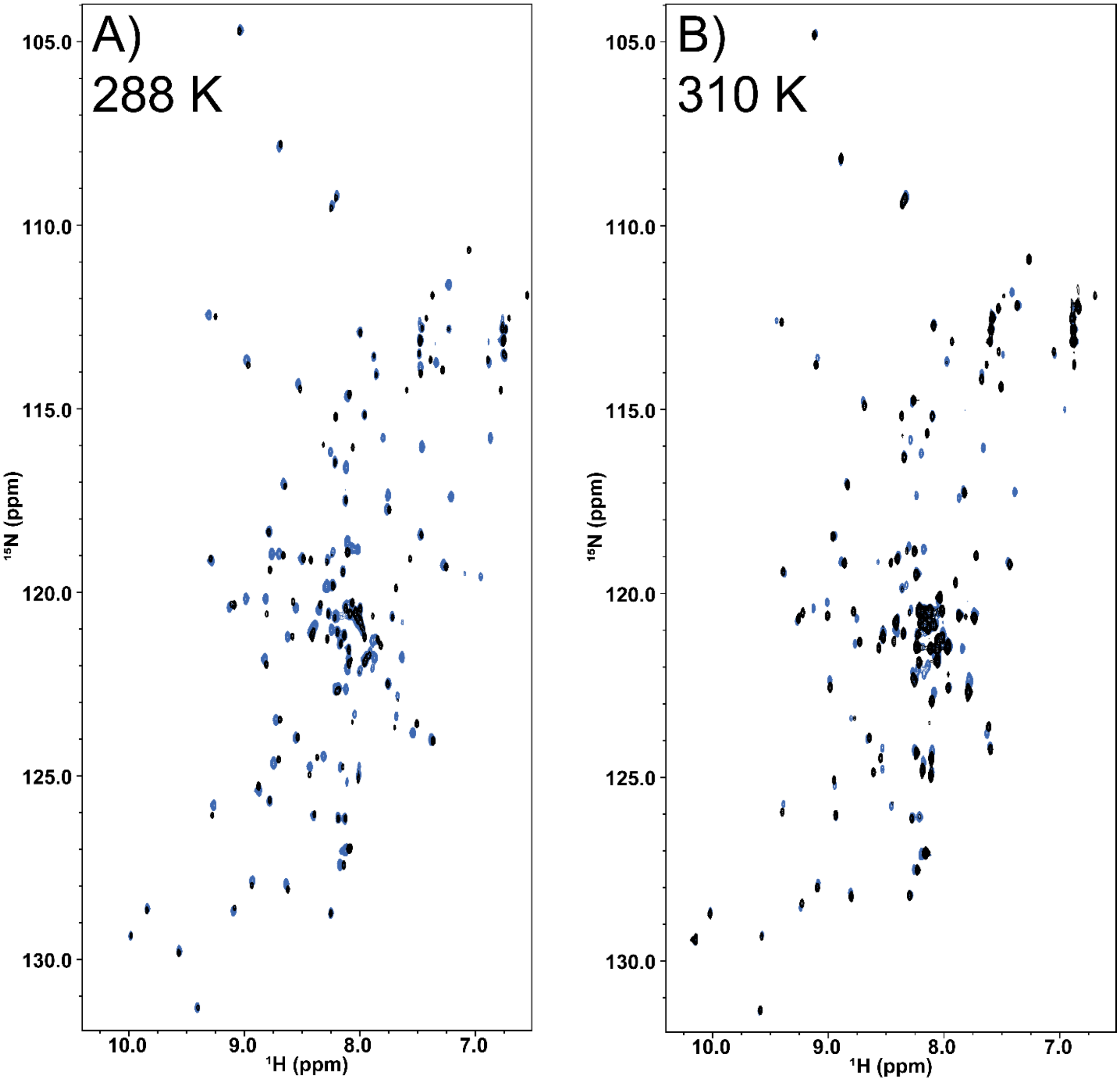


Figure S4.6. Fold comparison of WT SH3 and SH3 N51K. ^{15}N-^{1}H HSQC spectra of WT (black) and N51K SH3 (blue) in buffer at A) 288.15 K and B) 310.15 K.

Table S4.1. SH3 charge-change variants excluded from analysis.

Variant	Charge compared to WT	Reason for exclusion
S10K	+1	Too unstable for injection
R20N	-1	Too unstable for injection
Q23E	-1	Too unstable for injection
M30K	+1	Too unstable for injection
N51D	-1	Did not follow two-state folding in oocyte
E54K	+2	Too unstable for injection
E54Q	+1	Too unstable for injection
E54R	+2	Too unstable for injection
K56E	-2	Too unstable for injection

Table S4.2. Global fits of wild type SH3 and the N51K variant correlation times[240,241] (τ_C, ns·radian^{-1}) with varied τ_e and S^2 values. Uncertainties represent the standard deviation from Monte Carlo analysis.

	τ_e (ps)									
						1	3	5	7	1
	10	30	50	75	100	10	30	50	75	100
S^2	S		WT SH3 Buffer					WT SH3 in oocyte		
0.5	6.8 ± 0.3	6.9 ± 0.3	7.0 ± 0.3	7.0 ± 0.3	7.1 ± 0.3	15 ± 5	16 ± 4	16 ± 4	16 ± 4	16 ± 4
0.6	5.8 ± 0.2	5.8 ± 0.2	5.9 ± 0.2	5.9 ± 0.2	6.0 ± 0.2	14 ± 3	14 ± 3	14 ± 3	14 ± 3	14 ± 3
0.7	5.1 ± 0.2	5.1 ± 0.2	5.1 ± 0.2	5.1 ± 0.2	5.2 ± 0.2	12 ± 3	12 ± 3	13 ± 3	13 ± 3	13 ± 3
0.8	4.5 ± 0.1	4.5 ± 0.1	4.5 ± 0.1	4.5 ± 0.1	4.6 ± 0.1	11 ± 2	11 ± 2	11 ± 2	11 ± 2	11 ± 2
0.9	4.0 ± 0.1	4.1 ± 0.1	4.1 ± 0.1	4.1 ± 0.1	4.1 ± 0.1	10 ± 2	10 ± 2	10 ± 2	10 ± 2	11 ± 2
1.0	3.7 ± 0.1	3.7 ± 0.1	3.7 ± 0.1	3.7 ± 0.1	3.7 ± 0.1	10 ± 2	10 ± 2	10 ± 2	10 ± 2	10 ± 2
			SH3 N51K Buffer					SH3 N51K in oocyte		
0.5	8.8 ± 0.4	8.8 ± 0.4	8.8 ± 0.4	8.9 ± 0.4	8.9 ± 0.3	23 ± 0.6	23 ± 6	23 ± 6	23 ± 6	23 ± 5
0.6	7.3 ± 0.3	7.3 ± 0.3	7.3 ± 0.3	7.4 ± 0.2	7.4 ± 0.2	19 ± 5	19 ± 5	19 ± 5	19 ± 4	19 ± 4
0.7	6.2 ± 0.2	6.3 ± 0.2	6.3 ± 0.2	6.3 ± 0.2	6.3 ± 0.2	16 ± 4	16 ± 4	16 ± 4	16 ± 4	16 ± 4
0.8	5.5 ± 0.1	5.5 ± 0.1	5.5 ± 0.1	5.5 ± 0.1	5.5 ± 0.1	14 ± 3	14 ± 3	14 ± 3	15 ± 3	15 ± 3
0.9	4.9 ± 0.1	4.9 ± 0.1	4.9 ± 0.1	4.9 ± 0.1	4.9 ± 0.1	13 ± 3	13 ± 3	13 ± 3	13 ± 3	13 ± 3
1.0	4.44 ± 0.09	4.44 ± 0.09	4.44 ± 0.09	4.44 ± 0.09	4.44 ± 0.09	12 ± 2	12 ± 2	12 ± 2	12 ± 2	12 ± 2

OTHER DATA AND CONCLUSIONS

INITIAL CHARACTERIZATION OF A GB1 TETRAMER

Rationale: Our lab currently has three GB1 (B1 domain of *Streptococcal* protein G) model systems that we use to quantify the effects of crowding and as a protectant in desiccation studies.[35,270-273] The T2Q variant of GB1 is highly stable and monomeric, unless at very high concentrations (~2 mM) when it begins to self-associate. The A34F (colloquially the side-by-side dimer) dimerizes at 30 μM, forming two kissing spheres with identical surfaces.[274] The L5V/F30V/Y33F/A34F variant forms a domain-swapping dimer, forming a dimer with less surface area relative to the same volume occupied, dimerizing at 100 μM.[275] A project we considered was to expand the repertoire of GB1 models, because having a variety of systems from the same basic architecture would make for a better comparison, and so we looked for an additional GB1 variant. We found a tetrameric GB1 (L5V/A26F/F30V/Y33F/A34F) that could easily be made with an additional point mutation starting from the domain-swapping construct.[276] After making the tetramer variant by mutagenesis-PCR (Forward primer 5'-AGAACTGTTTGACAACTTTTTCGAAGGTAGCAGCGTCAACAGCTT-3', Reverse primer 5'-GAAGCTGTTGACGCTGCTACCTTCGAAAAAGTTGTCAAACAGTTCT-3'.

Expression/Purification: After confirming the mutation, I tested expression, and found that optimal induction time was 2 hours. Expression was identical to what is described for SH3 T22G in rich-media[190] and minimal media.[195] Purification was found to be optimal (for 2 L of cell pellets at a time) on a 5 mL Q column (Cytiva) using 20 mM

Tris (tris(hydroxymethyl)aminomethane) buffer pH 9.5. After loading and washing the column, a linear ramp to 20 mM Tris, 230 mM NaCl, pH 9.5 and then a subsequent hold for 50 mL gave sufficient resolution. The column can then be quickly ramped to 1 M NaCl, then 2 M for column regeneration. The fractions containing the tetramer can be concentrated and then polished using size exclusion chromatography (75 pg resin).

Preliminary results: Measurements were made on a Bruker Avance III HD spectrometer equipped with QCI cryoprobes, ^{1}H Larmor frequencies of 500, 470 MHz for ^{19}F. ^{19}F measurements comprised 512 scans of 4708 Hz sweep width, 8570 points per acquisition with a relaxation delay of 3 s. Data were processed using Topspin 3.6.2. Initial measurements were made of a 500 μM GB1 sample in 20 mM sodium phosphate, 10% D_2O, pH 7.5. Serial dilutions using the buffer were made, to acquire spectra at 250, 125, 62.5, and 31.3 μM GB1 (Figure 5.1). Also, equilibrium sedimentation was performed using an Analytical Ultra Centrifuge (Beckman Optima XL-I) at concentrations of 17.2, 31.2, 42.4, 59.5, 91.4, and 119 μM in 20 mM sodium phosphate, 150 mM NaCl, pH = 7.5. Data was analyzed using SEDFIT (version 16.36).[277] Estimates for $K_{D,Dimer}$ via AUC is very high, in the mM concentration range, and the affinity for the tetramer, $K_{D,Tetramer}$, is slightly lower (Figure 5.2).

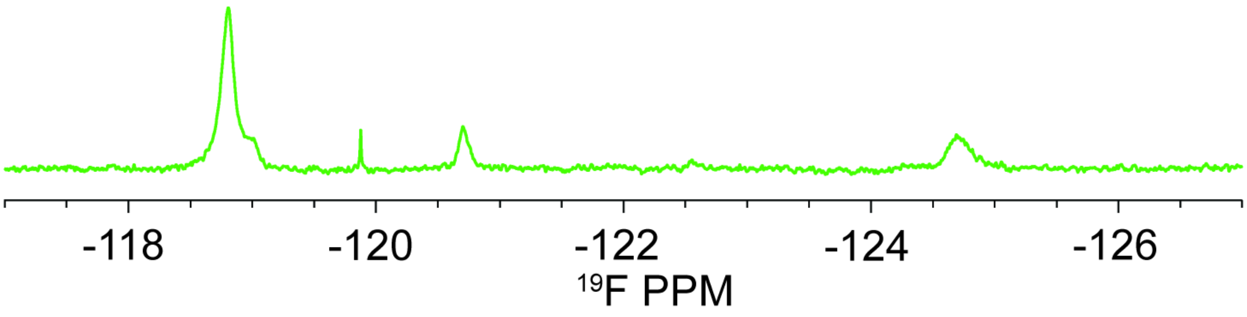


Figure 5.1. ^{19}F NMR spectra of 500 μM GB1 L5V/A26F/F30V/Y33F/A34F. Made at 298 K, in 20 mM sodium phosphate, 10% D_2O, pH = 7.5.

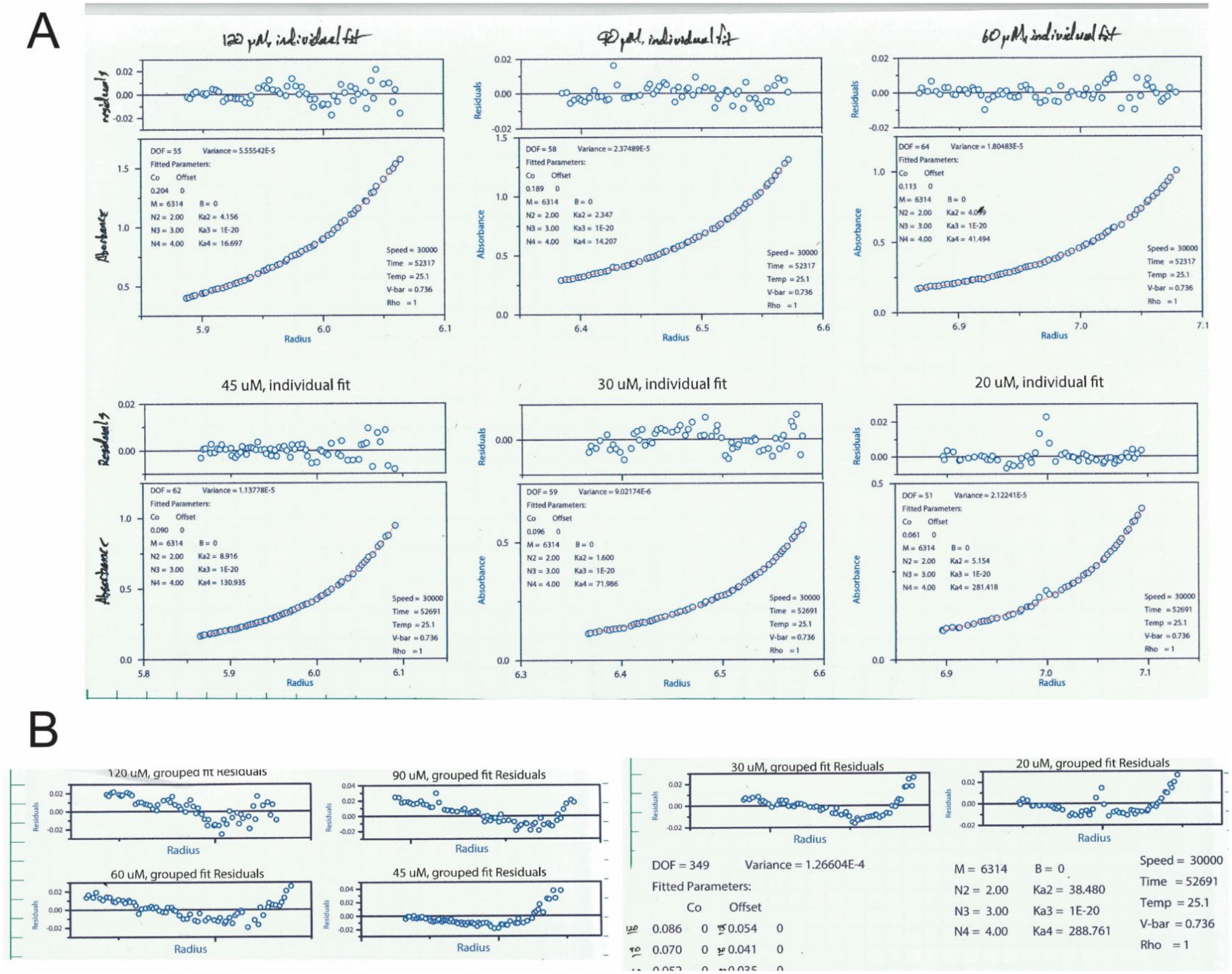

A
120 µM, individual fit
residuals
0.02
0.00
-0.02
Absorbance
1.5
1.0
0.5
DOF = 55 Variance = 5.55542E-5
Fitted Parameters:
Co Offset
0.204 0
M = 6314 B = 0
N2 = 2.00 Ka2 = 4.156
N3 = 3.00 Ka3 = 1E-20
N4 = 4.00 Ka4 = 16.697
Speed = 30000
Time = 52317
Temp = 25.1
V-bar = 0.736
Rho = 1
5.9
6.0
6.1
Radius
90 µM, individual fit
Residuals
0.02
0.00
-0.02
Absorbance
1.5
1.0
0.5
0.0
DOF = 58 Variance = 2.37489E-5
Fitted Parameters:
Co Offset
0.189 0
M = 6314 B = 0
N2 = 2.00 Ka2 = 2.347
N3 = 3.00 Ka3 = 1E-20
N4 = 4.00 Ka4 = 14.207
Speed = 30000
Time = 52317
Temp = 25.1
V-bar = 0.736
Rho = 1
6.4
6.5
6.6
Radius
60 µM, individual fit
Residuals
0.02
0.00
-0.02
Absorbance
1.0
0.5
0.0
DOF = 64 Variance = 1.80483E-5
Fitted Parameters:
Co Offset
0.113 0
M = 6314 B = 0
N2 = 2.00 Ka2 = 4.099
N3 = 3.00 Ka3 = 1E-20
N4 = 4.00 Ka4 = 41.494
Speed = 30000
Time = 52317
Temp = 25.1
V-bar = 0.736
Rho = 1
6.9
7.0
7.1
Radius
45 uM, individual fit
Residuals
0.02
0.00
Absorbance
1.0
0.5
0.0
DOF = 62 Variance = 1.13778E-5
Fitted Parameters:
Co Offset
0.090 0
M = 6314 B = 0
N2 = 2.00 Ka2 = 8.916
N3 = 3.00 Ka3 = 1E-20
N4 = 4.00 Ka4 = 130.935
Speed = 30000
Time = 52691
Temp = 25.1
V-bar = 0.736
Rho = 1
5.8
5.9
6.0
6.1
Radius
30 uM, individual fit
Residuals
0.00
Absorbance
0.5
0.0
DOF = 59 Variance = 9.02174E-6
Fitted Parameters:
Co Offset
0.096 0
M = 6314 B = 0
N2 = 2.00 Ka2 = 1.600
N3 = 3.00 Ka3 = 1E-20
N4 = 4.00 Ka4 = 71.986
Speed = 30000
Time = 52691
Temp = 25.1
V-bar = 0.736
Rho = 1
6.3
6.4
6.5
6.6
Radius
20 uM, individual fit
Residuals
0.02
0.00
Absorbance
0.5
0.0
DOF = 51 Variance = 2.12241E-5
Fitted Parameters:
Co Offset
0.061 0
M = 6314 B = 0
N2 = 2.00 Ka2 = 5.154
N3 = 3.00 Ka3 = 1E-20
N4 = 4.00 Ka4 = 281.418
Speed = 30000
Time = 52691
Temp = 25.1
V-bar = 0.736
Rho = 1
6.9
7.0
7.1
Radius
B
120 uM, grouped fit Residuals
90 uM, grouped fit Residuals
60 uM, grouped fit Residuals
45 uM, grouped fit Residuals
30 uM, grouped fit Residuals
20 uM, grouped fit Residuals
Residuals
0.04
0.02
0.00
-0.02
Radius
DOF = 349 Variance = 1.26604E-4
Fitted Parameters:
Co Offset
120 0.086 0 45 0.054 0
90 0.070 0 30 0.041 0
M = 6314 B = 0
N2 = 2.00 Ka2 = 38.480
N3 = 3.00 Ka3 = 1E-20
N4 = 4.00 Ka4 = 288.761
Speed = 30000
Time = 52691
Temp = 25.1
V-bar = 0.736
Rho = 1

Figure 5.2. Analytical Ultracentrifugation analysis of the GB1 L5V/A26F/F30V/Y33F/A34F tetramer variant. Ka2 represents the association constant for the dimeric state, and Ka4 is the association constant of the tetrameric state.

Final comments: The GB1 tetramer is likely not a good model for comparison to GB1, the side-by-side dimer, or the domain swapping dimer. The protein may have multiple oligomer states, and may be too complex for a simple analysis. There may be a future use for this system, so if one wants to improve the purification process, try adding glycerol or propylene glycol to the buffer to disrupt the higher-order states that are likely responsible for the long elution times off of the Q resin, as exchange rate between states is slow, as observed by ^{19}F NMR. This may help scaling up and allow purifying 6 L of cell lysates on a single larger column. If someone has reason to pick up this project, I solved the binding isotherm for the fraction monomer, fraction dimer, and fraction tetramer, which at the time I could not find published.

$$4M \underset{K_{D,D}}{\Longleftrightarrow} 2D \underset{K_{D,T}}{\Longleftrightarrow} T$$

$$0 = a\sqrt[4]{f_T} + b\sqrt{f_T} + cf_T - 1;\;\; a = \sqrt[4]{\frac{K_{D,D}^2 K_{D,T}}{[GB1]^3}},\; b = 2\sqrt{\frac{K_{D,T}}{[GB1]}},\;\; c = 4$$

$$0 = f_D + af_D + b\sqrt{f_D} - \frac{a}{2}, a = \frac{k_{D,DT}}{2[GB1]}, b = \frac{k_{D,DT\sqrt{k_{D,MD}}}}{4[GB1]^{\frac{3}{2}}}$$

$$0 = {f_M}^4 + b{f_M}^2 + \; af_M - a \;\; ; a = \frac{{K_{D,D}}^2 K_{D,T}}{4(GB1)^3} \;\; b = \frac{K_{D,D} K_{D,T}}{2(GB1)^2}$$

ADDITIONAL EXPERIMENTS NOT INCLUDED IN CHAPTER 4

^{31}P NMR

Rationale: The goal of in-cell NMR is to reliably measure the effects of the cytoplasm on a test protein and quantify the effects. As researchers, we want the cytoplasmic conditions to reflect, as best as possible, a healthy cell. A common method for determining the viability of cells is an ATP assay, which quantifies the amount ATP from a particular sample. A healthy cell controls the amount of cytoplasmic ATP,[278] which would make it a good control for determining if the cells in a given sample were healthy. A common method of measuring ATP is to use a luciferase assay, as luciferase uses ATP directly to generate light, which can be measured using a luminometer.[279] I attempted to do this in early zebrafish oocyte experiments, but found that getting a clean sample for the assay was neither quick or convenient given the locations of where I had to do zebrafish measurements and NMR measurements. It was particularly challenging to remove the chorion so that I could remove all of the surrounding buffer before homogenizing the cells.

The Bruker Avance III HD spectrometer is equipped with a QCI probe that has direct detect channels for ^{1}H, ^{15}N, ^{19}F and ^{31}P. Since I was already using this instrument for 19F measurements, switching to a different channel is trivial. My experimental plan was to quantify the ratio of ATP:ADP:AMP and use that as a measure of cell health since these ratios should be consistent for healthy cells.

Preliminary results: ATP, ADP, and AMP could indeed be detected by NMR, as well as other phosphorous-containing metabolites (Figure 5.3). Unfortunately, yolk proteins are heavily phosphorylated and drown out the signal of the metabolites. I tried to use 180° flips between excitation and detection, as the metabolites should have long R_2 relaxation constants, due to their size and correlation times (τ_c), while the phosphoproteins are much larger and should have much shorter R_2 constants, which would result in faster signal decay. I tried 2, 4, 6, and 8 rapid 180° flips, but none offered good resolution of the metabolite peaks. It was at this point that Gary pointed me to the GB1 K10H variant, which was very successful.

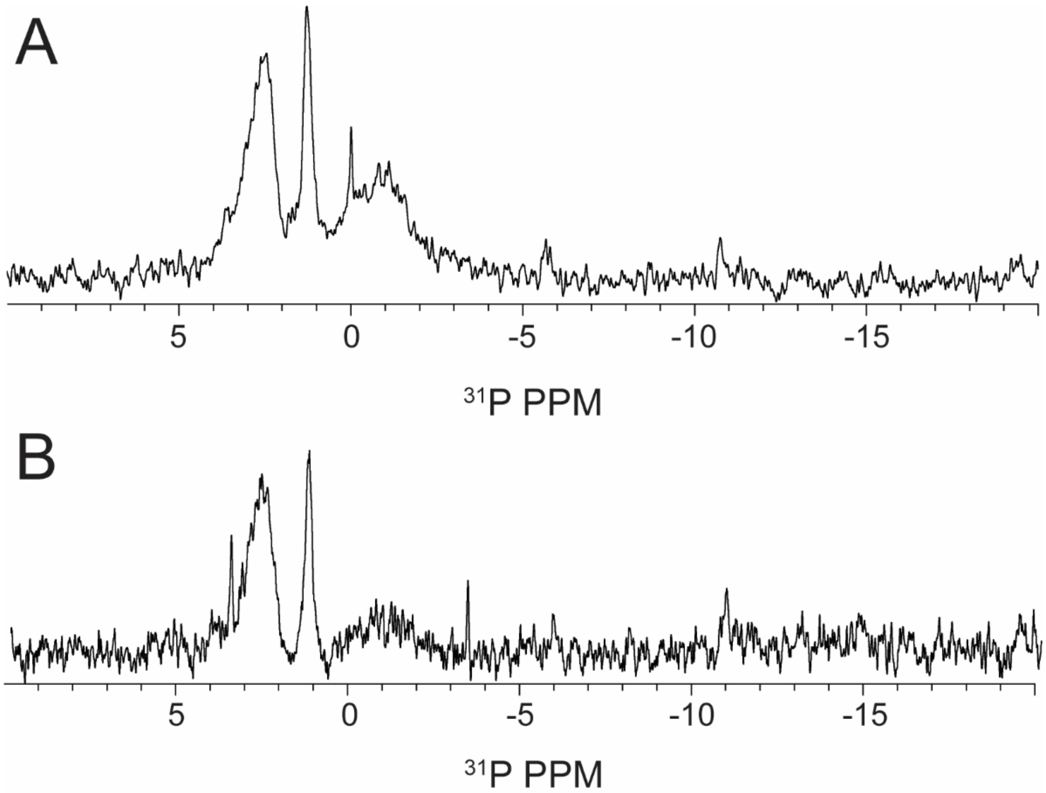


Figure 5.3. ^{31}P spectra of zebrafish oocytes. Made in 5 mM NaCl, 170 µM KCl, 330 µM CaCl2, 330 µM MgSO4, 0.00003% methylene blue, 1.5% w/v Ficoll 70 + 10% D_2O. A) 90° pulse, followed immediately by acquisition. B) 30° pulse, followed immediately by acquisition.

Additional SH3 surface mutant data

Rationale: My goal was to measure how the zebrafish oocyte cytoplasm changed the stability of SH3 as a function of surface charge. Most of the previously characterized mutants of the Downstream of receptor kinase N-terminal SH3 domain were associated with folding and changing the unfolded state,[280] no work on surface charge variants with this protein had been published. I chose residues that were residues that were not part of a charged-patch, unless the change was of the same charge. I also tried to avoid residues at the ends of β-sheets. I chose the following: S10K, S10E, R20N, Q23E, M30E, M30K, D42N, S50K, N51K, N51D, E54A, E54K, E54Q, E54R, K56E.

As described in Chapter 4, I used the in-cell NMR workflow to overexpress each variant, and then check for significant folded protein populations for quantification. (Figure 5.4) Several variants showed little to no folded state, and several showed non-two-state behavior with the appearance of additional peaks. The variants that were deemed to have sufficient folded populations (R20N, Q23E, M30K, N51D, N51K, E54Q) were then expressed and purified as described in Chapter 4, and then stabilities were quantified in the same manner as Chapter 4. N51D and N51K showed reasonable stability and were chosen to move forward because they were variants affecting the same residue, so we hoped that any changes would be a result of the difference in charge.

When it came time to carry out injections, N51K was sufficiently stable for in-cell measurements, but N51D slowly aggregated in the needle and after ~30 minutes, I could no longer carry out injections. The other charged variants were as stable, or less stable than N51D, so no other variants were pursued.

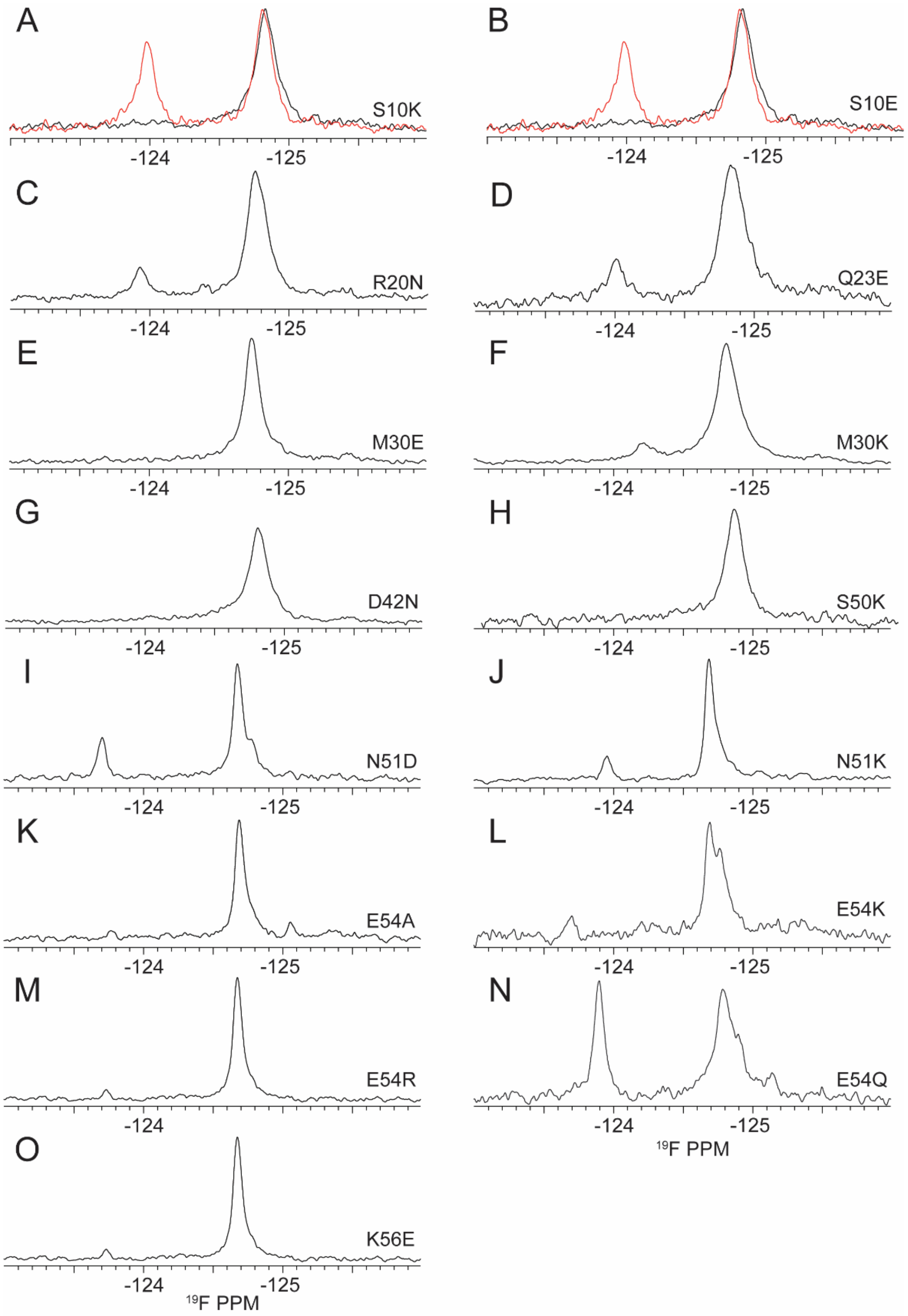


Figure 5.4. In-cell NMR spectra of SH3 surface charge variants in *E. coli*. A) and B) also show in-cell spectra of WT SH3 (red) as a reference. Folded state is at -124 PPM, unfolded state at -124.5 PPM.

R_1, R_2, and estimates for unfolded SH3

Rationale: In chapter 4, we used Trp 36 as a model for the entire folded state, and applied R_1 and R_2 relaxation rates to a model-free τ_c estimate.[240,241] This is a reasonable assumption because we expect that in the folded state, this residue is ordered ($S^2 > 0.6$) We do not have a good understanding of unfolded protein behavior, so it is likely unreasonable to try to use a single residue to represent the mix of ensembles that are likely. Therefore, we did not include the data in the manuscript. As part of the measurements, we made R_1 and R_2 measurements, and included them in the Monte Carlo estimates of τ_c.

Table 5.1. R_1 and R_2 relaxation values of Trp36 in the folded and unfolded state of SH3, *in vitro* and in oocytes.

		Trp36 Folded-state					Trp36 Unfolded-state				
		Buffer		In oocyte			Buffer		In oocyte		
	Larmore Frequency	R_1 (s^{-1})	R_2 (s^{-1})	R_1 (s^{-1})	R_2 (s^{-1})	Relative change in R_2	R_1 (s^{-1})	R_2 (s^{-1})	R_1 (s^{-1})	R_2 (s^{-1})	Relative change in R_2
WT SH3	500 MHz	2.46 ± 0.05	30.8 ± 0.4	3.2 ± 0.3	150 ± 20	4.8	1.9 ± 0.2	17 ± 1	1.7 ± 0.2	180 ± 20	10.6
	600 MHz	2.25 ± 0.06	41.6 ± 0.7	2.2 ± 0.2	159 ± 3	3.8	1.5 ± 0.09	22 ± 1	1.3 ± 0.2	90 ± 8	3.6
SH3 N51K	500 MHz	2.50 ± 0.03	35 ± 1	2.5 ± 0.1	105 ± 7	3.0	1.9 ± 0.2	29 ± 1	1.7 ± 0.2	180 ± 20	6.2
	600 MHz	2.20 ± 0.02	49.4 ± 0.4	1.9 ± 0.1	140 ± 10	2.8	1.65 ± 0.05	27 ± 2	1.3 ± 0.2	146 ± 8	5.4

Conclusion: Trp36 in the unfolded state has a relative increase in R_2 twice that of the folded state (Table 5.1). This makes sense, because in the unfolded state, there is more surface area that is able to interact with surrounding proteins in-cell.

ADDITIONAL DATA NOT INCUDED IN CHAPTER 3

Additional conditions tested for purification of Sos with comments

In determining the optimal purification conditions for the 25 kDa C-terminal domain of Son of sevenless, originally started with the loading buffer conditions of 15.1 mM Na_2HPO_4, 4.9 mM NaH_2PO_4, 20 mM imidazole, 300 mM NaCl, 2 mM tris(2-carboxyethyl)phosphine hydrochloride, 1 mM Ethylenediaminetetraacetic acid (EDTA), 10% v/v glycerol, pH 8.0. Cleavage buffer conditions were identical, except 100 µM phytic acid was added. Initially, I determined that the kinetics of the self-cleavage at 4°C is slow, although it does happen slowly even without addition of phytic acid. Therefore, I increased the phytic acid concentration to 300 µM. By happenstance, I found that EDTA, which is typically added to stop proteases in the cell lysate wasn’t doing much to protect the protein, so I removed it. I next tried denaturing conditions to see if I could alter the contaminant proteins in a way that made them not co-elute. I tried 2, 4, 6, and 8 M urea, as well as 2, 4, and 6 M guanidinium chloride. These helped to some extent, but I ended up losing a significant portion of the Sos that was bound to the nickel column. I then looked for non-ionic/zwitterionic detergents to disrupt the non-specific interactions that were causing Sos to elute with contaminants. Most have very low critical micelle concentrations, with 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS) having the highest concentration I could go. 6 mM CHAPS did improve results, but I realized that many of the remaining contaminants were actually related to Sos, probably degradation products.

To “polish” what my was able to purify, I tried size exclusion chromatography (SEC), hydrophobic interaction chromatography (HIC) and cation exchange chromatography. I found that SEC and HIC did very poorly, and resulted in long elutions where everything was mixed. Cation exchange on the other hand, was the most successful, and I tried a few pH values with 2-(N-morpholino)ethanesulfonic acid buffer—5.0, 5.5, and 6.0. Any higher and the charge of Sos got a little too close to 0. 6.0 seemed to work best, although there was still a large loss of Sos when contaminants eluted from the column, I was finally able to get pure protein. With speed on my mind, I used 3 kDa molecular weight cut-off spin concentrators to pool the nickel elution, concentrate, and then add my lower pH buffer to the sample. This also had the benefit of eliminating the high concentration of NaCl left over from the nickel method. However, the rapid pH change caused much of the sample to aggregate, so I added bis-tris propane to help buffer the transition from pH 8 to pH 6. This improved the aggregation, but was too salty for the cation exchange method. It was at this point that I added 4 M urea to the cation exchange buffer, to reduce aggregates—it did help. This worked, and much of the early work used this method.

It was at this point, where I performed early isothermal titration calorimetry (ITC) and found significant background binding, which explained the non-two state results that we saw via ^{19}F NMR. I designed the knockout construct, and quickly found that it did not express. I tried an extra C-terminal tag that was engineered to incorporate the Sos into an inclusion body to protect it from degradation (did not work), and then tried adding an N-terminal MBP-tag with the hypothesis that some N-terminal interaction was the source. This did work, but doesn’t falsify my hypothesis, so the cause is still unknown.

However, when I exposed the knockout construct with the MBP tag to 4 M urea, it aggregated, so I needed to find something that would break up the non-specific interactions. After reviewing other published methods, I saw others suggest that glycerol is typically added to purification buffers to break up non-specific interactions. I decided to do a side-by-side comparison of cation exchange chromatography with glycerol and a closely-related species—propylene glycol. The propylene glycol worked very well, and was further improved when the cation exchange chromatography was done at room temperature, compared to 4 °C, where I had done everything previously.

Because Sos is very sensitive to salt concentrations during cation exchange chromatography, I later tested the effect of salt on the elution step after overnight cleavage. I found that the MBP-tagged Sos elutes at 100 mM NaCl, which makes the dialysis into the pH 6 buffer a little simpler.

In the future, for anyone looking to tackle Sos, the addition of the MBP tag decreases Sos yield, so finding an alternative that works is very important, especially if specifically labeling to do ^{13}C NMR. Having two different affinity tags on the N and C terminus will allow you to select the correct species with high specificity. With hindsight, many of the early problems that I and others turned out to be related to the closely related degradation products. Perhaps a His-tag on the N-term and Strep-tag on the C-term with propylene glycol to prevent any carry-over will allow for more efficient expression and purification.

Fits of Sos site 2, site 4 and sites 2 and 4 together to one-site and two-site binding models using titrations of 19F SH3 T22G

In Chapter 3, we touched on the 19F NMR fits of the Sos-^{19}F enriched SH3 interaction. In collaboration with Chris Waudby, we tried to fit all the temperature data simultaneously, to a bidentate model, to constrain the parameter fitting by having a single global value for the enthalpy of binding, etc. This would have been a nice method to compare the peptide and full-length domain using comparable methods. The best fits were as follows:

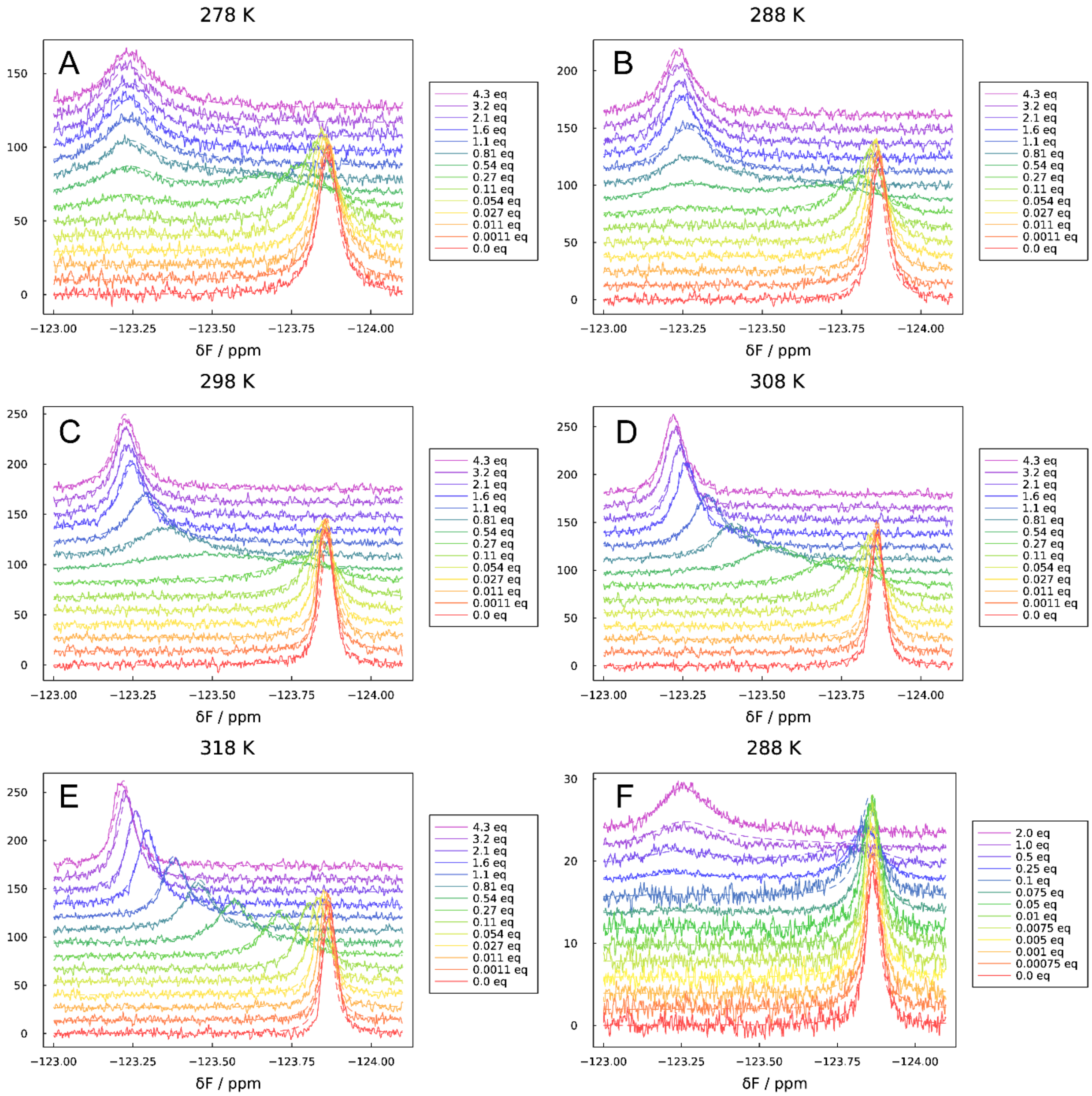


Figure 5.5. ^{19}F NMR fits of SH3-Sos site 4 binding. A-E performed at 470 MHz (^{19}F Larmor frequency), SH3 constant concentration of 270 μM. F performed at 564 MHz (^{19}F Larmor frequency), SH3 constant concentration of 50 μM.

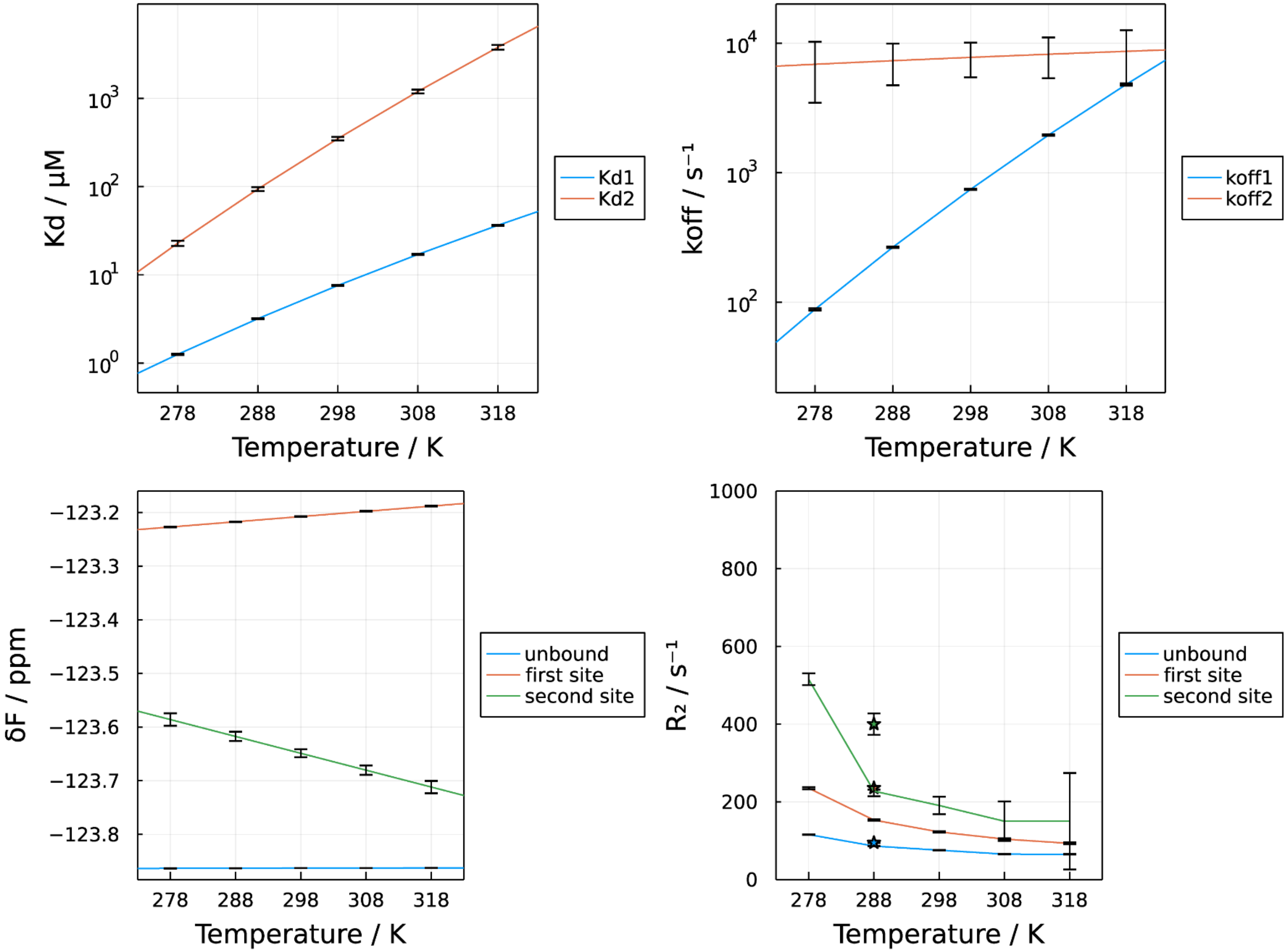


Figure 5.6. ^{19}F NMR parameter fits of SH3-Sos site 4 binding.

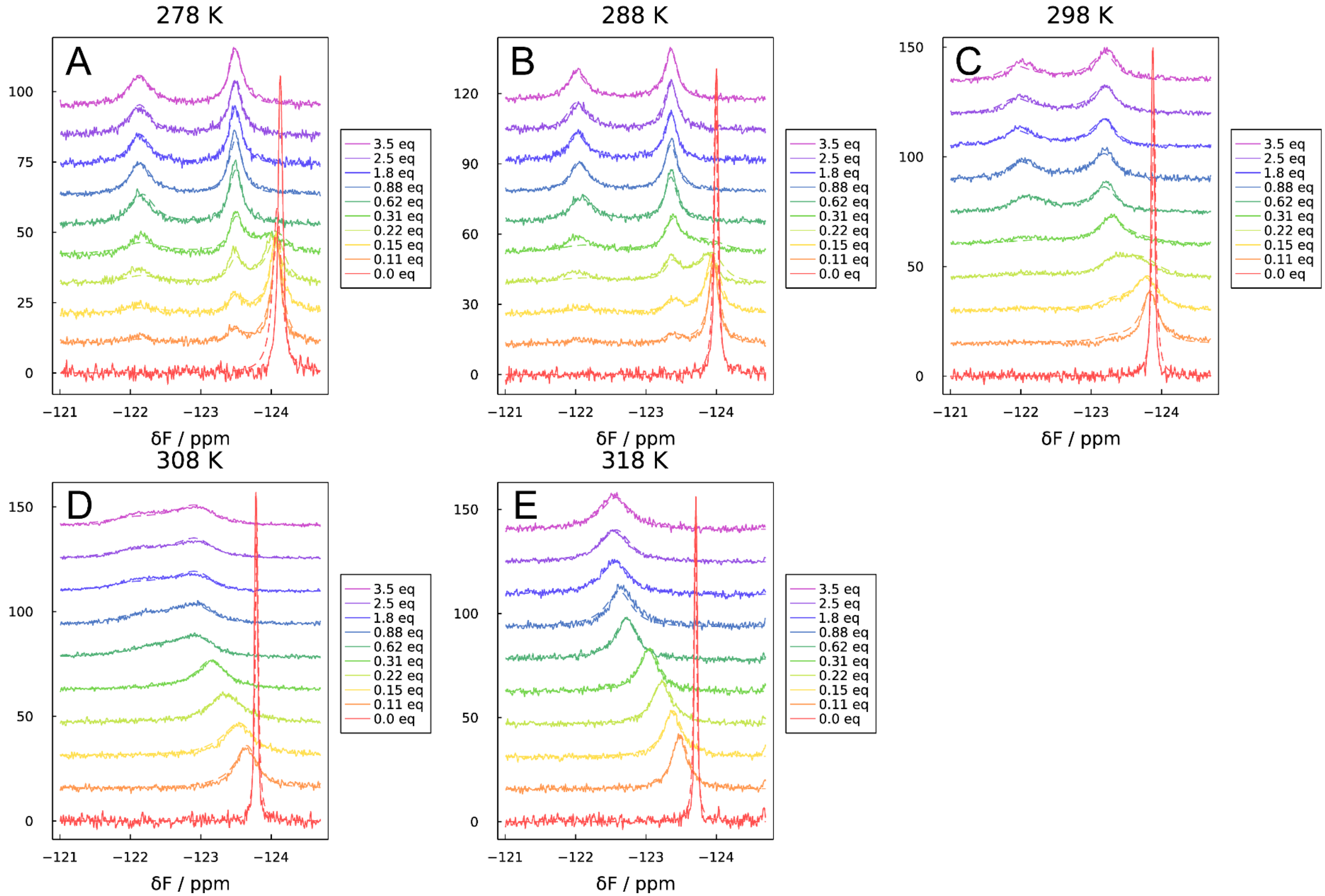


Figure 5.7. ^{19}F NMR fits of SH3-Sos sites 2&4 binding. Performed at 470 MHz (^{19}F Larmor frequency), SH3 constant concentration of 270 μM.

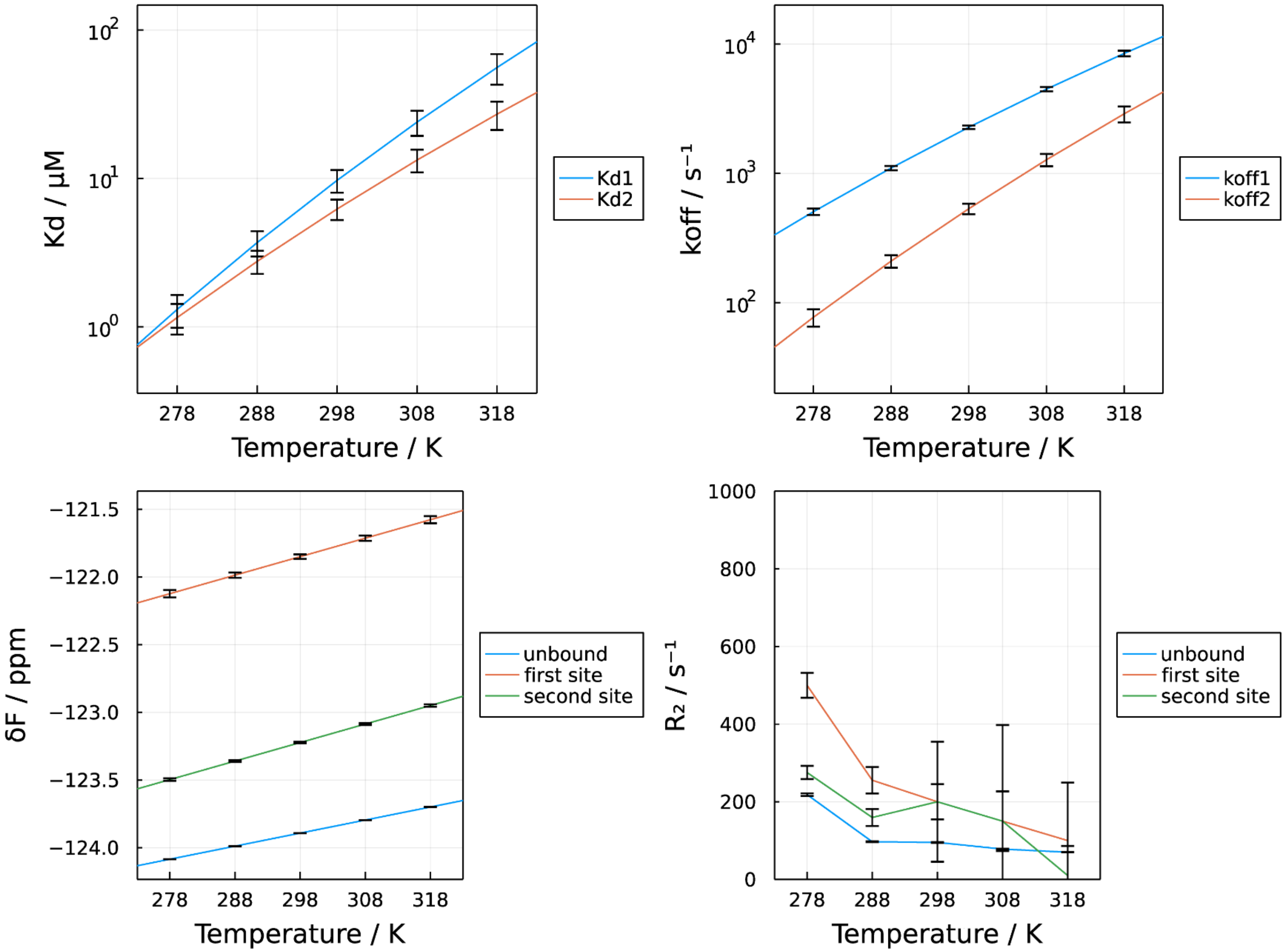


Figure 5.8. ^{19}F NMR parameter fits of SH3-Sos site 4 binding.

Conclusions: The ^{19}F data doesn't quite match up with the ^{13}C data presented in Chapter 3. The binding of Sos site 4 at 45 °C gives a $K_{D,app}$ of 45 μM, which makes sense, given that $K_{D,app}$ is a weighted average of the specific and non-specific binding. However, at low temperatures, it estimates a $K_{D,app}$ of 0.3 μM, which doesn't match what we have seen. The fits themselves look reasonable for site 4, we just likely need a more complex model to explain the data (Figures 5.5, 5.6).

However, the fits for the sites 2 and 4 construct (Figures 5.7, 5.8) are visibly poor, especially when you look at the free-state of SH3. Having a three-site mode is nice theoretically, but there are many more parameters that have to be added that make the power of analysis less precise.

Surface plasmon resonance experiments of Sos-SH3 binding

Rationale: We were struggling to figure out how to analyze the Sos-SH3 binding data via NMR, so we opted to try a different technique, one that we could theoretically use small sample volumes to apply crowding analysis. Surface plasmon resonance (SPR) is a well-used technique that might have some biological relevance for studying this interaction, as the disordered C-terminal tail is bound to several folded domains with a mass of 100 kDa. Additionally, it is likely to me bound to the membrane at the time of interaction, so having one end of the IDR bound might be reasonable.

Preliminary Data: At 25 °C, SPR estimated an $K_{D,app}$ of 49 ± 1 μM (Figure 5.9A), and when the SH3 analyte was added to a solution of 50 g/L GB1, the estimated affinity was 220 ± 20 μM (Figure 5.9B). Looking at the data in detail, we see that the addition of GB1 takes the sensor to ~50% saturation alone (Figure 5.9C), and the difference with the addition of SH3 (Figure 5.9D) is incredibly small (Figure 5.9B).

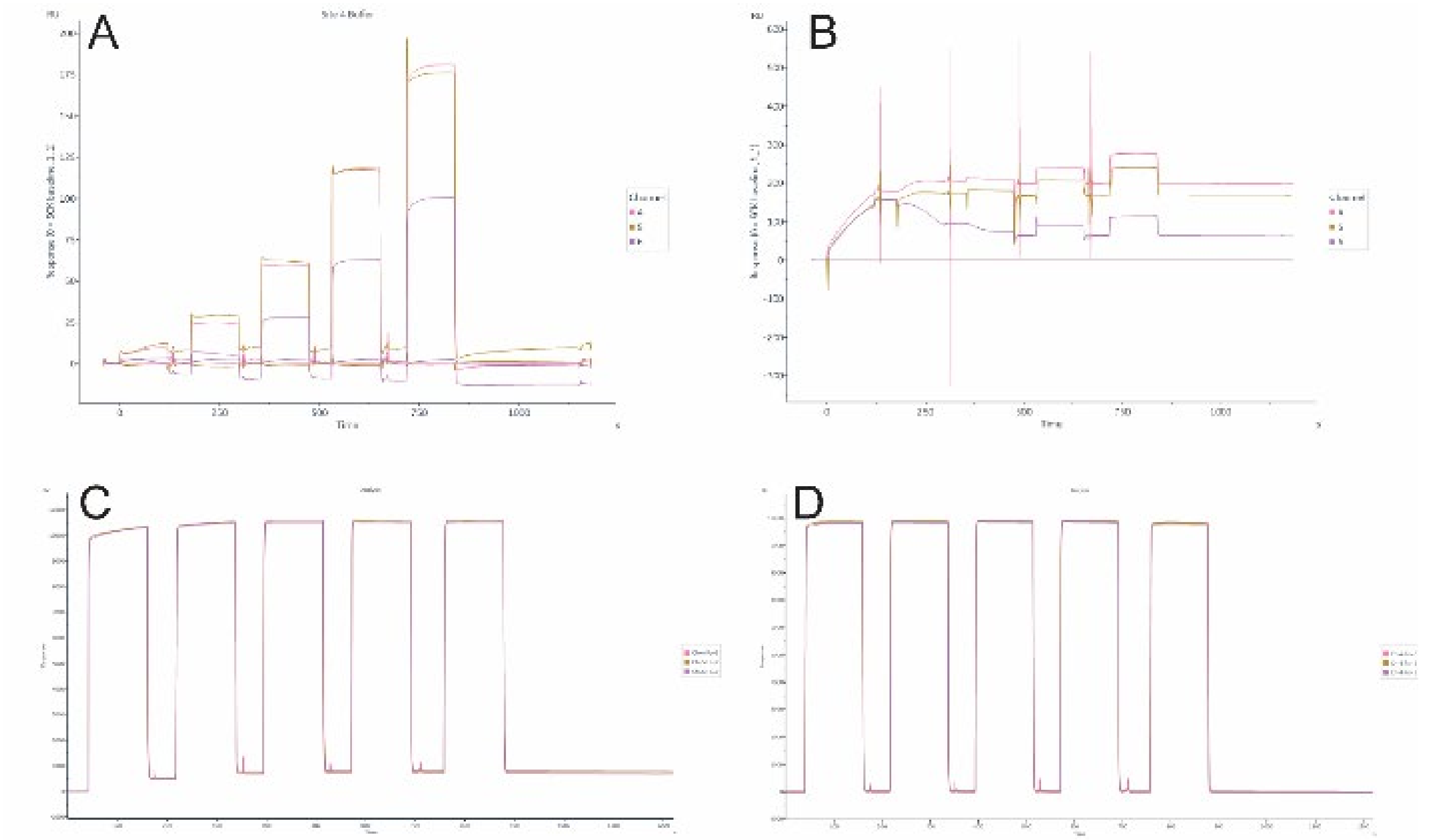


Figure 5.9. SPR titration of Sos and SH3 in the presence of GB1. Sos was the ligand, SH3 was the analyte. Measured with 50 mM HEPES, 50 mM bis-tris propane, 50 mM sodium acetate, pH = 7.5. A) Buffer, SH3 and Sos present. B) Buffer, SH3 and Sos present, 50 mg/mL GB1 present. (D with C subtracted) C) Raw readout of GB1 with no SH3. D) Raw readout of SH3 with GB1.

Conclusions: SPR readout is dependent on the mass of the species binding. SH3 is very small (6.3 kDa) and therefore naturally has a small readout. The addition of GB1 as a crowder pushes the signal very high, and I am unsure if the instrument has the dynamic range required to make the measurement. I also tried to add poly-ethylene glycol 8000 (PEG) to my background buffer, and the samples, but the viscosity was too high for the microfluidics of the system. Adding any other crowder to the background buffer would be unreasonable (you need > 1L).

C-TERMINAL CRKL DIMERIZATION

Rationale: Homo-dimer systems are of interest to macromolecular crowding studies because we can measure thermodynamic changes in the protein-complex thermodynamics, while assuming that the macromolecular crowders are affecting both species identically. While searching the literature, I found a paper that proposed the dimerization of the C-terminal SH3 domain of the protein Crk-L, which is involved in signal transduction.[281]

Preliminary results: I purified a His(X6)-tagged Crk-L C-terminal SH3 construct, and validated its purity by SDS-PAGE gel. I used dynamic light scattering to compare the hydrodynamic radius at 51.6, 31.3, 62.5, 125, 250, 500, 1000, and 2000 μM in 20 mM sodium phosphate, pH 7.5.

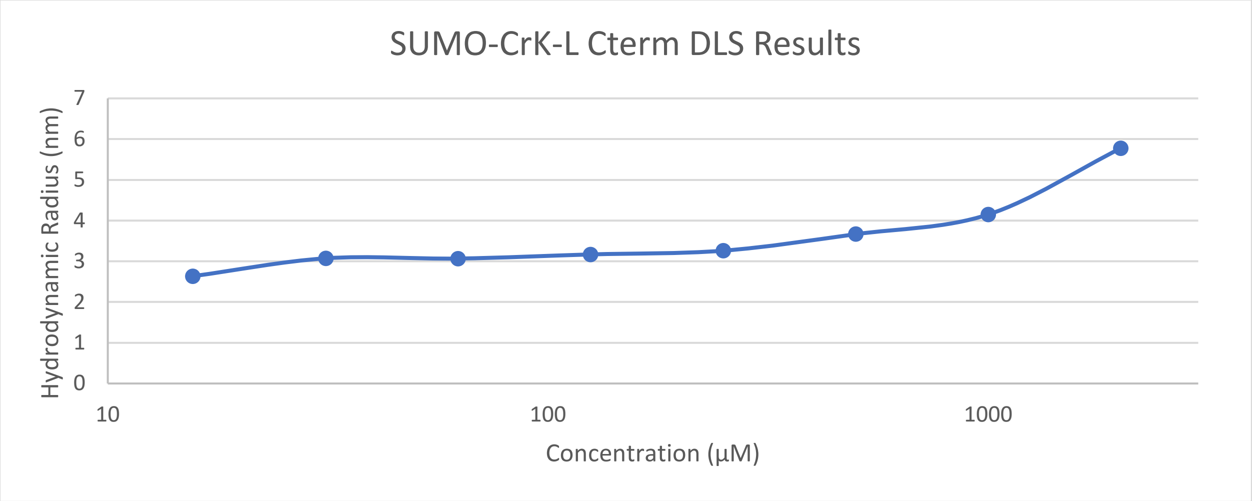


Figure 5.10. The hydrodynamic radius of Crk-L C-terminal SH3 as a function of concentration.

Conclusion: There appears to be measurable association around 500 μM (Figure 5.10). This interaction is likely not energetically favorable in a biological context, and probably is an artifact of high concentrations.

APPENDIX 3.1 CODE FOR FITTING 19F 1D DATA, ERROR ESTIMATES AND PLOTTING

```
clear all
close all
Temp = [5; 15; 25; 35; 45];
Kelvin = Temp + 273.5;
Temp = 1000./(Temp+273.15);
R = 0.001987204;

%3/24/22
Pep1KD = [6; 16; 33; 55.5; 104.4];
Pep1off = [4.8E2; 1.3E3; 3.6E3; 8.9E3; 2E4];

%3/14/22
Pep2KD = [10.6; 20; 35; 63.5; 123];
Pep2off = [4.9E2; 1.2E3; 3.5E3; 8.9E3; 1.6E4];

%7/26/22
Pep3KD = [9; 16; 29; 58.5; 92];
Pep3off = [4.9E2; 1.2E3; 3.5E3; 8.9E3; 1.6E4];

%collate KD and off vales
KD = [Pep1KD Pep2KD Pep3KD];
off = [Pep1off Pep2off Pep3off];


%generate on values
on = off./(KD.*1E-6);


%transform KD, off and on values
KD = log(1./(KD.*1E-6));
off = log((off*6.626E-34)./(Kelvin*1.380649E-23));
on = log((on*6.626E-34)./(Kelvin*1.380649E-23));

%Avgs and std devs
KDav = [mean(KD(1,:)); mean(KD(2,:)); mean(KD(3,:)); mean(KD(4,:)); mean(KD(5,:))];
KDerr = [std(KD(1,:)); std(KD(2,:)); std(KD(3,:)); std(KD(4,:)); std(KD(5,:))];


onav = [mean(on(1,:)); mean(on(2,:)); mean(on(3,:)); mean(on(4,:)); mean(on(5,:))];
onerr = [std(on(1,:)); std(on(2,:)); std(on(3,:)); std(on(4,:)); std(on(5,:))];


offav = [mean(off(1,:)); mean(off(2,:)); mean(off(3,:)); mean(off(4,:)); mean(off(5,:))];
offerr = [std(off(1,:)); std(off(2,:)); std(off(3,:)); std(off(4,:)); std(off(5,:))];

%% Sam's pep4 data
%Van't Hoff
InvTemp = [0.00360; 0.00347; 0.00335; 0.00325; 0.00314]; %For plotting averages with error
bars
InvTempLong = [0.00360; 0.00360; 0.00360;0.00347; 0.00347; 0.00347; 0.00335; 0.00335; 0.00335;
0.00335; 0.00335; 0.00335; 0.00325; 0.00325; 0.00325; 0.00314; 0.00314; 0.00314; 0.00314;
0.00314; 0.00314]; %For doing fit with all points

VHoff_all = [11.87; 10.45; 11.11; 11.33; 10.05; 9.85; 9.81; 9.48;9.88; 9.81;10.38; 9.17; 9.34;
9.00; 9.16; 9.11; 8.68; 8.19; 8.53; 8.34; 8.31]; %For doing van't Hoff analysis/fit
VHoff_averages = [11.14; 10.41; 9.76; 9.17; 8.53]; %ln(Ka), For plotting average value ln(Ka)
with error bars
```

```
VHoff_STD = [0.71; 0.80;0.41; 0.17;0.33]; %Standard deviation of buffer values
VHoff_SEM = [0.41; 0.46; 0.17; 0.10; 0.14]; %Standard error of mean for buffer values

%Fit Data to Van't Hoff equation
[xData, yData] = prepareCurveData( InvTempLong, VHoff_all);

ft = fittype('poly1');
[fit1, gof1] = fit(InvTempLong, VHoff_all, ft);
[fit6, gof6] = fit(Temp, VHoff_averages, ft);
%Display Data
fit1
gof1.rsquare

%Plot data points with standard errors and plot fit
e = errorbar(InvTemp, VHoff_averages, VHoff_SEM,
'o','MarkerSize',6,'LineWidth',2,'MarkerEdgeColor',[0.4392, 0.1882,
0.6275],'MarkerFaceColor',[0.4392, 0.1882, 0.6275]);
e.Color = [0.4392, 0.1882, 0.6275];
hold on
VHoff_fit = plot(fit1,'k--');
set(VHoff_fit,'LineWidth',1);
hold on

xlabel('1000/T (K^-^1)','fontsize', 14,'fontweight','bold');
ylabel('ln(K_A)','fontsize', 14,'fontweight','bold');
axis([0.00310 0.00365 8.0 12.0]);
legend('off');
set(gca,'fontsize',14,'LineWidth',2);


hold off

%Eyring kon
InvTemp = [0.00360; 0.00347; 0.00335; 0.00325; 0.00314]; %For plotting averages with error
bars
InvTempLong = [0.00360; 0.00360; 0.00360;0.00347; 0.00347; 0.00347; 0.00335; 0.00335; 0.00335;
0.00335; 0.00335; 0.00335; 0.00325; 0.00325; 0.00325; 0.00314; 0.00314; 0.00314; 0.00314;
0.00314; 0.00314]; %For doing fit with all points

Eon_all = [-12.27; -12.88; -12.65; -11.09; -12.78; -12.97; -11.79; -12.15; -11.88; -11.60; -
10.87; -12.11; -11.39; -11.87; -11.70; -11.11; -10.64; -11.34; -10.93; -11.04; -10.93]; %For
doing Eyring analysis/fit
Eon_averages = [-12.60; -12.28; -11.73; -11.65; -11.00]; %ln(konh/kBT), For plotting average
value ln(Ka) with error bars
Eon_STD = [0.31; 1.03; 0.47; 0.25; 0.23]; %Standard deviation of buffer values
Eon_SEM = [0.18; 0.60; 0.19; 0.14; 0.13]; %Standard error of mean for buffer values

%Fit Data to Eyring equation
[xData, yData] = prepareCurveData( InvTempLong, Eon_all);

ft = fittype('poly1');
[fit2, gof2] = fit(InvTempLong, Eon_all, ft);

[fit3, gof3] = fit(Temp, Eon_averages, ft);

%Display Data
fit2
gof1.rsquare

fit3
gof2.rsquare
```

```
%Plot data points with standard errors and plot fit
g = errorbar(InvTemp, Eon_averages, Eon_SEM,
'o','MarkerSize',6,'LineWidth',2,'MarkerEdgeColor',[0.7176, 0.5961,
0.7804],'MarkerFaceColor',[0.7176, 0.5961, 0.7804]);
g.Color = [0.7176, 0.5961, 0.7804];
hold on
Eon_fit = plot(fit2,'k--');
set(Eon_fit,'LineWidth',1);
hold on

xlabel('1000/T (K^-^1)','fontsize', 14,'fontweight','bold');
ylabel('ln(k_o_nh/k_BT)','fontsize', 14,'fontweight','bold');
axis([0.00310 0.00365 -14.0 -10.0]);
legend('off');
set(gca,'fontsize',14,'LineWidth',2);


hold off

%Eyring koff
InvTemp = [0.00360; 0.00347; 0.00335; 0.00325; 0.00314]; %For plotting averages with error
bars
InvTempLong = [0.00360; 0.00360; 0.00360;0.00347; 0.00347; 0.00347; 0.00335; 0.00335; 0.00335;
0.00335; 0.00335; 0.00335; 0.00325; 0.00325; 0.00325; 0.00314; 0.00314; 0.00314; 0.00314;
0.00314; 0.00314]; %For doing fit with all points

Eoff_all = [-24.14; -23.32; -23.75; -22.42; -22.83; -22.82; -21.59; -21.63; -21.76; -21.41; -
21.25; -21.28; -20.73; -20.88; -20.86; -20.22; -19.32; -19.52; -19.46; -19.38; -19.25]; %For
doing Eyring analysis/fit
Eoff_averages = [-23.74; -22.69; -21.49; -20.82; -19.53]; %ln(koffh/kBT), For plotting average
value with error bars
Eoff_STD = [0.41; 0.23; 0.206; 0.08; 0.35]; %Standard deviation of buffer values
Eoff_SEM = [0.24; 0.13; 0.084; 0.05; 0.14]; %Standard error of mean for buffer values

%Fit Data to Eyring equation
[xData, yData] = prepareCurveData( InvTempLong, Eoff_all);

ft = fittype('poly1');
[fit4, gof4] = fit(InvTempLong, Eoff_all, ft);

[fit5, gof5] = fit(Temp, Eoff_averages, ft);

%Display Data
fit4
gof1.rsquare

fit5
gof2.rsquare

%Plot data points with standard errors and plot fit
h = errorbar(InvTemp, Eoff_averages, Eoff_SEM,
'o','MarkerSize',6,'LineWidth',2,'MarkerEdgeColor',[0.2157, 0.1020,
0.3137],'MarkerFaceColor',[0.2157, 0.1020, 0.3137]);
h.Color = [0.2157, 0.1020, 0.3137];
hold on
Buffer_fit = plot(fit4,'k--');
set(Buffer_fit,'LineWidth',1);
hold on

xlabel('1000/T (K^-^1)','fontsize', 14,'fontweight','bold');
```

```
ylabel('ln(k_o_f_fh/k_BT)','fontsize', 14,'fontweight','bold');
axis([0.00310 0.00365 -25.0 -18.0]);
legend('off');
set(gca,'fontsize',14,'LineWidth',2);


hold off

%% Generate Bootrstrapped data
nsamples = 1000; %number of trials
numdata = length(Temp);

bootKD = zeros(5,nsamples);
bootoff = zeros(5,nsamples);
booton = zeros(5,nsamples);

for i=[1:numdata]
    bootKD(i,:)=datasample(KD(i,:),nsamples); %creates bootstraped dataset
end

for j=[1:numdata]
    bootoff(j,:)=datasample(off(j,:),nsamples); %creates bootstraped dataset
end

for i=[1:numdata]
    booton(i,:)=datasample(on(i,:),nsamples); %creates bootstraped dataset
end

%%
dH_vanHoff = zeros(nsamples,1);
dS_vanHoff = zeros(nsamples,1);
KDrsqr = zeros(nsamples,1);

linfit = fittype('a*T + b','independent',{'T'},'coefficients',{'a','b'});

for i=[1:nsamples] % This will fit the data for however many simulations you choose
    tmp = bootKD(:,i); % This reads the matrix row-by-row
    [fitKD,gof_KD]=fit(Temp,tmp,linfit,'startpoint',[1,1]); %Fits each column
    dH_vanHoff(i,1) = fitKD.a;
    dS_vanHoff(i,1) = fitKD.b;
    KDrsqr(i,1) = gof_KD.rsquare;
end %Then we loop back till i=nsamples

%% Eyring, association

dHA_Eyring = zeros(nsamples,1);
dSA_Eyring = zeros(nsamples,1);
onrsqr = zeros(nsamples,1);

linfit = fittype('a*T + b','independent',{'T'},'coefficients',{'a','b'});

for i=[1:nsamples] % This will fit the data for however many simulations you choose
    tmp = booton(:,i); % This reads the matrix row-by-row
    [fiton,gof_on]=fit(Temp,tmp,linfit,'startpoint',[1,1]); %Fits each column
    dHA_Eyring(i,1) = fiton.a;
    dSA_Eyring(i,1) = fiton.b;
    onrsqr(i,1) = gof_on.rsquare;
end %Then we loop back till i=nsamples

%% Eyring, dissociation
```

```
dHD_Eyring = zeros(nsamples,1);
dSD_Eyring = zeros(nsamples,1);
offrsqr = zeros(nsamples,1);

linfit = fittype('a*T + b','independent',{'T'},'coefficients',{'a','b'});

for i=[1:nsamples] % This will fit the data for however many simulations you choose
    tmp = bootoff(:,i); % This reads the matrix row-by-row
    [fitoff,gof_off]=fit(Temp,tmp,linfit,'startpoint',[1,1]); %Fits each column
    dHD_Eyring(i,1) = fitoff.a;
    dSD_Eyring(i,1) = fitoff.b;
    offrsqr(i,1) = gof_off.rsquare;
end %Then we loop back till i=nsamples

%% Processing fits

dH_vanHoff = -dH_vanHoff.*R*1000;
dS_vanHoff = dS_vanHoff*R*298.15;

dHA_Eyring = -dHA_Eyring.*R*1000;
dSA_Eyring = dSA_Eyring*R*298.15;

dHD_Eyring = -dHD_Eyring.*R*1000;
dSD_Eyring = dSD_Eyring*R*298.15;

Finalstats = [mean(dH_vanHoff) std(dH_vanHoff) mean(dS_vanHoff) std(dS_vanHoff);
mean(dHA_Eyring) std(dHA_Eyring) mean(dSA_Eyring) std(dSA_Eyring);    mean(dHD_Eyring)
std(dHD_Eyring) mean(dSD_Eyring) std(dSD_Eyring)]

%% export data

xlswrite('Pep2 analysis.xlsx',Finalstats,'Overview');
xlswrite('Pep2 analysis.xlsx',[dH_vanHoff dS_vanHoff KDrsqr],'Vant Hoff');
xlswrite('Pep2 analysis.xlsx',[dHA_Eyring dSA_Eyring onrsqr],'Eyring Association');
xlswrite('Pep2 analysis.xlsx',[dHD_Eyring dSD_Eyring offrsqr],'Eyring Dissociation');


%% Plot

figure
subplot(3,1,1)
hold on
errorbar(Temp,KDav,KDerr,'oblue','MarkerSize',5,'LineWidth',1)
a= polyfit(Temp,KDav,1);
b = polyval(a,Temp);
plot(Temp,b,'-blue');

hold on
f = errorbar(Temp, VHoff_averages, VHoff_SEM, 'o','MarkerSize',5,'LineWidth',1);
f.Color = [0.4392, 0.1882, 0.6275];
hold on
plot(fit6,'k--');
hold on

%errorbar(Temp,KDav4,KDerr4,'oblack','MarkerSize',5,'LineWidth',1)
%f= polyfit(Temp,KDav4,1);
%g = polyval(f,Temp);
%plot(Temp,g,'-black');

axis([3.13 3.61 8 13]);
xticks([3.1,3.2,3.3,3.4,3.5,3.6]);
```

```
xticklabels({'','','','','',''});
ylabel('ln({\itK_A})','fontsize',10);
legend('{\it\DeltaH_A}^{\circ\prime} = -11 \pm 1
kcal/mol','T{\it\DeltaS_{A,298.15K}}^{\circ\prime} = -5 \pm 1
kcal/mol','location','northwest');


subplot(3,1,2)
hold on
errorbar(Temp,onav,onerr,'oblack','MarkerSize',5,'LineWidth',1)
c= polyfit(Temp,onav,1);
d = polyval(c,Temp);
plot(Temp,d,'-black');


g = errorbar(Temp, Eon_averages, Eon_SEM, 'o','MarkerSize',5,'LineWidth',1);
g.Color = [0.7176, 0.5961, 0.7804];
hold on
Eon_fit = plot(fit3,'k--');
set(Eon_fit,'LineWidth',1);
hold on

%errorbar(Temp,on4,on4err,'oblack','MarkerSize',5,'LineWidth',1)
%h= polyfit(Temp,on4,1);
%i = polyval(h,Temp);
%plot(Temp,i,'-black');

axis([3.13 3.61 -14 -10]);
xticks([3.1,3.2,3.3,3.4,3.5,3.6]);
xticklabels({'','','','','',''});
ylabel('ln({\itk_{on}h/k_BT})','fontsize',10);
legend('{\it\DeltaH_A}^{\circ\prime} = 4 \pm 1
kcal/mol','T{\it\DeltaS_{A,298.15K}}^{\circ\prime} = -2 \pm 1
kcal/mol','location','northeast');

subplot(3,1,3)
hold on
errorbar(Temp,offav,offerr,'ored','MarkerSize',5,'LineWidth',1)
d= polyfit(Temp,offav,1);
e = polyval(d,Temp);
plot(Temp,e,'-red');

hold on
h = errorbar(Temp, Eoff_averages, Eoff_SEM, 'o','MarkerSize',5,'LineWidth',1);
h.Color = [0.2157, 0.1020, 0.3137];
hold on
Buffer_fit = plot(fit5,'k--');
set(Buffer_fit,'LineWidth',1);
hold on

%errorbar(Temp,off4,off4err,'oblue','MarkerSize',5,'LineWidth',1)
%j= polyfit(Temp,off4,1);
%k = polyval(j,Temp);
%plot(Temp,k,'-blue');

axis([3.13 3.61 -24.5 -19]);
xticks([3.1,3.2,3.3,3.4,3.5,3.6]);
xlabel('1000/T (K^{-1} )','fontsize', 10);
ylabel('ln({\itk_{off}h/k_BT})','fontsize',10);
```

```
legend('{\it\DeltaH_A}^{\circ\prime} = 15.5 \pm 0.4
kcal/mol','T{\it\DeltaS_{A,298.15K}}^{\circ\prime} = -2.8 \pm 0.4
kcal/mol','location','northeast');

%% looking at fit parameters
figure(2)
subplot(3,2,1)
histogram(dH_vanHoff);
title('{\it\DeltaH_A}^{\circ\prime}');

subplot(3,2,2)
histogram(dS_vanHoff);
title('T{\it\DeltaS_{A,298.15K}}^{\circ\prime}');

subplot(3,2,3)
histogram(dHA_Eyring)
title('{\it\DeltaH_A}^{\circ\prime}');

subplot(3,2,4)
histogram(dSA_Eyring);
title('T{\it\DeltaS_{A,298.15K}}^{\circ\prime}');

subplot(3,2,5)
histogram(dHD_Eyring);
title('{\it\DeltaH_D}^{\circ\prime}');

subplot(3,2,6)
histogram(dSD_Eyring);
title('T{\it\DeltaS_{D,298.15K}}^{\circ\prime}');
```

APPENDIX 4.1 CODE FOR BOOTSTRAPED ERROR ESTIMATES FOR IN-OOCYTE STABILITY MEASUREMENTS

```
clear all
close all

tempOocyte =[286.15;289.15;292.15;295.15;298.15;301.15]; %in-oocyte temperatures - kelvin
data = [616 644 571 557 521 456; 653 683 653 635 534 378; 642 700 693 631 528 426]; %measured
in-oocyte data (cal/mol). Each data set in individual rows.
nsamples = 1000; %number of trials
bootstrap = [datasample(data(:,1),nsamples), datasample(data(:,2),nsamples),
datasample(data(:,3),nsamples), datasample(data(:,4),nsamples),
datasample(data(:,5),nsamples), datasample(data(:,6),nsamples)]; %creates bootstraped dataset

TgCp = fittype('dH-(T*(dH/Tg))+ 970*(T-(Tg)-
(T*log(T/(Tg))))','independent',{'T'},'coefficients',{'dH','Tg'}); % Declaring the fit: a is
dH (20000), fixed Cp, t is Ts (50)
TsCp = fittype('dH+(970*((x-Ts)-
(x*log(x/Ts))))','independent',{'x'},'coefficients',{'dH','Ts'}); % Holding dCp constant
Tgm2 = fittype('dH*(1-(x/Tm))+970*((x-Tm)-
(x*log(x/(Tm))))','independent',{'x'},'coefficients',{'dH','Tm'}); %My equation from Michael's
paper
%%
dHTgfits = zeros(nsamples,1); %creates an 1000 x 1 empty vector that will allow you to fill in
the data as you make it
Tgfits = zeros(nsamples,1);

for i=[1:nsamples] % This will fit the data for however many simulations you choose
    boot = bootstrap(i,:)'; % This reads the matrix row-by-row and the ' converts it into a
column vector
    [fitx]=fit(tempOocyte,boot,TgCp,'startpoint',[20000,300],'algorithm','Trust-
Region','Robust','on','TolFun',1E-7,'TolX',1E-7,'MaxIter',1E5,'MaxFunEvals',1E5); %Fits each
row - Fixed Cp
    dHTgfits(i,1) = fitx.dH;
    Tgfits(i,1) = fitx.Tg;
    fprintf('%d ', i);
end %Then we loop back till i=nsamples


dHTsfits = zeros(nsamples,1);
Tsfits = zeros(nsamples,1);

for i=[1:nsamples] % This will fit the data for however many simulations you choose
    boot2 = bootstrap(i,:)'; % This reads the matrix row-by-row and the ' converts it into a
column vector
    [fitx]=fit(tempOocyte,boot2,TsCp,'startpoint',[1000,290],'algorithm','Trust-
Region','Robust','on','TolFun',1E-7,'TolX',1E-7,'MaxIter',1E5,'MaxFunEvals',1E5); %Fits each
row - Fixed Cp
    dHTsfits(i,1) = fitx.dH;
    Tsfits(i,1) = fitx.Ts;
    fprintf('%d ', i);
end %Then we loop back till i=nsamples

Tg_Fits = zeros(nsamples,2);
Tg_Fits(:,1) = dHTgfits;
Tg_Fits(:,2) = Tgfits;

Ts_Fits = zeros(nsamples,2);
Ts_Fits(:,1) = dHTsfits;
Ts_Fits(:,2) = Tsfits;
```

```
xlswrite('N51K In-oocyte Stability Bootstrap',Tg_Fits,'Tg_Fits');
xlswrite('N51K In-oocyte Stability Bootstrap',Ts_Fits,'Ts_Fits');

figure
hold on
tiledlayout(2,2)
%Tg_Estimate
nexttile
histogram(dHTgfits);
title('{\it\DeltaH_{U,Tm}\circ}')
xlabel('\it{kcal/mol}');
xticks([19000 19500 20000 20500 21000 21500]);
xticklabels({'19','','20','','21',''});
ylabel('counts');
mean1 = line([mean(dHTgfits), mean(dHTgfits)], ylim, 'LineWidth', 2, 'Color', 'r');
median1 = line([median(dHTgfits), median(dHTgfits)], ylim, 'LineWidth', 2, 'Color', 'g');
legend([mean1 median1], {'Mean', 'Median'}, 'Location', 'northwest')
nexttile
histogram(Tgfits)
title('T_{m}')
xlabel('Temperature (\it{K})');
ylabel('counts');
mean2 = line([mean(Tgfits), mean(Tgfits)], ylim, 'LineWidth', 2, 'Color', 'r');
median2 = line([median(Tgfits), median(Tgfits)], ylim, 'LineWidth', 2, 'Color', 'g');
legend([mean2 median2], {'Mean', 'Median'}, 'Location', 'northeast')
%Ts_Estimate
nexttile
histogram(dHTsfits)
title('{\it\DeltaH_{U,Ts}\circ}')
xlabel('\it{cal/mol_{Ts}}');
ylabel('counts');
mean3 = line([mean(dHTsfits), mean(dHTsfits)], ylim, 'LineWidth', 2, 'Color', 'r');
median3 = line([median(dHTsfits), median(dHTsfits)], ylim, 'LineWidth', 2, 'Color', 'g');
legend([mean3 median3], {'Mean', 'Median'}, 'Location', 'northwest')
nexttile
histogram(Tsfits)
title('T_{s}')
xlabel('Temperature (\it{K})');
ylabel('counts');
mean4 = line([mean(Tsfits), mean(Tsfits)], ylim, 'LineWidth', 2, 'Color', 'r');
median4 = line([median(Tsfits), median(Tsfits)], ylim, 'LineWidth', 2, 'Color', 'g');
legend([mean4 median4], {'Mean', 'Median'}, 'Location', 'northeast')

mean(dHTgfits)
std(dHTgfits)
mean(Tgfits)
std(Tgfits)

mean(dHTsfits)
std(dHTsfits)
mean(Tsfits)
std(Tsfits)
```

APPENDIX 4.2 CODE FOR VIOLIN PLOTS OF DISTRIBUTION COMPARISON

```
close all
clear all

dHTgfitsb = xlsread('Buffer_Bootstrap_output_042220','Bootstraped_dH_Tg', 'A1:A1000');
Tgfitsb = xlsread('Buffer_Bootstrap_output_042220','Bootstraped_Tg', 'A1:A1000');
dCp1 = xlsread('Buffer_Bootstrap_output_042220','Bootstraped_dCp_Tg', 'A1:A1000');
dHTsfitsb = xlsread('Buffer_Bootstrap_output_042220','Bootstraped_dH_Ts', 'A1:A1000');
Tsfitsb = xlsread('Buffer_Bootstrap_output_042220','Bootstraped_Ts', 'A1:A1000');

dHTgfits = xlsread('Incell_Bootstrap_output_042220','Bootstraped_dH_Tg', 'A1:A1000');
Tgfits = xlsread('Incell_Bootstrap_output_042220','Bootstraped_Tg', 'A1:A1000');
dHTsfits = xlsread('Incell_Bootstrap_output_042220','Bootstraped_dH_Ts', 'A1:A1000');
Tsfits = xlsread('Incell_Bootstrap_output_042220','Bootstraped_Ts', 'A1:A1000');

dHTgfitsbN51K = xlsread('N51K Buffer Stability Bootstrap','Tg_Fits', 'A1:A1000');
TgfitsbN51K = xlsread('N51K Buffer Stability Bootstrap','Tg_Fits', 'B1:B1000');
dCp1N51K = xlsread('N51K Buffer Stability Bootstrap','Tg_Fits', 'C1:C1000');
dHTsfitsbN51K = xlsread('N51K Buffer Stability Bootstrap','Ts_Fits', 'A1:A1000');
TsfitsbN51K = xlsread('N51K Buffer Stability Bootstrap','Ts_Fits', 'B1:B1000');


dHTgfitsN51K = xlsread('N51K In-oocyte Stability Bootstrap','Tg_Fits', 'A1:A1000');
TgfitsN51K = xlsread('N51K In-oocyte Stability Bootstrap','Tg_Fits', 'B1:B1000');
dHTsfitsN51K = xlsread('N51K In-oocyte Stability Bootstrap','Ts_Fits', 'A1:A1000');
TsfitsN51K = xlsread('N51K In-oocyte Stability Bootstrap','Ts_Fits', 'B1:B1000');
%%
figure(3)

dHTm = [dHTgfits,dHTgfitsb,dHTgfitsN51K,dHTgfitsbN51K];
Tm = [Tgfits,Tgfitsb,TgfitsN51K,TgfitsbN51K];
dHTs = [dHTsfits,dHTsfitsb,dHTsfitsN51K,dHTsfitsbN51K];
Ts = [Tsfits,Tsfitsb,TsfitsN51K,TsfitsbN51K];
Cp = [dCp1,dCp1N51K];
tiledlayout(3,2)


nexttile
a = distributionPlot([dHTgfits,dHTgfitsN51K],'widthDiv',[2
1],'histOri','left','color','b','showMM',5,'globalNorm',2);
b = distributionPlot(gca,[dHTgfitsb,dHTgfitsbN51K],'widthDiv',[2
2],'histOri','right','color','k','showMM',5,'globalNorm',2);
xticklabels({'WT SH3','SH3 N51K'});
ylabel('\Delta{\itH}\circ\prime_{U,{\itT}m} (kcal/mol)','fontweight','bold');
ylim([17000 25000])
yticks([18000 19000 20000 21000 22000 23000 24000 25000 26000]);
yticklabels({'18','','20','','22','','24','','26'});


nexttile
c = distributionPlot([Tgfits,TgfitsN51K],'widthDiv',[2
1],'histOri','left','color','b','showMM',5,'globalNorm',2);
d = distributionPlot(gca,[Tgfitsb,TgfitsbN51K],'widthDiv',[2
2],'histOri','right','color','k','showMM',5,'globalNorm',2);
xticklabels({'WT SH3','SH3 N51K'});
yticks([305 306 307 308 309 310 311 312 313 314 315 316 317 318]);
yticklabels({'','306','','308','','310','','312','','314','','316','','318'});
ylabel('{\itT}_{m} (K)','fontweight','bold');
```

```
nexttile
e = distributionPlot([dHTsfits,dHTsfitsN51K],'widthDiv',[2
1],'histOri','left','color','b','showMM',5,'globalNorm',2);
f = distributionPlot(gca,[dHTsfitsb,dHTsfitsbN51K],'widthDiv',[2
2],'histOri','right','color','k','showMM',5,'globalNorm',2);
xticklabels({'WT SH3','SH3 N51K'});
yticks([500 550 600 650 700 750 800 850 900 950]);
yticklabels({'0.5','','0.6','','0.7','','0.8','','0.9',''});
ylabel('\Delta{\itH}\circ\prime_{U,{\itT}s} (kcal/mol)','fontweight','bold');


nexttile
g = distributionPlot([Tsfits,Tsfitsb],'widthDiv',[2
1],'histOri','left','color','b','showMM',5,'globalNorm',2);
h = distributionPlot(gca,[TsfitsN51K,TsfitsbN51K],'widthDiv',[2
2],'histOri','right','color','k','showMM',5,'globalNorm',2);
xticklabels({'WT SH3','SH3 N51K'});
yticks([285 286 287 288 289 290 291 292 293 294 295 296]);
yticklabels({'','286','','288','','290','','292','','294','','296'});
ylabel('{\itT}_{s} (K)','fontweight','bold');


nexttile
i = distributionPlot(Cp,'showMM',5,'globalNorm',2,'histOpt',1);
xticklabels({'WT SH3','SH3 N51K'});
yticks([700 750 800 850 900 950 1000 1050 1100]);
yticklabels({'0.7','','0.8','','0.9','','1','','1.1'});
ylabel('{\it\DeltaC}_{P} (kcal/mol\cdotK)','fontweight','bold');
```

APPENDIX 4.3 CODE FOR CORRELATION TIME ESTIMATES

```
clear all;
close all;

global R1F R2F  %R1U R2U  % experimental R1
global calc_R1F calc_R2F %calc_R1U %calc_R2U   % obtained by simulations of R1
global err2 %err3                  % error
global r tm te S field
global MCR1F MCR2F

nsamples = 1000;
fileIDF = fopen('Foldedtm.dat','w');
fileIDU = fopen('Unfoldedtm.dat','w');
formatSpec = '%d %d %d\n'; %
field = [500.12 600.13]; %larmor freqs (1H)

R1F=[normrnd(2.45,0.07,nsamples,1) normrnd(2.26,0.08,nsamples,1)];  % R1 Folded field1 field2
R2F=[normrnd(30.1,0.4,nsamples,1) normrnd(39.7,0.6,nsamples,1)];  % R2 Folded
%R1U=[normrnd(1.4,0.2,nsamples,1) normrnd(1.4,0.2,nsamples,1)];  % R1 unfolded
%R2U=[normrnd(128,7,nsamples,1) normrnd(128,7,nsamples,1)];  % R2 unfolded

% ---------------- END OF INPUT -----------------
% --- FIT  ---
fprintf('\nFOLDED:\n')
for i=1:nsamples
    MCR1F = R1F(i,:);
    MCR2F = R2F(i,:);

    r = 2.0;
    tm = 10;
    te = 20e-12;
    S = 0.82;

    x0 = [r,tm];

    minopt = optimset('TolX',1e-14,'TolFun',1e-14,'MaxFunEvals',1e10,'MaxIter',1e10);

    x=fminsearch('F19_Folded_tc_equation_MC_global',x0,minopt);

    Ans= {x(1),x(2),err2};
    fprintf(fileIDF,formatSpec,Ans{1,:}); % prints to file
    %fprintf('%d ', i);
end

load Foldedtm.dat;
fprintf('\nF-H distance (A) = ');
mean(Foldedtm(:,1))
std(Foldedtm(:,1))
fprintf('\nCorrelation time (ns) = ');
mean(Foldedtm(:,2))
std(Foldedtm(:,2))

fprintf('\nInternal motion (ps) = ');
fprintf('%d\n',20);
fprintf('Order parameter = ');
fprintf('%d\n',0.82);
fprintf('Error = ');
fprintf('%d\n', err2);
```

```
fprintf('\n')

% fprintf('\nUNFOLDED:\n')
% for i=1:nsamples
%     MCR1U = R1U(i,:);
%     MCR2U = R2U(i,:);
%
% r = 2.5;
% tm = 10e-9;
% te = 1200e-12;
% S = 0.34;
%
% x0 = [r,tm];
%
% minopt = optimset('TolX',1e-14,'TolFun',1e-14,'MaxFunEvals',1e10,'MaxIter',1e10);
%
% x=fminsearch('F19_Unfolded_tc_equation_MC',x0,minopt);
%
% Ans= {x(1),x(2),err3};
% fprintf(fileIDU,formatSpec,Ans{1,:}); % prints to file
% %fprintf('%d ', i);
% end
%
% load Unfoldedtm.dat;
% fprintf('\nF-H distance (A) = ');
% mean(Unfoldedtm(:,1))
% std(Unfoldedtm(:,1))
% fprintf('\nCorrelation time (ns) = ');
% mean(Unfoldedtm(:,2))
% std(Unfoldedtm(:,2))
%
% fprintf('\nInternal motion (ps) = ');
% fprintf('%d\n',1200);
% fprintf('Order parameter = ');
% fprintf('%d\n',0.34);
% fprintf('Error = ');
% fprintf('%d\n', err3);
% fprintf('\n')
```

REFERENCES

1. Linderstrom-Lang, K. U., *Proteins and Enzymes: Lane Medical Lectures, 1951*. Stanford University Press: Stanford, CA, 1952.

2. Anfinsen, C. B., Principles That Govern the Folding of Protein Chains. *Science* **1973,** *181* (4096), 223-230.

3. Pauling, L.; Corey, R. B.; Branson, H. R., The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *Proc. Natl. Acad. Sci. USA* **1951,** *37* (4), 205-211.

4. Pauling, L.; Corey, R. B., Configurations of Polypeptide Chains with Favored Orientations around Single Bonds: Two New Pleated Sheets. *Proc. Natl. Acad. Sci. USA* **1951,** *37* (11), 729-40.

5. Rossmann, M. G.; Argos, P., Protein Folding. *Annu. Rev. Biochem.* **1981,** *50* (1), 497-532.

6. Cohen, R. D.; Pielak, G. J., A Cell Is More Than the Sum of Its (Dilute) Parts: A Brief History of Quinary Structure. *Protein Sci.* **2017,** *26* (3), 403-413.

7. Kendrew, J. C.; Dickerson, R. E.; Strandberg, B. E.; Hart, R. G.; Davies, D. R.; Phillips, D. C.; Shore, V. C., Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å. Resolution. *Nature* **1960,** *185* (4711), 422-427.

8. Blake, C. C. F.; Koenig, D. F.; Mair, G. A.; North, A. C. T.; Phillips, D. C.; Sarma, V. R., Structure of Hen Egg-White Lysozyme: A Three-Dimensional Fourier Synthesis at 2 Å Resolution. *Nature* **1965,** *206* (4986), 757-761.

9. Porter, L. L.; Looger, L. L., Extant Fold-Switching Proteins Are Widespread. *Proc. Natl. Acad. Sci. USA* **2018,** *115* (23), 5968-5973.

10. Dunker, A. K.; Babu, M. M.; Barbar, E.; Blackledge, M.; Bondos, S. E.; Dosztányi, Z.; Dyson, H. J.; Forman-Kay, J.; Fuxreiter, M.; Gsponer, J.; Han, K.-H.; Jones, D. T.; Longhi, S.; Metallo, S. J.; Nishikawa, K.; Nussinov, R.; Obradovic, Z.; Pappu, R. V.; Rost, B.; Selenko, P.; Subramaniam, V.; Sussman, J. L.; Tompa, P.; Uversky, V. N., What's in a Name? Why These Proteins Are Intrinsically Disordered. *Intrinsically Disordered Proteins* **2013,** *1* (1), e24157.

11. Dunker, A. K.; Garner, E.; Guilliot, S.; Romero, P.; Albrecht, K.; Hart, J.; Obradovic, Z.; Kissinger, C.; Villafranca, J. E., Protein Disorder and the Evolution of Molecular Recognition: Theory, Predictions and Observations. *Pac Symp Biocomput* **1998**, 473-84.

12. Radivojac, P.; Iakoucheva, L. M.; Oldfield, C. J.; Obradovic, Z.; Uversky, V. N.; Dunker, A. K., Intrinsic Disorder and Functional Proteomics. *Biophys. J.* **2007,** *92* (5), 1439-1456.

13. Brown, C. J.; Johnson, A. K.; Dunker, A. K.; Daughdrill, G. W., Evolution and Disorder. *Curr Opin Struct Biol* **2011,** *21* (3), 441-6.

14. Wright, P. E.; Dyson, H. J., Intrinsically Disordered Proteins in Cellular Signalling and Regulation. *Nat. Rev. Mol. Cell Biol.* **2015,** *16* (1), 18-29.

15. Basile, W.; Salvatore, M.; Bassot, C.; Elofsson, A., Why Do Eukaryotic Proteins Contain More Intrinsically Disordered Regions? *PLoS Comput. Biol.* **2019,** *15* (7), e1007186.

16. Zimmerman, S. B.; Trach, S. O., Estimation of Macromolecule Concentrations and Excluded Volume Effects for the Cytoplasm of *Escherichia coli*. *J. Mol. Biol.* **1991,** *222* (3), 599-620.

17. Zeskind, B. J.; Jordan, C. D.; Timp, W.; Trapani, L.; Waller, G.; Horodincu, V.; Ehrlich, D. J.; Matsudaira, P., Nucleic Acid and Protein Mass Mapping by Live-Cell Deep-Ultraviolet Microscopy. *Nat. Methods* **2007,** *4* (7), 567-9.

18. Eppig, J. J., Jr.; Dumont, J. N., Amino Acid Pools in Developing Oocytes of *Xenopus Laevis*. *Dev. Biol.* **1972,** *28* (3), 531-536.

19. Hallberg, R. L.; Smith, D. C., Ribosomal Protein Synthesis in *Xenopus Laevis* Oocytes. *Dev. Biol.* **1975,** *42* (1), 40-52.

20. Masui, Y., Oscillatory Activity of Maturation Promoting Factor (MPF) in Extracts of *Rana pipiens* Eggs. *J. Exp. Zool.* **1982,** *224* (3), 389-399.

21. Lohka, M. J.; Maller, J. L., Induction of Nuclear Envelope Breakdown, Chromosome Condensation, and Spindle Formation in Cell-Free Extracts. *J. Cell Biol.* **1985,** *101* (2), 518-523.

22. Taylor, M. A.; Smith, L. D., Accumulation of Free Amino Acids in Growing *Xenopus Laevis* Oocytes. *Dev. Biol.* **1987,** *124* (1), 287-290.

23. Berg, O. G., The Influence of Macromolecular Crowding on Thermodynamic Activity: Solubility and Dimerization Constants for Spherical and Dumbbell-Shaped Molecules in a Hard-Sphere Mixture. *Biopolymers* **1990,** *30* (11-12), 1027-1037.

24. Minton, A. P., Macromolecular Crowding and Molecular Recognition. *J. Mol. Recognit.* **1993,** *6* (4), 211-4.

25. Minton, A. P., The Influence of Macromolecular Crowding and Macromolecular Confinement on Biochemical Reactions in Physiological Media. *J. Biol. Chem.* **2001,** *276* (14), 10577-80.

26. Davis-Searles, P. R.; Saunders, A. J.; Erie, D. A.; Winzor, D. J.; Pielak, G. J., Interpreting the Effects of Small Uncharged Solutes on Protein-Folding Equilibria. *Annu. Rev. Biophys. Biomol. Struct.* **2001,** *30* (1), 271-306.

27. Sharp, K. A., Analysis of the Size Dependence of Macromolecular Crowding Shows That Smaller Is Better. *Proc. Natl. Acad. Sci. USA* **2015,** *112* (26), 7990-7995.

28. Asakura, S.; Oosawa, F., On Interaction between Two Bodies Immersed in a Solution of Macromolecules. *The Journal of Chemical Physics* **1954,** *22* (7), 1255-1256.

29. Asakura, S.; Oosawa, F., Interaction between Particles Suspended in Solutions of Macromolecules. *Journal of Polymer Science* **1958,** *33* (126), 183-192.

30. Speer, S. L.; Stewart, C. J.; Sapir, L.; Harries, D.; Pielak, G. J., Macromolecular Crowding Is More Than Hard-Core Repulsions. *Annu. Rev. Biophys.* **2022,** *51* (1), 267-300.

31. Sarkar, M.; Li, C.; Pielak, G. J., Soft Interactions and Crowding. *Biophys. Rev.* **2013,** *5* (2), 187-194.

32. Smith, A. E.; Zhou, L. Z.; Gorensek, A. H.; Senske, M.; Pielak, G. J., In-Cell Thermodynamics and a New Role for Protein Surfaces. *Proc. Natl. Acad. Sci. USA* **2016,** *113* (7), 1725-30.

33. Senske, M.; Smith, A. E.; Pielak, G. J., Protein Stability in Reverse Micelles. *Angew. Chem., Int. Ed.* **2016,** *55* (11), 3586-3589.

34. Stadmiller, S. S.; Gorensek-Benitez, A. H.; Guseman, A. J.; Pielak, G. J., Osmotic Shock Induced Protein Destabilization in Living Cells and Its Reversal by Glycine Betaine. *J. Mol. Biol.* **2017,** *429* (8), 1155-1161.

35. Guseman, A. J.; Perez Goncalves, G. M.; Speer, S. L.; Young, G. B.; Pielak, G. J., Protein Shape Modulates Crowding Effects. *Proc. Natl. Acad. Sci. USA* **2018,** *115* (43), 10965-10970.

36. Guseman, A. J.; Speer, S. L.; Perez Goncalves, G. M.; Pielak, G. J., Surface Charge Modulates Protein-Protein Interactions in Physiologically Relevant Environments. *Biochemistry* **2018,** *57* (11), 1681-1684.

37. Thole, J. F.; Fadero, T. C.; Bonin, J. P.; Stadmiller, S. S.; Giudice, J. A.; Pielak, G. J., Danio rerio Oocytes for Eukaryotic in-Cell NMR. *Biochemistry* **2021,** *60* (6), 451-459.

38. Speer, S. L.; Zheng, W.; Jiang, X.; Chu, I. T.; Guseman, A. J.; Liu, M.; Pielak, G. J.; Li, C., The Intracellular Environment Affects Protein-Protein Interactions. *Proc. Natl. Acad. Sci. USA* **2021,** *118* (11).

39. Fonin, A. V.; Darling, A. L.; Kuznetsova, I. M.; Turoverov, K. K.; Uversky, V. N., Intrinsically Disordered Proteins in Crowded Milieu: When Chaos Prevails within the Cellular Gumbo. *Cell. Mol. Life Sci.* **2018,** *75* (21), 3907-3929.

40. Borgia, A.; Borgia, M. B.; Bugge, K.; Kissling, V. M.; Heidarsson, P. O.; Fernandes, C. B.; Sottini, A.; Soranno, A.; Buholzer, K. J.; Nettels, D.; Kragelund, B. B.; Best, R. B.; Schuler, B., Extreme Disorder in an Ultrahigh-Affinity Protein Complex. *Nature* **2018,** *555* (7694), 61-66.

41. Fersht, A. R., Characterizing Transition States in Protein Folding: An Essential Step in the Puzzle. *Curr. Opin. Struct. Biol.* **1995,** *5* (1), 79-84.

42. Bugge, K.; Brakti, I.; Fernandes, C. B.; Dreier, J. E.; Lundsgaard, J. E.; Olsen, J. G.; Skriver, K.; Kragelund, B. B., Interactions by Disorder – a Matter of Context. *Front. Mol. Biosci.* **2020,** *7*, 110.

43. Mollica, L.; Bessa, L. M.; Hanoulle, X.; Jensen, M. R.; Blackledge, M.; Schneider, R., Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment. *Front. Mol. Biosci.* **2016,** *3*, 52.

44. Arai, M.; Sugase, K.; Dyson, H. J.; Wright, P. E., Conformational Propensities of Intrinsically Disordered Proteins Influence the Mechanism of Binding and Folding. *Proc. Natl. Acad. Sci. USA* **2015,** *112* (31), 9614-9619.

45. Kurcinski, M.; Pawel Ciemny, M.; Oleniecki, T.; Kuriata, A.; Badaczewska-Dawid, A. E.; Kolinski, A.; Kmiecik, S., Cabs-Dock Standalone: A Toolbox for Flexible Protein–Peptide Docking. *Bioinformatics* **2019,** *35* (20), 4170-4172.

46. Fersht, A. R.; Matouschek, A.; Serrano, L., The Folding of an Enzyme: I. Theory of Protein Engineering Analysis of Stability and Pathway of Protein Folding. *J. Mol. Biol.* **1992,** *224* (3), 771-782.

47. Rogers, J. M.; Oleinikovas, V.; Shammas, S. L.; Wong, C. T.; De Sancho, D.; Baker, C. M.; Clarke, J., Interplay between Partner and Ligand Facilitates the Folding and Binding of an Intrinsically Disordered Protein. *Proc. Natl. Acad. Sci. USA* **2014,** *111* (43), 15420-15425.

48. Bonetti, D.; Troilo, F.; Brunori, M.; Longhi, S.; Gianni, S., How Robust Is the Mechanism of Folding-Upon-Binding for an Intrinsically Disordered Protein? *Biophys. J.* **2018,** *114* (8), 1889-1894.

49. Sørensen, C. S.; Jendroszek, A.; Kjaergaard, M., Linker Dependence of Avidity in Multivalent Interactions between Disordered Proteins. *J. Mol. Biol.* **2019,** *431* (24), 4784-4795.

50. Berlow, R. B.; Dyson, H. J.; Wright, P. E., Multivalency Enables Unidirectional Switch-Like Competition between Intrinsically Disordered Proteins. *Proc. Natl. Acad. Sci. USA* **2022,** *119* (3).

51. Olsen, J. G.; Teilum, K.; Kragelund, B. B., Behaviour of Intrinsically Disordered Proteins in Protein–Protein Complexes with an Emphasis on Fuzziness. *Cell. Mol. Life Sci.* **2017,** *74* (17), 3175-3183.

52. Choi, J. M.; Holehouse, A. S.; Pappu, R. V., Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu Rev Biophys* **2020,** *49*, 107-133.

53. Bonucci, A.; Palomino-Schätzlein, M.; Malo de Molina, P.; Arbe, A.; Pierattelli, R.; Rizzuti, B.; Iovanna, J. L.; Neira, J. L., Crowding Effects on the Structure and Dynamics of the Intrinsically Disordered Nuclear Chromatin Protein Nupr1. *Front. Mol. Biosci.* **2021,** *8*.

54. Timm, D. E.; Vissavajjhala, P.; Ross, A. H.; Neet, K. E., Spectroscopic and Chemical Studies of the Interaction between Nerve Growth Factor (Ngf) and the Extracellular Domain of the Low Affinity Ngf Receptor. *Protein Sci.* **1992,** *1* (8), 1023-1031.

55. Uversky, V. N.; Li, J.; Fink, A. L., Evidence for a Partially Folded Intermediate in A-Synuclein Fibril Formation. *J. Biol. Chem.* **2001,** *276* (14), 10737-10744.

56. Uversky, V. N.; Permyakov, S. E.; Zagranichny, V. E.; Rodionov, I. L.; Fink, A. L.; Cherskaya, A. M.; Lyubov A.Wasserman, A.; Permyakov, E. A., Effect of Zinc and Temperature on the Conformation of the Γ Subunit of Retinal Phosphodiesterase: A Natively Unfolded Protein. *J. Proteome Res.* **2002,** *1* (2), 149-159.

57. Permyakov, S. E.; Millett, I. S.; Doniach, S.; Permyakov, E. A.; Uversky, V. N., Natively Unfolded C-terminal Domain of Caldesmon Remains Substantially Unstructured after the Effective Binding to Calmodulin. *Proteins: Struct., Funct., Genet.* **2003,** *53* (4), 855-Na.

58. Wetzler, D. E.; Fuchs Wightman, F.; Bucci, H. A.; Rinaldi, J.; Caramelo, J. J.; Iusem, N. D.; Ricardi, M. M., Conformational Plasticity of the Intrinsically

Disordered Protein Asr1 Modulates Its Function as a Drought Stress-Responsive Gene. *PLoS One* **2018,** *13* (8), e0202808.

59. Nettels, D.; Müller-Späth, S.; Küster, F.; Hofmann, H.; Haenni, D.; Rüegger, S.; Reymond, L.; Hoffmann, A.; Kubelka, J.; Heinz, B.; Gast, K.; Best, R. B.; Schuler, B., Single-Molecule Spectroscopy of the Temperature-Induced Collapse of Unfolded Proteins. *Proc. Natl. Acad. Sci. USA* **2009,** *106* (49), 20740-20745.

60. Ciasca, G.; Campi, G.; Battisti, A.; Rea, G.; Rodio, M.; Papi, M.; Pernot, P.; Tenenbaum, A.; Bianconi, A., Continuous Thermal Collapse of the Intrinsically Disordered Protein Tau Is Driven by Its Entropic Flexible Domain. *Langmuir* **2012,** *28* (37), 13405-13410.

61. Wuttke, R.; Hofmann, H.; Nettels, D.; Borgia, M. B.; Mittal, J.; Best, R. B.; Schuler, B., Temperature-Dependent Solvation Modulates the Dimensions of Disordered Proteins. *Proc. Natl. Acad. Sci. USA* **2014,** *111* (14), 5213-5218.

62. Battisti, A.; Ciasca, G.; Grottesi, A.; Tenenbaum, A., Thermal Compaction of the Intrinsically Disordered Protein Tau: Entropic, Structural, and Hydrophobic Factors. *Phys. Chem. Chem. Phys.* **2017,** *19* (12), 8435-8446.

63. Moretti, P.; Mariani, P.; Ortore, M. G.; Plotegher, N.; Bubacco, L.; Beltramini, M.; Spinozzi, F., Comprehensive Structural and Thermodynamic Analysis of Prefibrillar Wt A-Synuclein and Its G51d, E46k, and A53t Mutants by a Combination of Small-Angle X-Ray Scattering and Variational Bayesian Weighting. *J. Chem. Inf. Model.* **2020,** *60* (10), 5265-5281.

64. Bhuyan, A. K., Negative Thermal Expansion and Disorder-to-Order Collapse of an Intrinsically Disordered Protein under Marginally Denaturing Conditions. *J. Phys. Chem. B* **2022,** *126* (27), 5055-5065.

65. Zerze, G. H.; Best, R. B.; Mittal, J., Sequence- and Temperature-Dependent Properties of Unfolded and Disordered Proteins from Atomistic Simulations. *J. Phys. Chem. B* **2015,** *119* (46), 14622-14630.

66. Park, S.; Barnes, R.; Lin, Y.; Jeon, B.-J.; Najafi, S.; Delaney, K. T.; Fredrickson, G. H.; Shea, J.-E.; Hwang, D. S.; Han, S., Dehydration Entropy Drives Liquid-Liquid Phase Separation by Molecular Crowding. *Commun. Chem.* **2020,** *3* (1).

67. Phillip, Y.; Sherman, E.; Haran, G.; Schreiber, G., Common Crowding Agents Have Only a Small Effect on Protein-Protein Interactions. *Biophys. J.* **2009,** *97* (3), 875-885.

68. Lee, C. W.; Martinez-Yamout, M. A.; Dyson, H. J.; Wright, P. E., Structure of the P53 Transactivation Domain in Complex with the Nuclear Receptor Coactivator Binding Domain of Creb Binding Protein. *Biochemistry* **2010,** *49* (46), 9964-9971.

69. McDonald, C. B.; Seldeen, K. L.; Deegan, B. J.; Farooq, A., SH3 Domains of Grb2 Adaptor Bind to Pxpsipxr Motifs within the Sos1 Nucleotide Exchange Factor in a Discriminate Manner. *Biochemistry* **2009,** *48* (19), 4074-85.

70. Stadmiller, S. S.; Aguilar, J. S.; Waudby, C. A.; Pielak, G. J., Rapid Quantification of Protein-Ligand Binding Via ^{19}F NMR Lineshape Analysis. *Biophys. J.* **2020,** *118* (10), 2537-2548.

71. Stadmiller, S. S.; Aguilar, J. S.; Parnham, S.; Pielak, G. J., Protein-Peptide Binding Energetics under Crowded Conditions. *J. Phys. Chem. B* **2020,** *124* (42), 9297-9309.

72. Schnatwinkel, J.; Herrmann, C., The Interaction Strength of an Intrinsically Disordered Protein Domain with Its Binding Partner Is Little Affected by Very Different Cosolutes. *Phys. Chem. Chem. Phys.* **2020,** *22* (47), 27903-27911.

73. Gruber, T.; Lewitzky, M.; Machner, L.; Weininger, U.; Feller, S. M.; Balbach, J., Macromolecular Crowding Induces a Binding Competent Transient Structure in Intrinsically Disordered Gab1. *J. Mol. Biol.* **2022,** *434* (5), 167407.

74. Cino, E. A.; Karttunen, M.; Choy, W.-Y., Effects of Molecular Crowding on the Dynamics of Intrinsically Disordered Proteins. *PLoS ONE* **2012,** *7* (11), e49876.

75. Qin, S.; Zhou, H.-X., Effects of Macromolecular Crowding on the Conformational Ensembles of Disordered Proteins. *The Journal of Physical Chemistry Letters* **2013,** *4* (20), 3429-3434.

76. Banerjee, P. R.; Moosa, M. M.; Deniz, A. A., Two-Dimensional Crowding Uncovers a Hidden Conformation of A-Synuclein. *Angew. Chem., Int. Ed.* **2016,** *55* (41), 12789-12792.

77. Nguemaha, V.; Qin, S.; Zhou, H. X., Atomistic Modeling of Intrinsically Disordered Proteins under Polyethylene Glycol Crowding: Quantitative Comparison with Experimental Data and Implication of Protein-Crowder Attraction. *J Phys Chem B* **2018,** *122* (49), 11262-11270.

78. Zosel, F.; Soranno, A.; Buholzer, K. J.; Nettels, D.; Schuler, B., Depletion Interactions Modulate the Binding between Disordered Proteins in Crowded Environments. *Proc. Natl. Acad. Sci. USA* **2020,** *117* (24), 13480-13489.

79. Harmon, T. S.; Holehouse, A. S.; Rosen, M. K.; Pappu, R. V., Intrinsically Disordered Linkers Determine the Interplay between Phase Separation and Gelation in Multivalent Proteins. *Elife* **2017,** *6*.

80. Harmon, T. S.; Holehouse, A. S.; Pappu, R. V., Differential Solvation of Intrinsically Disordered Linkers Drives the Formation of Spatially Organized Droplets in Ternary Systems of Linear Multivalent Proteins. *New Journal of Physics* **2018,** *20* (4), 045002.

81. Boeynaems, S.; Holehouse, A. S.; Weinhardt, V.; Kovacs, D.; Van Lindt, J.; Larabell, C.; Van Den Bosch, L.; Das, R.; Tompa, P. S.; Pappu, R. V.; Gitler, A. D., Spontaneous Driving Forces Give Rise to Protein-RNA Condensates with Coexisting Phases and Complex Material Properties. *Proc. Natl. Acad. Sci. USA* **2019,** *116* (16), 7889-7898.

82. Mittag, T.; Pappu, R. V., A Conceptual Framework for Understanding Phase Separation and Addressing Open Questions and Challenges. *Mol Cell* **2022,** *82* (12), 2201-2214.

83. Mitrea, D. M.; Kriwacki, R. W., Phase Separation in Biology; Functional Organization of a Higher Order. *Cell Commun. Signaling* **2016,** *14* (1), 1.

84. Roden, C.; Gladfelter, A. S., RNA Contributions to the Form and Function of Biomolecular Condensates. *Nat. Rev. Mol. Cell Biol.* **2021,** *22* (3), 183-195.

85. Kim, T.-S.; Zhang, L.; Il Ahn, J.; Meng, L.; Chen, Y.; Lee, E.; Bang, J. K.; Lim, J. M.; Ghirlando, R.; Fan, L.; Wang, Y.-X.; Kim, B. Y.; Park, J.-E.; Lee, K. S., Molecular Architecture of a Cylindrical Self-Assembly at Human Centrosomes. *Nat. Commun.* **2019,** *10* (1), 1151.

86. Ahn, J. I.; Park, J.-E.; Meng, L.; Zhang, L.; Kim, T.-S.; Kruhlak, M. J.; Kim, B. Y.; Lee, K. S., Phase Separation of the Cep63•Cep152 Complex Underlies the Formation of Dynamic Supramolecular Self-Assemblies at Human Centrosomes. *Cell Cycle* **2020,** *19* (24), 3437-3457.

87. André, A. A. M.; Spruijt, E., Liquid–Liquid Phase Separation in Crowded Environments. *Int. J. Mol. Sci.* **2020,** *21* (16), 5908.

88. Itakura, A. K.; Futia, R. A.; Jarosz, D. F., It Pays to Be in Phase. *Biochemistry* **2018,** *57* (17), 2520-2529.

89. Alberti, S.; Gladfelter, A.; Mittag, T., Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **2019,** *176* (3), 419-434.

90. Leslie, M., Separation Anxiety. *Science* **2021,** *371* (6527), 336-338.

91. Poudyal, R. R.; Guth-Metzler, R. M.; Veenis, A. J.; Frankel, E. A.; Keating, C. D.; Bevilacqua, P. C., Template-Directed RNA Polymerization and Enhanced Ribozyme Catalysis inside Membraneless Compartments Formed by Coacervates. *Nat. Commun.* **2019,** *10* (1).

92. Huang, W. Y. C.; Alvarez, S.; Kondo, Y.; Lee, Y. K.; Chung, J. K.; Lam, H. Y. M.; Biswas, K. H.; Kuriyan, J.; Groves, J. T., A Molecular Assembly Phase Transition and Kinetic Proofreading Modulate Ras Activation by Sos. *Science* **2019,** *363* (6431), 1098-1103.

93. Peeples, W.; Rosen, M. K., Mechanistic Dissection of Increased Enzymatic Rate in a Phase-Separated Compartment. *Nat. Chem. Biol.* **2021,** *17* (6), 693-702.

94. Shin, Y.; Brangwynne, C. P., Liquid Phase Condensation in Cell Physiology and Disease. *Science* **2017,** *357* (6357), eaaf4382.

95. Banani, S. F.; Rice, A. M.; Peeples, W. B.; Lin, Y.; Jain, S.; Parker, R.; Rosen, M. K., Compositional Control of Phase-Separated Cellular Bodies. *Cell* **2016,** *166* (3), 651-663.

96. Sang, D.; Shu, T.; Pantoja, C. F.; Ibáñez De Opakua, A.; Zweckstetter, M.; Holt, L. J., Condensed-Phase Signaling Can Expand Kinase Specificity and Respond to Macromolecular Crowding. *Mol. Cell* **2022,** *82* (19), 3693-3711.e10.

97. Bah, A.; Forman-Kay, J. D., Modulation of Intrinsically Disordered Protein Function by Post-Translational Modifications. *J. Biol. Chem.* **2016,** *291* (13), 6696-6705.

98. Lavoie, H.; Gagnon, J.; Therrien, M., Erk Signalling: A Master Regulator of Cell Behaviour, Life and Fate. *Nat. Rev. Mol. Cell Biol.* **2020,** *21* (10), 607-632.

99. Widmann, C.; Gibson, S.; Jarpe, M. B.; Johnson, G. L., Mitogen-Activated Protein Kinase: Conservation of a Three-Kinase Module from Yeast to Human. *Physiol. Rev.* **1999,** *79* (1), 143-180.

100. Raabe, T., The Sevenless Signaling Pathway: Variations of a Common Theme. *Biochim. Biophys. Acta* **2000,** *1496* (2-3), 151-63.

101. Tena, G.; Asai, T.; Chiu, W.-L.; Sheen, J., Plant Mitogen-Activated Protein Kinase Signaling Cascades. *Curr. Opin. Plant Biol.* **2001,** *4* (5), 392-400.

102. Cohen, S.; Carpenter, G., Human Epidermal Growth Factor: Isolation and Chemical and Biological Properties. *Proc. Natl. Acad. Sci. USA* **1975,** *72* (4), 1317-21.

103. Richman, R. A.; Claus, T. H.; Pilkis, S. J.; Friedman, D. L., Hormonal Stimulation of DNA Synthesis in Primary Cultures of Adult Rat Hepatocytes. *Proc. Natl. Acad. Sci. USA* **1976,** *73* (10), 3589-93.

104. Cohen, S.; Ushiro, H.; Stoscheck, C.; Chinkers, M., A Native 170,000 Epidermal Growth Factor Receptor-Kinase Complex from Shed Plasma Membrane Vesicles. *J Biol Chem* **1982,** *257* (3), 1523-31.

105. Cohen, S.; Fava, R. A.; Sawyer, S. T., Purification and Characterization of Epidermal Growth Factor Receptor/Protein Kinase from Normal Mouse Liver. *Proc. Natl. Acad. Sci. USA* **1982,** *79* (20), 6237-41.

106. Bessman, Nicholas J.; Bagchi, A.; Ferguson, Kathryn M.; Lemmon, Mark A., Complex Relationship between Ligand Binding and Dimerization in the Epidermal Growth Factor Receptor. *Cell Reports* **2014,** *9* (4), 1306-1317.

107. Ullrich, A.; Schlessinger, J., Signal Transduction by Receptors with Tyrosine Kinase Activity. *Cell* **1990,** *61* (2), 203-212.

108. Hubbard, S. R.; Till, J. H., Protein Tyrosine Kinase Structure and Function. *Annu. Rev. Biochem.* **2000,** *69* (1), 373-398.

109. Li, W. X., Functions and Mechanisms of Receptor Tyrosine Kinase Torso Signaling: Lessons from *Drosophila* Embryonic Terminal Development. *Dev. Dyn.* **2005,** *232* (3), 656-672.

110. Bonfini, L.; Karlovich, C. A.; Dasgupta, C.; Banerjee, U., The Son of Sevenless Gene Product: A Putative Activator of Ras. *Science* **1992,** *255* (5044), 603-606.

111. Olivier, J. P.; Raabe, T.; Henkemeyer, M.; Dickson, B.; Mbamalu, G.; Margolis, B.; Schlessinger, J.; Hafen, E.; Pawson, T., A Drosophila SH2-SH3 Adaptor Protein Implicated in Coupling the Sevenless Tyrosine Kinase to an Activator of Ras Guanine Nucleotide Exchange, Sos. *Cell* **1993,** *73* (1), 179-91.

112. Chardin, P.; Camonis, J. H.; Gale, N. W.; van Aelst, L.; Schlessinger, J.; Wigler, M. H.; Bar-Sagi, D., Human Sos1: A Guanine Nucleotide Exchange Factor for Ras That Binds to Grb2. *Science* **1993,** *260* (5112), 1338-1343.

113. Chong, H.; Vikis, H. G.; Guan, K.-L., Mechanisms of Regulating the Raf Kinase Family. *Cellular Signalling* **2003,** *15* (5), 463-469.

114. Freeman, A. K.; Ritt, D. A.; Morrison, D. K., The Importance of Raf Dimerization in Cell Signaling. *Small GTPases* **2013,** *4* (3), 180-185.

115. Lenormand, P.; Sardet, C.; Pagès, G.; L'Allemain, G.; Brunet, A.; Pouysségur, J., Growth Factors Induce Nuclear Translocation of Map Kinases (P42mapk and

P44mapk) but Not of Their Activator Map Kinase Kinase (P45mapkk) in Fibroblasts. *J. Cell Biol.* **1993,** *122* (5), 1079-1088.

116. Zheng, Y.; Zhang, C.; Croucher, D. R.; Soliman, M. A.; St-Denis, N.; Pasculescu, A.; Taylor, L.; Tate, S. A.; Hardy, W. R.; Colwill, K.; Dai, A. Y.; Bagshaw, R.; Dennis, J. W.; Gingras, A.-C.; Daly, R. J.; Pawson, T., Temporal Regulation of Egf Signalling Networks by the Scaffold Protein Shc1. *Nature* **2013,** *499* (7457), 166-171.

117. Reddy, R. J.; Gajadhar, A. S.; Swenson, E. J.; Rothenberg, D. A.; Curran, T. G.; White, F. M., Early Signaling Dynamics of the Epidermal Growth Factor Receptor. *Proc. Natl. Acad. Sci. USA* **2016,** *113* (11), 3114-3119.

118. Kong, Q.; Ke, M.; Weng, Y.; Qin, Y.; He, A.; Li, P.; Cai, Z.; Tian, R., Dynamic Phosphotyrosine-Dependent Signaling Profiling in Living Cells by Two-Dimensional Proximity Proteomics. *J. Proteome Res.* **2022**.

119. Deribe, Y. L.; Pawson, T.; Dikic, I., Post-Translational Modifications in Signal Integration. *Nat. Struct. Mol. Biol.* **2010,** *17* (6), 666-72.

120. Dhillon, A. S.; Hagan, S.; Rath, O.; Kolch, W., Map Kinase Signalling Pathways in Cancer. *Oncogene* **2007,** *26* (22), 3279-90.

121. Kim, E. K.; Choi, E. J., Pathological Roles of Mapk Signaling Pathways in Human Diseases. *Biochim Biophys Acta* **2010,** *1802* (4), 396-405.

122. Roberts, A. E.; Allanson, J. E.; Tartaglia, M.; Gelb, B. D., Noonan Syndrome. *Lancet* **2013,** *381* (9863), 333-42.

123. Drosten, M.; Barbacid, M., Targeting the Mapk Pathway in Kras-Driven Tumors. *Cancer Cell* **2020,** *37* (4), 543-550.

124. Lee, S.; Rauch, J.; Kolch, W., Targeting Mapk Signaling in Cancer: Mechanisms of Drug Resistance and Sensitivity. *Int J Mol Sci* **2020,** *21* (3).

125. Tashiro, K.; Tsunematsu, T.; Okubo, H.; Ohta, T.; Sano, E.; Yamauchi, E.; Taniguchi, H.; Konishi, H., Garem, a Novel Adaptor Protein for Growth Factor Receptor-Bound Protein 2, Contributes to Cellular Transformation through the Activation of Extracellular Signal-Regulated Kinase Signaling*. *J. Biol. Chem.* **2009,** *284* (30), 20206-20214.

126. Ravichandran, K. S.; Lorenz, U.; Shoelson, S. E.; Burakoff, S. J., Interaction of Shc with Grb2 Regulates Association of Grb2 with Msos. *Mol Cell Biol* **1995,** *15* (2), 593-600.

127. Liu, Y.; Rohrschneider, L. R., The Gift of Gab. *FEBS Lett.* **2002,** *515* (1-3), 1-7.

128. Gill, K.; Macdonald-Obermann, J. L.; Pike, L. J., Epidermal Growth Factor Receptors Containing a Single Tyrosine in Their C-terminal Tail Bind Different Effector Molecules and Are Signaling-Competent. *J. Biol. Chem.* **2017,** *292* (50), 20744-20755.

129. Nishino, T.; Matsunaga, R.; Konishi, H., Functional Relationship between Cabit, Sam and 14-3-3 Binding Domains of Garem1 That Play a Role in Its Subcellular Localization. *Biochem. Biophys. Res. Commun.* **2015,** *464* (2), 616-621.

130. Brummer, T.; Larance, M.; Abreu, M. T. H.; Lyons, R. J.; Timpson, P.; Emmerich, C. H.; Fleuren, E. D.; Lehrbach, G. M.; Schramek, D.; Guilhaus, M.; James, D. E.; Daly, R. J., Phosphorylation-Dependent Binding of 14-3-3 Terminates Signalling by the Gab2 Docking Protein. *EMBO J.* **2008,** *27* (17), 2305-2316.

131. Kaushansky, A.; Gordus, A.; Chang, B.; Rush, J.; MacBeath, G., A Quantitative Study of the Recruitment Potential of All Intracellular Tyrosine Residues on Egfr, Fgfr1 and Igf1r. *Mol Biosyst* **2008,** *4* (6), 643-53.

132. Jones, R. B.; Gordus, A.; Krall, J. A.; Macbeath, G., A Quantitative Protein Interaction Network for the Erbb Receptors Using Protein Microarrays. *Nature* **2006,** *439* (7073), 168-174.

133. Lemmon, M. A.; Ferguson, K. M.; O'Brien, R.; Sigler, P. B.; Schlessinger, J., Specific and High-Affinity Binding of Inositol Phosphates to an Isolated Pleckstrin Homology Domain. *Proc. Natl. Acad. Sci. USA* **1995,** *92* (23), 10472-6.

134. Pannone, L.; Bocchinfuso, G.; Flex, E.; Rossi, C.; Baldassarre, G.; Lissewski, C.; Pantaleoni, F.; Consoli, F.; Lepri, F.; Magliozzi, M.; Anselmi, M.; Delle Vigne, S.; Sorge, G.; Karaer, K.; Cuturilo, G.; Sartorio, A.; Tinschert, S.; Accadia, M.; Digilio, M. C.; Zampino, G.; De Luca, A.; Cavé, H.; Zenker, M.; Gelb, B. D.; Dallapiccola, B.; Stella, L.; Ferrero, G. B.; Martinelli, S.; Tartaglia, M., Structural, Functional, and Clinical Characterization of a Novel Ptpn11 Mutation Cluster Underlying Noonan Syndrome. *Hum Mutat* **2017,** *38* (4), 451-459.

135. Kaimachnikov, N. P.; Kholodenko, B. N., Toggle Switches, Pulses and Oscillations Are Intrinsic Properties of the Src Activation/Deactivation Cycle. *The FEBS Journal* **2009,** *276* (15), 4102-4118.

136. Lee, Y. K.; Low-Nam, S. T.; Chung, J. K.; Hansen, S. D.; Lam, H. Y. M.; Alvarez, S.; Groves, J. T., Mechanism of Sos Pr-Domain Autoinhibition Revealed by Single-Molecule Assays on Native Protein from Lysate. *Nat. Commun.* **2017,** *8* (1), 15061.

137. Gureasko, J.; Galush, W. J.; Boykevisch, S.; Sondermann, H.; Bar-Sagi, D.; Groves, J. T.; Kuriyan, J., Membrane-Dependent Signal Integration by the Ras Activator Son of Sevenless. *Nat. Struct. Mol. Biol.* **2008,** *15* (5), 452-461.

138. Gureasko, J.; Kuchment, O.; Makino Debora, L.; Sondermann, H.; Bar-Sagi, D.; Kuriyan, J., Role of the Histone Domain in the Autoinhibition and Activation of the Ras Activator Son of Sevenless. *Proc. Natl. Acad. Sci. USA* **2010,** *107* (8), 3430-3435.

139. Huang, W. Y. C.; Alvarez, S.; Kondo, Y.; Kuriyan, J.; Groves, J. T., Relating Cellular Signaling Timescales to Single-Molecule Kinetics: A First-Passage Time Analysis of Ras Activation by Sos. *Proc. Natl. Acad. Sci. USA* **2021,** *118* (45), e2103598118.

140. Yu, H.; Chen, J. K.; Feng, S.; Dalgarno, D. C.; Brauer, A. W.; Schrelber, S. L., Structural Basis for the Binding of Proline-Rich Peptides to SH3 Domains. *Cell* **1994,** *76* (5), 933-945.

141. Liao, T.-J.; Jang, H.; Nussinov, R.; Fushman, D., High-Affinity Interactions of the Nsh3/Csh3 Domains of Grb2 with the C-Terminal Proline-Rich Domain of Sos1. *J. Am. Chem. Soc.* **2020,** *142* (7), 3401-3411.

142. Corbalan-Garcia, S.; Yang, S. S.; Degenhardt, K. R.; Bar-Sagi, D., Identification of the Mitogen-Activated Protein Kinase Phosphorylation Sites on Human Sos1 That Regulate Interaction with Grb2. *Mol Cell Biol* **1996,** *16* (10), 5674-82.

143. Saha, M.; Carriere, A.; Cheerathodi, M.; Zhang, X.; Lavoie, G.; Rush, J.; Roux, P. P.; Ballif, B. A., Rsk Phosphorylates Sos1 Creating 14-3-3-Docking Sites and Negatively Regulating Mapk Activation. *Biochem J* **2012,** *447* (1), 159-66.

144. Reich, A.; Sapir, A.; Shilo, B., Sprouty Is a General Inhibitor of Receptor Tyrosine Kinase Signaling. *Development* **1999,** *126* (18), 4139-4147.

145. Gross, I.; Bassit, B.; Benezra, M.; Licht, J. D., Mammalian Sprouty Proteins Inhibit Cell Growth and Differentiation by Preventing Ras Activation. *J. Biol. Chem.* **2001,** *276* (49), 46460-46468.

146. Edwin, F.; Anderson, K.; Ying, C.; Patel, T. B., Intermolecular Interactions of Sprouty Proteins and Their Implications in Development and Disease. *Mol. Pharmacol.* **2009,** *76* (4), 679-691.

147. Leeksma, O. C.; Van Achterberg, T. A. E.; Tsumura, Y.; Toshima, J.; Eldering, E.; Kroes, W. G. M.; Mellink, C.; Spaargaren, M.; Mizuno, K.; Pannekoek, H.; De Vries, C. J. M., Human Sprouty 4, a New Ras Antagonist on 5q31, Interacts

with the Dual Specificity Kinase Tesk1. *Eur. J. Biochem.* **2002,** *269* (10), 2546-2556.

148. Rubin, C.; Zwang, Y.; Vaisman, N.; Ron, D.; Yarden, Y., Phosphorylation of Carboxyl-Terminal Tyrosines Modulates the Specificity of Sprouty-2 Inhibition of Different Signaling Pathways. *J. Biol. Chem.* **2005,** *280* (10), 9735-9744.

149. Brady, S. C.; Coleman, M. L.; Munro, J.; Feller, S. M.; Morrice, N. A.; Olson, M. F., Sprouty2 Association with B-Raf Is Regulated by Phosphorylation and Kinase Conformation. *Cancer Research* **2009,** *69* (17), 6773-6781.

150. Ozaki, K.-I.; Miyazaki, S.; Tanimura, S.; Kohno, M., Efficient Suppression of Fgf-2-Induced Erk Activation by the Cooperative Interaction among Mammalian Sprouty Isoforms. *J. Cell Sci.* **2005,** *118* (24), 5861-5871.

151. Hanafusa, H.; Torii, S.; Yasunaga, T.; Nishida, E., Sprouty1 and Sprouty2 Provide a Control Mechanism for the Ras/Mapk Signalling Pathway. *Nature Cell Biology* **2002,** *4* (11), 850-858.

152. Lao, D.-H.; Chandramouli, S.; Yusoff, P.; Fong, C. W.; Saw, T. Y.; Tai, L. P.; Yu, C. Y.; Leong, H. F.; Guy, G. R., A Src Homology 3-Binding Sequence on the C Terminus of Sprouty2 Is Necessary for Inhibition of the Ras/Erk Pathway Downstream of Fibroblast Growth Factor Receptor Stimulation. *J. Biol. Chem.* **2006,** *281* (40), 29993-30000.

153. Edwin, F.; Patel, T. B., A Novel Role of Sprouty 2 in Regulating Cellular Apoptosis. *J. Biol. Chem.* **2008,** *283* (6), 3181-3190.

154. Therrien, M.; Michaud, N. R.; Rubin, G. M.; Morrison, D. K., Ksr Modulates Signal Propagation within the Mapk Cascade. *Genes Dev* **1996,** *10* (21), 2684-95.

155. Roy, F.; Laberge, G.; Douziech, M.; Ferland-McCollough, D.; Therrien, M., Ksr Is a Scaffold Required for Activation of the Erk/Mapk Module. *Genes Dev* **2002,** *16* (4), 427-38.

156. Witzel, F.; Maddison, L.; Blüthgen, N., How Scaffolds Shape Mapk Signaling: What We Know and Opportunities for Systems Approaches. *Frontiers in Physiology* **2012,** *3*.

157. Casar, B.; Crespo, P., Erk Signals: Scaffolding Scaffolds? *Frontiers in Cell and Developmental Biology* **2016,** *4*.

158. Dhanasekaran, D. N.; Kashef, K.; Lee, C. M.; Xu, H.; Reddy, E. P., Scaffold Proteins of Map-Kinase Modules. *Oncogene* **2007,** *26* (22), 3185-3202.

159. Consortium, T. U., Uniprot: The Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* **2020,** *49* (D1), D480-D489.

160. Erdős, G.; Dosztányi, Z., Analyzing Protein Disorder with Iupred2a. *Current Protocols in Bioinformatics* **2020,** *70* (1), e99.

161. McKay, M. M.; Ritt, D. A.; Morrison, D. K., Signaling Dynamics of the Ksr1 Scaffold Complex. *Proc. Natl. Acad. Sci. USA* **2009,** *106* (27), 11022-11027.

162. Teilum, K.; Olsen, J.; Kragelund, B. B., Globular and Disordered – the Non-Identical Twins in Protein-Protein Interactions. *Front. Mol. Biosci.* **2015,** *2*, 40.

163. Garbett, D.; Bretscher, A., The Surprising Dynamics of Scaffolding Proteins. *Mol. Biol. Cell* **2014,** *25* (16), 2315-2319.

164. Printen, J. A.; Sprague, G. F., Jr, Protein-Protein Interactions in the Yeast Pheromone Response Pathway: Ste5p Interacts with All Members of the Map Kinase Cascade. *Genetics* **1994,** *138* (3), 609-619.

165. Good, M. C.; Zalatan, J. G.; Lim, W. A., Scaffold Proteins: Hubs for Controlling the Flow of Cellular Information. *Science* **2011,** *332* (6030), 680-686.

166. Wright, P. E.; Dyson, H. J., Intrinsically Unstructured Proteins: Re-Assessing the Protein Structure-Function Paradigm. *J. Mol. Biol.* **1999,** *293* (2), 321-331.

167. Jin, F.; Gräter, F., How Multisite Phosphorylation Impacts the Conformations of Intrinsically Disordered Proteins. *PLoS Comput. Biol.* **2021,** *17* (5), e1008939.

168. Rieloff, E.; Skepö, M., The Effect of Multisite Phosphorylation on the Conformational Properties of Intrinsically Disordered Proteins. *Int J Mol Sci* **2021,** *22* (20).

169. Liu, N.; Guo, Y.; Ning, S.; Duan, M., Phosphorylation Regulates the Binding of Intrinsically Disordered Proteins Via a Flexible Conformation Selection Mechanism. *Commun. Chem.* **2020,** *3* (1), 123.

170. Kulkarni, P.; Jolly, M. K.; Jia, D.; Mooney, S. M.; Bhargava, A.; Kagohara, L. T.; Chen, Y.; Hao, P.; He, Y.; Veltri, R. W.; Grishaev, A.; Weninger, K.; Levine, H.; Orban, J., Phosphorylation-Induced Conformational Dynamics in an Intrinsically Disordered Protein and Potential Role in Phenotypic Heterogeneity. *Proc. Natl. Acad. Sci. USA* **2017,** *114* (13), E2644-E2653.

171. Theillet, F. X.; Binolfi, A.; Bekei, B.; Martorana, A.; Rose, H. M.; Stuiver, M.; Verzini, S.; Lorenz, D.; van Rossum, M.; Goldfarb, D.; Selenko, P., Structural Disorder of Monomeric Alpha-Synuclein Persists in Mammalian Cells. *Nature* **2016,** *530* (7588), 45-50.

172. Xue, B.; Dunker, A. K.; Uversky, V. N., Orderly Order in Protein Intrinsic Disorder Distribution: Disorder in 3500 Proteomes from Viruses and the Three Domains of Life. *J. Biomol. Struct. Dyn.* **2012,** *30* (2), 137-149.

173. Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z., Intrinsically Disordered Protein. *J. Mol. Graphics Modell.* **2001,** *19* (1), 26-59.

174. Bondos, S. E.; Dunker, A. K.; Uversky, V. N., Intrinsically Disordered Proteins Play Diverse Roles in Cell Signaling. *Cell Commun. Signaling* **2022,** *20* (1), 20.

175. van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; Kim, P. M.; Kriwacki, R. W.; Oldfield, C. J.; Pappu, R. V.; Tompa, P.; Uversky, V. N.; Wright, P. E.; Babu, M. M., Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **2014,** *114* (13), 6589-6631.

176. Li, J.; White, J. T.; Saavedra, H.; Wrabl, J. O.; Motlagh, H. N.; Liu, K.; Sowers, J.; Schroer, T. A.; Thompson, E. B.; Hilser, V. J., Genetically Tunable Frustration Controls Allostery in an Intrinsically Disordered Transcription Factor. *eLife* **2017,** *6*, e30688.

177. Raabe, T.; Olivier, J. P.; Dickson, B.; Liu, X.; Gish, G. D.; Pawson, T.; Hafen, E., Biochemical and Genetic Analysis of the Drk SH2/SH3 Adaptor Protein of Drosophila. *EMBO J.* **1995,** *14* (11), 2509-18.

178. Kumar, J. P., The Fly Eye: Through the Looking Glass. *Dev. Dyn.* **2018,** *247* (1), 111-123.

179. Yadav Kamlesh, K.; Bar-Sagi, D., Allosteric Gating of Son of Sevenless Activity by the Histone Domain. *Proc. Natl. Acad. Sci. USA* **2010,** *107* (8), 3436-3440.

180. Hwang, T.; Parker, S. S.; Hill, S. M.; Grant, R. A.; Ilunga, M. W.; Sivaraman, V.; Mouneimne, G.; Keating, A. E., Native Proline-Rich Motifs Exploit Sequence Context to Target Actin-Remodeling Ena/Vasp Protein Enah. *eLife* **2022,** *11*, e70680.

181. Kragelj, J.; Orand, T.; Delaforge, E.; Tengo, L.; Blackledge, M.; Palencia, A.; Jensen, M. R., Enthalpy–Entropy Compensation in the Promiscuous Interaction of an Intrinsically Disordered Protein with Homologous Protein Partners. *Biomolecules* **2021,** *11* (8).

182. Lumry, R.; Rajender, S., Enthalpy-Entropy Compensation Phenomena in Water Solutions of Proteins and Small Molecules: A Ubiquitous Property of Water. *Biopolymers* **1970,** *9* (10), 1125-227.

183. Biancucci, M.; Dolores, J. S.; Wong, J.; Grimshaw, S.; Anderson, W. F.; Satchell, K. J. F.; Kwon, K., New Ligation Independent Cloning Vectors for Expression of Recombinant Proteins with a Self-Cleaving Cpd/6xhis-Tag. *BMC Biotechnol.* **2017,** *17* (1), 1.

184. Gardner, K. H.; Kay, L. E., Production and Incorporation of 15N, 13C, 2H (1H-Δ1 Methyl) Isoleucine into Proteins for Multidimensional NMR Studies. *J. Am. Chem. Soc.* **1997,** *119* (32), 7599-7600.

185. Goto, N. K.; Gardner, K. H.; Mueller, G. A.; Willis, R. C.; Kay, L. E., A Robust and Cost-Effective Method for the Production of Val, Leu, Ile (Delta 1) Methyl-Protonated 15N-, 13C-, 2H-Labeled Proteins. *J. Biomol. NMR* **1999,** *13* (4), 369-374.

186. Contreras-Martos, S.; Nguyen, H. H.; Nguyen, P. N.; Hristozova, N.; Macossay-Castillo, M.; Kovacs, D.; Bekesi, A.; Oemig, J. S.; Maes, D.; Pauwels, K.; Tompa, P.; Lebrun, P., Quantification of Intrinsically Disordered Proteins: A Problem Not Fully Appreciated. *Front. Mol. Biosci.* **2018,** *5*, 83-83.

187. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S. E.; Wilkins, M. R.; Appel, R. D.; Bairoch, A., Protein Identification and Analysis Tools on the Expasy Server. In *The Proteomics Protocols Handbook*, Humana Press: 2005; pp 571-607.

188. Zhang, O.; Forman-Kay, J. D., NMR Studies of Unfolded States of an SH3 Domain in Aqueous Solution and Denaturing Conditions. *Biochemistry* **1997,** *36* (13), 3959-70.

189. Bezsonova, I.; Singer, A.; Choy, W. Y.; Tollinger, M.; Forman-Kay, J. D., Structural Comparison of the Unstable Drkn SH3 Domain and a Stable Mutant. *Biochemistry* **2005,** *44* (47), 15550-60.

190. Piszkiewicz, S.; Gunn, K. H.; Warmuth, O.; Propst, A.; Mehta, A.; Nguyen, K. H.; Kuhlman, E.; Guseman, A. J.; Stadmiller, S. S.; Boothby, T. C.; Neher, S. B.; Pielak, G. J., Protecting Activity of Desiccated Enzymes. *Protein Sci.* **2019,** *28* (5), 941-951.

191. Ollerenshaw, J. E.; Tugarinov, V.; Kay, L. E., Methyl Trosy: Explanation and Experimental Verification. *Magn. Reson. Chem.* **2003,** *41* (10), 843-852.

192. Waudby, C. A.; Ramos, A.; Cabrita, L. D.; Christodoulou, J., Two-Dimensional NMR Lineshape Analysis. *Sci. Rep.* **2016,** *6*, 24826.

193. Waudby, C. A.; Christodoulou, J., NMR Lineshape Analysis of Intrinsically Disordered Protein Interactions. In *Intrinsically Disordered Proteins*, Springer US: 2020; Vol. 2141, pp 477-504.

194. Pierce, M. M.; Raman, C. S.; Nall, B. T., Isothermal Titration Calorimetry of Protein-Protein Interactions. *Methods* **1999,** *19* (2), 213-21.

195. Stadmiller, S. S.; Pielak, G. J., Enthalpic Stabilization of an SH3 Domain by D_2O. *Protein Sci.* **2018,** *27* (9), 1710-1716.

196. Veer; Zeng, D.; Krieger, I.; James; Cho, J.-H., Binding Mechanism of the N-Terminal SH3 Domain of Crkii and Proline-Rich Motifs in Cabl. *Biophys. J.* **2016,** *110* (12), 2630-2641.

197. Ferreon, J. C.; Hilser, V. J., Thermodynamics of Binding to SH3 Domains: The Energetic Impact of Polyproline Ii Helix Formation. *Biochemistry* **2004,** *43* (24), 7787-7797.

198. Candel, A. M.; van Nuland, N. A. J.; Martin-Sierra, F. M.; Martinez, J. C.; Conejero-Lara, F., Analysis of the Thermodynamics of Binding of an SH3 Domain to Proline-Rich Peptides Using a Chimeric Fusion Protein. *J. Mol. Biol.* **2008,** *377* (1), 117-135.

199. Demers, J.-P.; Mittermaier, A., Binding Mechanism of an SH3 Domain Studied by NMR and Itc. *J. Am. Chem. Soc.* **2009,** *131* (12), 4355-4367.

200. Langridge, T. D.; Tarver, M. J.; Whitten, S. T., Temperature Effects on the Hydrodynamic Radius of the Intrinsically Disordered N-terminal Region of the P53 Protein. *Proteins* **2014,** *82* (4), 668-678.

201. Gao, A.; Shrinivas, K.; Lepeudry, P.; Suzuki, H. I.; Sharp, P. A.; Chakraborty, A. K., Evolution of Weak Cooperative Interactions for Biological Specificity. *Proc. Natl. Acad. Sci. USA* **2018,** *115* (47), E11053-E11060.

202. Strickland, M.; Kale, S.; Strub, M.-P.; Schwieters, C. D.; Liu, J.; Peterkofsky, A.; Tjandra, N., Potential Regulatory Role of Competitive Encounter Complexes in Paralogous Phosphotransferase Systems. *J. Mol. Biol.* **2019,** *431* (12), 2331-2342.

203. Kale, S.; Strickland, M.; Peterkofsky, A.; Liu, J.; Tjandra, N., Model of a Kinetically Driven Crosstalk between Paralogous Protein Encounter Complexes. *Biophys. J.* **2019,** *117* (9), 1655-1665.

204. Monteith, W. B.; Pielak, G. J., Residue Level Quantification of Protein Stability in Living Cells. *Proc. Natl. Acad. Sci. USA* **2014,** *111* (31), 11335-40.

205. Feng, R.; Gruebele, M.; Davis, C. M., Quantifying Protein Dynamics and Stability in a Living Organism. *Nat. Commun.* **2019,** *10* (1), 1179.

206. Bao, H.-L.; Xu, Y., Investigation of Higher-Order RNA G-Quadruplex Structures in Vitro and in Living Cells by ^{19}F NMR Spectroscopy. *Nat. Protocols* **2018,** *13* (4), 652-665.

207. Broft, P.; Dzatko, S.; Krafcikova, M.; Hansel-Hertsch, R.; Wacker, A.; Doetsch, V.; Trantirek, L.; Schwalbe, H., In-Cell NMR Spectroscopy of Functional Riboswitch Aptamers in Eukaryotic Cells. *Angew. Chem., Int. Ed.* **2021,** *published on line* (*published on line*).

208. Graziano, G., Shape Effect on Non-Covalent Dimer Stability Using Classic Scaled Particle Theory. *Chem. Phys. Lett.* **2020,** *743*, 137176.

209. Guseman, A. J.; Pielak, G. J., Chapter 12 Protein Stability and Weak Intracellular Interactions. In *In-Cell NMR Spectroscopy: From Molecular Sciences to Cell Biology*, The Royal Society of Chemistry: 2020; pp 188-206.

210. Konopka, M. C.; Shkel, I. A.; Cayley, S.; Record, M. T.; Weisshaar, J. C., Crowding and Confinement Effects on Protein Diffusion in Vivo. *J. Bacteriol.* **2006,** *188* (17), 6115-23.

211. Shi, X.; Foo, Y. H.; Sudhaharan, T.; Chong, S.-W.; Korzh, V.; Ahmed, S.; Wohland, T., Determination of Dissociation Constants in Living Zebrafish Embryos with Single Wavelength Fluorescence Cross-Correlation Spectroscopy. *Biophys. J.* **2009,** *97* (2), 678-686.

212. Komatsubara, A. T.; Goto, Y.; Kondo, Y.; Matsuda, M.; Aoki, K., Single-Cell Quantification of the Concentrations and Dissociation Constants of Endogenous Proteins. *J. Biol. Chem.* **2019,** *294* (15), 6062-6072.

213. Lerner, E.; Cordes, T.; Ingargiola, A.; Alhadid, Y.; Chung, S.; Michalet, X.; Weiss, S., Toward Dynamic Structural Biology: Two Decades of Single-Molecule Förster Resonance Energy Transfer. *Science* **2018,** *359* (6373), eaan1133.

214. Sukenik, S.; Salam, M.; Wang, Y.; Gruebele, M., In-Cell Titration of Small Solutes Controls Protein Stability and Aggregation. *J. Am. Chem. Soc.* **2018,** *140* (33), 10497-10503.

215. Serber, Z.; Keatinge-Clay, A. T.; Ledwidge, R.; Kelly, A. E.; Miller, S. M.; Dötsch, V., High-Resolution Macromolecular NMR Spectroscopy inside Living Cells. *J. Am. Chem. Soc.* **2001,** *123* (10), 2446-7.

216. 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. USA* **2006,** *103* (32), 11904-9.

217. Inomata, K.; Ohno, A.; Tochio, H.; Isogai, S.; Tenno, T.; Nakase, I.; Takeuchi, T.; Futaki, S.; Ito, Y.; Hiroaki, H.; Shirakawa, M., High-Resolution Multi-Dimensional NMR Spectroscopy of Proteins in Human Cells. *Nature* **2009,** *458* (7234), 106-9.

218. Xu, G.; Ye, Y.; Liu, X.; Cao, S.; Wu, Q.; Cheng, K.; Liu, M.; Pielak, G. J.; Li, C., Strategies for Protein NMR in *Escherichia coli*. *Biochemistry* **2014,** *53* (12), 1971-1981.

219. Stadmiller, S. S.; Pielak, G. J., The Expanding Zoo of in-Cell Protein NMR. *Biophys. J.* **2018,** *115* (9), 1628-1629.

220. Ye, Y.; Wu, Q.; Zheng, W.; Jiang, B.; Pielak, G. J.; Liu, M.; Li, C., Quantification of Size Effect on Protein Rotational Mobility in Cells by ^{19}F NMR Spectroscopy. *Anal. Bioanal. Chem.* **2018,** *410* (3), 869-874.

221. Tanaka, T.; Ikeya, T.; Kamoshida, H.; Suemoto, Y.; Mishima, M.; Shirakawa, M.; Güntert, P.; Ito, Y., High-Resolution Protein 3d Structure Determination in Living Eukaryotic Cells. *Angew. Chem., Int. Ed.* **2019,** *58* (22), 7284-7288.

222. Li, C.; Charlton, L. M.; Lakkavaram, A.; Seagle, C.; Wang, G.; Young, G. B.; Macdonald, J. M.; Pielak, G. J., Differential Dynamical Effects of Macromolecular Crowding on an Intrinsically Disordered Protein and a Globular Protein: Implications for in-Cell NMR Spectroscopy. *J. Am. Chem. Soc.* **2008,** *130* (20), 6310-6311.

223. Kyne, C.; Jordon, K.; Filoti, D. I.; Laue, T. M.; Crowley, P. B., Protein Charge Determination and Implications for Interactions in Cell Extracts. *Protein Sci.* **2017,** *26* (2), 258-267.

224. Mu, X.; Choi, S.; Lang, L.; Mowray, D.; Dokholyan, N. V.; Danielsson, J.; Oliveberg, M., Physicochemical Code for Quinary Protein Interactions in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **2017,** *114* (23), E4556-E4563.

225. Binolfi, A.; Limatola, A.; Verzini, S.; Kosten, J.; Theillet, F.-X.; May Rose, H.; Bekei, B.; Stuiver, M.; van Rossum, M.; Selenko, P., Intracellular Repair of Oxidation-Damaged A-Synuclein Fails to Target C-terminal Modification Sites. *Nat. Commun.* **2016,** *7* (1), 10251.

226. Bertrand, K.; Reverdatto, S.; Burz, D. S.; Zitomer, R.; Shekhtman, A., Structure of Proteins in Eukaryotic Compartments. *J. Am. Chem. Soc.* **2012,** *134* (30), 12798-12806.

227. Banci, L.; Barbieri, L.; Bertini, I.; Luchinat, E.; Secci, E.; Zhao, Y.; Aricescu, A. R., Atomic-Resolution Monitoring of Protein Maturation in Live Human Cells by NMR. *Nat. Chem. Biol.* **2013,** *9* (5), 297-299.

228. Ogino, S.; Kubo, S.; Umemoto, R.; Huang, S.; Nishida, N.; Shimada, I., Observation of NMR Signals from Proteins Introduced into Living Mammalian Cells by Reversible Membrane Permeabilization Using a Pore-Forming Toxin, Streptolysin O. *J. Am. Chem. Soc.* **2009,** *131* (31), 10834-10835.

229. Speer, S. L.; Guseman, A. J.; Patteson, J. B.; Ehrmann, B. M.; Pielak, G. J., Controlling and Quantifying Protein Concentration in *Escherichia coli*. *Protein Sci.* **2019,** *28* (7), 1307-1311.

230. Chu, I. T.; Speer, S. L.; Pielak, G. J., Rheostatic Control of Protein Expression Using Tuner Cells. *Biochemistry* **2020,** *59* (6), 733-735.

231. Nebreda, A. R.; Ferby, I., Regulation of the Meiotic Cell Cycle in Oocytes. *Curr. Opin. Cell Biol.* **2000,** *12* (6), 666-675.

232. Tunquist, B. J.; Maller, J. L., Under Arrest: Cytostatic Factor (CSF)-Mediated Metaphase Arrest in Vertebrate Eggs. *Genes. Dev.* **2003,** *17* (6), 683-710.

233. Sánchez-López, C.; Labadie, N.; Lombardo, V. A.; Biglione, F. A.; Manta, B.; Jacob, R. S.; Gladyshev, V. N.; Abdelilah-Seyfried, S.; Selenko, P.; Binolfi, A., An NMR-Based Biosensor to Measure Stereospecific Methionine Sulfoxide Reductase Activities in Vitro and in Vivo. *Chem. - Eur. J.* **2020,** *26* (65), 14838-14843.

234. Gorensek-Benitez, A. H.; Smith, A. E.; Stadmiller, S. S.; Perez Goncalves, G. M.; Pielak, G. J., Cosolutes, Crowding, and Protein Folding Kinetics. *J. Phys. Chem. B* **2017,** *121* (27), 6527-6537.

235. Lau, Y. K.; Baytshtok, V.; Howard, T. A.; Fiala, B. M.; Johnson, J. M.; Carter, L. P.; Baker, D.; Lima, C. D.; Bahl, C. D., Discovery and Engineering of Enhanced Sumo Protease Enzymes. *J. Biol. Chem.* **2018,** *293* (34), 13224-13233.

236. Westerfield, M., *The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio)*. 5th ed.; University of Oregon Press: Eugene, 2007.

237. Fuentes, R.; Fernández, J., Ooplasmic Segregation in the Zebrafish Zygote and Early Embryo: Pattern of Ooplasmic Movements and Transport Pathways. *Dev. Dyn.* **2010,** *239* (8), 2172-89.

238. Cohen, R. D.; Guseman, A. J.; Pielak, G. J., Intracellular pH Modulates Quinary Structure. *Protein Sci.* **2015,** *24* (11), 1748-1755.

239. James, G., Witten, D., Haistie, T., Tibshirani, R., *An Introduction to Statistical Learning*. 1 ed.; Springer-Verlag New York: 2013; p 426.

240. Lipari, G.; Szabo, A., Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules 2. Analysis of Experimental Results. *J. Am. Chem. Soc.* **1982,** *104* (17), 4559-4570.

241. Lipari, G.; Szabo, A., Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J. Am. Chem. Soc.* **1982,** *104* (17), 4546-4559.

242. Farrow, N. A.; Zhang, O.; Forman-Kay, J. D.; Kay, L. E., Comparison of the Backbone Dynamics of a Folded and an Unfolded SH3 Domain Existing in Equilibrium in Aqueous Buffer. *Biochemistry* **1995,** *34* (3), 868-878.

243. Serber, Z.; Ledwidge, R.; Miller, S. M.; Dötsch, V., Evaluation of Parameters Critical to Observing Proteins inside Living *Escherichia coli* by in-Cell NMR Spectroscopy. *J. Am. Chem. Soc.* **2001,** *123* (37), 8895-901.

244. Barnes, C. O.; Monteith, W. B.; Pielak, G. J., Internal and Global Protein Motion Assessed with a Fusion Construct and in-Cell NMR Spectroscopy. *ChemBioChem* **2011,** *12* (3), 390-391.

245. Theillet, F.-X.; Binolfi, A.; Frembgen-Kesner, T.; Hingorani, K.; Sarkar, M.; Kyne, C.; Li, C.; Crowley, P. B.; Gierasch, L.; Pielak, G. J.; Elcock, A. H.; Gershenson, A.; Selenko, P., Physicochemical Properties of Cells and Their Effects on Intrinsically Disordered Proteins (IDPs). *Chem. Rev.* **2014,** *114* (13), 6661-6714.

246. Crowley, P. B.; Chow, E.; Papkovskaia, T., Protein Interactions in the *Escherichia coli* Cytosol: An Impediment to in-Cell NMR Spectroscopy. *ChemBioChem* **2011,** *12* (7), 1043-1048.

247. Tuinstra, R. L.; Peterson, F. C.; Kutlesa, S.; Elgin, E. S.; Kron, M. A.; Volkman, B. F., Interconversion between Two Unrelated Protein Folds in the Lymphotactin Native State. *Proc. Natl. Acad. Sci. USA* **2008,** *105* (13), 5057-5063.

248. Crowley, P. B.; Kyne, C.; Monteith, W. B., Simple and Inexpensive Incorporation of ^{19}F-Tryptophan for Protein NMR Spectroscopy. *Chem. Commun.* **2012,** *48* (86), 10681-3.

249. Welte, H.; Zhou, T.; Mihajlenko, X.; Mayans, O.; Kovermann, M., What Does Fluorine Do to a Protein? Thermodynamic, and Highly-Resolved Structural Insights into Fluorine-Labelled Variants of the Cold Shock Protein. *Sci. Rep.* **2020,** *10* (1), 2640.

250. Evanics, F.; Bezsonova, I.; Marsh, J.; Kitevski, J. L.; Forman-Kay, J. D.; Prosser, R. S., Tryptophan Solvent Exposure in Folded and Unfolded States of an SH3 Domain by ^{19}F and 1H NMR. *Biochemistry* **2006,** *45* (47), 14120-14128.

251. Becktel, W. J.; Schellman, J. A., Protein Stability Curves. *Biopolymers* **1987,** *26* (11), 1859-77.

252. Prabhu, N. V.; Sharp, K. A., Heat Capacity in Proteins. *Annu. Rev. Phys. Chem.* **2005,** *56*, 521-48.

253. Wang, Y.; Sarkar, M.; Smith, A. E.; Krois, A. S.; Pielak, G. J., Macromolecular Crowding and Protein Stability. *J. Am. Chem. Soc.* **2012,** *134* (40), 16614-16618.

254. Sarkar, M.; Smith, A. E.; Pielak, G. J., Impact of Reconstituted Cytosol on Protein Stability. *Proc. Natl. Acad. Sci. USA* **2013,** *110* (48), 19342.

255. Ge, C.; Lu, W.; Chen, A., Quantitative Proteomic Reveals the Dynamic of Protein Profile During Final Oocyte Maturation in Zebrafish. *Biochem. Biophys. Res. Commun.* **2017,** *490* (3), 657-663.

256. Kozlowski, L. P., Proteome-Pi: Proteome Isoelectric Point Database. *Nucleic Acids Res.* **2017,** *45* (D1), D1112-D1116.

257. Wang, Y.; Li, C.; Pielak, G. J., Effects of Proteins on Protein Diffusion. *J. Am. Chem. Soc.* **2010,** *132* (27), 9392-9397.

258. Ye, Y.; Liu, X.; Zhang, Z.; Wu, Q.; Jiang, B.; Jiang, L.; Zhang, X.; Liu, M.; Pielak, G. J.; Li, C., ^{19}F NMR Spectroscopy as a Probe of Cytoplasmic Viscosity and Weak Protein Interactions in Living Cells. *Chem. - Eur. J.* **2013,** *19* (38), 12705-12710.

259. Nawrocki, G.; Wang, P.-h.; Yu, I.; Sugita, Y.; Feig, M., Slow-Down in Diffusion in Crowded Protein Solutions Correlates with Transient Cluster Formation. *J. Phys. Chem. B* **2017,** *121* (49), 11072-11084.

260. Nawrocki, G.; Karaboga, A.; Sugita, Y.; Feig, M., Effect of Protein–Protein Interactions and Solvent Viscosity on the Rotational Diffusion of Proteins in Crowded Environments. *Phys. Chem. Chem. Phys.* **2019,** *21* (2), 876-883.

261. Lagadic-Gossmann, D.; Huc, L.; Lecureur, V., Alterations of Intracellular pH Homeostasis in Apoptosis: Origins and Roles. *Cell Death Differ.* **2004,** *11* (9), 953-961.

262. Casey, J. R.; Grinstein, S.; Orlowski, J., Sensors and Regulators of Intracellular pH. *Nat. Rev. Mol. Cell Biol.* **2010,** *11* (1), 50-61.

263. Bhattacharya, S.; Lecomte, J. T., Temperature Dependence of Histidine Ionization Constants in Myoglobin. *Biophys. J.* **1997,** *73* (6), 3241-3256.

264. Molich, A.; Heisler, N., Determination of pH by Microfluorometry: Intracellular and Interstitial pH Regulation in Developing Early-Stage Fish Embryos (*Danio rerio*). *J. Exp. Biol.* **2005,** *208* (Pt 21), 4137-49.

265. Sakai, T.; Tochio, H.; Tenno, T.; Ito, Y.; Kokubo, T.; Hiroaki, H.; Shirakawa, M., In-Cell NMR Spectroscopy of Proteins inside *Xenopus Laevis* Oocytes. *J. Biomol. NMR* **2006,** *36* (3), 179-88.

266. Veldman, M. B.; Lin, S., Zebrafish as a Developmental Model Organism for Pediatric Research. *Pediatr. Res.* **2008,** *64* (5), 470-476.

267. Blum, M.; Ott, T., Xenopus: An Undervalued Model Organism to Study and Model Human Genetic Disease. *Cells Tissues Organs* **2018,** *205* (5-6), 303-313.

268. Schultz, T. W.; Dawson, D. A., Housing and Husbandry of *Xenopus* for Oocyte Production. *Lab. Anim.* **2003,** *32* (2), 34-39.

269. Kurtzman, M. S.; Craig, M. P.; Grizzle, B. K.; Hove, J. R., Sexually Segregated Housing Results in Improved Early Larval Survival in Zebrafish. *Lab. Anim.* **2010,** *39* (6), 183-189.

270. Crilly, C. J.; Brom, J. A.; Warmuth, O.; Esterly, H. J.; Pielak, G. J., Protection by Desiccation-Tolerance Proteins Probed at the Residue Level. *Protein Sci* **2022,** *31* (2), 396-406.

271. Crilly, C. J.; Brom, J. A.; Kowalewski, M. E.; Piszkiewicz, S.; Pielak, G. J., Dried Protein Structure Revealed at the Residue Level by Liquid-Observed Vapor Exchange NMR. *Biochemistry* **2021,** *60* (2), 152-159.

272. Guseman, A. J.; Pielak, G. J., Cosolute and Crowding Effects on a Side-by-Side Protein Dimer. *Biochemistry* **2017,** *56* (7), 971-976.

273. Monteith, W. B.; Cohen, R. D.; Smith, A. E.; Guzman-Cisneros, E.; Pielak, G. J., Quinary Structure Modulates Protein Stability in Cells. *Proc. Natl. Acad. Sci. USA* **2015,** *112* (6), 1739-42.

274. Jee, J.; Byeon, I. J.; Louis, J. M.; Gronenborn, A. M., The Point Mutation A34f Causes Dimerization of GB1. *Proteins* **2008,** *71* (3), 1420-31.

275. Byeon, I. J.; Louis, J. M.; Gronenborn, A. M., A Protein Contortionist: Core Mutations of GB1 That Induce Dimerization and Domain Swapping. *J Mol Biol* **2003,** *333* (1), 141-52.

276. Kirsten Frank, M.; Dyda, F.; Dobrodumov, A.; Gronenborn, A. M., Core Mutations Switch Monomeric Protein GB1 into an Intertwined Tetramer. *Nat Struct Biol* **2002,** *9* (11), 877-85.

277. Schuck, P., Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophys J* **2000,** *78* (3), 1606-19.

278. Takaine, M.; Imamura, H.; Yoshida, S., High and Stable Atp Levels Prevent Aberrant Intracellular Protein Aggregation in Yeast. *eLife* **2022,** *11*, e67659.

279. Morciano, G.; Imamura, H.; Patergnani, S.; Pedriali, G.; Giorgi, C.; Pinton, P., Chapter 8 - Measurement of Atp Concentrations in Mitochondria of Living Cells Using Luminescence and Fluorescence Approaches. In *Methods in Cell Biology*, Pon, L. A.; Schon, E. A., Eds. Academic Press: 2020; Vol. 155, pp 199-219.

280. Mok, Y.-K.; Elisseeva, E. L.; Davidson, A. R.; Forman-Kay, J. D., Dramatic Stabilization of an SH3 Domain by a Single Substitution: Roles of the Folded and Unfolded States11edited by C. R. Matthews. *J. Mol. Biol.* **2001,** *307* (3), 913-928.

281. Harkiolaki, M.; Gilbert, R. J. C.; E; Feller, S. M., The C-Terminal SH3 Domain of Crkl as a Dynamic Dimerization Module Transiently Exposing a Nuclear Export Signal. *Structure* **2006,** *14* (12), 1741-1753.