DEVELOPMENT AND APPLICATION OF LIQUID-OBSERVED VAPOR EXCHANGE NMR TO STUDY DEHYDRATED PROTEIN STRUCTURE AND PROTECTION AT THE RESIDUE-LEVEL

Candice J. Crilly

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
2021

Approved by:

Gary J. Pielak

Dorothy Erie

Amy Gladfelter

Sergei Sheiko

Marcey Waters

ABSTRACT

Candice J. Crilly: Development and application of liquid-observed vapor exchange NMR to study dehydrated protein structure and protection at the residue-level
(Under the direction of Gary J. Pielak)

Life on Earth evolved in the oceans, and therefore the stability, dynamics, and function of proteins – the molecules that carry out the majority of life's processes – are intricately linked to the properties of liquid water. However, despite decades of research, our understanding of the fundamental interaction between proteins and water remains surprisingly limited, in part because we have been unable to study, at adequate resolution, what happens to proteins when water is removed. Such technological boundaries have also hindered our understanding of how molecules known as excipients protect proteins from the destructive effects of dehydration, which is important in the context of protein-based drugs such as insulin and vaccines. For my dissertation work, I sought to help fill these knowledge gaps by developing Liquid-Observed Vapor Exchange (LOVE) NMR, a solution NMR technique that provides residue-level information on the structure and interactions of dehydrated proteins. My dissertation begins with a brief overview and contextualization of what we do and do not know about dehydrated protein structure and mechanisms of dehydration protection (Chapter 1), and then goes into the initial proof-of-concept experiments for LOVE NMR, in which I showed that LOVE NMR reports on the fraction of dried protein population for which a given residue is protected from exchange with D_2O vapor, and that this fraction is related to the

amount of local structure and/or inter-molecular interactions in the dry state (Chapter 2). I then applied LOVE NMR to uncover water's variable role in the mutation-induced (de)stabilization of two different proteins (Chapter 3) and to probe the importance of surface-area, charge-patterning, and electrostatic interactions in protein dehydration protection by two distinct disordered proteins from desiccation-tolerant animals (Chapter 4). Finally, Chapter 5 serves as a guide for future developers of LOVE NMR by providing suggestions for how to improve the method and deepen our understanding of the results it produces. Overall, the results presented in this dissertation demonstrate the potential of LOVE NMR to provide new insights into protein-water and protein-excipient interactions.

To my dad, Jonathon R. Crilly.
For your endless love, support, and belief in me.
Your pride kept me going.

ACKNOWLEDGEMENTS

My decision to pursue (and my ability to complete) a PhD in biological chemistry was the culmination of years of guidance and support from dozens of wonderful people, some of whom I would like to acknowledge here. First and foremost, I would like to thank my parents – Kathryn J. Moore and Jonathon R. Crilly – for instilling in me the value of education and investing so much in my own. Mom, your imagination and love of science rubbed off on me, and I'm grateful for all your support during my early education. Dad – your (seemingly sincere) belief that I could do anything I put my mind to gave me the confidence and resilience I needed to overcome the many failures one faces when conducting research – this made a huge impact on my ability to complete a five-year PhD program. I also want to thank my sister, Jenny Hutcherson, for encouraging me to "complete all my education" in one fell swoop and go straight into a PhD program after my undergraduate studies. And my brother, Lucas Kellner – thank you for teaching me how a car engine works, nerding out with me about calculus over Thanksgiving, and for always being there for our family. Finally, I'd like to thank the Waterbury family, and especially Sandy Waterbury, for their love and support over the years.

I would also like to thank the various academic mentors I've had over the course of my academic journey. First and foremost, I would like to thank Mr. Alan Tanaka for being a great math teacher in middle school and, through that, making

me consider a career in STEM. I would also like to thank Professor Mike Hill for allowing me to do research in his lab at Occidental College – that first research experience in your lab helped me realize that good science is accomplished by teams of people with diverse backgrounds and attitudes. I'd also like to thank Occidental Professors Shana Goffredi for instilling in me an appreciation of basic science research, Gary Schindelman for checking in on me during a tough time, and Chris Craney for your support as my academic advisor – your glowing recommendation letter to the Amgen Scholars program may have been what got me into the program, and for that I am forever grateful. Thank you, Dr. Scott McIsaac and Professor Frances Arnold, for taking me into your lab as an Amgen Scholar during the summer of 2014 – that summer I learned how truly exciting (and occasionally posh) the world of scientific research could be and that you could "engineer to discover", which solidified my decision to pursue graduate studies.

I would not have been able to complete my PhD studies without my UNC academic mentors, Dr. Stuart Parnham and Professor Gary J. Pielak. Stu, thank you for always being willing to teach me how to run complicated NMR experiments, listen to (and improve) my project pitches, and advise me on how to handle tough situations – your help and guidance made a big difference in my PhD journey. And Gary – although we've had our disagreements, I am grateful for everything you have taught me over the past five years. Thank you for taking me into your research group and teaching me how to write and review papers, how to anticipate everything "taking longer than you think", and how to finish a research project. I still have a lot

to learn, but I know that working with you has made me a better scientist, colleague, and person.

I'd also like to thank all the wonderful friends I've made while in Chapel Hill. Thank you, Dr. Eric Friedlander, for recruiting me (and Adam) to attend UNC instead of the University of Minnesota, and for being our friend ever since. Thank you, Pielak lab – especially Dr. Sam Stadmiller and Joey Thole – for being exceptionally supportive and collaborative scientists and friends. Thank you, Rachel Johnson, for always making our bar crawls down Franklin Street interesting (hey, it's me the moon!) and for sharing the struggle of trying to get organized with me. Thank you, Marg Fam, for all the hilarious game nights and for the occasional nerd-out. Lastly, thank you Dr. Marie Düker, Dr. Jack Prothero, and Wayana Dolan for being my non-chemistry-grad-student friends – our hang outs were always fun and refreshing. Ya'll better visit me in California, ya hear?

Last but certainly not least, I would like to thank my lovely flatmates Teddy and Helen, and my fiancé, Adam Waterbury, for their endless support over the past five years. Thank you, Adam, for tolerating me when I was stressed about an experiment going wrong, for making amazing dinners without me ever having to ask, and for helping me relax after a long day. Grad school would have been far less fun and far more difficult without your love and support.

PREFACE

At the time this dissertation was submitted, the contents of Chapter 2 had been published in Volume 60 of the peer-reviewed journal *Biochemistry* in 2021 with text and title nearly verbatim as it appears here. The contents of Chapter 3-4 were submitted to peer-reviewed journals with text and title verbatim as it is presented here. The contents of Chapter 1 are intended for publication as a mini-review in the near future.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

^{1}H	hydrogen
1D	one-dimensional
2D	two-dimensional
^{13}C	carbon-13
^{15}N	nitrogen-15
$\Delta G°'_{op}$	modified standard state opening free energy
$\Delta\Delta G°'_{op}$	change in modified standard state opening free energy
ħ	Planck's constant
κ	charge-patterning parameter
BLAST	basic local alignment search tool
BSA	bovine serum albumin
CAHS	cytosolic-abundant heat soluble
CI2	chymotrypsin inhibitor 2
CSP	chemical shift perturbation
D_2O	deuterium oxide
DSC	differential scanning calorimetry
E. coli	*Escheria coli*
FTIR	Fourier-transform Infrared spectroscopy
g	gram
GB1	B1 domain of streptococcal protein G
h	hours
H_2O	water
HDX	hydrogen-deuterium exchange
HSQC	heteronuclear single quantum coherence

IDP	intrinsically disordered protein
IPTG	isopropyl β-D-1-thiogalactopyranoside
K	Kelvin
kDa	kilodalton
k_{int}	intrinsic exchange rate
k_{obs}	observed exchange rate
L	liter
LEA	late embryogenesis abundant
LOVE	liquid-observed vapor exchange
M	molar
min	minutes
NMR	nuclear magnetic resonance
OD_{600}	optical density at 600 nm
PDB	Protein Data Bank
QC	quench-correction
R	gas constant
R_2	transverse relaxation rate
RH	relative humidity
ssHDX-MS	solid-state hydrogen-deuterium exchange mass spectrometry
t	time
T	absolute temperature
T_1	longitudinal relaxation time
T_g	glass transition temperature
TGA	thermogravimetric analysis
V	peak volume
w/w	weight for weight

CHAPTER 1: BIOPHYSICAL CONSEQUENCES OF PROTEIN DEHYDRATION AND PROTECTION AGAINST ITS ADVERSE EFFECTS

INTRODUCTION

Kauzman's 1955 review of the hydrophobic effect[1] alerted the scientific community to the integral role played by liquid water on protein structure, stability, and dynamics.[2-4] Yet the life-giving reactivity and flexibility that liquid water lends proteins also contribute to various degradation pathways, including hydrolysis, deamidation, and agitation-induced aggregation.[5] The plethora of water-mediated protein degradation pathways could explain why organisms from all domains of life prepare for long-term storage via anhydrobiosis, a latent, dehydrated state where cellular water is less than 5% w/w (compared to the normal value of 60-70%) and metabolism slows to nearly undetectable levels. [6-7] Organisms capable of anhydrobiosis (Fig. 1.1) can persist in this dehydrated state for decades and, upon contact with liquid water, reanimate within minutes. Interestingly, organisms in this state are also resilient to a broad range of extreme environmental stresses. For instance, microscopic animals known as tardigrades can survive up to 8 h at -272 °C, 15 min at 151 °C, and 570,000 r of acute X-radiation while in the anhydrobiotic state.

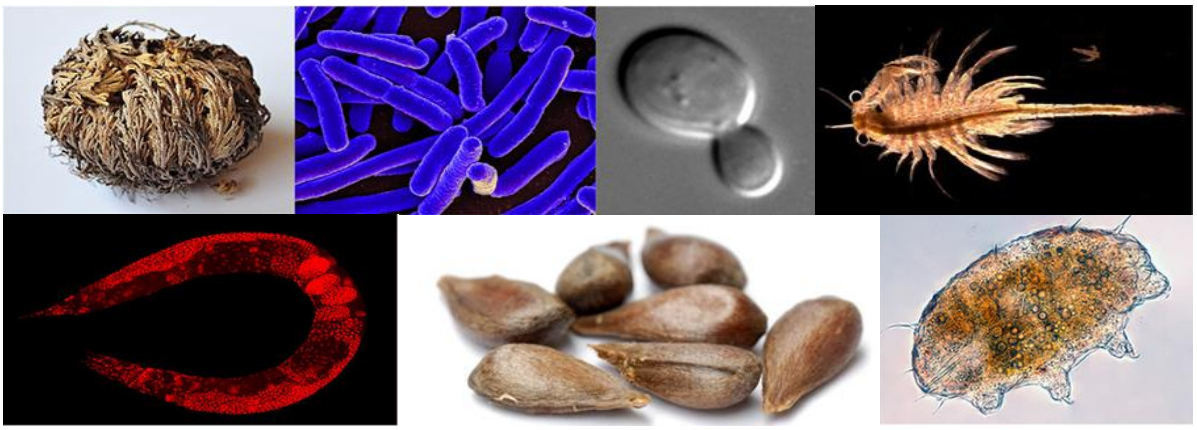


Figure 1.1. Some organisms that undergo anhydrobiosis. Clockwise from upper left: resurrection plant, bacteria, budding yeast, brine shrimp (top right), tardigrade (bottom right), plant seeds, and a nematode worm (bottom left). Figure adapted from Boothby & Pielak[8] with permission from Wiley and Sons.

In addition to enabling the long-term storage and resilience of whole organisms, dehydration can also improve the shelf life and thermostability of biomolecules *in vitro*.[9] Thus, there is demand for dehydrated formulations of biologic drugs, which must be produced, transported, and stored at low temperatures (<8 °C) in their aqueous form – a costly logistical challenge that, if not well-managed, can be ruinous.[10-11] However, given the key role of water in determining protein structure and stability, it comes as no surprise that many proteins succumb to dehydration. To prevent these destructive effects, which include unfolding and irreversible aggregation, many biologic drugs are formulated with protective molecules known as excipients to prevent unfolding during drying.[12] These molecules include trehalose, a non-reducing sugar found in many anhydrobiotic organisms. Unfortunately, the efficacy of dehydration-protective molecules depends on the client biomolecule, and their mode of action is poorly understood. This lack of understanding of dehydration protection by excipients means that dehydrated formulations must be developed through costly trial-and-error processes.[12]

Our poor understanding of protein dehydration protection stems largely from our limited knowledge of dehydration effects on protein structure. For decades, Fourier-Transform Infrared (FTIR) spectroscopy was one of few techniques able to probe dehydrated protein structure, albeit at the bulk secondary-structure level.[13] Local changes to structure remained invisible until the relatively recent advent of solid-state hydrogen-deuterium exchange mass spectrometry (ssHDX-ms) and liquid-observed vapor exchange (LOVE) NMR,[14-15] which permit peptide- and residue- level study of dehydrated protein structure, respectively.

The aim of this chapter is to synthesize and contextualize current knowledge and theory surrounding protein dehydration and dehydration-protection. We begin by analyzing the effects of mild and extreme dehydration from a macromolecular perspective, and then relate the physicochemical effects of dehydration to models of dehydration protection. We intend for this chapter to serve as a jumping-off point for future studies on dehydrated biomolecular structure and molecular mechanisms of dehydration protection.

MACROMOLECULAR CONSEQUENCES OF MILD DEHYDRATION

To understand how dehydration affects protein structure and function, one must consider the initial environment of the protein and how that environment changes over the entire process of dehydration, from initial stages where there is only a slight volume reduction all the way through to rehydration from a completely desiccated state.

Mild dehydration, which occurs through passive water evaporation or osmotic shock, is characterized by the *partial* loss of bulk water. In humans, a 4-5% loss in body water is tolerated with minimal side effects, but a loss of 15-25% can be lethal.[16] The immediate effects of mild dehydration, both *in vitro* and *in vivo,* are a decrease in system volume and a concomitant increase in macromolecular crowding.[17-18] A substantial increase in macromolecule concentration can alter the stability and structure of proteins and their complexes by increasing the frequency and strength of both hard-core steric repulsions and soft chemical interactions.[19] An increase in hard-core steric repulsions should result in entropic destabilization of the more voluminous state, thus increasing the population of the more compact species, i.e. the folded state of the protein. Increasing the frequency of chemical interactions with neighboring macromolecules, on the other hand, can be either stabilizing or destabilizing depending on the biomolecule(s) and their environment.

For instance, in *E. coli* cells, where the normal macromolecular concentration can exceed 300 g/L,[20] protein homodimers that exhibit different surface charges are differentially stabilized by the cellular environment – a result attributed to repulsive electrostatic interactions between the negatively charged monomer surface and negatively charged cellular environment.[21] By contrast, the small, metastable protein domain, drkN SH3, is less stable in osmotically shocked *E. coli* cells than it is in unshocked cells, despite the smaller cell volume of the shocked cells. This result can be attributed to an increase in attractive chemical interactions between the unfolded state of the protein and the cellular environment.[22]

As these results suggest, a dehydration-induced increase in macromolecular crowding can have myriad effects on protein structure, stability, and interactions. Although some proteins unfold and irreversibly aggregate under crowded conditions,[23-25] others – especially intrinsically disordered proteins[26] – will adopt new folds and/or form functional, supramolecular complexes.[27] For example, a late-embryogenesis abundant (LEA) protein involved in *Arabidopsis thaliana* desiccation tolerance transitions from mostly disordered in solution to half α-helical under even slightly dehydrating conditions;[28] and a different LEA protein in the brine shrimp *Artemia franciscana* phase-separates to form membraneless organelles that differentially sequester client proteins based on their net surface charge.[29] Observations like these support the idea that macromolecular crowding-induced changes to biomolecular structure are responsible for cellular dehydration sensation and response.[30]

MACROMOLECULAR CONSEQUENCES OF EXTREME DEHYDRATION

Further drying leads to complete loss of bulk water, at which point there is barely enough water to cover the protein surface (<0.38 g H_2O/g protein).[31] Such extreme dehydration occurs through prolonged passive evaporation or, for protein-based drugs, through accelerated drying methods such as lyophilization and spray-drying.[32] When extreme dehydration is reached via passive evaporation, proteins are gradually exposed to an increasingly crowded environment, and the prolonged time in such conditions increases the likelihood of irreversible aggregation. Lyophilization avoids this problem by first locking (freezing) the protein in place, and then subliming

the water at a relatively rapid pace. The shorter dwell time in a crowded, flexible environment could explain why lyophilization generally preserves protein structure better than passive evaporation, despite the fact that freeze-drying exerts the additional stress of freezing. The outcomes of extreme protein dehydration described below, however, are expected to occur regardless of drying method.

Loss of mobility: Numerous experiments and molecular dynamics simulations suggest that functional protein conformational changes are directly coupled to water motions.[33-35] Thus, it comes as no surprise that protein solids containing <0.20 g H_2O/g protein are significantly less flexible and functional than a fully hydrated protein.[36-37] It is tempting to speculate, then, that the retardation of large protein conformational changes (e.g, catalytic loop migrations, global unfolding) upon extreme dehydration contributes to the reduced chemical instability and extended shelf-life of many lyophilized proteins.

Loss of the hydrophobic effect: The tendency of nonpolar moieties to minimize interactions with water,[38] i.e., the hydrophobic effect,[1] drives globular protein folding by facilitating collapse of the hydrophobic core.[39-40] Given the key role of the hydrophobic effect in protein folding, one might expect extreme dehydration to "undo" the results of the hydrophobic effect, i.e. cause catastrophic global unfolding because there is no longer the entropic penalty of surface water re-organization around a hydrophobic solute. However, my work suggests that loss of the hydrophobic effect is not the primary cause of protein unfolding in the dry state, as the proteins studied maintain a substantial amount of structure in the dry state. Moreover, others have observed hydrophobic intermolecular interactions in the gas

phase.[41] Observations such as these suggest that once formed, many of the enthalpically favorable interactions found in the folded state of proteins and their complexes are quite stable in the absence of water.

Loss of stabilizing hydrogen bonds with water: In addition to the intramolecular H-bonds found in the folded state of proteins,[42] intermolecular H-bonds between the surfaces of globular proteins and water are essential to protein structure and stability. For instance, it has been shown that H-bonds with water enthalpically stabilize the folded state of proteins[43] and mediate mid-to-long range intra- and intermolecular interactions critical to native protein dynamics and function.[44-45] Under the conditions of extreme dehydration, the surface waters mediating these intra- and intermolecular interactions are lost, resulting in unfolding and structural rearrangement. In particular, dehydrated proteins often witness some degree of unfolding and exhibit reduced α-helical- and increased β-sheet-character.[46] This increase in β-sheet-character is attributed to the observation that α-helices tend to share more H-bonds with water than β-sheets, and are thus more perturbed by dehydration.[47]

MACROMOLECULAR CONSEQUENCES OF REHYDRATION

Finally, in both biological and pharmaceutical contexts, dry proteins must be resuspended in aqueous solution to regain their active form. Rehydration is therefore an important step to consider in protein dehydration and serves as an additional source of concern. Given that a substantial portion of a protein is unfolded in the dehydrated state,[15] the risk of irreversible aggregation is perhaps highest at the moment water contacts the dehydrated protein solid. At this moment, two competing

reactions, both driven by the hydrophobic effect, are initiated: the first-order reaction of protein refolding and the second-order reaction of protein aggregation.[48] The relative rates of these competing reactions thus determine the relative populations of folded vs. and aggregated protein.

PROPOSED MECHANISMS OF DEHYDRATION PROTECTION

Ever since organisms left the sea for the land, nature has evolved solutions to problems posed by periodic dehydration. In addition to the adaptations that allow organisms to retain or replenish cellular water, such as stomata in plants or dew-condensing microstructures on desert beetle exteriors,[49-50] many organisms also accumulate or synthesize protective molecules that prevent irreversible aggregation caused by dehydration. The modes of action by these protective molecules are not well-understood; however, several mechanisms to explain their protection have been proposed.

Stabilize the protein in solution. The initial response of an organism to osmotic stress is to produce or accumulate osmolytes such as glycine-betaine or trehalose;[22, 51-53] many of these osmolytes act by stabilizing biomolecules through the mechanism of preferential exclusion from the hydrophobic protein backbone, i.e. minimizing protein surface area:volume ratio.[54-55] Minimizing the population of unfolded protein in solution helps prevent irreversible aggregation typically promoted by the dehydration-induced increase in macromolecular crowding; however, this strategy is relatively ineffective at preventing the unfolding that occurs upon extreme dehydration.

Prevent aggregation by acting as a molecular shield or "holdase". Under the aforementioned model of aggregation kinetics, the aggregation reaction can be slowed by decreasing the population of the reactants (unfolded proteins) or by inhibiting complex formation (blocking the interaction). The latter – inhibition of complex formation – is the idea behind the relatively recent "molecular shield" hypothesis.[56-58] Under this hypothesis, the protective molecule acts as a shield that blocks the interaction between unfolded proteins; it can do this by "being in the way" (many desiccation-tolerant organisms produce large amounts of protective molecules) or by "holding" the exposed hydrophobic surface of the unfolded proteins, competitively occupying the so-called binding site. Given the concentration-dependence of the molecular shield mechanism (more inhibitor = reduced complex formation), this mechanism is expected to be effective at preventing aggregation during rehydration.

Replacement of H-bonds with water. Decades of FTIR studies on dehydrated protein structure suggest that extreme dehydration perturbs protein structure primarily through the loss of stabilizing surface H-bonds. It follows that replacing these H-bonds should prevent unfolding during extreme dehydration. This "water replacement" hypothesis is supported by FTIR studies of proteins dehydrated in the presence of trehalose, a disaccharide that can donate eight H-bonds. These efforts show trehalose preserves folded protein structure in the dry state.[59-60]

Immobilize the protein in a glassy matrix. A property common among dehydration-protective molecules is their ability to form glasses, i.e. to vitrify. This commonality, and the observation that such molecules lose their protective ability

above the glass transition temperature, led to the vitrification hypothesis.[61] The idea behind this hypothesis is that such protective molecules preserve structure by encapsulating the protein in a glassy matrix, thus stabilizing the protein by preventing large motions like protein unfolding.

Catalyze refolding. Finally, as has been mentioned, the rehydration step serves as a major potential source of aggregation. Therefore, molecules that catalyze refolding will also reduce irreversible aggregation.[62] This mechanism may explain why molecules that stabilize proteins in solution, but don't prevent dehydration-induced unfolding, behave synergistically with other molecules thought to behave as molecular shields.[63]

SUMMARY AND OUTLOOK

Some of the mechanisms for dehydration protection listed have more experimental support than others, but realistically, dehydration protection likely arises from a combination of mechanisms. Given the complexity of protein chemistry and, until recently, the paucity of techniques to study dehydrated protein structure at high resolution, it is unsurprising that despite decades of effort, our understanding of dehydration protection remains in its infancy. Our development of Liquid-Observed Vapor Exchange NMR (Chapter 2), a solution NMR technique that provides residue-level information on dehydrated protein structure and the effect of excipients, we hope to gain insight into the effects of extreme dehydration on protein structure and mechanisms of dehydration protection.

REFERENCES

1. Kauzmann, W., Some Factors in the Interpretation of Protein Denaturation. *Adv. Protein Chem.* **1959,** *14*, 1-63.

2. Bellissent-Funel, M. C.; Hassanali, A.; Havenith, M.; Henchman, R.; Pohl, P.; Sterpone, F.; van der Spoel, D.; Xu, Y.; Garcia, A. E., Water Determines the Structure and Dynamics of Proteins. *Chem. Rev.* **2016,** *116* (13), 7673-97.

3. Ball, P., Water is an active matrix of life for cell and molecular biology. *Proc. Natl. Acad. Sci. U.S.A.* **2017,** *114* (51), 13327-13335.

4. Ratkova, E. L.; Dawidowski, M.; Napolitano, V.; Dubin, G.; Fino, R.; Ostertag, M. S.; Sattler, M.; Popowicz, G.; Tetko, I. V., Water envelope has a critical impact on the design of protein–protein interaction inhibitors. *Chem. Commun.* **2020,** *56* (31), 4360-4363.

5. Patel, J.; Kothari, R.; Tunga, R.; Ritter, N.; Tunga, B., Stability considerations for biopharmaceuticals: overview of protein and peptide degradation pathways. *BioProcess International* **2011,** *9*, 2-11.

6. Crowe, J. H., Anhydrobiosis: an unsolved problem. *The American Naturalist* **1971,** *105* (946), 563-573.

7. J H Crowe; F A Hoekstra, a.; Crowe, L. M., Anhydrobiosis. *Annu. Rev. Physiol.* **1992,** *54* (1), 579-599.

8. Boothby, T. C.; Pielak, G. J., Intrinsically Disordered Proteins and Desiccation Tolerance: Elucidating Functional and Mechanistic Underpinnings of Anhydrobiosis. *BioEssays* **2017,** *39* (11), 1700119.

9. Pikal, M. J.; Dellerman, K. M., Stability testing of pharmaceuticals by high-sensitivity isothermal calorimetry at 25°C: cephalosporins in the solid and aqueous solution states. *Int. J. Pharm.* **1989,** *50* (3), 233-252.

10. Barrowclough, N. The Cost of a Broken Cold Chain in the Pharmaceutical Industry. https://pharma-mon.com/drug-storage-monitoring/the-cost-of-a-broken-cold-chain-in-the-pharmaceutical-industry/.

11. Yu, Y. B.; Briggs, K. T.; Taraban, M. B.; Brinson, R. G.; Marino, J. P., Grand Challenges in Pharmaceutical Research Series: Ridding the Cold Chain for Biologics. *Pharmaceutical Research* **2021,** *38* (1), 3-7.

12. Piszkiewicz, S.; Pielak, G. J., Protecting Enzymes from Stress-Induced Inactivation. *Biochemistry* **2019,** *58* (37), 3825-3833.

13. Moorthy, B. S.; Iyer, L. K.; Topp, E. M., Characterizing Protein Structure, Dynamics and Conformation in Lyophilized Solids. *Curr. Pharm. Des.* **2015,** *21* (40), 5845-5853.

14. Sinha, S.; Li, Y.; Williams, T. D.; Topp, E. M., Protein Conformation in Amorphous Solids by FTIR and by Hydrogen/Deuterium Exchange with Mass Spectrometry. *Biophysical Journal* **2008,** *95* (12), 5951-5961.

15. 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.

16. Ashcroft, F., *Life at the extremes: the science of survival*. Univ of California Press: 2002.

17. Costill, D.; Branam, L.; Eddy, D.; Fink, W., Alterations in red cell volume following exercise and dehydration. *Journal of applied physiology* **1974,** *37* (6), 912-916.

18. Mansell, J. L.; Clegg, J. S., Cellular and molecular consequences of reduced cell water content. *Cryobiology* **1983,** *20* (5), 591-612.

19. Daniel Harries, S. S., Claire Stewart, Liel Sapir, and Gary J. Pielak, Crowding Effects on Protein- and Protein- Complex Stability. *Annu. Rev. Biophys.* **2021,** *TBD* (TBD).

20. Cayley, S.; Lewis, B. A.; Guttman, H. J.; Record, M. T., Characterization of the cytoplasm of Escherichia coli K-12 as a function of external osmolarity: Implications for protein-DNA interactions in vivo. *J. Mol. Biol.* **1991,** *222* (2), 281-300.

21. 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. U.S.A.* **2021,** *118* (11), e2019918118.

22. 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.

23. Ellis, R. J.; Minton, A. P., Protein aggregation in crowded environments. **2006,** *387* (5), 485-497.

24. Mukherjee, S.; Chowdhury, P.; Gai, F., Effect of dehydration on the aggregation kinetics of two amyloid peptides. *The journal of physical chemistry. B* **2009,** *113* (2), 531-535.

25. Moronetti Mazzeo, L. E.; Dersh, D.; Boccitto, M.; Kalb, R. G.; Lamitina, T., Stress and aging induce distinct polyQ protein aggregation states. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109* (26), 10587-10592.

26. 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.

27. Wang, Y.; Sukenik, S.; Davis, C. M.; Gruebele, M., Cell Volume Controls Protein Stability and Compactness of the Unfolded State. *J. Phys. Chem. B* **2018,** *122* (49), 11762-11770.

28. Popova, A. V.; Hundertmark, M.; Seckler, R.; Hincha, D. K., Structural transitions in the intrinsically disordered plant dehydration stress protein LEA7 upon drying are modulated by the presence of membranes. *Biochimica et Biophysica Acta, Biomembranes* **2011,** *1808* (7), 1879-1887.

29. Belott, C.; Janis, B.; Menze, M. A., Liquid-liquid phase separation promotes animal desiccation tolerance. *Proc. Natl. Acad. Sci. U.S.A.* **2020,** *117* (44), 27676-27684.

30. Model, M. A.; Hollembeak, J. E.; Kurokawa, M., Macromolecular Crowding: a Hidden Link Between Cell Volume and Everything Else. *Cell Physiol Biochem* **2021,** *55* (S1), 25-40.

31. Rupley, J. A.; Gratton, E.; Careri, G., Water and globular proteins. *Trends Biochem. Sci.* **1983,** *8* (1), 18-22.

32. Abdul-Fattah, A. M.; Kalonia, D. S.; Pikal, M. J., The challenge of drying method selection for protein pharmaceuticals: Product quality implications. *J. Pharm. Sci.* **2007,** *96* (8), 1886-1916.

33. Schirò, G.; Fichou, Y.; Gallat, F.-X.; Wood, K.; Gabel, F.; Moulin, M.; Härtlein, M.; Heyden, M.; Colletier, J.-P.; Orecchini, A.; Paciaroni, A.; Wuttke, J.; Tobias, D. J.; Weik, M., Translational diffusion of hydration water correlates with functional motions in folded and intrinsically disordered proteins. *Nat. Commun.* **2015,** *6* (1), 6490.

34. Oroguchi, T.; Nakasako, M., Changes in hydration structure are necessary for collective motions of a multi-domain protein. *Sci. Rep.* **2016,** *6* (1), 26302.

35. Atamas, N.; Bardik, V.; Bannikova, A.; Grishina, O.; Lugovskoi, E.; Lavoryk, S.; Makogonenko, Y.; Korolovych, V.; Nerukh, D.; Paschenko, V., The effect of water dynamics on conformation changes of albumin in pre-denaturation state: photon correlation spectroscopy and simulation. *Journal of Molecular Liquids* **2017,** *235*, 17-23.

36. Rupley, J. A.; Yang, P. H.; Tollin, G., Thermodynamic and Related Studies of Water Interacting with Proteins. In *Water in Polymers*, AMERICAN CHEMICAL SOCIETY: 1980; Vol. 127, pp 111-132.

37. Poole, P.; Finney, J., Hydration-induced conformational and flexibility changes in lysozyme at low water content. *Int. J. Biol. Macromol.* **1983,** *5* (5), 308-310.

38. Xi, E.; Patel, A. J., The hydrophobic effect, and fluctuations: The long and the short of it. *Proc. Natl. Acad. Sci. U.S.A.* **2016,** *113* (17), 4549-4551.

39. Dill, K. A., Dominant forces in protein folding. *Biochemistry* **1990,** *29* (31), 7133-7155.

40. Agashe, V. R.; Shastry, M. C. R.; Udgaonkar, J. B., Initial hydrophobic collapse in the folding of barstar. *Nature* **1995,** *377* (6551), 754-757.

41. Liu, L.; Bagal, D.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S., Hydrophobic Protein−Ligand Interactions Preserved in the Gas Phase. *J. Am. Chem. Soc.* **2009,** *131* (44), 15980-15981.

42. 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. U.S.A.* **1951,** *37* (4), 205-211.

43. Thanki, N.; Umrania, Y.; Thornton, J. M.; Goodfellow, J. M., Analysis of protein main-chain solvation as a function of secondary structure. *J. Mol. Biol.* **1991,** *221* (2), 669-691.

44. Watenpaugh, K. D.; Margulis, T. N.; Sieker, L. C.; Jensen, L. H., Water structure in a protein crystal: Rubredoxin at 1.2 Å resolution. *J. Mol. Biol.* **1978,** *122* (2), 175-190.

45. Kuffel, A., How water mediates the long-range interactions between remote protein molecules. *Phys. Chem. Chem. Phys.* **2017,** *19* (7), 5441-5448.

46. Griebenow, K.; Klibanov, A. M., Lyophilization-induced reversible changes in the secondary structure of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1995,** *92* (24), 10969-10976.

47. Morris, A. S.; Thanki, N.; Goodfellow, J. M., Hydration of amino acid side chains: dependence on secondary structure. *Protein Eng., Des. Sel.* **1992,** *5* (8), 717-728.

48. Kiefhaber, T.; Rudolph, R.; Kohler, H.-H.; Buchner, J., Protein Aggregation in vitro and in vivo: A Quantitative Model of the Kinetic Competition between Folding and Aggregation. *Bio/Technology* **1991,** *9* (9), 825-829.

49. Malik, F.; Clement, R.; Gethin, D.; Krawszik, W.; Parker, A., Nature's moisture harvesters: a comparative review. *Bioinspiration & biomimetics* **2014,** *9* (3), 031002.

50. Sachs, F.; Sivaselvan, M. V., Cell volume control in three dimensions: Water movement without solute movement. *J. Gen. Physiol.* **2015,** *145* (5), 373-380.

51. Yancey, P. H.; Clark, M. E.; Hand, S. C.; Bowlus, R. D.; Somero, G. N., Living with water stress: evolution of osmolyte systems. *Science* **1982,** *217* (4566), 1214-22.

52. Csonka, L. N., Physiological and genetic responses of bacteria to osmotic stress. *Microbiological reviews* **1989,** *53* (1), 121-147.

53. Record, M. T.; Courtenay, E. S.; Cayley, S.; Guttman, H. J., Biophysical compensation mechanisms buffering E. coli protein–nucleic acid interactions against changing environments. *Trends Biochem. Sci.* **1998,** *23* (5), 190-194.

54. Xie, G.; Timasheff, S. N., The thermodynamic mechanism of protein stabilization by trehalose. *Biophys. Chem.* **1997,** *64* (1-3), 25-43.

55. Mukherjee, M.; Mondal, J., Unifying the Ambivalent Mechanisms of Protein-Stabilising Osmolytes. *bioRxiv* **2020**, 2020.02.22.960856.

56. Chakrabortee, S.; Boschetti, C.; Walton, L. J.; Sarkar, S.; Rubinsztein, D. C.; Tunnacliffe, A., Hydrophilic protein associated with desiccation tolerance exhibits broad protein stabilization function. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104* (46), 18073-18078.

57. Das, U.; Hariprasad, G.; Ethayathulla, A. S.; Manral, P.; Das, T. K.; Pasha, S.; Mann, A.; Ganguli, M.; Verma, A. K.; Bhat, R.; Chandrayan, S. K.; Ahmed, S.; Sharma, S.; Kaur, P.; Singh, T. P.; Srinivasan, A., Inhibition of Protein Aggregation: Supramolecular Assemblies of Arginine Hold the Key. *PLoS ONE* **2007,** *2* (11), e1176.

58. Chakrabortee, S.; Tripathi, R.; Watson, M.; Schierle, G. S. K.; Kurniawan, D. P.; Kaminski, C. F.; Wise, M. J.; Tunnacliffe, A., Intrinsically disordered proteins as molecular shields. *Mol. BioSyst.* **2012,** *8* (1), 210-219.

59. Crowe, J. H.; Clegg, J. S.; Crowe, L. M., Anhydrobiosis: the water replacement hypothesis. *The properties of water in foods ISOPOW 6* **1998**, 440-455.

60. Jain, N. K.; Roy, I., Effect of trehalose on protein structure. *Protein science : a publication of the Protein Society* **2009,** *18* (1), 24-36.

61. Crowe, J. H.; Carpenter, J. F.; Crowe, L. M., The role of vitrification in anhydrobiosis. *Annu. Rev. Physiol.* **1998,** *60* (1), 73-103.

62. Zhang, M. Z.; Wen, J.; Arakawa, T.; Prestrelski, S. J., A New Strategy for Enhancing the Stability of Lyophilized Protein: The Effect of the Reconstitution Medium on Keratinocyte Growth Factor. *Pharm. Res.* **1995,** *12* (10), 1447-1452.

63. Kim, S. X.; Çamdere, G.; Hu, X.; Koshland, D.; Tapia, H., Synergy between the small intrinsically disordered protein Hsp12 and trehalose sustain viability after severe desiccation. *eLife* **2018,** *7*, e38337.

CHAPTER 2: DRIED PROTEIN STRUCTURE REVEALED AT THE RESIDUE LEVEL BY LIQUID-OBSERVED VAPOR EXCHANGE NMR

Edited from: Crilly et al. *Biochemistry* **(2021)** 152-159.

INTRODUCTION

It is well-established that liquid water is necessary for proteins to realize their native structure and function,[1] yet uncovering how protein-water interactions contribute to protein stability and structure remains an ongoing endeavor.[2-3] Our understanding of this fundamental interaction has been limited in part by the technological inability to observe how water removal affects local protein structure.[4] The same restrictions make it challenging to understand how protective molecules, collectively known as excipients,[5-6] prevent dehydration-induced protein damage.

Our limited understanding of dehydrated protein structure poses a hurdle for distributing protein-based therapeutics such as vaccines, antibodies and other biologic drugs, for which dried formulations are in high demand due to their enhanced thermostability and shelf life.[7] The advent of new methods to study dry protein structure at high resolution would inform and accelerate formulation,[8] allowing more protein-based therapeutics to reach the market in freeze-dried form, reducing cold-chain costs.[9]

Studies performed with solid-state hydrogen-deuterium exchange mass spectrometry (ssHDX-MS),[10] a technique that provides peptide-level information on dried protein structure, demonstrate the predictive power afforded by high resolution data.[11-12] However, residue-level information is essential for gaining a thorough understanding of protein-water interactions and how they relate to the mechanisms of dehydration-induced unfolding and cosolute-mediated dehydration protection.

We developed Liquid-Observed Vapor Exchange (LOVE) NMR (Fig. 2.1) to enable the study of dehydrated protein structure and protection at the residue level.

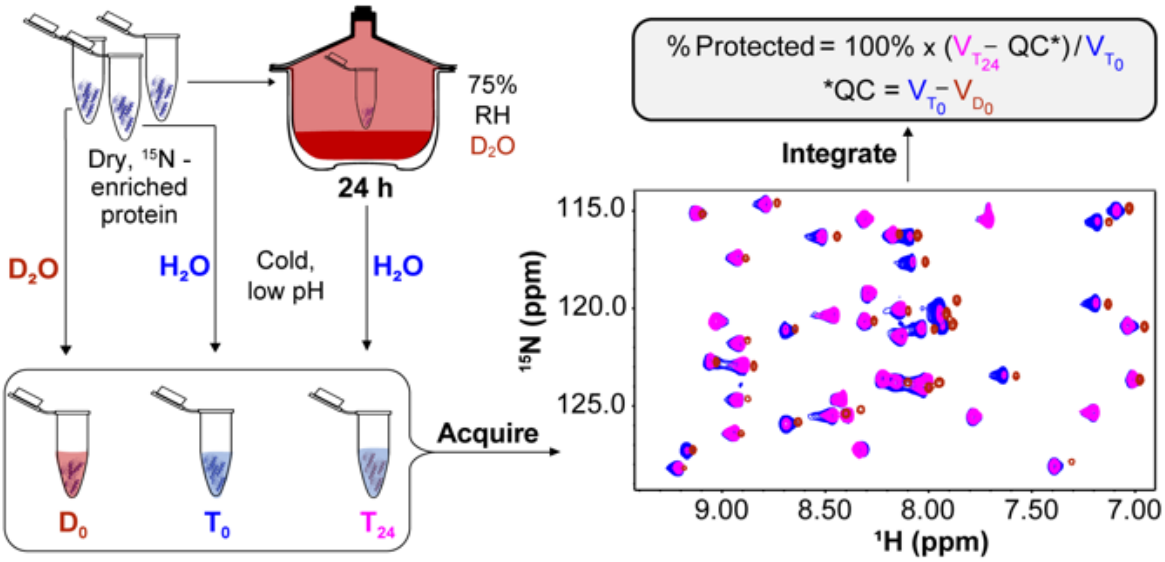


Figure 2.1. Original LOVE NMR workflow. Three identical samples of ^{15}N-enriched protein dried alone or in the presence of a cosolute are resuspended in cold, acidic buffer before (D_0, T_0) or after (T_{24}) 24 h exposure to D_2O vapor at 75% relative humidity (RH). Amide protons that are unprotected by an H-bond in the dry state will exchange with deuterons from the vapor, resulting in smaller cross peak volumes in the T_{24} ^{15}N-^{1}H HSQC spectrum relative to the T_0 spectrum (T_0 and T_{24} cross peaks are shown in blue and pink, respectively). A third sample, D_0, is resuspended in D_2O quench buffer (red cross peaks). The difference in volumes between corresponding peaks in the T_0 and D_0 spectra ($V_{T0} – V_{D0}$) reflects quench-labelling and is used as a quench correction (QC, see text), which is subtracted from the volume of the corresponding peak in the T_{24} spectrum (V_{T24}). The difference is divided by V_{T0} and multiplied by 100% to yield %Protected.

Inspired by the report of Desai et al.,[13] LOVE NMR uses solution NMR spectroscopy to quantify the extent of hydrogen-deuterium exchange (HDX) between D_2O vapor and the unprotected amide protons of a dried protein. Based on the well-established principle that amide protons are less likely to exchange with deuterons from D_2O if involved in intra- or inter-molecular H-bonds,[14-16] we expect residues in structured regions of the protein, or in regions of the protein that hydrogen-bond with cosolutes, to be more protected from exchange than residues in unstructured, exposed regions.

MATERIALS AND METHODS

Materials. Ampicillin (Sigma Aldrich), trehalose, and urea (Thermo Fisher) were used without further purification. H_2O with a resistivity >17 MΩ cm^{-1} was used to prepare buffers. pH values are direct readings, uncorrected for the deuterium isotope effect.[17] The pET11a plasmid containing the gene encoding the T2Q variant of the immunoglobulin G binding domain of streptococcal G was provided by Leonard D. Spicer's laboratory at Duke University (Durham, North Carolina). This variant, which we call GB1 throughout the manuscript, was chosen because the mutation prevents N-terminal deamidation.[18] A constant relative humidity (RH) of 75 ± 5%, as measured by a digital hygrometer (Fisherbrand TraceableGO™ Bluetooth Datalogging Hygrometer), was created by sealing a 0.5 L chamber containing 200 mL of >99% D_2O (Cambridge Isotope Labs) saturated with anhydrous $Co(II)Cl_2$ (Agros Organics).[19-20]

Protein expression and purification. ^{15}N-enriched GB1 was expressed in Agilent BL21 Gold (DE3) *E. coli* in minimal media.[21-22] Following 2-3 h of expression,

cells from each 1-L culture were harvested via centrifugation at 4,000*g*, the supernatant discarded, and the pellets stored at -20 °C. Cell pellets were thawed at room temperature, resuspended in 5 mL of 20 mM Tris pH 7.5, and lysed via sonication for 8 min at 20% amplitude with a 33% duty cycle using a Fisher Scientific Sonic Dismembrator Model 500. Lysates were clarified by centrifugation at 15,000*g* for 1 h. Clarified lysates were passed through a 0.45 µm filter (Millipore) and purified via liquid chromatography.[21] The concentration of purified protein was determined from the absorbance at 280 nm (A_{280}) (Nanodrop One, Thermo Fisher) using an extinction coefficient of 9530 M^{-1} cm^{-1}.[23] Purity was confirmed by Q-TOF mass spectrometry (ThermoScientific, Q Exactive HF-X) in the UNC Mass Spectrometry Chemical Research and Teaching Core Laboratory (6290.32 Da expected, 6290 Da observed). Purified protein was exchanged into H_2O by dialysis (ThermoScientific Snakeskin™ dialysis tubing, 3500 Da molecular weight cutoff), and divided into aliquots such that resuspension in 650 µL gives a protein concentration of 500 µM. Aliquots were flash frozen, lyophilized by exposure to pressures <0.3 mBar for 24 h (LABCONCO FreeZone 1 Liter Benchtop Freeze Dry System), and stored at -20 °C.

NMR. Experiments were performed in triplicate on Bruker Avance III HD spectrometers with cryogenic QCI probes at ^{1}H Larmor frequencies of 600 MHz for LOVE experiments and 850 MHz for solution amide-proton exchange experiments. For LOVE experiments, ^{15}N-^{1}H heteronuclear single-quantum coherence (HSQC) spectra were acquired in ~20 min (128 increments in the ^{15}N dimension, 8 scans per increment) with sweep widths of 3041 Hz in the ^{15}N dimension and 8418 Hz in the ^{1}H dimension. For residues that exchange quickly in solution, Band-selective

Excitation Short-Transient (BEST) ^{15}N-^{1}H HSQC experiments were acquired in ~3 min (128 increments in the ^{15}N dimension, 4 scans per increment) with sweep widths of 3016 Hz in the ^{15}N dimension and 13587 Hz in the ^{1}H dimension.[24] For all other solution exchange experiments, traditional ^{15}N-^{1}H HSQC spectra were acquired in ~20 min (128 increments in the ^{15}N dimension, 8 scans per increment) with sweep widths of 4308 Hz in the ^{15}N dimension and 11904 Hz in the ^{1}H dimension. Spectra were processed with NMRPipe.[25] Cross peaks were integrated using NMRViewJ.[26]

Backbone resonances were assigned (pH 4.5, 4 °C and pH 7.5, 22 °C, Figure S1, Table S1) using isotopically-enriched GB1 T2Q expressed in minimal media containing ^{13}C D-glucose and ^{15}N NH_4Cl (Cambridge Isotope Labs) as the sole sources of carbon and nitrogen, respectively, and purified as described above. HNCACB spectra were acquired with 10% sampling in the indirect dimensions using a Poisson gap scheduling scheme (http://gwagner.med.harvard.edu/intranet/hmsIST/gensched_new.html).[27-28] 3D spectra were reconstructed using the SMILE algorithm and processed in NMRpipe.[29]

Solution hydrogen-deuterium exchange. Lyophilized aliquots of purified, ^{15}N-enriched GB1 were resuspended in 650 µL of 7.5-mM HEPES, pH 6.5 with or without 100 g/L trehalose or urea, flash frozen and lyophilized. After 24 h, samples were removed, resuspended in 650 µL 99% D_2O and immediately used to acquire serial NMR HSQC spectra at 22 °C. To obtain data for residues that fully exchange in <1 h (fast regime), Band-selective Excitation Short-Transient HSQC experiments were used to acquire ~10 spectra in the first 30 min with a deadtime of ~3 min.[24] To capture decay curves for residues that completely exchange in 2-24 h (intermediate

regime), traditional ^{15}N-^{1}H HSQC experiments were used to acquire 10-12 spectra over ~24 h. For slowly exchanging residues (>24 h), samples were resuspended, and following acquisition of the first ^{15}N-^{1}H HSQC spectrum (0 h timepoint), placed in an incubator at 22 °C. Samples were removed from the incubator every 1-3 days for spectrum acquisition. Sample pH was measured at the end of exchange; all samples possessed a pH of ~ 7.5.

The rate analysis tool in NMRViewJ was used to fit peak volumes as a function of time to the 3-parameter equation $V = Ae^{-Bt} + C$, where V is peak volume in arbitrary units, t is time in s, and B is the observed rate constant (k_{obs}). For each residue, k_{obs} was divided by the estimated intrinsic rate constant of exchange (k_{int}) at pH 7.5, 22 °C to approximate the opening equilibrium constant, K_{op}. Values of k_{int} were obtained using the online Server Program for Hydrogen Exchange Rate Estimation, SPHERE.[30] To ensure accurate k_{obs} values, only crosspeak volumes that decayed to ~30% of their initial value were analyzed. Opening free energies ($\Delta G°'_{op}$ values) were calculated as:

$$\mathbf{\Delta G_{op}}^{\circ\prime} = -RTln(K_{op}) = -RTln(\frac{k_{obs}}{k_{int}})$$

where R is the gas constant and T is the absolute temperature.[15, 31]

Liquid Observed Vapor Exchange (LOVE) NMR. For each experiment, three identical aliquots of pure, lyophilized, ^{15}N-enriched GB1 were resuspended in 650 μL of 1.5-mM HEPES pH 6.5 with or without 20 g/L trehalose or urea to a final protein concentration of 500 uM, flash-frozen, and exposed to pressures of <0.3 mBar on a standard LABCONCO FreeZone 1 Liter Benchtop Freeze Dry System for

24 h. Following lyophilization, two samples, designated T_0 and D_0, were immediately resuspended in 650 µL of cold quench buffer (100 mM citrate buffer, pH 4.5, 90% H_2O/10% D_2O for T_0 or >99% D_2O for D_0) and transferred to an NMR spectrometer for spectrum acquisition at 4 °C. The third sample, designated T_{24}, was placed, with the cap opened, in a chamber with a controlled relative humidity of ~75% (D_2O), prepared as described above. After 24 h, the T_{24} sample was resuspended in 650 µL of cold quench buffer and an HSQC spectrum was acquired using the same parameters as the T_0 and D_0 samples. The time between resuspension and initiation of spectrum acquisition was 10 min, ~8 min of which were spent at 4 °C. For the vapor exchange time course, GB1 samples dried from 650 µL 1.5-mM HEPES, pH 6.5, were stored in the constant humidity chamber for the times provided in the caption of Figure 2.

To ensure that the change in cross peak volumes originates exclusively from exchange with D_2O vapor, differences in GB1 concentration between the T_0, D_0, and T_{24} samples were determined post-experiment via the absorbance at 280 nm using the extinction coefficient provided above. The A_{280} values were used to normalize each cross peak volume across the three datasets. Using concentration-normalized cross peak volumes V_{T0}, V_{D0}, and V_{T24}, the percent of the dried protein population for which a given amide proton is protected from vapor exchange is calculated as:

$$\%Protected = 100 \times (V_{T24} - QC)/ V_{T0}$$

where $QC = V_{T0} - V_{D0}$. See results section for additional information on the QC value.

The change in %Protected from drying in the presence of a cosolute is calculated as: $\Delta\%Protected = \%Protected_{buffer + cosolute} - \%Protected_{buffer\ only}$.

Uncertainties from triplicate analysis and propagation of error analysis are discussed in the text and figure captions.

Thermogravimetric analysis. Aliquots of purified, unenriched GB1 (650 µL, 500 µM in 1.5-mM HEPES buffer, pH 6.5) were flash-frozen and exposed to pressures <0.3 mBar for 24 h. Samples were then placed, without caps, in a chamber with a controlled relative humidity of 75 ±5%, created with >99% D_2O saturated with $Co(II)Cl_2$ as described above. Individual tubes were removed after 0, 1, 2, 4, 6, 12, 24, 48, and 72 h and immediately analyzed using a TA Instruments Thermogravimetric Analyzer 550. The samples were loaded onto an open Pt pan and ramped from 25 °C to 200 °C at 4 °C/min under N_2(g) sample purge of 60 mL/min and balance purge of 40 mL/min. The well-defined mass loss ending at 140 °C was used to quantify H_2O+D_2O content.[32-33]

RESULTS

Quantifying vapor exchange with LOVE NMR. To quantify deuterium incorporation in the solid state, two dried ^{15}N-enriched protein samples, one exposed (T_{24}) and one not exposed (T_0) to D_2O vapor for 24 h, are dissolved in cold acidic buffer, which slows solution HDX and enables the immediate acquisition of a solution ^{15}N-^{1}H heteronuclear single-quantum coherence (HSQC) NMR spectrum (Fig. S2.1A). Cross peak volumes (V) from the assigned resonances (Table S2.1) are directly proportional to the concentration of amide protons, meaning that the volumes of corresponding cross peaks in the pre- and post-exchange spectra can be

compared to determine the percent of amide protons protected from exchange (i.e. cross peaks from residues that exchange with D_2O vapor will be smaller than those from residues protected from exchange).

This approach can be applied to any protein for which one can obtain a well-dispersed HSQC spectrum (<40 kDa), as long as the protein is stable (i.e. it folds much faster than the rate of solution HDX for an unprotected amide proton). Meeting the latter condition ensures that solution HDX does not significantly alter the difference in ^{1}H-^{15}N signal between T_0 and T_{24}. We know that solution HDX was minimal during the 30 minute period between resuspension and complete HSQC spectrum acquisition for the protein used here, the T2Q variant of the 6-kDa immunoglobulin binding domain B1 of streptococcal protein G (GB1) (Table S2.2).[34-35] For proteins more susceptible to HDX than GB1, time spent in solution can be reduced by modifying NMR acquisition parameters or by using an alternative HSQC pulse sequence.

While necessary for the acquisition of a well-dispersed solution NMR spectrum, resuspension in aqueous buffer causes a complication called quench-labelling, where labile protons or deuterons from the quench buffer immediately exchange with a fraction of surface-exposed amides upon resuspension. The degree of quench-labelling varies by residue position, with residues in structured areas witnessing less quench-labelling than those in areas lacking structure (Fig. S2.2). Quench-labelling can mask some or all of the amide proton signal loss from deuterium incorporation in the dry state, and must therefore be taken into account.

To correct for quench-labelling artifacts, a spectrum is acquired for a third, non-vapor-exchanged sample resuspended in a D_2O quench buffer (D_0). The isotope effect on quench-labelling is small; for most residues, the signal lost due to quench-labelling by D_2O ($V_{T0} – V_{D0}$) is nearly identical to the signal gained by a dried, deuterium-exchanged protein due to quench-labelling by H_2O (Fig. S2.2). The signal lost upon resuspension in D_2O is thus used as a quench correction (QC), which is subtracted from the volume of the corresponding peak in the T_{24} spectrum (V_{T24}).

Finally, each quench-corrected peak volume is normalized by dividing by the corresponding peak volume from the non-vapor-exchanged, H_2O-quenched sample (V_{T0}), and multiplied by 100% to yield the percent of the protein population for which a given amide proton is protected from vapor exchange (%Protected).

Interpreting LOVE NMR measurements. The abilities of solution HDX and LOVE NMR to detect protein structure are derived from the same principle: H-bonding prevents exchange. However, these methods differ in how they measure protection from exchange. With solution HDX, all amide protons eventually exchange to completion (Fig. 2.2A, dotted lines), but at different rates that, under certain conditions,[14, 31] can be used to estimate protection factors and opening free energies ($\Delta G^{o'}_{op}$).[36] In contrast, the extent of vapor exchange measured by LOVE NMR varies by residue (Fig. 2.2A, solid lines) while the rate for all residues mirrors that of vapor sorption (Fig. 2.2B). Although mechanisms of protein folding and HDX in the solid state are poorly understood, the observation that vapor exchange plateaus at different levels for different residues suggests that LOVE NMR does not

report on an equilibrium folding process as solution HDX does. Rather, this observation suggests that each plateau value measured by LOVE NMR reports on the fraction of the dried protein population that became trapped in a conformation where a given residue is protected from vapor exchange.

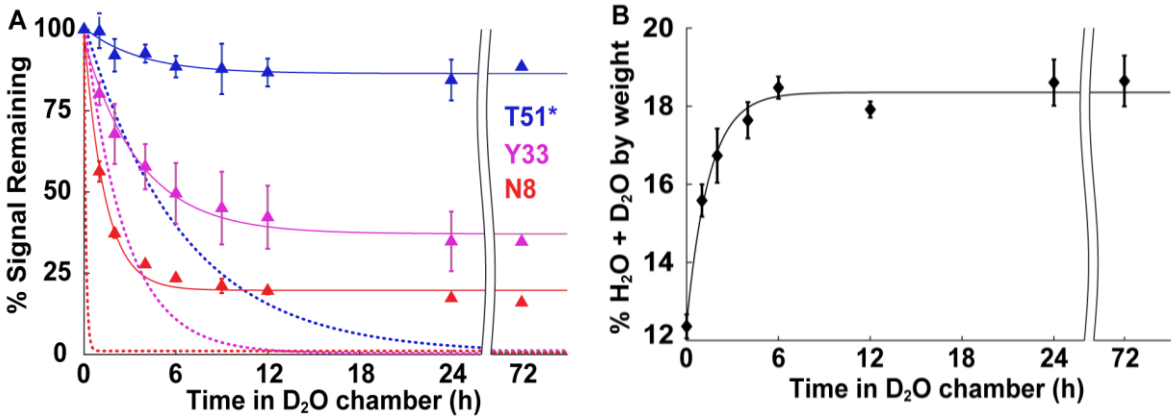


Figure 2.2. Time courses of solution HDX, vapor exchange, and vapor sorption. (A) Signal remaining as a function of time in D_2O liquid (dotted lines) or vapor (75% RH, triangles) for three GB1 residues representing different solution exchange rate regimes: slow (blue, complete in >24 h), intermediate (pink, 2 – 24 h), and fast (red, < 2 h). *For T51 in solution, the abscissa is compressed tenfold (e.g. 6 h corresponds to 60 h). Solution exchange curves are derived from the average observed rate constants from three independent experiments in 7.5 mM HEPES, pH 7.5 (>99% D_2O). For the vapor exchange time course, identical GB1 samples were freeze-dried in 1.5 mM HEPES, pH 6.5, for 24 h and resuspended in H_2O-based quench buffer after 0, 1, 2, 4, 6, 9 ,12, 24 and 72 h of vapor exchange at 75% RH. Spectra were acquired at 4 °C immediately upon resuspension. Vapor data are uncorrected for quench-labelling. (B) Average % H_2O + D_2O content (w/w) of freeze-dried GB1 as a function of time spent at 75% RH in a D_2O chamber. At 75% RH, the ratio of exchangeable deuterons in the vapor to protons in the solid is ~ 7:1, and GB1 solvent-accessible surface area covered by H_2O+D_2O ≤ 43% at 72 h (Table S2.3). For vapor exchange and sorption, curves are of no theoretical significance. Error bars represent the standard deviation of the mean from three independent experiments, except for the 72-h vapor-exchange data in panel A, which are from a single experiment.

The similarity between the kinetics of vapor exchange and vapor sorption suggests that D_2O vapor concentration may affect LOVE NMR data. Measurements made after exposing freeze-dried GB1 to 85% RH (D_2O) for 24 h confirm this idea, with all residues witnessing a ~10% reduction in signal relative to measurements

made after exposure to 75% RH (Fig. S2.3). The observation that additional signal loss at 85% RH is distributed almost evenly across the protein sequence suggests that decreased vapor exchange in the reverse direction (N-D→N-H),[37] rather than humidity-induced changes to dried protein structure, is the source of signal reduction. The preservation of protection trends at different humidities reinforces the idea that protection from vapor exchange is provided by dehydrated protein structure and demonstrates that LOVE NMR can be performed at different relative humidities. However, this result also suggests that the fine structure of the LOVE profile in the most protected protein regions might not be revealed by experiments performed at low relative humidities.

Assessing protein-wide trends. Plotting %Protected against residue number yields the LOVE profile of a protein. Comparing the LOVE profile of GB1 to a secondary structure map of the protein (Fig. 2.3A) shows that regions with stable secondary structure in solution tend to be more protected in the dry state. The similar levels of protection experienced by regions that form tertiary contacts in the native structure (e.g. β-sheets 1 and 4)[38] suggest that some GB1 molecules possess native or near-native tertiary structure in the dry state. Residues that undergo solution HDX only upon global unfolding ("global unfolders")[39] experience the most protection, yet witness 40% signal loss on average, implying that a subpopulation of GB1 unfolds substantially during drying.

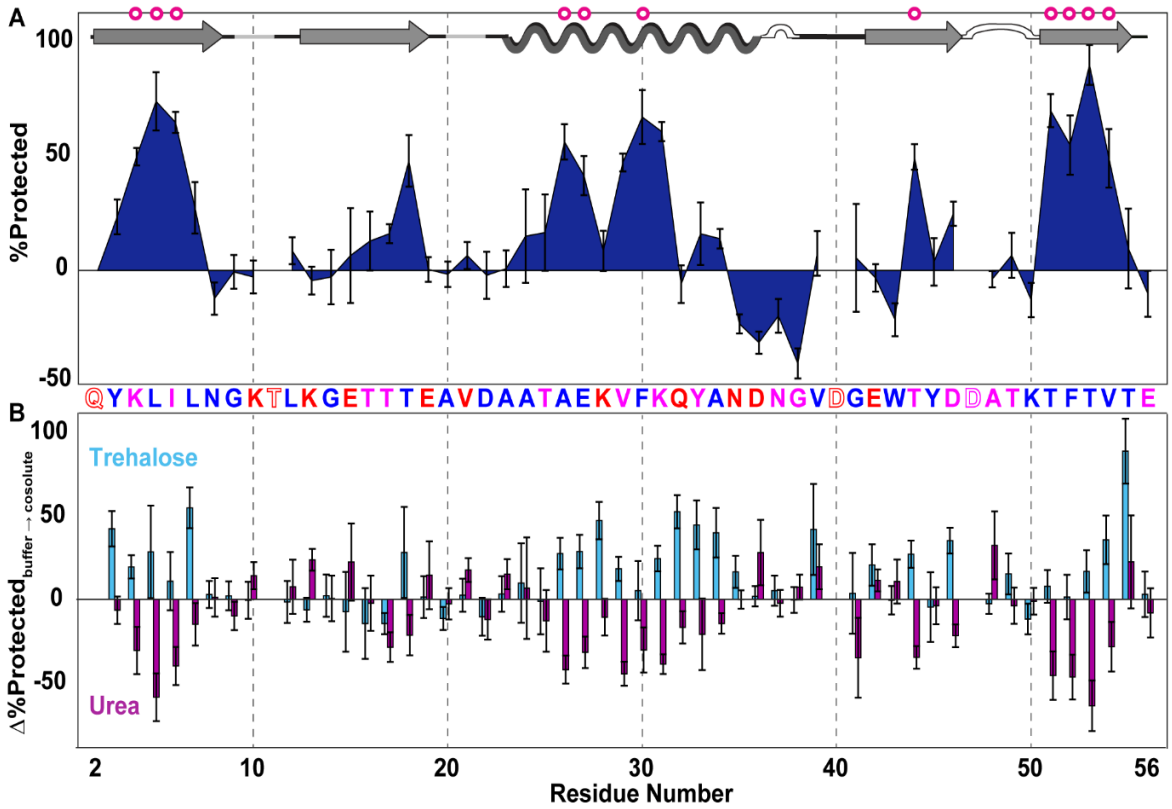


Figure 2.3. Dry-state protection of GB1 freeze-dried alone and in the presence of cosolutes. (A) LOVE profile of GB1 freeze-dried in 1.5-mM HEPES, pH 6.5. (B) Change in %Protected (%Protected$_{cosolute}$ – %Protected$_{buffer}$) due to freeze-drying in 1.5-mM HEPES plus 20 g/L -trehalose or -urea, pH 6.5. GB1 primary structure is shown between the panels. Letters are colored by the total solvent-accessible surface area of the residues in the solution structure, as predicted by the online web application ProtSA for PDB structure 2QMT (blue, 0 – 50 Å^2; pink 51 – 100 Å^2; red, > 100 Å^2). Open letters indicate residues with undefined dry-state protection because they are 100% quench-labelled. GB1 secondary structures (as defined in PDB entry 2QMT; arrows, β-strands; undulations, helix; white bumps, turns; gray lines, bends) are shown at top, with magenta circles indicating solution global unfolding residues. Error bars represent uncertainty propagated from standard deviations of the mean from triplicate analysis. For the cosolute data, D_0 spectra of GB1 freeze-dried with cosolute were used to calculate QC values (Fig. S2.4).

One region, residues 35-38, exhibits negative %Protected values, indicating that the D_0 sample experiences more quench-labelling than the T_{24} sample. The "quench-label profile" of a sample fully exchanged into D_2O before drying and then resuspended in H_2O quench buffer shows similar quench-labelling in this region (Fig. S2.2), signifying that a solvent isotope effect is not responsible. Moreover, the degree of quench-labelling experienced by a sample exposed to H_2O vapor for 24 h followed by resuspension in D_2O quench buffer is the same as that of the D_0 sample

(Fig. S2.2). These data suggest that exchange with D_2O vapor in the dry state is the primary source of the quench-labelling artifact.

The nuances of quench-labelling are not fully understood, but we suspect that interactions with D_2O vapor cause this region to fold into an alternative H-bonded conformation in the dry state, thus preventing quench-labelling. This explanation is consistent with the observations that these residues possess small, positive %Protected values before quench correction (Fig. S2.5) and that this region of GB1 can adopt an alternative conformation in solution.[40]

In addition to providing insight into dry protein structure, LOVE NMR can reveal residue-level effects of drying in the presence of cosolutes. We quantified the change in %Protected due to freeze-drying GB1 with 20 g/L -urea or -trehalose by subtracting the quench-corrected LOVE profile of GB1 dried in buffer from that of GB1 dried in buffer plus cosolute (Fig. 2.3B). Drying with urea, an osmolyte that destabilizes proteins in solution via preferential interactions with protein backbone,[41-42] decreases dry-state protection in general, with global unfolding residues experiencing the largest effect (Fig. 2.3B). Unexpectedly, there are a few residues near the termini of secondary structures that witness a small but significant protective effect from drying with urea.

Drying with trehalose, an osmolyte that is thought to increase solution stability via preferential exclusion from protein backbone,[43] increases dry-state protection of global unfolders but exerts a more pronounced effect on residues in regions that are solvent-exposed or near the termini of secondary structures (e.g. K28 & Q32, L7 & T55). The observation that a denaturant generally decreases protection and a

stabilizer increases protection provides further confidence that LOVE NMR data reflect the presence of protein structure in the dry state.

Comparing dry-state protection to solution stability.

To understand how dry-state protection relates to solution stability, we measured solution HDX in buffer and in buffer plus 100 g/L -urea or -trehalose and quantified the opening free energies ($\Delta G°'_{op}$, Table S2.4).

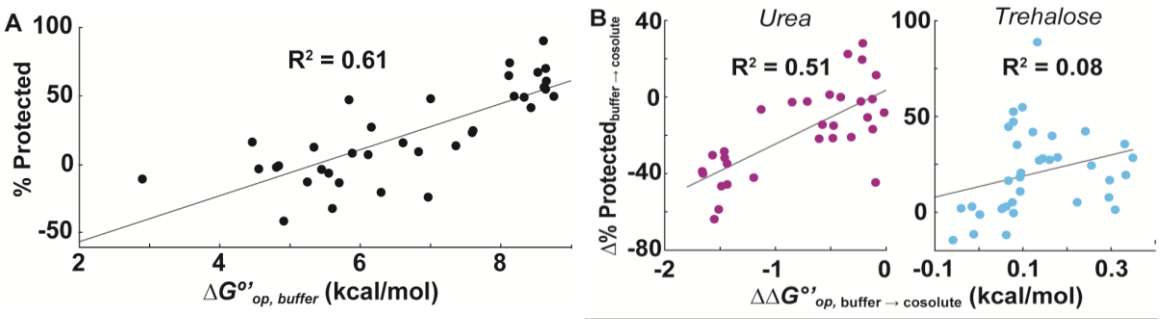


Figure 2.4. Correlations of solution- and dry- state protection. A) %Protected after freeze-drying in buffer vs. average opening free energy ($\Delta G°'_{op}$) in buffer for 34 GB1 residues. B) Average Δ%Protected due to drying with 20 g/L urea or trehalose vs. change in opening free energy ($\Delta\Delta G°'_{op}$) due to the presence of 100 g/L of the same cosolute ($\Delta\Delta G°'_{op} = \Delta G°'_{op, cosolute} - \Delta G°'_{op, buffer}$). $\Delta G°'_{op}$- and %Protected values are reported with their uncertainties in Tables S2.4 and S2.5, respectively.

For GB1 dried in buffer alone, %Protected correlates positively with $\Delta G°'_{op}$ in solution (p-value <.00001; Fig. 2.4A), indicating that native structure is a major source of dry-state protection. Intermolecular H-bonds with opening free energies <6 kcal/mol experience little to no dry-state protection, suggesting that freeze-drying may lead to complete loss of native structure in those regions and highlighting the importance of water to the maintenance of secondary structure.[44]

Comparing the urea-induced change in %Protected to the change in opening free energy ($\Delta\Delta G°'_{op, buffer \rightarrow urea}$) in solution also reveals a positive correlation (p-value <0.00001, Fig. 2.4B). This correlation corroborates the idea that the regional stability of a protein

before drying influences which conformations become trapped in the dry state. Notably, the correlation between solution- and dry-state- protection by trehalose is not significant (p-value >0.10, Fig. 2.4B); this lack of correlation indicates that increased solution stability is not the primary source of additional dehydration protection and suggests that the protective mechanism exerted by trehalose during freeze-drying differs substantially from that in solution.

DISCUSSION

The observation that trehalose seems to target different residues in the solution- and dry- states is consistent with the observation that trehalose is one of the few osmolytes that, in addition to protecting organisms from milder osmotic stresses, also protects against desiccation.[45] It therefore makes sense that the mechanism of protection exerted by trehalose in dehydrating conditions would differ from that which it and other osmolytes use in solution.

There are two main ideas about how trehalose and similar sugars protect proteins from dehydration-induced damage: vitrification and water-replacment.[46-47] The vitrification hypothesis posits that the sugar confines the protein in a glassy matrix, preventing large motions such as global unfolding. The water-replacement hypothesis proposes that the sugar maintains native structure by replacing the stabilizing H-bonds usually provided by water.If trehalose was protecting GB1 via vitrification, we would expect global-unfolding residues to benefit most, which is generally not the case (Fig. 2.3B). Instead, trehalose-induced protection is largest at or near the termini of secondary structures, or in areas with high solvent-accessibility, such as at the center of the α-helix (Figs. 2.3B and 2.5A) – an observation that is more consistent with the water-replacement hypothesis.

However, if the stabilization of native structure was the sole source of trehalose-induced protection, as suggested by the water-replacement hypothesis, we would expect

native H-bond donor-acceptor pairs (double arrows in Fig. 2.5B) to experience similar degrees of trehalose-induced protection. Yet there are several instances where such pairs exhibit a large difference in $\Delta\%Protected_{buffer \rightarrow trehalose}$ (e.g. E42/T55, L7/G14), indicating that some protection may arise from H-bonding outside of native contacts.

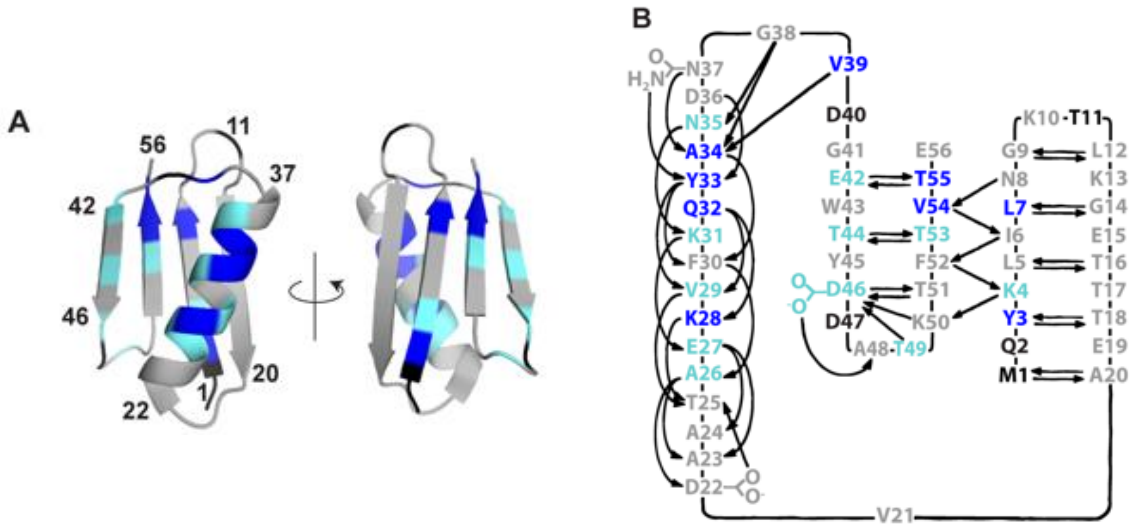


Figure 2.5. Trehalose protection mapped onto GB1 structure. A) Cartoon representation of GB1 (PDB 2QMT) with residues colored by $\Delta\%Protected_{buffer \rightarrow trehalose}$. B) Native H-bonding pattern of GB1. Residues colored dark blue exhibit a Δ%Protected >1 standard deviation (S.D.) above average; light-blue, within 1 S.D. and not within uncertainty of zero; gray, <1 S.D. or within uncertainty of zero; black, undefined protection (i.e. 100% quench-labelled).

Taken together, these data suggest that protection by trehalose results from a complex combination of native structure preservation and blocking of exchange, the latter perhaps arising from H-bonding between protein backbone and sugar hydroxyls. The multi-faceted nature of vapor-exchange protection by trehalose is confirmed by the observation that plotting $\Delta\%Protected_{buffer \rightarrow trehalose}$ values as a function of ~40 residue properties, including solvent-accessible surface area and

amino acid sidechain transfer free energies, did not yield a strong linear correlation (all R^2 <0.16, Table S2.6).

CONCLUSIONS

In summary, LOVE NMR data on the dehydrated structure of the model protein GB1 confirm the notion that a protein's structure in the dry state is heavily influenced by its solution stability, but also demonstrate that models of cosolute-mediated dehydration protection require refinement. The application of high-resolution methods such as ssHDX-MS and LOVE NMR to a broad range of proteins will enable more nuanced models of dehydration protection to be proposed and tested. Such models can in turn be used to streamline new formulations for freeze-dried protein-based therapeutics, reducing costs and increasing accessibility to these life-saving medicines.

SUPPLEMENTAL INFORMATION

Supplementary figures

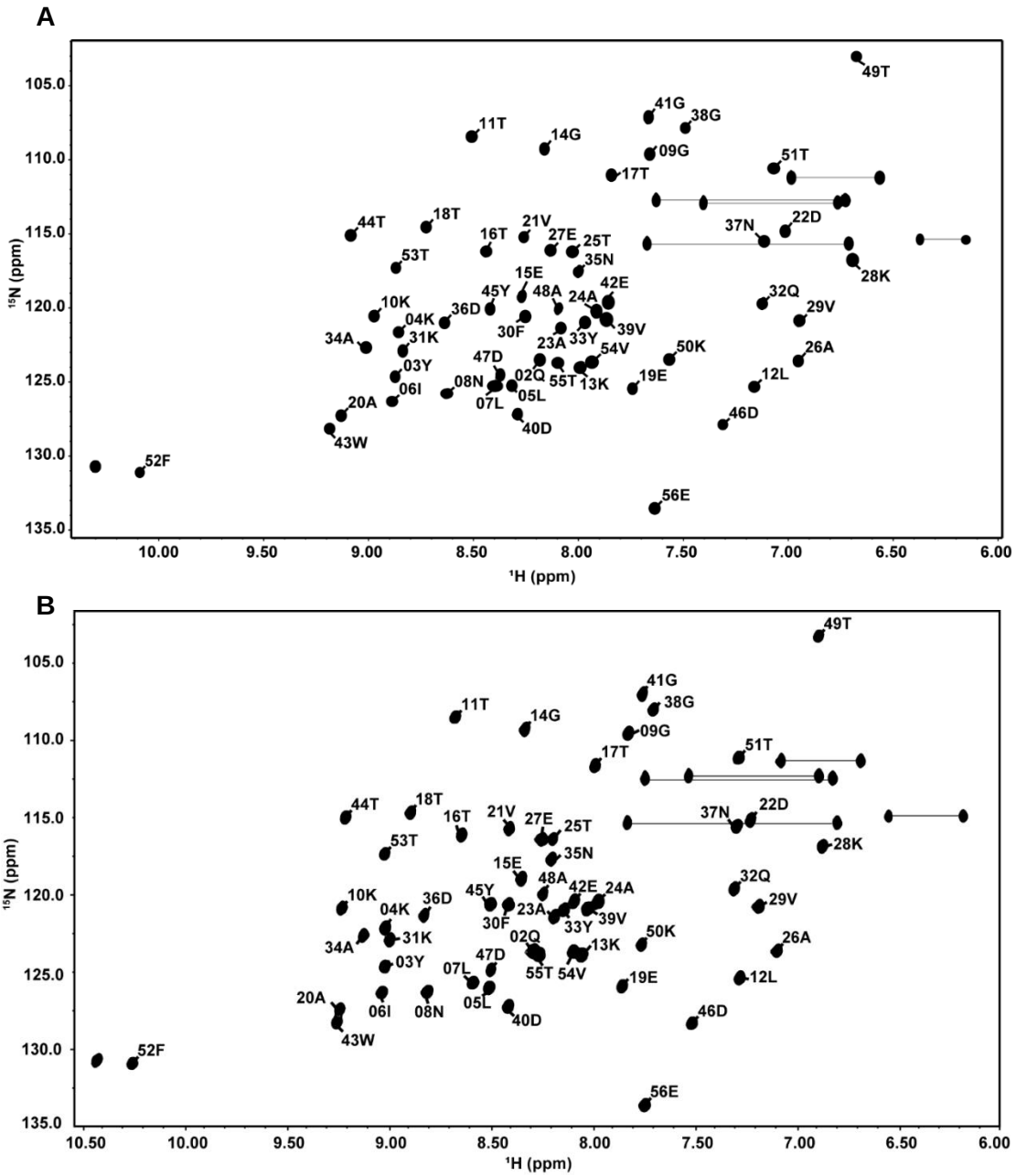


Figure S2.1. Backbone resonance assignments of GB1 T2Q at A) 4 °C in 100-mM citrate buffer (10% D_2O), pH 4.5 and B) 22 °C in 7.5 mM HEPES (10% D_2O) pH 7.5.

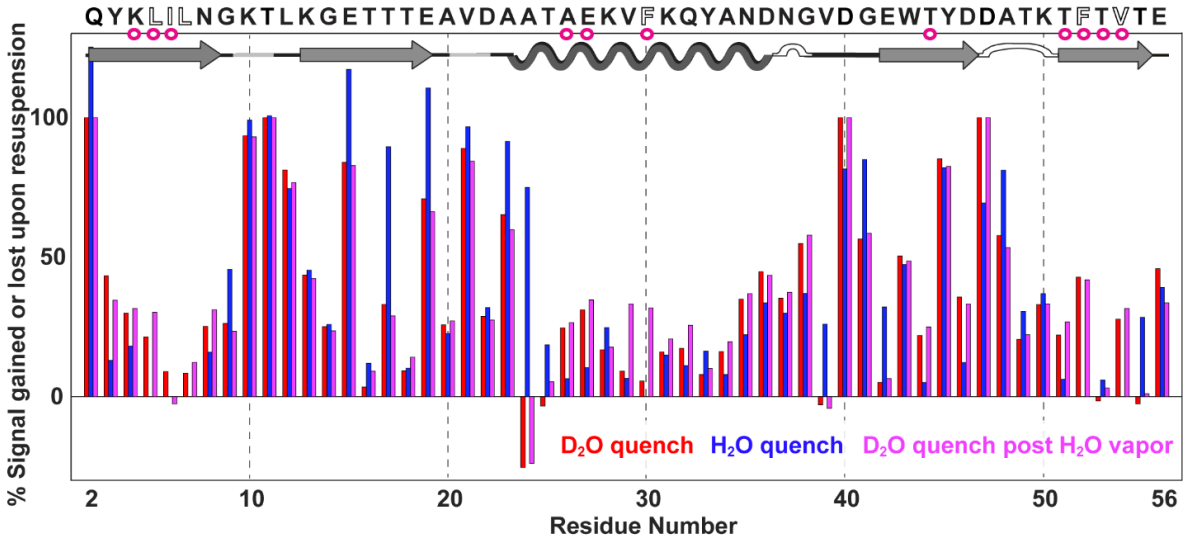


Figure S2.2. Quench-label profiles of GB1 exposed to different vapor and buffer conditions. The %quench-labelling experienced by a protonated (unexchanged) protein upon resuspension in D_2O-based quench buffer ("D_2O quench" values) is equivalent to the %signal lost upon resuspension, acquired and calculated as described in the main text. The %quench-labelling experienced by a sample of freeze-dried GB1 resuspended in D_2O quench buffer after 24-h incubation in a H_2O chamber at 75% relative humidity (achieved using a saturated NaCl solution – "D_2O quench post H_2O vapor" values) was determined in the same manner. To determine the %signal gained by a deuterium-exchanged protein upon resuspension in a H_2O-quench buffer ("H_2O quench" values), an aliquot of GB1 was twice exchanged into D_2O, lyophilized for 24 h, and resuspended in cold H_2O quench buffer before immediate HSQC spectrum acquisition; peak volumes were divided by the corresponding peak volumes of a fully protonated, non-exchanged sample (V_{T0}) and multiplied by 100%. The primary and secondary structures of GB1 are plotted at the top, with open letters indicating residues for which no signal is detected for the H_2O quenched sample. Measurements were made once for each condition.

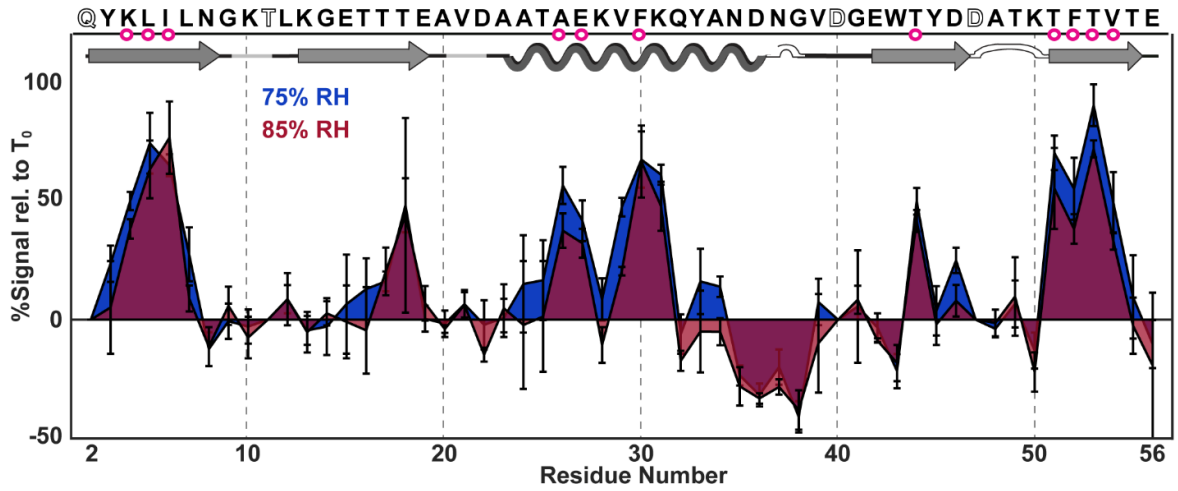


Figure S2.3. %Signal remaining as a function of GB1 residue after exposure to 75% - and 85%-RH (D_2O) for 24 h. GB1 was lyophilized in 1.5 mM HEPES, pH 6.5 for 24 h and then placed at 75- or 85- % RH (D_2O) for 24 h before spectrum acquisition. Data were corrected for quench-labelling and normalized as described in Fig. 2.1. Primary and secondary structures are plotted at top, with pink circles indicating global unfolding residues and open letters indicating residues with undefined %signal remaining because they are 100% quench-labelled. Error bars represent uncertainty propagated from standard deviations of the mean from triplicate analysis.

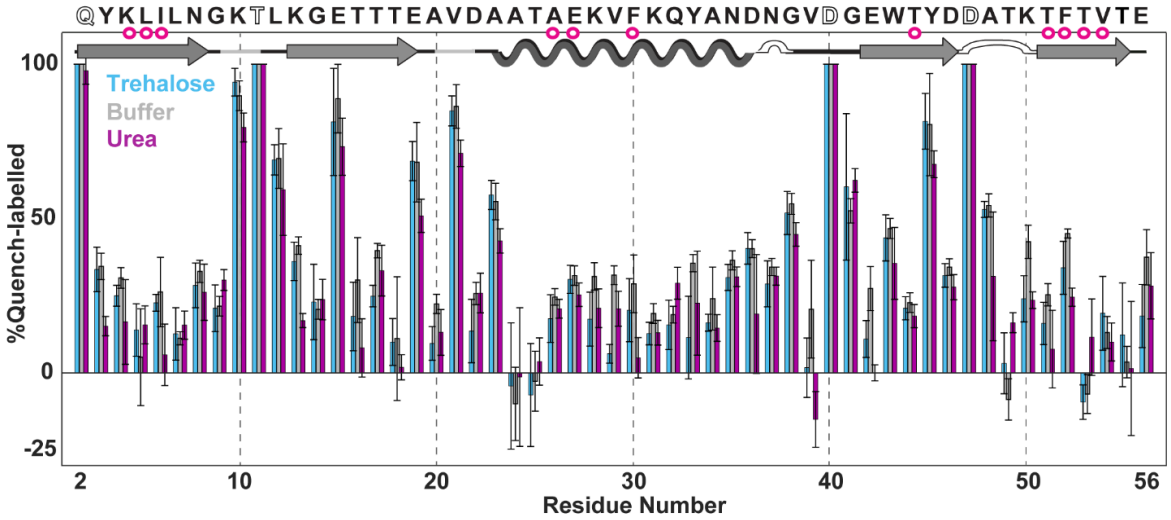


Figure S2.4. %Quench-labelled as a function of residue and cosolute plotted beneath a secondary structure map of GB1. ^{15}N-enriched GB1 was lyophilized in 1.5-mM HEPES + 20 g/L trehalose or urea and analyzed (Fig. 2.1). Percent quench-labelled was calculated by dividing the quench correction value for each residue by the respective T_0 volume (V_{T0}). Error bars represent uncertainties propagated from standard deviations of the mean from triplicate analysis. Other details are given in the caption to Figure 2.3.

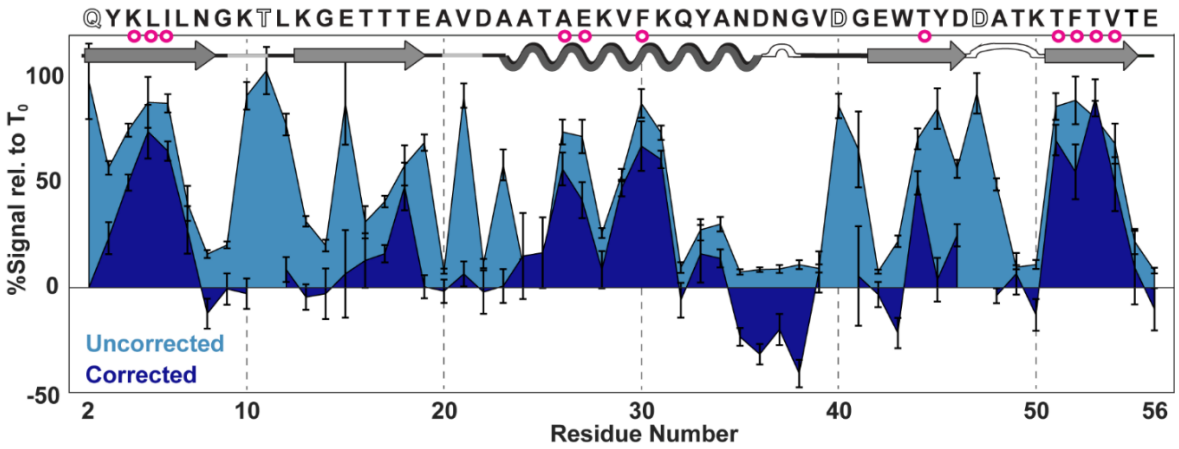


Figure S2.5. LOVE profiles of GB1 before and after applying the quench-correction , as described in the main text. Error bars represent uncertainty propagated from standard deviations of the mean from triplicate analysis. Other details are given in the caption to Figure 2.3.

Supplementary tables

Table S2.1. ^{15}N and ^{1}H chemical shifts of GB1 backbone amides at pH 4.5, 4 °C (blue, left) and pH 7.5, 22 °C (pink, right). .

Residue	δ ^{15}N (ppm)	δ ^{1}H (ppm)	δ ^{15}N (ppm)	δ ^{1}H (ppm)
02Q	123.5	8.2	123.6	8.3
03Y	124.6	8.9	124.6	9.0
04K	121.6	8.9	122.1	9.0
05L	125.2	8.3	126.0	8.5
06I	126.3	8.9	126.3	9.0
07L	125.3	8.4	125.7	8.6
08N	125.8	8.6	126.3	8.8
09G	109.6	7.7	109.5	7.8
10K	120.5	9.0	120.9	9.2
11T	108.4	8.5	108.5	8.7
12L	125.3	7.2	125.4	7.3
13K	124.0	8.0	123.9	8.0
14G	109.3	8.2	109.3	8.3
15E	119.2	8.3	119.0	8.3
16T	116.2	8.4	116.1	8.6
17T	111.0	7.8	111.6	8.0
18T	114.6	8.7	114.7	8.9
19E	125.4	7.7	125.9	7.9
20A	127.3	9.1	127.4	9.2
21V	115.2	8.3	115.7	8.4
22D	114.8	7.0	115.2	7.2
23A	121.4	8.1	121.4	8.2
24A	120.2	7.9	120.5	8.0
25T	116.2	8.0	116.4	8.2
26A	123.6	7.0	123.6	7.1
27E	116.1	8.1	116.4	8.2
28K	116.7	6.7	116.9	6.9
29V	120.8	6.9	120.8	7.2
30F	120.6	8.3	120.6	8.4
31K	122.9	8.8	122.9	9.0
32Q	119.7	7.1	119.6	7.3
33Y	121.0	8.0	121.0	8.1
34A	122.7	9.0	122.6	9.1
35N	117.6	8.0	117.7	8.2
36D	121.0	8.6	121.4	8.8
37N	115.5	7.1	115.6	7.3
38G	107.8	7.5	108.0	7.7
39V	120.8	7.9	120.9	8.0
40D	127.2	8.3	127.2	8.4
41G	107.1	7.7	107.0	7.8
42E	119.6	7.9	120.4	8.1
43W	128.2	9.2	128.3	9.2
44T	115.1	9.1	115.0	9.2
45Y	120.1	8.4	120.6	8.5
46D	127.9	7.3	128.3	7.5
47D	124.5	8.4	124.8	8.5
48A	120.0	8.1	120.0	8.2
49T	103.0	6.7	103.3	6.9
50K	123.5	7.6	123.2	7.8
51T	110.6	7.1	111.1	7.3

52F	131.1	10.1	130.9	10.3
53T	117.3	8.9	117.3	9.0
54V	123.7	7.9	123.7	8.1
55T	123.7	8.1	123.8	8.3
56E	133.5	7.6	133.6	7.7

Table S2.2. Signal change of T_{24} sample due to solution HDX during LOVE NMR spectrum acquisition*

.

Residue[†]	%Signal 0 h[‡]	%Signal 12 h[‡]	%Change in 12 h[‡]	%Signal change during expt[§]
N8	19	60	41	2
G9	27	78	51	6
K13	42	84	42	10
G14	17	51	34	2
T17	30	47	18	13
E19	83	96	12	24
D22	14	51	37	2
A23	71	92	21	18
A24	23	88	65	4
D36	12	35	23	1
N37	11	25	14	1
G38	12	47	35	2
G41	68	79	11	26
W43	31	85	54	4
A48	73	94	22	24
T49	13	56	43	2
E56	11	44	33	2

**Determined by taking serial HSQCs of T_{24} over 12 h period, at pH 4.5 and 4 °C. The majority of signal gain by T_{24} is accounted for by the quench correction, because D_0 witnesses a similar degree of signal loss due to solution HDX.*

[†]Only residues that witness a change in signal >10% in 12 h are shown

[‡]Percent is defined relative to average T_0 signal for that residue, and is not corrected for quench-labelling

[§]Calculated from the experimentally-determined observed rate constant (k_{obs}) obtained by fitting normalized data to the equation $\%Signal = 100*(1-e^{-kobs * t})$, where t is time between resuspension and completion of spectrum acquisition (30 minutes for all LOVE NMR experiments performed for this manuscript)

Table S2.3. Water content of freeze-dried protein samples. .

Condition	Initial mass[†] (mg)	Final mass[†] (mg)	%H_2O at 0 h (w/w)	%(H_2O + D_2O) at 72 h (w/w)	%SASA covered by H_2O + D_2O at 72 h[§]
Buffer	2.6	2.8	10	19	40
+Urea	15.6	16.0	2	5	60
+Trehalose	16.1	17.4	5	12	170

[†]Assumes no sample lost during lyophilization or transfer to TGA instrument.

$$Mass = Total\ mass\ measured\ by\ TGA\ x \frac{Theoretical\ dry\ mass}{Measured\ dry\ mass}$$

[§]Calculated by multiplying the molar ratio of bound H_2O to protein by the average amount of protein surface covered by a water molecule (20 Å^2)[48] and dividing by the surface area of the native solution structure of GB1 (3727 Å^2), as determined by the PyMOL get_area function for PDB 2QMT.

Table S2.4. Average %Protected values of GB1 lyophilized in buffer only or with cosolutes.

.

Residue	%Protected ± SEM†		
	Buffer	+ 20 g/L trehalose	+ 20 g/L urea
02Q	N/A	N/A	N/A
03Y	20 ± 10	*66 ± 7	17 ± 3
04K	50 ± 4	*69 ± 6	20 ± 10
05L	70 ± 10	*100 ± 30	16 ± 7
06I	65 ± 5	*80 ± 20	30 ± 10
07L	30 ± 10	82 ± 5	13 ± 5
08N	-12 ± 7	-9 ± 4	-11 ± 9
09G	-1 ± 7	1 ± 4	‡-11 ± 4
10K	-3 ± 7	-3 ± 8	‡11 ± 4
11T	N/A	N/A	N/A
12L	9 ± 6	7 ± 10	20 ± 20
13K	-5 ± 6	-11 ± 4	‡19 ± 2
14G	0 ± 10	-1 ± 4	‡-2 ± 7
15E	10 ± 20	0 ± 10	‡30 ± 10
16T	10 ± 10	0 ± 12	‡10 ± 10
17T	16 ± 4	2 ± 5	-13 ± 8
18T	50 ± 10	*80 ± 30	27 ± 4
19E	0 ± 5	2 ± 7	‡20 ± 20
20A	-2 ± 6	-13 ± 4	‡-4 ± 8
21V	7 ± 6	9 ± 7	‡24 ± 4
22D	0 ± 10	-13 ± 5	‡-14 ± 6
23A	1 ± 8	4 ± 6	‡16 ± 4
24A	20 ± 20	*30 ± 10	‡20 ± 20
25T	20 ± 20	15 ± 6	‡4 ± 8
26A	60 ± 10	*84 ± 5	14 ± 3
27E	40 ± 10	*70 ± 5	10 ± 4
28K	9 ± 9	56 ± 7	‡-2 ± 7
29V	47 ± 4	*66 ± 6	‡3 ± 6
30F	70 ± 10	*70 ± 10	37 ± 7
31K	61 ± 5	*85 ± 7	22 ± 4
32Q	-6 ± 8	46 ± 6	‡-23 ± 5
33Y	20 ± 10	*61 ± 5	0 ± 20
34A	14 ± 4	50 ± 20	-1 ± 4
35N	-23 ± 4	-7 ± 9	-24 ± 4
36D	-32 ± 5	-30 ± 4	0 ± 20
37N	-20 ± 8	-15 ± 5	-22 ± 3
38G	-41 ± 7	-41 ± 4	-34 ± 4
39V	10 ± 10	50 ± 30	27 ± 9
40D	N/A	N/A	N/A
41G	0 ± 20	9 ± 5	‡-30 ± 3
42E	-3 ± 6	20 ± 10	8 ± 3
43W	-22 ± 7	-22 ± 5	-10 ± 10
44T	50 ± 6	*77 ± 6	15 ± 4
45Y	0 ± 10	0 ± 10	‡0 ± 4
46D	25 ± 5	*60 ± 6	‡3 ± 4
47D	N/A	N/A	N/A
48A	-4 ± 4	-6 ± 5	30 ± 20
49T	10 ± 10	22 ± 6	3 ± 5
50K	-13 ± 8	-25 ± 3	-14 ± 3
51T	70 ± 9	*78 ± 7	30 ± 10
52F	60 ± 10	*56 ± 3	9 ± 3
53T	90 ± 10	*107 ± 8	30 ± 10

54V	50 ± 10	*85 ± 6	21 ± 7
55T	10 ± 20	*98 ± 9	‡30 ± 20
56E	-10 ± 10	-7 ± 8	‡-20 ± 10

*Residues that met their maximum observable value of %Protected (Uncorrected T_{24} = 100% of uncorrected T_0).
‡Residues that met "minimum value" of %Protected (Quench-corrected T_{24} from buffer OR urea dataset within error of 0% of T_0).
†Standard error of the mean propagated from standard errors of the mean from triplicate analysis of each dataset used to calculate %Protected (T_0, D_0, and T_{24}).

Table S2.5. Average free energies of opening in solution (ΔG_{op}°')[§]. .

Residue	ΔG_{op}°' ± SEM‡ (kcal/mol) Buffer	+ 100 g/L trehalose	+ 100 g/L urea
02Q	N/A	N/A	N/A
03Y	7.59 ± 0.01	7.83 ± 0.02	6.46 ± 0.01
04K	8.19 ± 0.01	8.52 ± 0.01	N/A
05L	8.13 ± 0.00	8.48 ± 0.01	6.61 ± 0.01
06I	8.11 ± 0.02	8.21 ± 0.02	6.46 ± 0.03
07L	6.16 ± 0.04	6.26 ± 0.03	5.68 ± 0.01
08N	5.25 ± 0.01	5.23 ± 0.02	4.74 ± 0.01
09G	4.84 ± 0.03	4.80 ± 0.02**	N/A
10K	N/A	N/A	N/A
11T	N/A	N/A	N/A
12L	N/A	N/A	N/A
13K	N/A	N/A	N/A
14G	4.55 ± 0.03	4.61 ± 0.02**	N/A
15E	N/A	N/A	N/A
16T	5.34 ± 0.02	5.28 ± 0.08	4.627 ± 0.004
17T	N/A	N/A	N/A
18T	7.00 ± 0.04	7.15 ± 0.02	6.52 ± 0.01
19E	N/A	N/A	N/A
20A	4.81 ± 0.01	4.79 ± 0.04	3.96 ± .02**
21V	N/A	N/A	N/A
22D	N/A	N/A	N/A
23A	N/A	N/A	N/A
24A	N/A	N/A	N/A
25T	4.46 ± 0.02	4.46 ± 0.01	N/A
26A	8.62 ± 0.01	8.78 ± 0.00	7.42 ± 0.01
27E	8.43 ± 0.01	8.61 ± 0.00	6.97 ± 0.01
28K	5.89 ± 0.03	5.96 ± 0.03	5.72 ± 0.01
29V	5.84 ± 0.03	5.93 ± 0.04	5.74 ± 0.01
30F	8.53 ± 0.01	8.75 ± 0.00	6.96 ± 0.01
31K	8.65 ± 0.00	8.90 ± 0.01	6.99 ± 0.01
32Q	5.55 ± 0.04	5.63 ± 0.03	5.43 ± 0.00
33Y	6.61 ± 0.04	6.67 ± 0.02	6.29 ± 0.01
34A	7.36 ± 0.02	7.526 ± 0.003	6.78 ± 0.01
35N	6.97 ± 0.04	7.03 ± 0.05	6.55 ± 0.01
36D	5.60 ± 0.04	5.65 ± 0.02	5.39 ± 0.02
37N	6.30 ± 0.03	6.37 ± 0.03	6.07 ± 0.02
38G	4.91*	4.99 ± 0.04**	N/A
39V	6.11 ± 0.03	6.23 ± 0.02	5.89 ± 0.01
40D	3.47 ± 0.03	3.49 ± 0.03**	N/A
41G	N/A	N/A	N/A
42E	5.45 ± 0.04	5.54 ± 0.03	5.36 ± 0.01
43W	N/A	N/A	N/A

44T	8.755 ± 0.003	8.89 ± 0.01	7.32 ± 0.01
45Y	N/A	N/A	N/A
46D	7.60 ± 0.01	7.69 ± 0.01	7.00 ± 0.01
47D	N/A	N/A	N/A
48A	N/A	N/A	N/A
49T	N/A	N/A	N/A
50K	5.70 ± 0.03	5.76 ± 0.03	5.58 ± 0.01
51T	8.64 ± 0.01	8.93 ± 0.00	7.20 ± 0.01
52F	8.64 ± 0.01	8.95 ± 0.01	7.15 ± 0.01
53T	8.61 ± 0.00	8.90 ± 0.01	7.05 ± 0.01
54V	8.33 ± 0.00	8.66 ± 0.00	6.87 ± 0.02
55T	6.83 ± 0.04	6.96 ± 0.03	6.48 ± 0.01
56E	2.90 ± 0.03	2.96 ± 0.03	2.88 ± 0.01

Footnotes

§ ΔG_{op}°′ values from NMR-detected solution H-D exchange at 22 °C, pH 7.5.

‡ Standard error of the mean from triplicate analysis. Stars mark residues for which only *one or **two measurements were made.

Table S2.6. Correlations between residue-specific predictors and Δ%Protected$_{buffer→cosolute}$.

Predictor†	**Δ%Protected$_{buffer→trehalose}$**		**Δ%Protected$_{buffer→urea}$**	
	Coeff.	***R^2***	***Coeff.***	***R^2***
$SASA_{backbone}$	-	0.15	+	0.25
$CASA_{backbone}$	-	0.14	+	0.25
Δ $SASA_{U-->F}$	-	0.08	+	0.25
# H-bonds	+	0.07	-	0.14
# H-bonds with water	-	0.06	+	0.08
# Weak H-bonds with water	-	0.06	+	0.01
k_{int} (pH 4.5)	-	0.05	+	0.03
# Highly frustrated contacts	+	0.04	-	0.00
# Ionic interactions	-	0.03	+	0.00
# Minimally frustrated contacts	+	0.03	-	0.03
# Aromatic contacts	-	0.02	-	0.03
# Neutral frustrated contacts	+	0.02	-	0.02
$CASA_{apolar}$	-	0.02	+	0.12
k_{int} (pH 7.0)	-	0.02	+	0.00
# Weak H-bonds	+	0.02	-	0.08
$CASA_{total}$	-	0.02	+	0.22
$SASA_{total}$	-	0.02	+	0.21
$\Delta G^\circ_{tr,\ Octanol\ -->\ H2O}$	+	0.02	-	0.06
$SASA_{polar}$	-	0.01	+	0.16
# Carbonyl interactions	+	0.01	-	0.18
SASA ratio (polar:nonpolar)	-	0.01	+	0.03
$SASA_{apolar}$	-	0.01	+	0.11
$\Delta G^\circ_{tr,\ Cyclohexane\ -->\ H2O}$	+	0.01	-	<.01
$\Delta G^\circ_{tr,\ NMA\ -->\ H2O}$ (normalized)	-	0.01	+	<.01
$\Delta G^\circ_{tr,\ EtOH\ -->\ H2O}$ (normalized)	-	< 0.01	+	<.01
$\Delta G^\circ_{tr,\ EtOH\ -->\ H2O}$	+	< 0.01	-	0.07
$\Delta G^\circ_{tr,\ vapor\ -->\ H2O}$ (normalized)	+	< 0.01	+	<.01
$\Delta G^\circ_{tr,\ NMA\ -->\ H2O}$	+	< 0.01	-	0.03
Net frustration parameter	+	< 0.01	-	0.02
# Hydrophobic contacts	+	< 0.01	-	0.09
$\Delta G^\circ_{tr,\ Cyclohexane\ -->\ H2O}$ (normalized)	-	< 0.01	-	0.04
$CASA_{polar}$	-	< 0.01	+	0.15
$CASA_{sidechain}$	-	< 0.01	+	0.10
$\Delta G^\circ_{tr,\ vapor\ -->\ H2O}$	-	< 0.01	+	0.01
$\Delta G^\circ_{tr,\ Octanol\ -->\ H2O}$ (normalized)	+	< 0.01	+	<.01
$SASA_{sidechain}$	+	< 0.01	+	0.09

†Predictor values are calculated using the native structure of GB1 (PDB 2QMT), with the exception of non-normalized transfer free energies (ΔG°_{tr}). The specific types of intra- and inter-molecular contacts (H-bonds, carbonyl interactions, etc.) were obtained from the Arpeggio server.[49] Solvent-accessible

surface areas (SASA), cosolute-accessible surface areas (CASA), and change in surface area upon folding ($\Delta SASA_{U\text{-}>F}$) were obtained from the ProtSA server using a probe radius of 1.40-, 4.00-, and 1.85- Å for water, trehalose, and urea, respectively.[50] Frustration parameters are from the Frustratometer server.[51] Transfer free energies of amino acid side chains are from.[52-53] Normalized transfer free energies were calculated by multiplying literature values by the fraction of sidechain area accessible to solvent in the native structure.

REFERENCES

1. Kauzmann, W., Some Factors in the Interpretation of Protein Denaturation. *Adv. Protein Chem.* **1959,** *14*, 1-63.

2. Bellissent-Funel, M. C.; Hassanali, A.; Havenith, M.; Henchman, R.; Pohl, P.; Sterpone, F.; van der Spoel, D.; Xu, Y.; Garcia, A. E., Water Determines the Structure and Dynamics of Proteins. *Chem. Rev.* **2016,** *116* (13), 7673-97.

3. Ball, P., Water is an active matrix of life for cell and molecular biology. *Proc. Natl. Acad. Sci. U.S.A.* **2017,** *114* (51), 13327-13335.

4. Moorthy, B. S.; Iyer, L. K.; Topp, E. M., Characterizing Protein Structure, Dynamics and Conformation in Lyophilized Solids. *Curr. Pharm. Des.* **2015,** *21* (40), 5845-5853.

5. Ohtake, S.; Kita, Y.; Arakawa, T., Interactions of formulation excipients with proteins in solution and in the dried state. *Adv. Drug Delivery Rev.* **2011,** *63* (13), 1053-73.

6. Piszkiewicz, S.; Pielak, G. J., Protecting Enzymes from Stress-Induced Inactivation. *Biochemistry* **2019,** *58* (37), 3825-3833.

7. Crommelin, D. J. A.; Hawe, A.; Jiskoot, W., Formulation of Biologics Including Biopharmaceutical Considerations. In *Pharmaceutical Biotechnology: Fundamentals and Applications*, 5th ed.; Crommelin, D. J. A.; Sindelar, R. D.; Meibohm, B., Eds. Springer International Publishing: Cham, 2019; pp 83-103.

8. Pandya, A.; Howard, M. J.; Zloh, M.; Dalby, P. A., An Evaluation of the Potential of NMR Spectroscopy and Computational Modelling Methods to Inform Biopharmaceutical Formulations. *Pharmaceutics* **2018,** *10* (4), 165.

9. Hill, A. B.; Kilgore, C.; McGlynn, M.; Jones, C. H., Improving global vaccine accessibility. *Curr. Opin. Biotechnol.* **2016,** *42*, 67-73.

10. Li, Y.; Williams, T. D.; Schowen, R. L.; Topp, E. M., Characterizing protein structure in amorphous solids using hydrogen/deuterium exchange with mass spectrometry. *Anal. Biochem.* **2007,** *366* (1), 18-28.

11. Moorthy, B. S.; Schultz, S. G.; Kim, S. G.; Topp, E. M., Predicting protein aggregation during storage in lyophilized solids using solid state amide hydrogen/deuterium exchange with mass spectrometric analysis (ssHDX-MS). *Mol. Pharm.* **2014,** *11* (6), 1869-1879.

12. Moorthy, B. S.; Zarraga, I. E.; Kumar, L.; Walters, B. T.; Goldbach, P.; Topp, E. M.; Allmendinger, A., Solid-State Hydrogen–Deuterium Exchange Mass Spectrometry: Correlation of Deuterium Uptake and Long-Term Stability of Lyophilized Monoclonal Antibody Formulations. *Mol. Pharm.* **2018,** *15* (1), 1-11.

13. Desai, U. R.; Osterhout, J. J.; Klibanov, A. M., Protein Structure in the Lyophilized State: A Hydrogen Isotope Exchange/NMR Study with Bovine Pancreatic Trypsin Inhibitor. *J. Am. Chem. Soc.* **1994,** *116* (21), 9420-9422.

14. Hvidt, A.; Nielsen, S. O., Hydrogen exchange in proteins. *Adv. Prot. Chem.* **1966,** *21*, 287-386.

15. Englander, S. W.; Kallenbach, N. R., Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys.* **1983,** *16* (4), 521-655.

16. Percy, A. J.; Rey, M.; Burns, K. M.; Schriemer, D. C., Probing protein interactions with hydrogen/deuterium exchange and mass spectrometry-a review. *Anal. Chim. Acta* **2012,** *721*, 7-21.

17. Glasoe, P. K.; Long, F. A., Use of Glass Electrodes to Measure Acidities in Deuterium Oxide. *J. Phys. Chem.* **1960,** *64* (1), 188-190.

18. Smith, C. K.; Withka, J. M.; Regan, L., A thermodynamic scale for the beta-sheet forming tendencies of the amino acids. *Biochemistry* **1994,** *33* (18), 5510-7.

19. Rockland, L. B., Saturated Salt Solutions for Static Control of Relative Humidity between 5° and 40° C. *Anal. Chem.* **1960,** *32* (10), 1375-1376.

20. Jakli, G.; Van Hook, W. A., Vapor pressure of heavy water at 283-363 K. *J. Chem. Eng. Data* **1981,** *26* (3), 243-245.

21. Monteith, W. B.; Pielak, G. J., Residue level quantification of protein stability in living cells. *Proc. Natl. Acad. Sci. U.S.A.* **2014,** *111* (31), 11335-11340.

22. Correction for Monteith and Pielak, Residue level quantification of protein stability in living cells. *Proc. Natl. Acad. Sci. U.S.A.* **2015,** *112* (50), E7031.

23. Gill, S. C.; von Hippel, P. H., Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **1989,** *182* (2), 319-26.

24. Schanda, P.; Van Melckebeke, H.; Brutscher, B., Speeding Up Three-Dimensional Protein NMR Experiments to a Few Minutes. *J. Am. Chem. Soc.* **2006,** *128* (28), 9042-9043.

25. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995,** *6* (3), 277-93.

26. Johnson, B. A.; Blevins, R. A., NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **1994,** *4* (5), 603-14.

27. Hyberts, S. G.; Takeuchi, K.; Wagner, G., Poisson-Gap Sampling and Forward Maximum Entropy Reconstruction for Enhancing the Resolution and Sensitivity of Protein NMR Data. *J. Am. Chem. Soc.* **2010,** *132* (7), 2145-2147.

28. Miljenović, T. M.; Jia, X.; Mobli, M., Nonuniform Sampling in Biomolecular NMR. In *Modern Magnetic Resonance*, 2nd ed.; Webb, G. A., Ed. Springer International Publishing: Cham, 2017; pp 1-21.

29. Ying, J.; Delaglio, F.; Torchia, D. A.; Bax, A., Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J. Biomol. NMR* **2017,** *68* (2), 101-118.

30. Zhang, Y.-Z. Protein and peptide structure and interactions studied by hydrogen exchange and NMR. PhD thesis. , University of Pennsylvania, Philadelphia, 1995.

31. Miklos, A. C.; Li, C.; Pielak, G. J., Using NMR-Detected Backbone Amide ^{1}H Exchange to Assess Macromolecular Crowding Effects on Globular-Protein Stability. *Methods Enzymol.* **2009,** *466*, 1-18.

32. Joan C. May, E. G., Roscoe M. Wheeler, and Jerry Westy, Determination of residual moisture in freeze-dried viral vaccines: Karl Fischer, gravimetric and

thermogravimetric methodologies. *Journal of Biological Standardization* **1982,** *10*, 249-259.

33. Joan C. May, R. M. W., and Elizabeth Grim, The Gravimetric Method for the Determination of Residual Moisture in Freeze-Dried Biological Products. *Cryobiology* **1989,** *26*, 277-284.

34. Smith, C. K.; Withka, J. M.; Regan, L., A Thermodynamic Scale for the β-Sheet Forming Tendencies of the Amino Acids. *Biochemistry* **1994,** *33* (18), 5510-5517.

35. Morrone, A.; Giri, R.; Toofanny, R. D.; Travaglini-Allocatelli, C.; Brunori, M.; Daggett, V.; Gianni, S., GB1 is not a two-state folder: identification and characterization of an on-pathway intermediate. *Biophys. J.* **2011,** *101* (8), 2053-60.

36. Marmorino, J. L.; Auld, D. S.; Betz, S. F.; Doyle, D. F.; Young, G. B.; Pielak, G. J., Amide proton exchange rates of oxidized and reduced *Saccharomyces cerevisiae* iso-1-cytochrome *c*. *Protein Sci.* **1993,** *2* (11), 1966-74.

37. Kammari, R.; Topp, E. M., Solid-State Hydrogen–Deuterium Exchange Mass Spectrometry (ssHDX-MS) of Lyophilized Poly-d,l-Alanine. *Mol. Pharm.* **2019,** *16* (7), 2935-2946.

38. Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M., A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* **1991,** *253* (5020), 657-61.

39. Orban, J.; Alexander, P.; Bryan, P.; Khare, D., Assessment of stability differences in the protein G B1 and B2 domains from hydrogen-deuterium exchange: comparison with calorimetric data. *Biochemistry* **1995,** *34* (46), 15291-300.

40. Blanco, F. J.; Ortiz, A. R.; Serrano, L., Role of a nonnative interaction in the folding of the protein G B1 domain as inferred from the conformational analysis of the α-helix fragment. *Folding Des.* **1997,** *2* (2), 123-33.

41. Yancey, P. H.; Clark, M. E.; Hand, S. C.; Bowlus, R. D.; Somero, G. N., Living with water stress: evolution of osmolyte systems. *Science* **1982,** *217* (4566), 1214-22.

42. Rydeen, A. E.; Brustad, E. M.; Pielak, G. J., Osmolytes and Protein-Protein Interactions. *J. Am. Chem. Soc.* **2018,** *140* (24), 7441-7444.

43. Xie, G.; Timasheff, S. N., The thermodynamic mechanism of protein stabilization by trehalose. *Biophys. Chem.* **1997,** *64* (1-3), 25-43.

44. Barlow, D. J.; Poole, P. L., The hydration of protein secondary structures. *FEBS Lett.* **1987,** *213* (2), 423-427.

45. Kosar, F.; Akram, N. A.; Sadiq, M.; Al-Qurainy, F.; Ashraf, M., Trehalose: A Key Organic Osmolyte Effectively Involved in Plant Abiotic Stress Tolerance. *J. Plant Growth Regul.* **2019,** *38* (2), 606-618.

46. Crowe, J. H.; Carpenter, J. F.; Crowe, L. M., The role of vitrification in anhydrobiosis. *Annu. Rev. Physiol.* **1998,** *60* (1), 73-103.

47. Mensink, M. A.; Frijlink, H. W.; van der Voort Maarschalk, K.; Hinrichs, W. L., How sugars protect proteins in the solid state and during drying (review): Mechanisms of stabilization in relation to stress conditions. *Eur. J. Pharm. Biopharm.* **2017,** *114*, 288-295.

48. Rupley, J. A.; Gratton, E.; Careri, G., Water and globular proteins. *Trends Biochem. Sci.* **1983,** *8* (1), 18-22.

49. Jubb, H. C.; Higueruelo, A. P.; Ochoa-Montano, B.; Pitt, W. R.; Ascher, D. B.; Blundell, T. L., Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *J. Mol. Biol.* **2017,** *429* (3), 365-371.

50. Estrada, J.; Bernado, P.; Blackledge, M.; Sancho, J., ProtSA: a web application for calculating sequence specific protein solvent accessibilities in the unfolded ensemble. *BMC Bioinf.* **2009,** *10*, 104.

51. Jenik, M.; Parra, R. G.; Radusky, L. G.; Turjanski, A.; Wolynes, P. G.; Ferreiro, D. U., Protein frustratometer: a tool to localize energetic frustration in protein molecules. *Nucleic Acids Res.* **2012,** *40* (Web Server issue), W348-51.

52. Damodaran, S.; Song, K. B., The role of solvent polarity in the free energy of transfer of amino acid side chains from water to organic solvents. *J. Biol. Chem.* **1986,** *261* (16), 7220-2.

53. Radzicka, A.; Wolfenden, R., Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry* **1988,** *27* (5), 1664-1670.

CHAPTER 3: WATER'S VARIABLE ROLE IN PROTEIN STABILITY UNCOVERED BY LIQUID-OBSERVED VAPOR EXCHANGE NMR.[1]

INTRODUCTION

Kauzmann's 1955 review titled "Some factors in the interpretation of protein denaturation" illuminated the key role water plays in the determination of the native structure, stability, and function of proteins.[1] Yet, despite over six decades of research into protein-water interactions, our ability to predict, let alone experimentally assess, the contribution of water to local protein structure and stability remains limited.

Our inadequate understanding of how protein-water interactions shape and stabilize native protein structure stems in part from our technological inability to experimentally observe the impact of dehydration on local protein structure. For years, Fourier-Transform Infrared (FTIR) was one of few techniques capable of assessing dehydrated protein structure, yet this and most other techniques can only provide information on the global secondary structure composition of a protein.[2] Fortunately, with the recent development of liquid-observed vapor exchange nuclear magnetic resonance (LOVE NMR),[3] it is now possible to localize, at the residue-level, protein regions that lose structure upon dehydration.

---

[1]*Submitted to a peer-reviewed journal with the same title and the following author list: Candice J. Crilly, Jonathon E. Eicher, Owen Warmuth, Joanna Atkin, and Gary J. Pielak.*

Based on the well-established principle that amide protons are less likely to exchange with deuterons from the local environment if involved in intra- or intermolecular H-bonds, LOVE NMR uses solution NMR spectroscopy to quantify the extent of hydrogen–deuterium exchange (HDX) between D_2O vapor and the amide protons of a dried protein. The percent of amide-proton signal remaining after vapor exchange (%Protected) reflects the fraction of dry protein population for which a given residue is involved in an inter- or intramolecular H-bond. Thus, LOVE NMR can be used to pinpoint where, and to what degree, dehydration induces protein unfolding, i.e. which protein regions are most dependent on interactions with water.

Here, we use a modified version of LOVE NMR (Fig. S3.1), FTIR spectroscopy, and NMR-detected solution HDX to link mutation-induced changes in dehydrated protein structure to changes in solution stability. We apply this methodology to two proteins – the 6-kDa B1 domain of staphylococcal protein G (GB1) and 7-kDa chymotrypsin inhibitor 2 (CI2) from barley (Fig. 3.1), as well as on two variants thereof – the I6L variant of GB1 and the I20V variant of CI2.

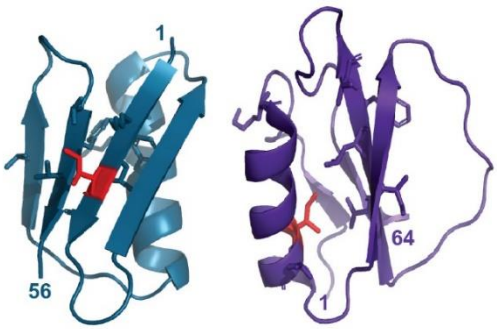


Figure 3.1. Structures of GB1 (left, PDB 2QMT) and CI2 (PDB 2CI2). Sidechains are shown for global unfolding residues, with mutated residues in red.

These variants were selected because they both perturb a “global unfolding residue”, i.e. a residue whose opening free energy is comparable to the global stability of the protein.[4-6]

MATERIALS AND METHODS

Materials. Ampicillin, kanamycin sulfate (Sigma Aldrich), trisodium citrate (Agros Organics), citric acid monohydrate and HEPES (Thermo Fisher) were used without further purification. H_2O with a resistivity >17 MΩ cm^{-1} was used to prepare buffers. pH values are direct readings, uncorrected for the deuterium isotope effect.[7] For LOVE NMR experiments, a constant relative humidity of 75 ±5% (measured with a Fisherbrand TraceableGO™ Bluetooth datalogging digital hygrometer) was created as described.[3]

Protein expression and purification

pET11a plasmids (Novagen) containing the genes for the T2Q and T2Q, I6L variants of the immunoglobulin G binding domain of streptococcal G (WT GB1 and I6L GB1, respectively) were obtained as described.[8] The T2Q mutation prevents N-terminal deamidation.[9] The pET28a plasmid (Novagen) containing the gene for truncated wild-type barley chymotrypsin inhibitor 2 (WT CI2), was provided by Dr. Andrew Lee’s laboratory at UNC-Chapel Hill. The first 19 residues of full length CI2 were not included in the construct because they are disordered, and therefore exchange too quickly for observation by NMR. Residue 20 of full-length CI2 is

referred to as Residue 1. To create the I20V variant (I20L CI2), site-directed mutagenesis was performed using the following primers (mutated codon in bold).
Forward: 5' A GAA GCG AAA AAA GTG **GTG** CTG CAG GAT AAA C 3'
Reverse: 5' C CGG TTT ATC CTG CAG **CAC** CAC TTT TTT C 3'.

^{15}N-enriched WT GB1, I6L GB1, WT CI2, and I20V CI2 were expressed in Agilent BL21 Gold (DE3) *E. coli* in minimal media and purified as described,[8, 10] with the following modifications for GB1 I6L: Minimal media was supplemented with 2 g/L glycerol, and expression was carried out for ~16 h at 18 °C.
The concentration of purified proteins was determined from the absorbance at 280 nm (A_{280}) (Nanodrop One, Thermo Fisher) using an extinction coefficient of 9530 M^{-1} cm^{-1} for GB1 variants and 7040 M^{-1} cm^{-1} for CI2 variants.[11] Purity was confirmed by Q-TOF mass spectrometry (ThermoScientific, Q Exactive HF-X) in the UNC Mass Spectrometry Chemical Research and Teaching Core Laboratory. Purified protein was exchanged into H_2O by dialysis (ThermoScientific Snakeskin™ dialysis tubing, 3500 Da molecular weight cutoff), and divided into aliquots such that resuspension in 650 µL gives a protein concentration of 500 µM. Aliquots were flash frozen, lyophilized, and stored at -20 °C.

NMR

Unless noted, experiments were performed in triplicate on Bruker Avance III HD spectrometers with cryogenic QCI probes at ^{1}H Larmor frequencies of 600 MHz for LOVE experiments and 850 MHz for solution amide-proton exchange

experiments. For LOVE NMR experiments, sensitivity-enhanced ^{15}N-^{1}H heteronuclear single-quantum coherence (HSQC) spectra were acquired in ~10 min (128 increments in the ^{15}N dimension, 4 scans per increment) with sweep widths of 3041 Hz in the ^{15}N dimension and 8418 Hz in the ^{1}H dimension. For solution exchange experiments, ^{15}N-^{1}H HSQC spectra were acquired in ~10 min (128 increments in the ^{15}N dimension, 4 scans per increment) with sweep widths of 4308 Hz in the ^{15}N dimension and 11904 Hz in the ^{1}H dimension. Spectra were processed with NMRPipe.[12] Cross peaks were integrated using NMRViewJ.[13]

Backbone resonances of WT GB1, WT CI2, and CI2 I20V at pH 4.5, 4 °C were assigned using isotopically-enriched protein expressed in minimal media containing ^{13}C D-glucose and ^{15}N NH_4Cl (Cambridge Isotope Labs) as the sole sources of carbon and nitrogen, respectively, and then purified as described above. HNCACB spectra were acquired with 10% sampling in the indirect dimensions using a Poisson gap scheduling scheme (http://gwagner.med.harvard.edu/intranet/hmsIST/gensched_new.html).[14-15]

Spectra were processed using NMRpipe and reconstructed with the SMILE algorithm.[16] Backbone resonances for WT GB1 and I20V CI2 at pH 7.5, 22 °C were assigned in the same manner. Other assignments (WT CI2 at pH 7.5 and 22 °C, I6L GB1 under both conditions) were transferred. Assignments and labeled spectra are provided in Tables S1-S4 and Figures S3.1-S3.4, respectively.

Solution hydrogen-deuterium exchange

Lyophilized aliquots of purified, ^{15}N-enriched protein were resuspended in 650 µL of 7.5-mM HEPES, pH 6.5, flash frozen and lyophilized. After ~24 h, the samples were removed, resuspended in 650 µL 99% D_2O and immediately used to acquire serial NMR HSQC spectra at 22 °C. To capture decay curves for residues that completely exchange in 2 - 24 h (intermediate regime), 10 - 12 sensitivity ^{15}N-^{1}H HSQC spectra were acquired serially over ~12 h. For slowly exchanging residues (>24 h), samples were resuspended, and following acquisition of the first spectrum (0 h timepoint), placed in an incubator at 22 °C. Samples were removed from the incubator every 1 - 3 days for spectrum acquisition.

The rate analysis tool in NMRViewJ was used to fit peak volumes as a function of time to the 3-parameter equation $V = Ae^{-Bt}$, where V is peak volume in arbitrary units, A is a scaling factor, t is time in s, and B is the observed rate constant (k_{obs}). For each residue, k_{obs} was divided by the estimated intrinsic rate constant of exchange (k_{int}) at pH 7.5, 22 °C to approximate the opening equilibrium constant, K_{op}. Values of k_{int} were obtained using the online Server Program for Hydrogen Exchange Rate Estimation, SPHERE.[17] To ensure accurate k_{obs} values, only crosspeak volumes that decayed to ~30% of their initial value were analyzed. Opening free energies, $\Delta G°'_{op}$ values, were calculated using the equation

$$\mathbf{\Delta G_{op}}^{\circ\prime} = -RTln(K_{op}) = -RTln(\frac{k_{obs}}{k_{int}})$$

where R is the gas constant and T is the absolute temperature.[18-19]

Liquid Observed Vapor Exchange (LOVE) NMR

For each experiment, two identical aliquots of pure, lyophilized, ^{15}N-enriched protein were resuspended in 650 μL of 1.5-mM HEPES, pH 6.5, to a final protein concentration of 500 μM, flash-frozen, and lyophilized (LABCONCO FreeZone 1 Liter Benchtop Freeze Dry System) for 24 h. Following lyophilization, one sample, designated T_0, was immediately resuspended in 650 µL of cold quench buffer (100 mM citrate buffer, pH 4.5, 90% H_2O/10% D_2O) and transferred to an NMR spectrometer for spectrum acquisition at 4°C. The second sample, designated T_{24}, was placed, with the cap open, in a chamber with a controlled relative humidity of ~75%, prepared as described.[3] After 24 h, the T_{24} sample was resuspended in 650 µL of cold quench buffer and an HSQC spectrum obtained with the same acquisition parameters used for the T_0 sample.

To enable back-exchange correction, the T_{24} sample was left in the spectrometer at 4°C for ~ 12 h, during which time an additional 10-12 spectra were acquired. The time between resuspension and initiation of the first T_{24} spectrum acquisition was 10 min, ~8 min of which were spent at 4 °C.

Identifying residues that back-exchange completely before spectrum acquisition

To differentiate residues that are highly protected from back-exchange from those that back-exchange completely before the first T_{24} spectrum was acquired, a sample of ^{15}N-enriched protein was resuspended to 500 µM in 1.5 mM HEPES, lyophilized for 24 h, resuspended in cold D_2O-based quench buffer (100 mM citrate,

pH 4.5 >98% D_2O) and immediately transferred to a spectrometer at 4°C for acquisition of a ^{15}N-^{1}H HSQC spectrum using the same acquisition parameters used for the T_0 and T_{24} samples. Resonances present in H_2O but not in D_2O are presumed to back-exchange completely before the first T_{24} spectrum is acquired.

Given that an exchange rate cannot be estimated for these residues, we cannot approximate the pre-back-exchange signal, and, therefore, these residues are omitted from the dataset (see figure captions for lists). We call these resonances "quench labeled".

Estimating V_{T24}, the peak volume of T_{24} sample pre-back-exchange

For non-quench-labeled T_{24} resonances that exhibited a ≤5% increase in peak volume over ~12 h, back-exchange during the time it takes to acquire a HSQC spectrum (20 min here) is presumed to be negligible. Therefore, for these residues, $\mathbf{V_{T24}}$ is equivalent to the peak volume obtained from integrating the initial T_{24} spectrum.

For T_{24} resonances that exhibited a >5% increase in peak volume over ~12 h, peak volumes from the 10 - 12 spectra acquired serially for the T_{24} sample were plotted as a function of time and fit, using the nonlinear least-squares algorithm in MATLAB, to the 3-parameter equation $V(t) = A(1-e^{-Bt}) + C$, where t is the time between protein resuspension and signal acquisition, V(t) is the peak volume at time t, A is the maximum possible change in peak volume, B is the observed rate constant, and C is a constant equivalent to the initial peak volume before back-

exchange. For these residues, the fitted value for C was used to estimate the pre-back-exchange peak volume, V_{T24}.

The estimated deadtime was 20 min (based on the 10 min between resuspension and initiation of spectrum acquisition plus 10 min for spectrum acquisition). In practice, this estimation means timepoints are shifted by 20 min, e.g. the first spectrum acquired for T_{24} corresponds to the 20-minute timepoint.

Calculating average %Protected from vapor exchange

To determine the average percent of the dried protein population for which a given amide proton is protected from vapor exchange, the average value of V_{T24} (obtained as described above) was divided by the average peak volume of the corresponding crosspeak in the non-vapor-exchanged protein sample (T_0) and multiplied by 100%, i.e. $\overline{\%\boldsymbol{Protected}} = \mathbf{100}\%\, \boldsymbol{x} \frac{\overline{V_{T24}}}{\overline{V_{T0}}}$. Uncertainties were obtained using triplicate analysis and standard propagation of error.

Fourier transform infrared (FTIR) spectroscopy

ATR-FTIR spectra were recorded on a BioTools Prota-3S spectrophotometer equipped with a HgCdTe detector and a ZnSe-diamond attenuated total-internal-reflectance crystal. Spectra comprise 400 scans and were acquired from 805 to 5500 cm^{-1} at 4 cm^{-1} resolution with dry-air purging. Background, buffer, and vapor spectra were acquired at 30 ^{o}C using a Pike Peltier temperature controller. Empty, path-, buffer-, and sample- spectra were acquired and preprocessed for each sample using Prota3s (BioTools) software. The background spectrum was

subtracted from both the buffer- and sample- spectra to produce theabsorbance spectra. The buffer absorbance spectrum (7.5 mM HEPES at 30 ^{o}C) was then subtracted from the sample absorbance spectrum in the same proportion for all samples. To ensure full buffer subtraction, the procedure was performed such that the region at 3750 cm^{-1} is non-negative and the region from 2000 cm^{-1} to 1800 cm^{-1} is flat.

Spectra were processed using an Orange Datamining workflow,[20] and smoothed using 4-component principle-component-analysis denoising.[21] Smoothed bands were processed using a positive rubber band baseline-correction, vector-normalization, and then fit to Voigt profiles through a non-linear least-squares regression.[22] Peaks were assigned and their areas normalized against the total peak area to obtain the percentage of each type of resonance (Table S3.1); percent secondary structure was calculated from the total peak area excluding the contribution from aromatics. Second derivative spectra were calculated using a Savitzky-Golay filter with a window size of nine points and a second-order polynomial.[23]

RESULTS

To enable direct comparison of hydrated and dehydrated structure and provide a reference dataset for the LOVE NMR data, we used Attenuated Total Reflection FTIR (ATR-FTIR) spectroscopy. Curve-fitting of the amide 1 band (Fig. S3.2) was performed to derive secondary structure information (Table S3.1).

Upon lyophilization, all samples exhibited an increase in β-sheet character and a decrease in α-helical character (Fig. 3.2A & C). Inspection of the second derivative plots (Fig. 3.2B & D) reveals that upon dehydration, all proteins exhibit a frequency increase in the 1680 cm^{-1} β-sheet region, indicating an increase in amide-bond strength and/or a change in bond orientation.[24]

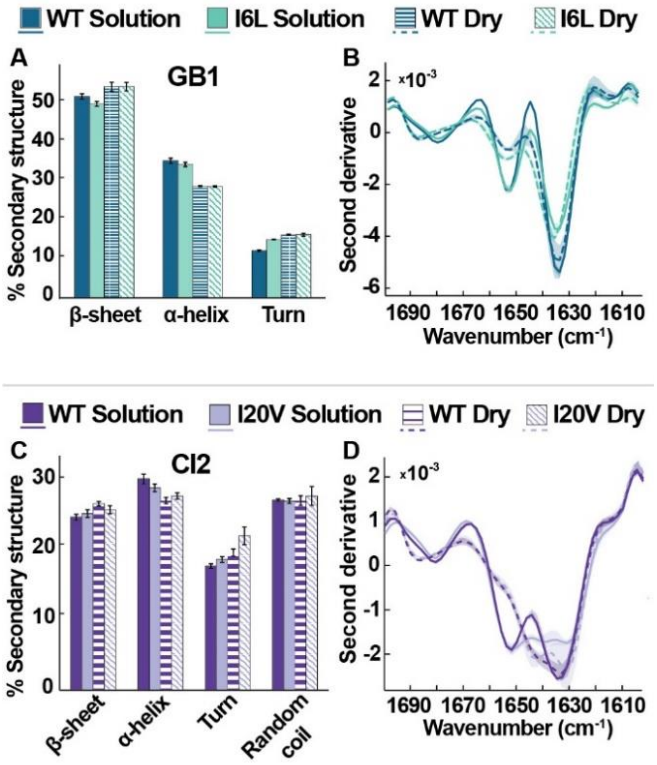


Figure 3.2. FTIR of wild-type and variant proteins in solution (7.5 mM HEPES, pH 6.5) and the lyophilized state. A and C, average secondary structures of GB1 and CI2 variants. Error bars represent standard deviations from the mean. B and D, averaged second derivative spectra. Shadows represent the range of data from three independent experiments.

In solution, I6L GB1 exhibits slightly less β-character and more turn character than the wild-type protein (Fig 3.2A), but these differences are lost upon lyophilization. Inspection of the second derivative data (Fig. 3.2B) reveals that in the dry state, both GB1 variants exhibit a pronounced intensity decrease at 1652 cm^{-1}

helix region, which suggests a broader variety of helix conformations in the lyophilized form.[22, 25-26]

In solution, WT and I20V CI2 possess similar secondary structure profiles, with I20V exhibiting only a slight decrease in α-helical character relative to WT (Fig. 3.2C). Second derivative spectra reveal that, in solution, I20V exhibits reduced intensity at 1630 cm^{-1} (Fig. 3.2D), which suggests a difference in the β-sheet H-bond strength and/or configuration.[22] Differences between WT and I20V CI2 become more apparent in the dry state, with I20V exhibiting slightly less β-sheet character and more turn character than WT (Fig. 3.2C). Like the GB1 variants, second derivative spectra of the CI2 variants reveal that dehydration decreases in the intensity at 1655 cm^{-1} (Fig. 3.2D), again indicative of a broader variety of helix conformations. Relative to the GB1, however, the CI2 variants appear to be less sensitive to dehydration, as they exhibit smaller secondary structure changes between the solution and dry states (Fig. 3.2C).

To gain more information on how mutation affects the solution- and solid-states of GB1 and CI2, we used HDX to acquire residue-level opening free energies ($\Delta G_{op}^{o\prime}$) in solution and LOVE NMR to acquire %Protected values in the dry state. For GB1 in solution at 22 °C and pH 7.5, the I6L substitution is generally destabilizing, reducing the average measurable $\Delta G_{op}^{o\prime}$ by 0.3 ±0.4 kcal/mol and the $\Delta G_{op}^{o\prime}$ of global unfolding residues by 0.7 ±0.2 kcal/mol (Fig. 3.3A). Residues preceding or following a global-unfolder experience small stability increases, with N7 witnessing the largest increase (0.52 ±0.04 kcal/mol). Despite these changes in solution stability, however,

the GB1 variants exhibit nearly identical LOVE profiles (Fig. 3.3B), in agreement with our FTIR analysis.

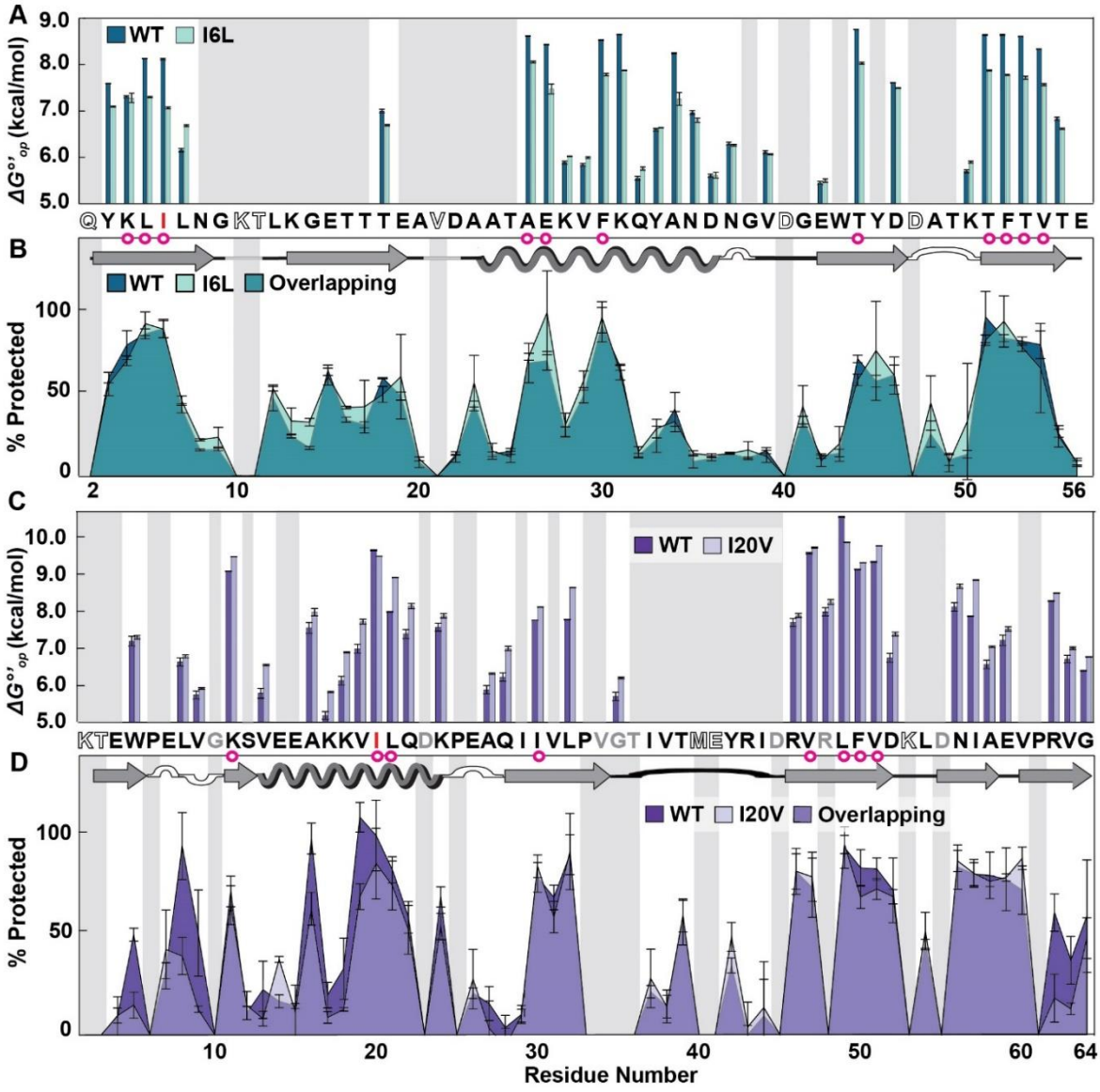


Figure 3.3. Residue-level solution stabilities and dry state structures of wild-type (WT) and variant proteins. (A) Opening free energies (7.5 mM HEPES, pH 6.5, 22 °C). (B) Overlaid LOVE profiles of WT and I6L GB1 freeze-dried in 1.5-mM HEPES pH 6.5. (C) Opening free energies CI2. (D) Overlaid LOVE profiles of WT and I20V CI2 freeze-dried in 1.5-mM HEPES pH 6.5. Primary- and secondary- structures of the wild-type proteins are shown between panels. In both primary structures, red letters indicate which residue was mutated. Magenta circles indicate solution global-unfolding residues of WT proteins. Gray boxes indicate data that are missing due to rapid back exchange (open letters in primary structure) and/or the inability to reliably integrate peak volumes due to overlapping resonances (gray letters in primary structure). Error bars represent standard deviations of the mean calculated from triplicate analysis.

For CI2 in solution, the I20V substitution is marginally stabilizing, increasing the average measurable ($\Delta G_{op}^{o\prime}$) by 0.4 ±0.3 kcal/mol and the average ($\Delta G_{op}^{o\prime}$) of global unfolders by 0.2 ±0.5 kcal/mol (Fig. 3.3C). Only the site of the mutation (V20) and L49, a residue within the hydrophobic core of CI2, are less stable in the I20V variant.

Although the I20V mutation is marginally stabilizing in solution, LOVE NMR data show that the mutation reduces dry-state protection for certain regions (Fig 3.3D). Specifically, the terminal β-sheets of dehydrated I20V CI2 are less than half as protected as they are in the wild-type protein. In addition, V19, the residue preceding the mutation; A16, the native H-bond acceptor of I20; and L8, the H-bond donor of one of the residues in the N-terminal β-sheet, all experience reduced protection in the dry state.

DISCUSSION

The results of this study indicate that changes in dehydrated protein structure cannot be predicted from changes in solution stability alone, pointing toward the variable role of water in maintaining local protein structure and stability. Although both the I6L and I20V substitutions perturb global unfolding residues, there are two key structural differences. First, in GB1, the sidechain of isoleucine 6 faces outward, while in CI2, that of isoleucine 20 faces inward toward the hydrophobic core (Fig. 3.1). Second, the I6L substitution occurs in a *β*-strand, while the I20V substitution occurs in an α-helix. These distinguishing structural characteristics may help explain the contrasting LOVE NMR data.

The observation that the I6L substitution destabilizes GB1 in solution but has almost no effect on dry-state structure suggests that water is primarily responsible for the reduced stability of the variant. Given the water-facing location of the position-6 sidechain, one mechanism of water-mediated destabilization could be reduced “sidechain blocking” of hydration by leucine relative to isoleucine. This mechanism of *β*-sheet destabilization, which was suggested by Bai and Englander,[27] would affect the protein in solution but not affect the dry state.

The observation that the I20V substitution slightly stabilizes CI2 in solution but makes the protein more susceptible to dehydration suggests that water plays an enhanced stabilizing role in I20V, particularly at the terminal β-sheets. The source of this instability is unclear, but the results are consistent with those of Ladurner *et al.*, who found strain in the α-helix of WT CI2 from non-optimal packing in the hydrophobic core.[28] Perhaps alleviating some of the steric strain in the core *via* a volume-reducing mutation allows tighter core packing and an enhanced hydrophobic effect. Alternatively, the mutation could lead to an increase in the number and/or strength of stabilizing hydrogen bonds with water.

CONCLUSIONS

In summary, comparisons of mutation-induced changes in protein solution stability to changes in dehydrated protein structure suggest that water plays a dominant role in the destabilization of the I6L variant of GB1 and a more nuanced, stabilizing role in the I20V variant of CI2. Our results illustrate the complex nature of protein-water interactions and demonstrate the ability of LOVE NMR to localize structural changes in the dry state. We envision that the residue-level information on

the dehydrated structure provided by LOVE NMR can be combined with other experimental and computational techniques to gain a deeper understand water's role in determining protein structure and stability. Understanding the fundamental interaction between proteins and water will in turn improve our ability to predict protein structure and stability, enabling the *de novo* design of protein-based therapeutics.

SUPPLEMENTAL INFORMATION

Supplementary figures

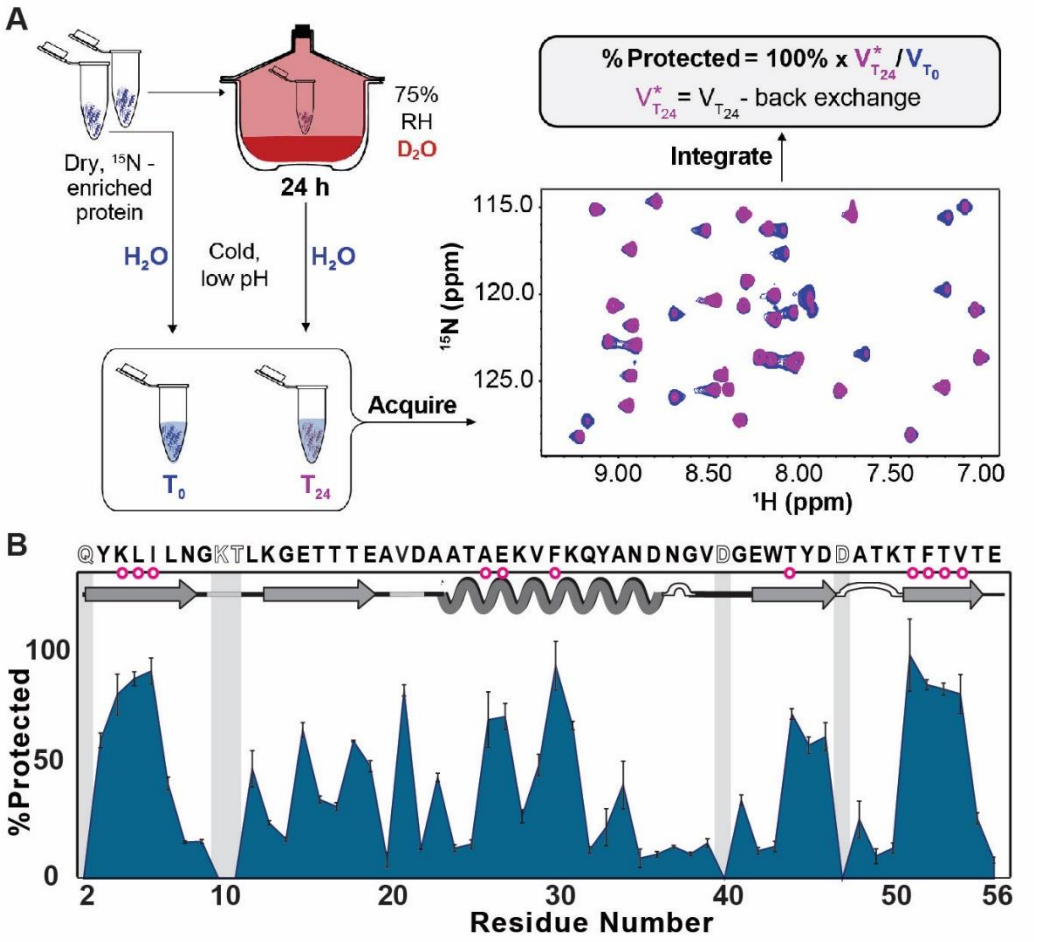


Figure S3.1. Updated LOVE NMR workflow and output. (A) LOVE NMR workflow. Identical samples of ^{15}N-enriched protein dried alone or in the presence of a cosolute are resuspended in cold, acidic buffer before (T_0) or after (T_{24}) 24 h exposure to D_2O vapor at 75% relative humidity (RH). Amide protons unprotected in the dry state exchange with deuterons from the vapor, resulting in smaller cross peak volumes in the T_{24} ^{15}N-^{1}H HSQC spectrum relative to the T_0 spectrum (pink and blue cross peaks, respectively). To correct for solution back-exchange that occurs before and during spectrum acquisition, serial HSQC spectra are obtained for the T_{24} sample, integrated, and fit to the equation $V_{T24}(t) = A(1-e^{-bt})+V_{T24}^*$, where $V_{T24}$is peak volume, t is time since resuspension, A is a scaling factor, b is the observed rate of exchange, and V_{T24}^* the peak volume before any back exchange (see Materials and Methods). The fitted V_{T24}^* value is then divided by the maximum possible peak volume, V_{T0}, and multiplied by 100 to obtain %Protected. (B) LOVE profile of model protein GB1 freeze-dried in 1.5 mM HEPES, pH 6.5. Open letters indicate residues with undefined dry-state protection because they are 100% quench-labelled. Secondary structure (arrows, β-strands; undulations, helix; white bumps, turns; gray lines, bends) is shown at top, with magenta circles indicating solution global unfolding residues. Gray areas indication the absence of data. Error bars represent uncertainty propagated from standard deviations of the mean from triplicate analysis.

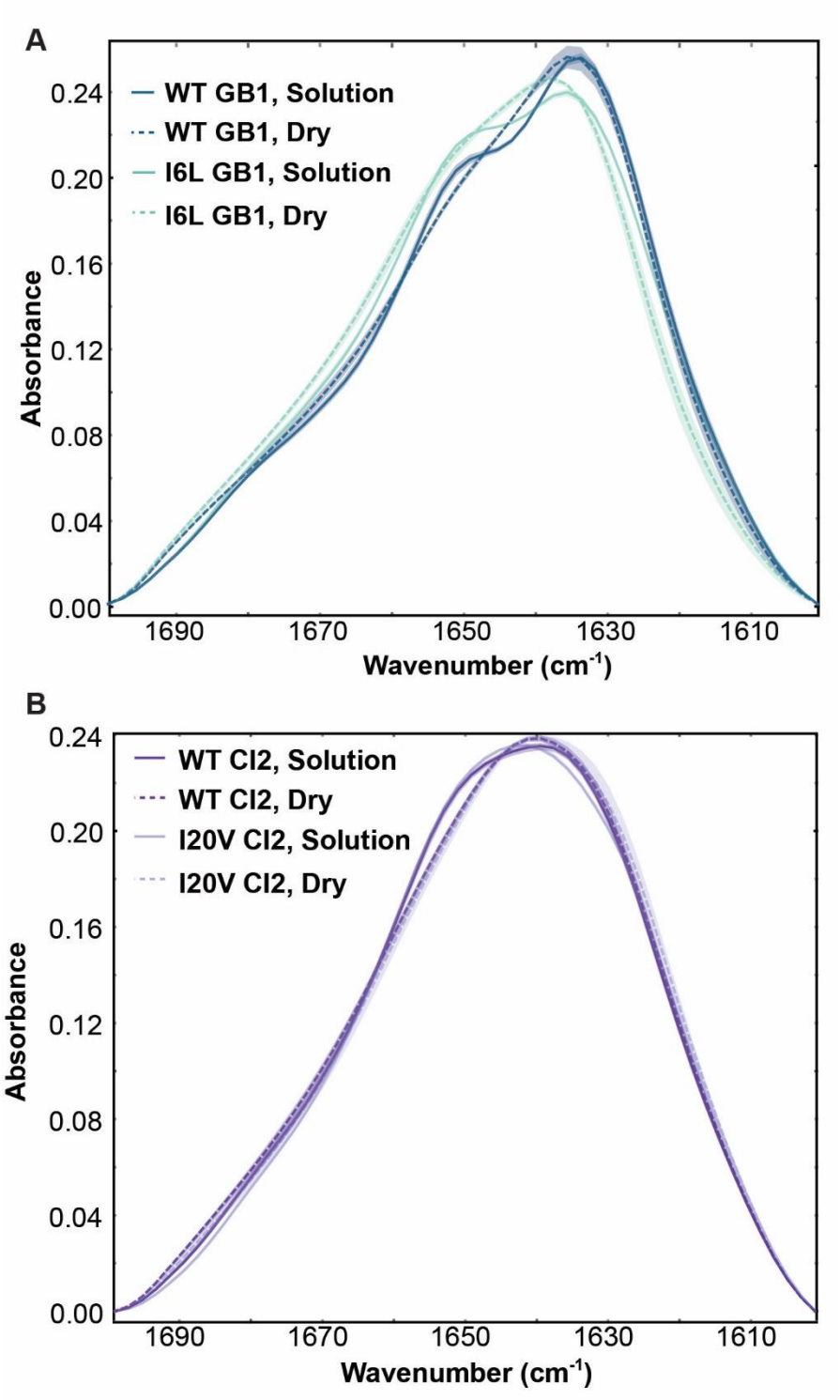


Figure S3.2. Average ATR-FTIR spectra of GB1 and CI2 variants. Amide-I regions of the averaged spectra of **(A)** GB1 variants and **(B)** CI2 variants in the solution and dry states. Averages are from triplicate analysis. Shaded areas represent the range.

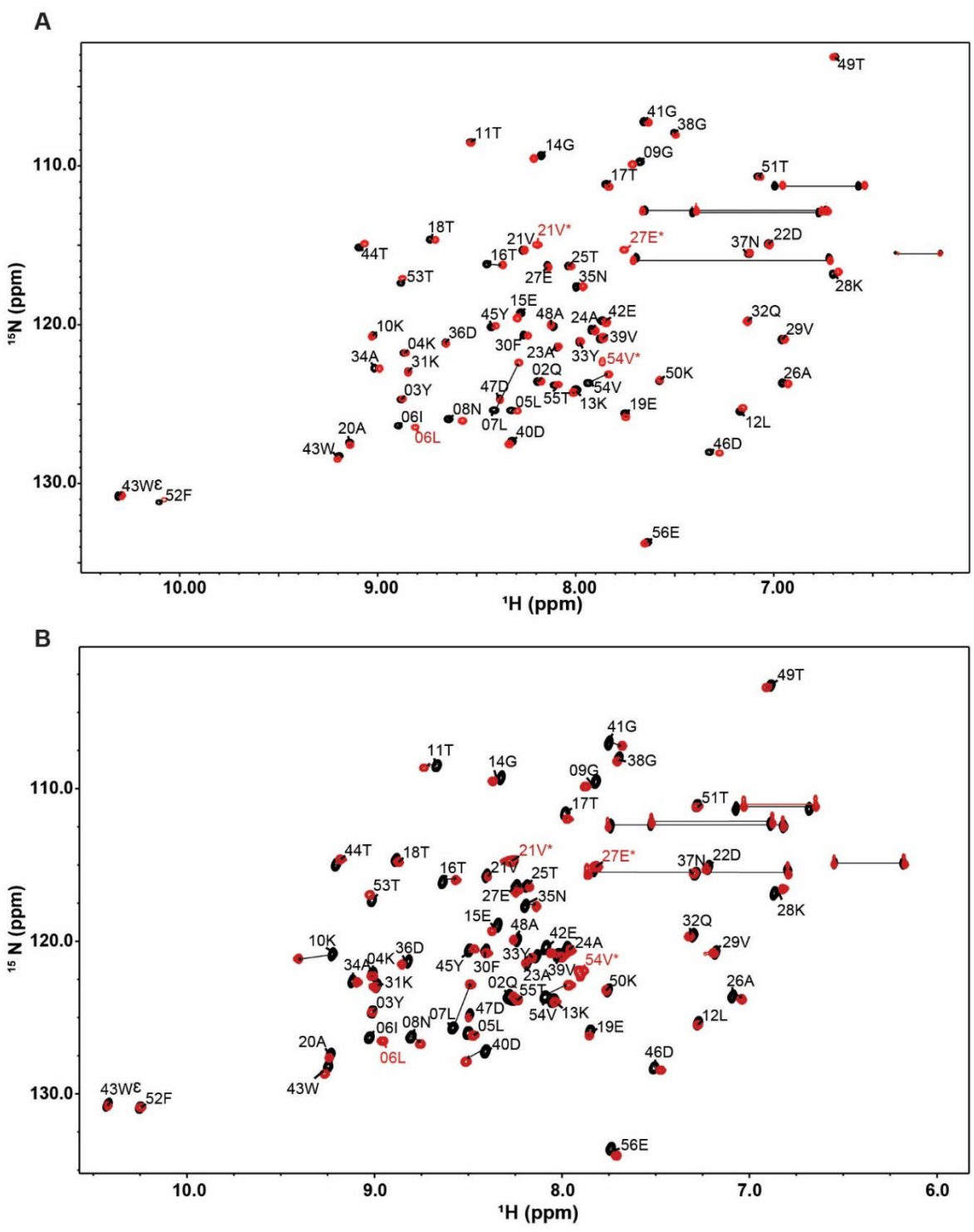


Figure S3.3. Assigned ^{1}H-^{15}N HSQC spectra of GB1 variants. A) Under spectrum acquisition conditions used for LOVE NMR experiments (100 mM citrate in 90%H_2O, 10% D_2O, pH 4.5, 4°C). B) Under conditions similar to those used for solution hydrogen-deuterium exchange (7.5 mM HEPES in 95%H_2O, 5% D_2O, pH 7.5, 22°C. WT GB1 resonances shown in black, I6L in red. *Starred resonances arise from an alternative conformation.

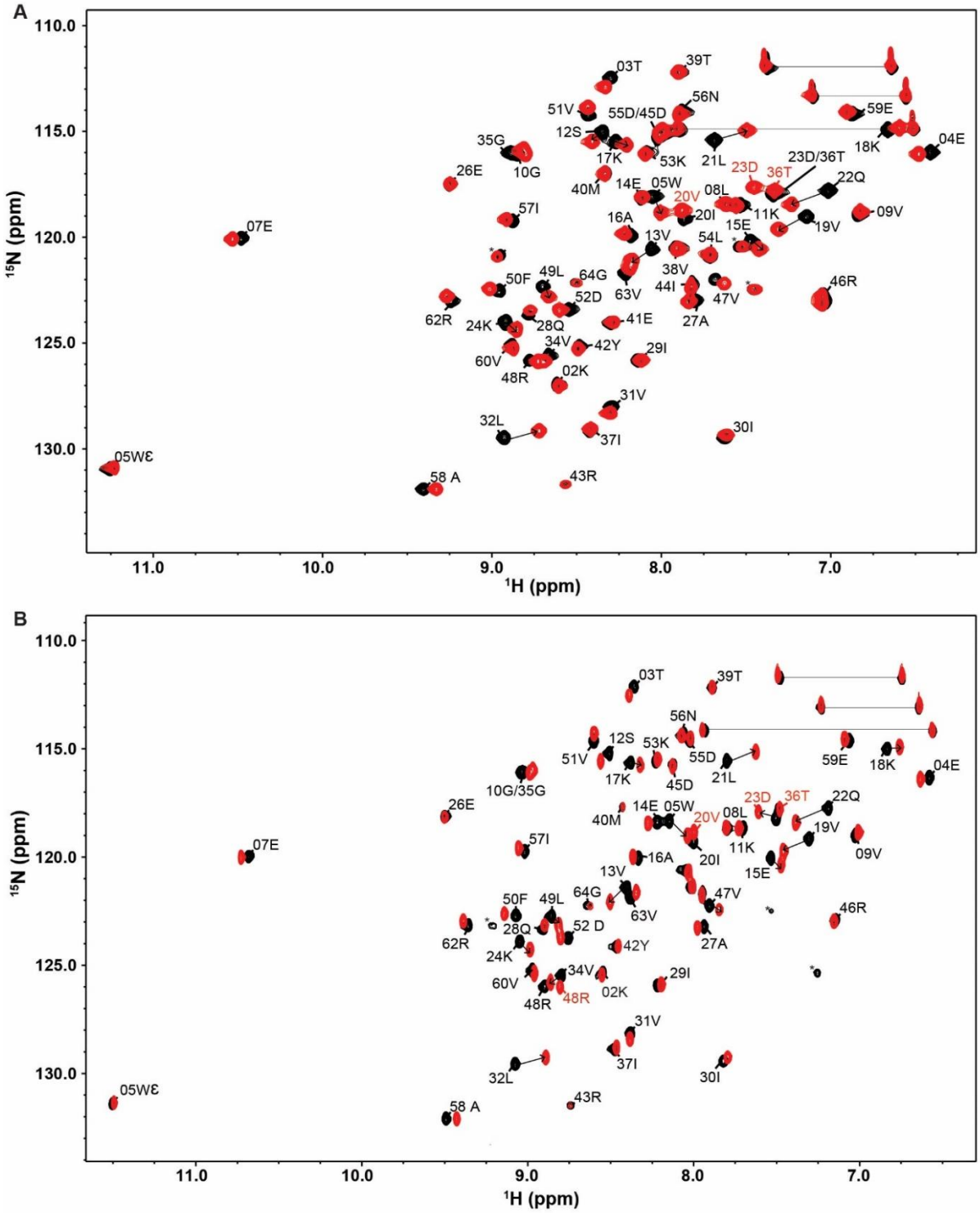


Figure S3.4. Assigned ^{1}H-^{15}N HSQC spectra of CI2 variants. A) Under spectrum acquisition conditions used for LOVE NMR experiments (100 mM citrate in 90%H_2O, 10% D_2O, pH 4.5, 4°C). B) Under conditions similar to those used for solution hydrogen-deuterium exchange (7.5 mM HEPES in 95%H_2O, 5% D_2O, pH 7.5, 22°C. WT CI2 resonances shown in black, I20V in red. *Starred resonances arise from an alternate conformation.

Supplementary tables

Table S3.1. FTIR peak locations and secondary structure assignments. .

GB1 variants in solution						
	Aromatics	**β-sheet**	**β-sheet**	**α-helix**	**Turn**	**β-sheet**
WT	1617.2 ±0.3 cm^{-1} 4.9 ± 0.2%	1626.9 ±0.2 cm^{-1} 29.1 ± 0.3%	1635.6 ±0.2 cm^{-1} 17.2 ± 0.4%	1650.0 ±0.1 cm^{-1} 35.6 ± 0.5%	1665.2 ±0.1 cm^{-1} 12.2 ± 0.3%	1678.8 ±0.1 cm^{-1} 5.9 ± 0.2%
I6L	1617.0 ±0.2 cm^{-1} 4.9 ± 0.2%	1626.9 ±0.3 cm^{-1} 27.4 ± 0.2%	1636.2 ±0.1 cm^{-1} 16.1 ± 0.4%	1649.9 ±0.1 cm^{-1} 34.6%	1663.9 ±0.3 cm^{-1} 15.1%	1677.9 ±0.2 cm^{-1} 6.8%
GB1 variants, lyophilized						
	Aromatics	**β-sheet**	**β-sheet**	**α-helix**	**Turn**	**Aromatics**
WT	1619.0 ±0.8 cm^{-1} 6.2 ± 0.2%	1628.9 ±0.7 cm^{-1} 26.6 ± 0.1%	1637.8 ±0.5 cm^{-1} 22 ± 1%	1650.0 ±0.6 cm^{-1} 28.9 ± 0.3%	1664.5 ±0.7 cm^{-1} 16.3 ± 0.8%	1619.0 ±0.8 cm^{-1} 6.4 ± 0.4%
I6L	1621 ±1 cm^{-1} 6.3 ± 0.3%	1631 ±1 cm^{-1} 26.1 ± 0.1%	1640 ±1 cm^{-1} 22 ± 1%	1652 ±1 cm^{-1} 28.9 ± 0.3%	1666±1 cm^{-1} 16.4 ± 0.6%	1621 ±1 cm^{-1} 6.2 ± 0.4%
CI2 variants in solution						
	Aromatics	**β-sheet**	**Random Coil**	**α-helix**	**Turn**	**β-sheet**
WT	1617.1 ±0.2 cm^{-1} 5.4 ± 0.1%	1627.1 ±0.2 cm^{-1} 17.6 ± 0.3%	1637.3 ±0.2 cm^{-1} 27.2 ± 0.1%	1649.5 ±0.2 cm^{-1} 30.2 ± 0.4%	1661.8 ±0.3 cm^{-1} 17.8 ± 0.3%	1676.4 ±0.3 cm^{-1} 7.2 ± 0.1%
I20V	1616.7 ±0.2 cm^{-1} 5.7 ± 0.1%	1626.8 ±0.2 cm^{-1} 18.3 ± 0.3%	1637.5±0.1 cm^{-1} 27.1 ± 0.3%	1649.1 ±0.2 cm^{-1} 29.0 ± 0.4%	1660.9 ±0.3 cm^{-1} 18.7 ± 0.3%	1675.3 ±0.3 cm^{-1} 7.0 ± 0.1%
CI2 variants, lyophilized						
	Aromatics	**β-sheet**	**Random Coil**	**α-helix**	**Turn**	**β-sheet**
WT	1617.5 ±0.2 6.1 ± 0.1%	1627.7 ±0.3 19.1 ± 0.3%	1637.9 ±0.4 26.5 ± 0.6%	1649.0 ±0.5 27.1 ± 0.4%	1661.9 ±0.5 19.7 ± 0.4%	1677.0 ±0.3 7.6 ± 0.2%
I20V	1616.7 ±0.2 5.8 ± 0.2%	1626.8 ±0.2 17.6 ± 0.5%	1637.5 ±0.1 24 ± 1%	1649.1 ±0.2 27.8 ± 0.4%	1660.9 ±0.3 22 ± 1%	1675.3 ±0.3 8.3 ± 0.3%

Assignments made using literature values[22-23, 29] and PDB files 2CI2 and 2QMT. Fits returned χ^2 values of ~2 x 10^{-5}.

Table S3.2. Opening free energies of WT and I6L GB1 at pH 7.5, 22°C. .

Residue[§]	$\Delta G_{op}°' \pm STD^{‡}$ (kcal/mol)					
	WT GB1			I6L GB1		
03Y	7.59	±	0.01	7.09	±	0.01
04K	8.19	±	0.01	7.3	±	0.1
05L	8.126	±	0.004	7.3	±	0.02
06I/L	8.11	±	0.02	7.07	±	0.02
07L	6.16	±	0.04	6.68	±	0.02
18T	7.00	±	0.04	6.69	±	0.02
26A	8.62	±	0.01	8.06	±	0.02
27E	8.43	±	0.01	7.5	±	0.1
28K	5.89	±	0.03	6.02	±	0.01
29V	5.84	±	0.03	5.99	±	0.02
30F	8.53	±	0.01	7.78	±	0.03
31K	8.648	±	0.004	7.87	±	0.01
32Q	5.55	±	0.04	5.76	±	0.03
33Y	6.61	±	0.04	6.64	±	0.01
34A	7.36	±	0.02	7.3	±	0.1
35N	6.97	±	0.04	6.80	±	0.05
36D	5.60	±	0.04	5.61	±	0.06
37N	6.30	±	0.03	6.26	±	0.02
39V	6.11	±	0.03	6.07	±	0.01
42E	5.45	±	0.04	5.50	±	0.04
44T	8.755	±	0.003	8.03	±	0.02
46D	7.60	±	0.01	7.50	±	0.01
50K	5.70	±	0.03	5.90	±	0.02
51T	8.64	±	0.01	7.87	±	0.02
52F	8.64	±	0.01	7.78	±	0.01
53T	8.606	±	0.005	7.72	±	0.03
54V	8.333	±	0.004	7.57	±	0.02
55T	6.83	±	0.04	6.62	±	0.02
56E	2.90	±	0.03	2.73	±	0.03

Footnotes

[§] $\Delta G_{op}°'$ values for residues that exchange slowly enough to be measured.

[‡] Sample standard deviation calculated from three independent experiments.

Table S3.3. Opening free energies of WT and I20V CI2 at pH 7.5, 22°C. .

Residue[§]	ΔG_{op}°′ ±STD[‡] (kcal/mol) WT CI2			I20V CI2		
05W	7.2	±	0.1	7.32	±	0.05
08L	6.6	±	0.1	6.80	±	0.04
09V	5.7	±	0.1	5.93	±	0.02
11K	9.114	±	0.004	9.515	±	0.003
13V	5.8	±	0.1	6.56	±	0.02
16A	7.6	±	0.1	8.0	±	0.1
17K	5.2	±	0.1	5.83	±	0.02
18K	6.1	±	0.1	6.91	±	0.02
19V	7.0	±	0.1	7.74	±	0.07
20I	9.69	±	0.01	9.526	±	0.002
21L	8.01	±	0.01	8.945	±	0.005
22Q	7.4	±	0.1	8.17	±	0.07
24K	7.6	±	0.1	7.90	±	0.06
27A	5.9	±	0.1	6.33	±	0.02
28Q	6.2	±	0.1	7.01	±	0.06
30I	7.782	±	0.002	8.142	±	0.005
32L	7.798	±	0.004	8.672	±	0.004
35G	5.7	±	0.1	6.21	±	0.02
46R	7.7	±	0.1	7.91	±	0.05
47V	9.61	±	0.02	9.76	±	0.01
48R	8.0	±	0.1	8.28	±	0.07
49L	10.592	±	0.009	9.901	±	0.008
50F	9.166	±	0.008	9.350	±	0.003
51V	9.37	±	0.01	9.810	±	0.003
52D	6.8	±	0.1	7.41	±	0.05
56N	8.1	±	0.1	8.71	±	0.06
57I	7.89	±	0.01	8.88	±	0.008
58A	6.6	±	0.1	7.06	±	0.02
59E	7.2	±	0.1	7.55	±	0.06
62R	8.31	±	0.01	8.523	±	0.006
63Y	6.7	±	0.1	7.02	±	0.04
64G	6.40	±	0.01	6.785	±	0.006

Footnotes
[§] ΔG_{op}°′ values from for non-proline residues without overlapping peaks that exchange slowly enough to be measured.
[‡] Sample standard deviation calculated from three independent experiments.

Table S3.4. Average %Protected values of WT and I6L GB1 freeze-dried in 1.5 mM HEPES pH 6.5. .

Residue	%Protected ±STD‡ WT GB1			I6L GB1		
3	61	±	3	57	±	7
4	81	±	9	72	±	3
5	88	±	3	95	±	7
6	92	±	6	91	±	5
7	42	±	3	47	±	3
8	16	±	1	23	±	1
9	16	±	1	24	±	6
12	49	±	8	54	±	1
13	25	±	1	34	±	8
14	17	±	1	33	±	2
15	66	±	3	62	±	5
16	35	±	2	42	±	1
17	32	±	2	40	±	20
18	60.7	±	0.4	51	±	5
19	50	±	2	60	±	30
20	8	±	3	11	±	1
22	13.1	±	0.3	12	±	3
23	45	±	2	60	±	20
24	13	±	1	15	±	5
25	15	±	2	12	±	3
26	70	±	10	74	±	2
27	72	±	6	100	±	30
28	28	±	3	32	±	7
29	50	±	5	59	±	6
30	90	±	10	99	±	10
31	68	±	2	64	±	5
32	13	±	1	15	±	3
33	23	±	8	29	±	6
34	41	±	10	34	±	6
35	9	±	4	14	±	2
36	11	±	1	13	±	1
37	14	±	1	14	±	0
38	11	±	1	16	±	5
39	16	±	2	12	±	4
41	35	±	2	43	±	13
42	12	±	1	9	±	4
43	14	±	2	20	±	10
44	73	±	2	60	±	4
45	59	±	4	80	±	30
46	63	±	6	60	±	10
48	26	±	8	50	±	20
49	10	±	3	8	±	6
50	13	±	2	30	±	40
51	100	±	20	84	±	3
52	86	±	2	100	±	20
53	84	±	3	80	±	2
54	81	±	9	70	±	30
55	27	±	2	25	±	7
56	8	±	1	8	±	3

Table S3.5. Average %Protected values of WT and I20V CI2 freeze-dried in 1.5 mM HEPES pH 6.5. .

Residue	%Protected ±STD‡ WT CI2			I20V CI2		
4	10	±	10	10	±	4
5	51	±	3	15	±	7
7	33	±	4	40	±	20
8	100	±	20	40	±	10
9	50	±	20	13	±	2
11	73	±	8	70	±	10
12	11	±	3	15	±	7
13	20	±	20	8	±	4
14	17	±	3	38	±	2
15	15	±	1	10	±	10
16	100	±	8	64	±	8
17	20	±	7	9	±	3
18	30	±	20	13	±	1
19	111	±	8	70	±	7
20	100	±	20	90	±	20
21	84	±	7	80	±	10
22	60	±	8	50	±	10
24	70	±	5	55	±	7
26	21	±	3	30	±	20
27	16	±	9	0	±	10
28	4	±	6	0	±	10
29	10	±	5	10	±	3
30	80	±	8	86	±	6
31	70	±	6	60	±	8
32	90	±	10	90	±	20
37	22	±	3	30	±	20
38	14	±	8	14	±	6
39	60	±	10	61	±	8
42	35	±	4	50	±	7
43	0	±	5	0	±	10
44	10	±	20	10	±	20
46	83	±	9	80	±	10
47	80	±	20	80	±	10
49	100	±	10	97	±	5
50	85	±	9	70	±	6
51	84	±	6	74	±	5
52	70	±	20	70	±	4
54	46	±	2	50	±	10
56	90	±	10	88	±	5
57	82	±	6	82	±	7
58	80	±	10	78	±	3
59	80	±	20	80	±	2
60	70	±	10	90	±	6
62	60	±	10	20	±	10
63	40	±	10	13	±	3
64	60	±	30	50	±	10

REFERENCES

1. Kauzmann, W., Some Factors in the Interpretation of Protein Denaturation. *Adv. Protein Chem.* **1959,** *14*, 1-63.

2. Moorthy, B. S.; Iyer, L. K.; Topp, E. M., Characterizing Protein Structure, Dynamics and Conformation in Lyophilized Solids. *Curr. Pharm. Des.* **2015,** *21* (40), 5845-5853.

3. 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.

4. Smith, C. K.; Withka, J. M.; Regan, L., A thermodynamic scale for the beta-sheet forming tendencies of the amino acids. *Biochemistry* **1994,** *33* (18), 5510-7.

5. Itzhaki, L. S.; Neira, J. L.; Fersht, A. R., Hydrogen exchange in chymotrypsin inhibitor 2 probed by denaturants and temperature. *J. Mol. Biol.* **1997,** *270* (1), 89-98.

6. Huyghues-Despointes, B. M. P.; Scholtz, J. M.; Pace, C. N., Protein conformational stabilities can be determined from hydrogen exchange rates. *Nat. Struct. Biol.* **1999,** *6* (10), 910-912.

7. Glasoe, P. K.; Long, F. A., Use of glass electrodes to measure acidities in deuterium oxide. *J. Phys. Chem.* **1960,** *64* (1), 188-190.

8. Monteith, W. B.; Pielak, G. J., Residue Level Quantification of Protein Stability in Living Cells. *Proc. Natl. Acad. Sci. USA* **2014,** *111* (31), 11335-11340.

9. Smith, C. K.; Withka, J. M.; Regan, L., A thermodynamic scale for theβ-sheet forming tendencies of the amino acids. *Biochemistry* **1994,** *33* (18), 5510-5517.

10. Charlton, L. M.; Barnes, C. O.; Li, C.; Orans, J.; Young, G. B.; Pielak, G. J., Residue-Level Interrogation of Macromolecular Crowding Effects on Protein Stability. *J. Am. Chem. Soc.* **2008,** *130* (21), 6826-6830.

11. Gill, S. C.; von Hippel, P. H., Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **1989,** *182* (2), 319-26.

12. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995,** *6* (3), 277-93.

13. Johnson, B. A.; Blevins, R. A., NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **1994,** *4* (5), 603-14.

14. Hyberts, S. G.; Takeuchi, K.; Wagner, G., Poisson-Gap Sampling and Forward Maximum Entropy Reconstruction for Enhancing the Resolution and Sensitivity of Protein NMR Data. *J. Am. Chem. Soc.* **2010,** *132* (7), 2145-2147.

15. Miljenović, T. M.; Jia, X.; Mobli, M., Nonuniform Sampling in Biomolecular NMR. In *Modern Magnetic Resonance*, 2nd ed.; Webb, G. A., Ed. Springer International Publishing: Cham, 2017; pp 1-21.

16. Ying, J.; Delaglio, F.; Torchia, D. A.; Bax, A., Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J. Biomol. NMR* **2017,** *68* (2), 101-118.

17. Zhang, Y.-Z. Protein and peptide structure and interactions studied by hydrogen exchange and NMR. PhD thesis. , University of Pennsylvania, Philadelphia, 1995.

18. Englander, S. W.; Kallenbach, N. R., Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys.* **1983,** *16* (4), 521-655.

19. Miklos, A. C.; Li, C.; Pielak, G. J., Using NMR-Detected Backbone Amide ^{1}H Exchange to Assess Macromolecular Crowding Effects on Globular-Protein Stability. *Methods Enzymol.* **2009,** *466*, 1-18.

20. Demšar, J.; Curk, T.; Erjavec, A.; Gorup, Č.; Hočevar, T.; Milutinovič, M.; Možina, M.; Polajnar, M.; Toplak, M.; Starič, A., Orange: data mining toolbox in Python. *Journal of Machine Learning Research* **2013,** *14* (1), 2349-2353.

21. Koziol, P.; Raczkowska, M. K.; Skibinska, J.; Urbaniak-Wasik, S.; Paluszkiewicz, C.; Kwiatek, W.; Wrobel, T. P., Comparison of spectral and spatial

denoising techniques in the context of High Definition FT-IR imaging hyperspectral data. *Sci. Rep.* **2018,** *8* (1), 14351.

22. Sadat, A.; Joye, I. J., Peak Fitting Applied to Fourier Transform Infrared and Raman Spectroscopic Analysis of Proteins. *Appl. Sci.* **2020,** *10* (17).

23. Gautam, R.; Vanga, S.; Ariese, F.; Umapathy, S., Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Techniques and Instrumentation* **2015,** *2* (1), 8.

24. Pézolet, M.; Bonenfant, S.; Dousseau, F.; Popineau, Y., Conformation of wheat gluten proteins Comparison between functional and solution states as determined by infrared spectroscopy. *FEBS Letters* **1992,** *299* (3), 247-250.

25. Allison, S. D.; Chang, B.; Randolph, T. W.; Carpenter, J. F., Hydrogen Bonding between Sugar and Protein Is Responsible for Inhibition of Dehydration-Induced Protein Unfolding. *Arch. Biochem. Biophys.* **1999,** *365* (2), 289-298.

26. Athokpam, B.; Ramesh, S. G.; McKenzie, R. H., Effect of hydrogen bonding on the infrared absorption intensity of OH stretch vibrations. *Chemical Physics* **2017,** *488-489*, 43-54.

27. Bai, Y.; Englander, S. W., Hydrogen bond strength and β-sheet propensities: The role of a side chain blocking effect. *Proteins* **1994,** *18* (3), 262-266.

28. Ladurner, A. G.; Itzhaki, L. S.; Fersht, A. R., Strain in the folding nucleus of chymotrypsin inhibitor 2. *Folding Des.* **1997,** *2* (6), 363-368.

29. Martinez, G.; Millhauser, G., FTIR Spectroscopy of Alanine-Based Peptides: Assignment of the Amide I′ Modes for Random Coil and Helix. *J. Struct. Biol.* **1995,** *114* (1), 23-27.

CHAPTER 4: PROTECTION BY DESICCATION-TOLERANCE PROTEINS PROBED AT THE RESIDUE LEVEL[1]

INTRODUCTION

Protein-based therapeutics (i.e., biologics) such as insulin, vaccines, and antibodies are among the most precise and effective drugs on the market. Yet the relatively short shelf-lives of proteins in aqueous solution, along with the high costs of refrigerated transport and storage, hinder their widespread use.[1-3] To increase their stability and mitigate challenges associated with the so-called “cold chain”, many biologics and industrial enzymes are dehydrated. However, given the key role water plays in protein structure and function, many proteins cannot withstand dehydration, irreversibly unfolding and aggregating during the process.[3-6]

To protect proteins from dehydration-induced damage, additives known as excipients are added before drying.[3-4, 6-7] Yet, despite decades of research, our understanding of the mechanisms of protein dehydration protection is limited, in part because we were unable to study the effects of dehydration on protein structure at high resolution.[8] In practice, our inadequate understanding means that an excipient formulation must be developed empirically for each protein, a time- and resource-intensive process that often fails.[3-4, 6, 9] Gaining a deeper understanding of

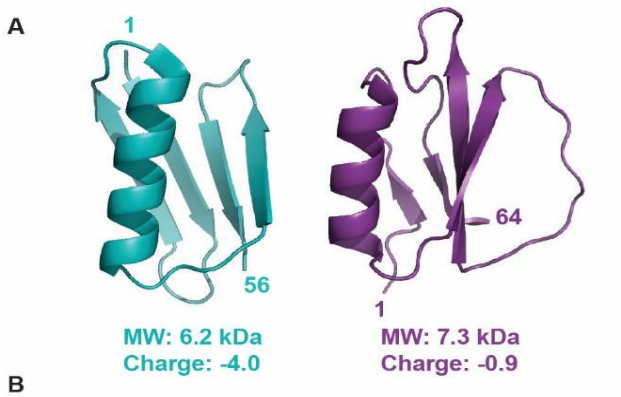
[1]*Submitted to a peer-reviewed journal with the same title and the following author list: Candice J. Crilly*, Julia A. Brom*, Owen Warmuth, Harrison J. Esterly, and Gary J. Pielak. *Co-first authors.*

dehydration protection and how it relates to client protein properties would enable the prediction of effective excipient formulations for dehydrated biologics, in turn making these life-saving drugs more affordable and accessible.[10]

To this end, we developed Liquid-Observed Vapor Exchange (LOVE) NMR, a solution NMR technique enabling the study of dehydrated protein structure and protection at the residue level.[11] Based on the well-established principle that amide protons are less likely to exchange with deuterons from the environment if involved in intra- or intermolecular H-bonds,[12-14] LOVE NMR uses solution NMR to quantify the extent of hydrogen–deuterium exchange between D_2O vapor and the amide protons of a dried protein. The percent of amide-proton signal remaining after vapor exchange (%Protected) reflects the fraction of the dry protein population for which a given residue is involved in an inter- or intramolecular H-bond.[11] Thus, LOVE NMR can determine where, and to what degree, excipients interact with dehydrated proteins and/or prevent dehydration-induced unfolding. Such information can be used to elucidate mechanisms of dehydration protection and explain why a given excipient will work for some client proteins but not others.

Here, we apply LOVE NMR to study the dehydration protection of two model proteins (Fig. 1A) – the B1 domain of staphylococcal protein G (GB1) and chymotrypsin inhibitor 2 (CI2) – by two classes of desiccation-tolerance proteins – the Cytosolic Abundant Heat Soluble (CAHS) proteins and Late Embryogenesis Abundant (LEA) proteins[15-16]— and two controls — bovine serum albumin (BSA) and gelatin. CAHS proteins comprise a family of intrinsically disordered proteins (IDPs)

unique to tardigrades, microscopic animals well-known for their ability to survive extreme stresses in their dehydrated “anhydrobiotic” form.[17-19]



B

Protectant	MW (kDa)	pI	Net Charge	κ
CAHS D	25.5	6.4	-0.7	.08
PvLEA4	16.2	5.6	-2.3	.05
Gelatin	10-250	7-9	--	.13*
BSA	66.4	5.3	-8.1	.13

Figure 4.1. Proteins used in this study. A) Structures of client proteins GB1 (left, PDB 2QMT) and CI2 (right, PDB 2CI2). B) Properties of protectant proteins. Net charge is calculated at pH 6.5. *κ calculated for porcine collagen α-1 chain preprotein.

CAHS proteins are necessary for tardigrades to survive desiccation and sufficient to protect heterologously-expressing cells and enzymes from desiccation damage.[20-21] LEA proteins, which are better studied, are IDPs implicated in the desiccation tolerance of many plants and animals and are thought to protect against dehydration-induced aggregation by acting as “molecular shields”. [22-26]

Although CAHS- and LEA- proteins have no sequence similarity (Fig. S4.1), they share several properties, including a relatively low molecular weight, a high degree of disorder and extensive charge patterning [represented by a near-zero κ value (Fig. 1B)].[27] Unlike LEA proteins, however, CAHS proteins form reversible, concentration-dependent gels.[28] To investigate the importance of disorder and

gelation in dehydration protection, we therefore also tested protection by the globular protein BSA and the disordered, gelling protein mixture, gelatin.

MATERIALS AND METHODS

Materials. Ampicillin, kanamycin sulfate, bovine serum albumin (Sigma Aldrich), trisodium citrate (Agros Organics), citric acid monohydrate and HEPES (Thermo Fisher) were used without further purification. H_2O with a resistivity >17 MΩ cm^{-1} was used to prepare buffers. Unflavored porcine gelatin (Knox) was dialyzed against H_2O to remove small molecules (ThermoScientific Snakeskin™ dialysis tubing, 3500 Da molecular weight cutoff), and lyophilized before resuspension in 1.5 mM HEPES to a final concentration of 5 g/L (w/w). pH values are direct readings, uncorrected for the deuterium isotope effect.[29] For LOVE NMR experiments, a constant relative humidity of 75 ± 5% (measured with a Fisherbrand TraceableGO™ Bluetooth datalogging digital hygrometer) was created as described.[11]

Expression and purification of client proteins. ^{15}N-enriched GB1 and CI2 were expressed and purified as described in the materials and methods section of Chapter 3.[30-31]

Expression and purification of CAHS D and PvLEA4. pET28b plasmids encoding the genes for CAHS D or PvLEA4, both fused to a N-terminal hexahistadine (His-) tag and a TEV protease cleavage site, were ordered from Gene Universal Inc. Vectors were transformed into Agilent BL21 Gold (DE3) E. coli as described [32]. A single colony was used to inoculate 100 mL of Lennox Luria broth (Fisher, 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) supplemented with the antibiotic kanamycin to a final concentration of 60 μg/mL. The culture was shaken at

37°C overnight (New Brunswick Scientific I26 incubator, 225 rpm). Ten mL of the overnight culture was used to inoculate 1 L of kanamycin-supplemented LB. One-L cultures were shaken at 37°C until they reached an optical density at 600 nm of 0.6 – 0.8, at which point protein expression was induced by adding isopropyl β-D-1-thiogalactopyranoside to the cultures (1 mM final concentration). Three hours after induction, cells were harvested via centrifugation at 4,000g. The cell pellet from each 1L culture was resuspended in 10 mL of 20 mM Tris, pH 7.5, and stored at -20 °C.

On the day of purification, frozen cell slurries from 3L worth of culture were thawed at room temperature and lysed via heat shock at 95°C for 15 min. Lysates were clarified by centrifugation at 15,000xg for 45 min, passed through a 0.45 µm filter, and purified via affinity chromatography on a nickel (Ni)-NTA column (5 mL GE HisTrap HP) using a GE AKTA Start FPLC.

To prevent proteins from gelling on the column, filtered lysates were mixed with an equal volume of 3 M urea containing 10 mM imidazole, 500 mM NaCl, and 20 mM sodium phosphate, pH 7.4 (HUA). The urea in HUA and in other urea solutions was deionized with 5 g/L Dowex® MB Mixed Ion Exchange Resin (Sigma) before removing the resin with a 0.22 µm filter (Corning Inc. 431161, Corning, NY, USA) and adding buffer salts.

Affinity chromatography was performed at room temperature in HUA. After sample loading, the Ni-NTA column was washed with 3 column volumes of HUA followed by a 29-column-volume gradient of 10 – 150 mM imidazole for CAHS D and 10 – 100 mM imidazole for PvLEA4. SDS-PAGE (Bio-Rad 4–20% Criterion™ TGX™ Gels) was used to identify fractions containing pure, His-tagged protein.

To remove the His-tag, the fractions were pooled, dialyzed at room temperature against 50 mM Tris (pH 8.1) containing 500 μM EDTA and 1 mM DTT for 4 h (ThermoScientific Snakeskin^{TM} dialysis tubing, 3500 Da molecular weight cutoff), and then incubated with 1 mg TEV protease at room temperature with gentle shaking for 16 – 24 h. After TEV digestion, the His-tag and protease were removed via incubation with loose Ni-NTA resin for 12 – 16 h.

To remove remaining buffer salts, the protein solution was subjected to 6 additional rounds of dialysis (>4 h each) at room temperature against deionized H_2O (ThermoScientific Snakeskin^{TM} dialysis tubing, 3500 Da molecular weight cutoff). Finally, the solution was heat-shocked to resolubilize any protein that precipitated during dialysis (both CAHS D and PvLEA4 are heat-soluble), passed through a 0.45 μm filter, flash-frozen, lyophilized and stored at -20°C.

Thermogravimetric Analysis (TGA). Aliquots of purified, lyophilized, ^{15}N-enriched GB1 were resuspended in 650 μL of 1.5-mM HEPES pH 6.5 with or without 5 g/L protectant to a final client protein concentration of 500 μM, flash-frozen, and lyophilized (LABCONCO FreeZone 1 Liter Benchtop Freeze Dry System) for 24 h. T_0 samples were analyzed immediately, while T_{24} samples were analyzed after 24 hours in a chamber with a controlled relative humidity of 75 ± 5% created as described above. Samples weighing ~0.5 to 1.5 mg were loaded into a TA Instruments model 550 thermogravimetric analyzer on an open Pt pan and heated from 25 to 170°C at a rate of 5°C/min under a N_2(g) sample purge of 60 mL/min and a balance purge of 40 mL/min. The well-defined mass loss ending around 130°C

was used to quantify the content of H_2O and D_2O (Fig. S4.4).[33-34] Thermograms were analyzed using Trios software to determine the weight change.

Differential Scanning Calorimetry (DSC). Aliquots of purified, lyophilized, ^{15}N-enriched GB1 were resuspended in 650 µL of 1.5-mM HEPES pH 6.5 with or without 5 g/L protectant to a final client protein concentration of 500 µM, flash-frozen, and lyophilized as described above. T_0 samples were analyzed immediately, while T_{24} samples were analyzed after 24 hours in a chamber with a controlled relative humidity of 75 ± 5% created as described above. Samples weighing ~1 to 2.5 mg were sealed in Tzero Hermetic Aluminum pans and then loaded into a TA Instruments DSC 250 equipped with a TA Instruments Refrigerated Cooling System 90. An identical, empty pan was used as a reference in all measurements. The sample cell was under 50 mL/min nitrogen purge.

To eliminate differing thermal history effects on the reversible glass transition, samples were heated at 7.5°C/min to ~5°C above their glass transition temperature, and then cooled to 10°C, where they remained for 1 minute.[35-38] Samples were again heated at 7.5°C/min to about 15°C above their denaturation temperature, and from this scan both glass transition temperature and denaturation temperature were measured. Samples were cooled to 10°C, and once more heated at 7.5°C/min to confirm the irreversibility of denaturation, as expected in proteins with water contents as in these samples [37]. Thermograms were analyzed using Trios V5.1.0.56403 software, with the midpoint of the endothermal shift in the baseline on the thermogram reported as the glass transition temperature (T_g) (Fig. S4.5),[35-36, 38-39]

and the minimum of the denaturation endothermic peak reported as the denaturation temperature (T_m) (Fig S4.6).[40-42] Samples were analyzed in duplicate.

NMR. Unless noted, LOVE NMR experiments were performed in triplicate on a Bruker Avance III HD spectrometer with a cryogenic QCI probe at a ^{1}H Larmor frequency of 600 MHz. Sensitivity-enhanced ^{15}N-^{1}H heteronuclear single-quantum coherence (HSQC) spectra were acquired in ~10 min (128 increments in the ^{15}N dimension, 4 scans per increment) with sweep widths of 3041, 2311 Hz in the ^{15}N dimension and 8418, 10822 Hz in the ^{1}H dimension for GB1 and CI2 (GB1, CI2), respectively. Spectra were processed with NMRPipe.[43] Crosspeaks were integrated using the ellipse integration tool in NMRViewJ.[44]

Liquid Observed Vapor Exchange (LOVE) NMR. For each experiment, two identical aliquots of purified, lyophilized, ^{15}N-enriched protein were resuspended in 650 µL of 1.5-mM HEPES pH 6.5 with or without 5 g/L protectant to a final client protein concentration of 500 µM, flash-frozen, and lyophilized (LABCONCO FreeZone 1 Liter Benchtop Freeze Dry System) for 24 h. Following lyophilization, one sample, designated T_0, was immediately resuspended in 650 µL of cold quench buffer (100 mM citrate buffer, pH 4.5, 90% H_2O/10% D_2O) and transferred to an NMR spectrometer for spectrum acquisition at 4°C. The second sample, designated T_{24}, was placed, with the cap open, in a chamber with a controlled relative humidity of ~75%, prepared as described.[11]

After 24 h, the T_{24} sample was resuspended in 650 µL of cold quench buffer and a spectrum obtained using the same parameters that were used for the T_0 sample. To enable back-exchange correction, the T_{24} sample was left in the spectrometer at 4°C

for ~12 h, during which time an additional 10 - 12 spectra were acquired. The time between resuspension and initiation of the first T_{24} spectrum acquisition was 10 min, ~8 min of which were spent at 4 °C. Spectra were processed with NMRPipe,[43] crosspeaks were integrated using the ellipse volume integration tool in NMRViewJ,[44] and %Protected for each condition was determined as described in the Materials and Methods section of Chapter 3.

RESULTS

Water content and glass transition temperature.

To determine if differences in protection are related to a cosolute's ability to retain water and/or form glasses, we performed thermogravimetric analysis (TGA) and differential scanning calorimetry (DSC) on GB1 samples lyophilized alone or in the presence of 5 g/L CAHS D, PvLEA4, gelatin, or BSA (Fig. 4.2). The samples possess 7% to 8% water by mass, which accounts for less than one layer of surface water on GB1 (Table S4.1). After exposure to 75% relative humidity for 24 h at room temperature, samples absorb a similar amount of water, increasing the water content to ~14% (Fig. S4.2),[11, 40] which still comprises less than a single layer of water on the client protein surface. The glass transition temperatures (T_g) of all samples are also similar, ranging from 54 °C to 56 °C (Fig 4.2B).

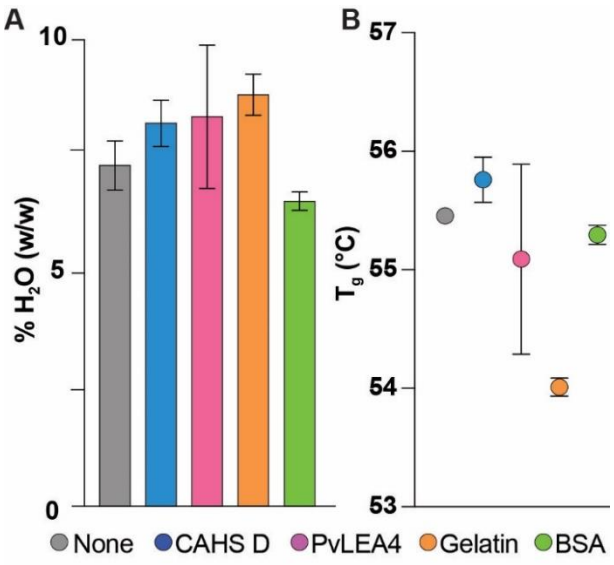


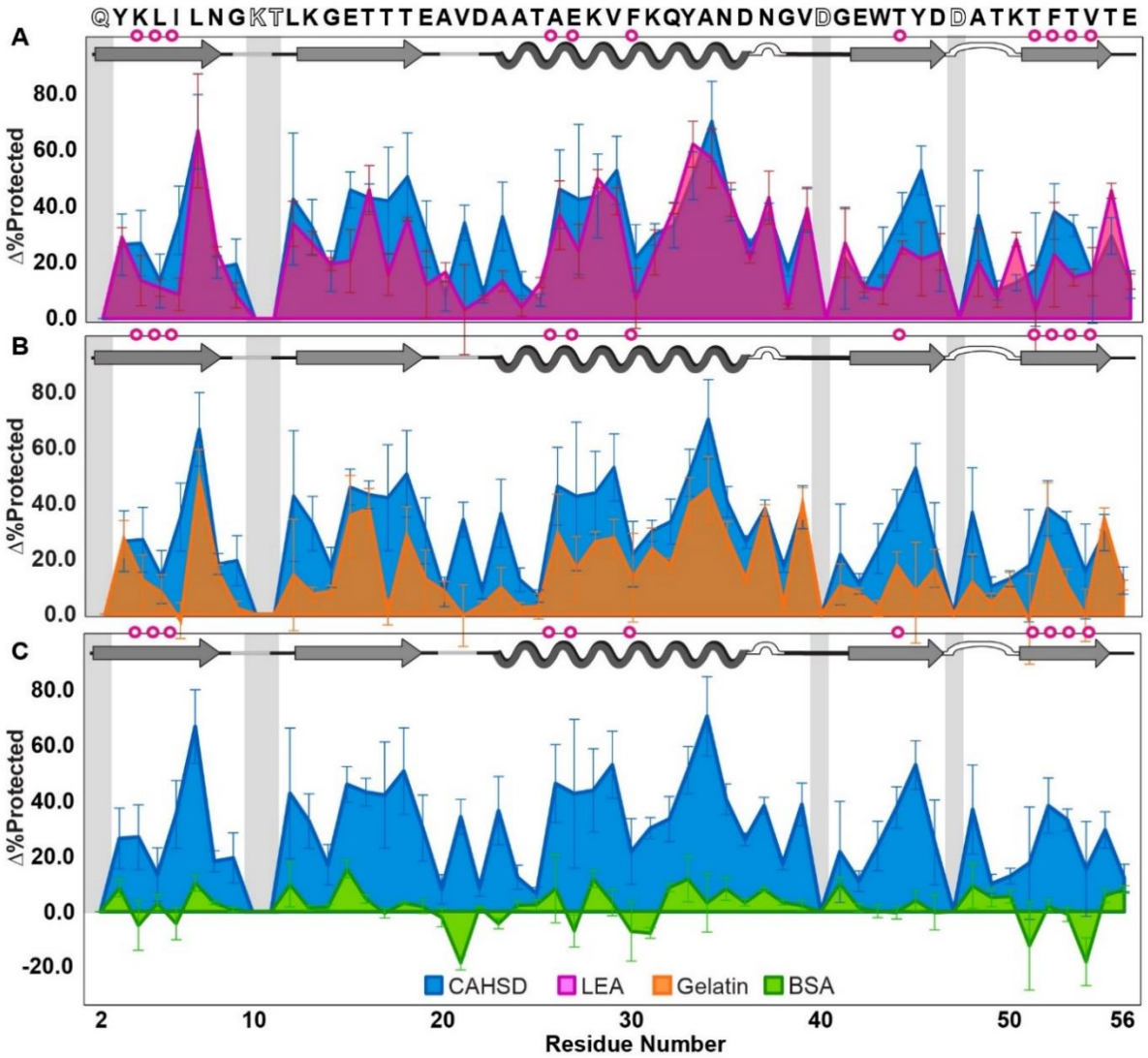
Figure 4. 2. Percent H_2O by mass (A) and glass transition temperature (T_g) (B) of lyophilized GB1 with indicated protectants. Samples (650 µL) comprising 500 µM (~3 g/L) GB1 alone or with 5 g/L indicated protectant were lyophilized for 24 h, and immediately analyzed. Uncertainties in $\%H_2O$ are the standard deviation from 3 independent measurements for GB1 with CAHS D, 4 independent measurements for GB1 with PvLEA4, and the range of 2 independent measurements for the other proteins.

Residue-level dehydration protection.

To compare where and how well each protein protects clients from dehydration-induced unfolding, we performed LOVE NMR on GB1 and CI2 alone and in the presence of the CAHS D, PvLEA4, gelatin or BSA. As with the TGA and DSC experiments, LOVE NMR experiments were performed on lyophilized samples that, before drying, comprised 500 µM (~3 g/L) client protein alone or with 5 g/L protectant in a total volume of 650 µL. %Protected values (Tables S4.2 and S4.3) of the dried proteins were subtracted from those dried in the presence of cosolute to give the change in dry-state protection (Δ%Protected).

Drying in the presence of 5 g/L CAHS D, LEA, or gelatin increases the average %Protected value of GB1 by 30 ±20%, 20 ±20%, and 20 ±10%, respectively, where the uncertainty is the standard deviation from the mean (Fig. 4.3 A and B). By

contrast, drying in 5 g/L BSA brings minimal additional protection, increasing the average %Protected by 2 ±7% (Fig. 4.3C).

Figure 4.3. Change in GB1 dry-state protection due to freeze-drying in CAHS D versus LEA (A), gelatin (B), and BSA (C). 

Δ%Protected = $\%Protected_{+cosolute} - \%Protected_{-cosolute}$. Primary structure of GB1 (PDB 2QMT) is shown at the top. Secondary-structure is shown atop each panel. Magenta circles indicate solution global-unfolding residues. Gray boxes indicate missing data from rapid back exchange (open letters in primary structure). Error bars represent standard deviations from the mean propagated from triplicate analysis.

CAHS D, PvLEA4, and gelatin give similar profiles for GB1 (Fig. 4.3A and B), with those of LEA and CAHS D appearing most similar. CAHS D consistently out-performs the other protein cosolutes at residues I6, V21, A23, T44, Y45, and T51, all but two of which (V21 and Y45) are, or immediately neighbor, global-unfolding residues.[45]. However, PvLEA4 outperforms CAHS D at residues K50 and T55.

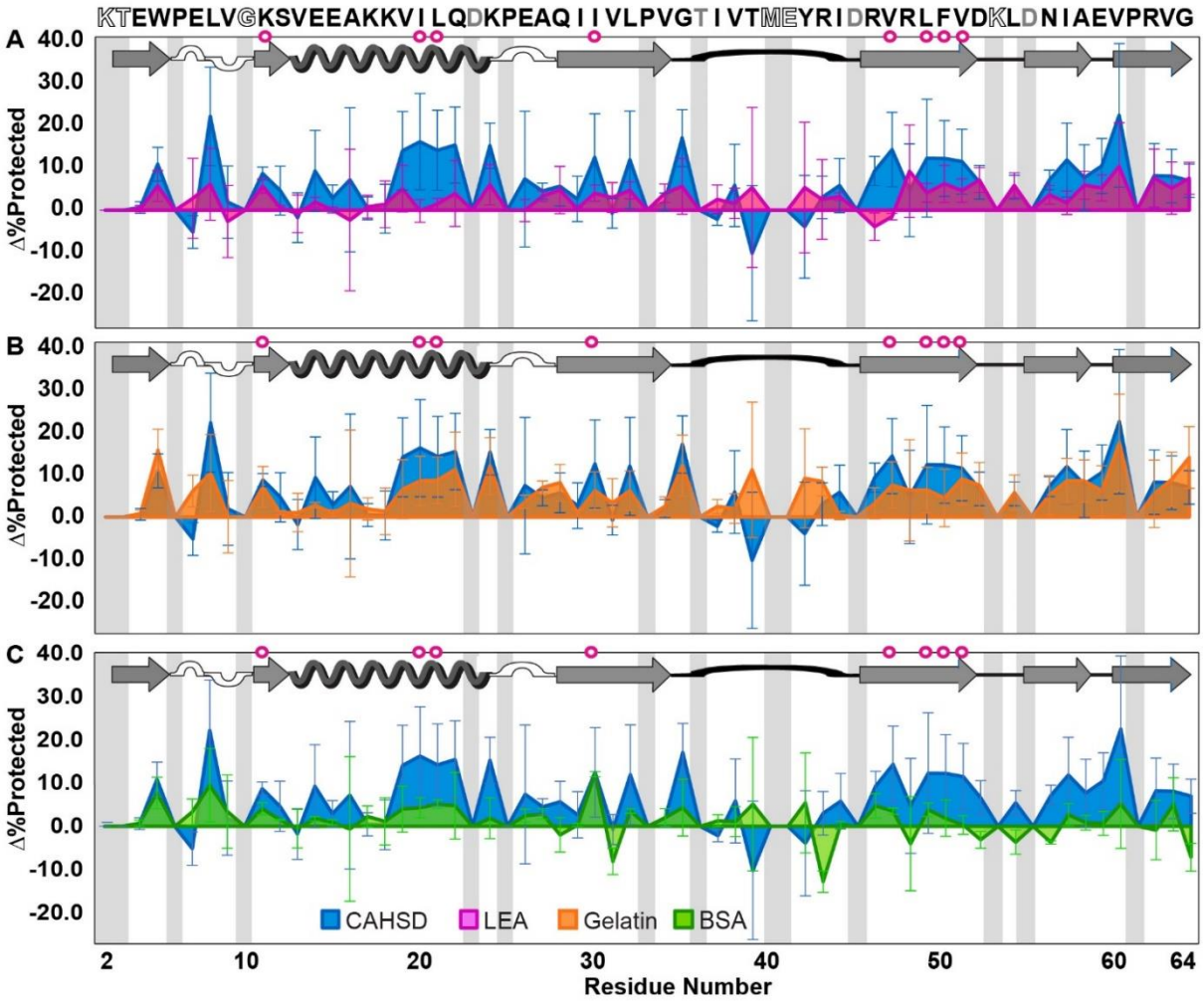


Figure 4.4. Change in CI2 dry-state protection due to freeze-drying in CAHS D versus LEA (A), gelatin (B), and BSA (C). Δ%Protected = %Protected$_{+cosolute}$ – %Protected$_{-cosolute}$. Primary structure of CI2 (PDB 2CI2) is shown at the top. Secondary- structure is shown atop each panel. Magenta circles indicate solution global-unfolding residues. Gray boxes indicate missing data from rapid back exchange (open letters in primary structure) and/or overlapping crosspeaks (gray letters in primary structure). Error bars represent standard deviations from the mean propagated from triplicate analysis.

CAHS D, PvLEA4, and gelatin also provide CI2 with additional protection (Fig. 4.4), but the average Δ%Protected values, 7 ±7%, 3 ±3%, and 6 ±4%, respectively, are smaller than those for GB1. Like GB1, CI2 is least protected by BSA, with a 2 ±4% increase. Although all protein cosolutes offer a similar level of protection for most regions of CI2, inspection suggests that with this client protein, CAHS D behaves most like gelatin, especially in the α-helix and C-terminal β-sheets (Fig. 4.4B). This result contrasts with the protection profiles of GB1, which show CAHS D behaving most similarly to PvLEA4.

DISCUSSION

The observation that CAHS D and PvLEA4 both protect GB1 better than CI2 is consistent with the idea that dehydration protection is client-protein-dependent. Moreover, the observation that the disordered proteins generally protect the same regions of GB1 suggests they operate via a similar mechanism.

There are three main models of cosolute-mediated dehydration protection: the vitrification-, water-replacement-, and “molecular shield”- hypotheses. The vitrification hypothesis poses that cosolutes protect against dehydration damage by enveloping the client protein in a glassy matrix, which in turn inhibits large motions such as global unfolding; this idea is supported by studies showing a positive correlation between T_g and protective ability.[46] The water replacement hypothesis posits that cosolutes protect proteins from dehydration-induced unfolding by replacing water-mediated H-bonds,[9, 47] and is supported by the observation that the protective ability of certain sugars correlates with their ability to maintain the position

of the amide II band, which is sensitive to changes in secondary structure and hydrogen-bonding with water.[48] Finally, the “molecular shield” hypothesis envisions protectants forming barriers between partially-unfolded client proteins, thus preventing irreversible aggregation.[25, 49] Support for this mechanism arises from the observation that a LEA protein inhibits aggregation but fails to prevent intramolecular changes upon drying.[50]

The observation that CAHS D generally performs better than PvLEA4 despite having a lower protectant-to-client mole ratio (~2:5 versus 3:5) implies that most interactions between the desiccation-protective proteins and client proteins are non-specific. It is unlikely, however, that dehydration protection arises from enhanced water retention or the properties of the glassy matrix because the dried mixtures protect to different degrees yet possess nearly the same water content and T_g (Fig. 4.2). These observations, in turn, suggest that vitrification is not the main source of the differences in protection.

The difference between the average Δ%Protected experienced by GB1 and CI2 may provide a window into the mechanism of dehydration-protective proteins. Although both client proteins are similarly stable in solution and possess a 4β + 1α architecture, GB1 possesses several-fold more net charge than CI2 (Fig. 4.1A) under our conditions, suggesting that electrostatic interactions are important in protection. Given that both client proteins are negatively charged and that BSA has the most negative charge of all the protectants (Fig. 4.1B), it is unlikely that repulsive electrostatic interactions, which stabilize proteins in solution,[51] are responsible for protection. Perhaps electrostatic interactions play a role in sequestering client

proteins into the matrix of protectants, a phenomenon observed with other dehydration-protective proteins.[52]

It is intriguing that the residue-level pattern of GB1 protection by CAHS D and PvLEA4 resemble that of GB1 chemical shift perturbations caused by the ionic liquid 1-butyl-3-methylimidazolium bromide (Fig. S4.3, $p<.05$),[53] suggesting that direct interaction with alternating charges is involved in dehydration protection. Moreover, the observation that disordered proteins generally perform best as dehydration protectants suggests that extended conformations,[27] which provide ample surface for intermolecular interactions, are important for dehydration protection. The apparent importance of charge patterning and surface interactions in dehydration protection aligns well with both the water-replacement and molecular shield hypotheses. A role for expanded surface area is also consistent with the observation that most dehydration-protective proteins are disordered.[54-56]

The observation that CAHS D prevents dehydration-induced exposure of global-unfolding residues better than PvLEA4 suggests that it is better at maintaining native tertiary structure. Additionally, CAHS D performing most like gelatin in the protection of CI2 suggests that gelation of CAHS D supplements the dehydration protection offered by electrostatic interactions. Perhaps gelation prevents global unfolding via confinement in a gel matrix[57] whose structure is maintained upon drying. Finally, the observation that PvLEA4 protects the polar and/or charged GB1 residues K50 and T55 better than CAHS D, yet generally performs worse than CAHS D in protecting the more neutral client protein CI2, suggests that PvLEA4 protects almost exclusively via electrostatic interactions.

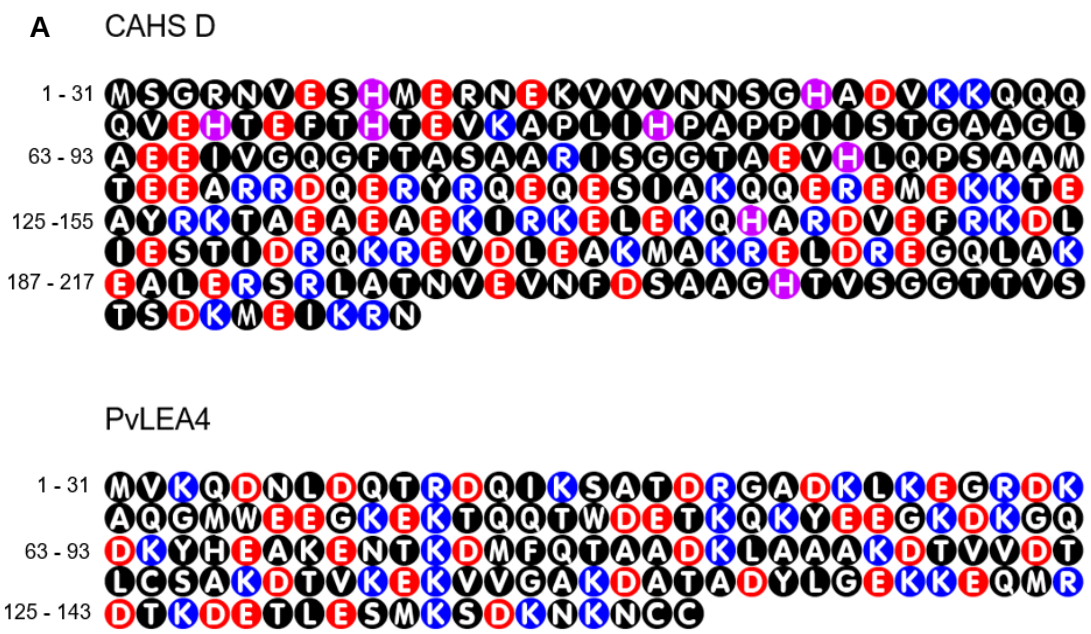
CONCLUSIONS

Our residue-level study of dehydration protection of two client proteins by two IDPs involved in animal desiccation tolerance confirm that protection is client-protein dependent and that desiccation-tolerance proteins generally protect better than those unrelated to dehydration stress response. Comparing protection trends of CAHS D and PvLEA4 to those of BSA and gelatin suggest that disorder and charge patterning are integral to dehydration protection, supporting both the molecular shield and water-replacement hypotheses. Application of LOVE NMR to other client proteins and protectants will further advance our understanding of dehydration protection, knowledge that can in turn be used to streamline the production, storage and shipping of protein products, making these valuable and lifesaving products less expensive and more accessible.

SUPPLEMENTAL INFORMATION

Supplemental figures

A



B

Job Title	**CAHS D**
RID	JK899JMW11N *Search expires on 08-29 10:36 am* Download All
Program	Blast 2 sequences Citation
Query ID	lcl|Query_84359 (amino acid)
Query Descr	CAHS D
Query Length	227
Subject ID	lcl|Query_84361 (amino acid)
Subject Descr	PvLEA4
Subject Length	143

⚠ No significant similarity found. For reasons why, click here

Figure S4.1. Primary structure comparisons of CAHS D and PvLEA4. A) Amino acid sequences of CAHS D and PvLEA4, where red circles indicate negatively charged residues, blue positively charged, gray neutral, and purple histidines. B) Results of attempted sequence alignment of CAHS D and PvLEA4 using protein BLAST with default algorithm parameters.[58]

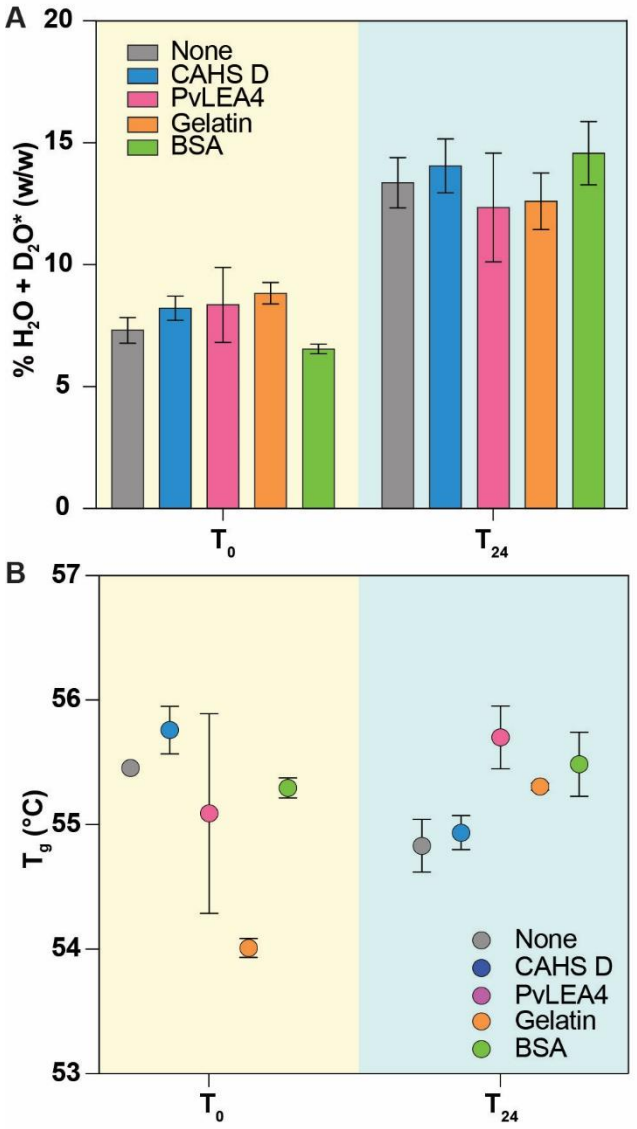


Figure S4.2. Water content and glass transition temperature of dehydrated protein mixtures before and after exposure to 75% relative humidity (D_2O). A) Percent *H_2O (T_0) or H_2O + D_2O (T_{24}) by weight of samples lyophilized from 650 μL of 500 μM GB1 and 5 g/L of indicated protectants before and after incubation in 75% RH D_2O chamber, T_0 and T_{24}, respectively. Error bars are the standard deviation from 3 independent measurements for GB1 with CAHS D, 4 independent measurements for GB1 with PvLEA4, and represent the range of 2 independent measurements for all other data plotted. B) Glass transition temperature, T_g, of formulations of lyophilized 650 μL aliquots of 500 μM GB1 and 5 g/L of indicated protectants before and after incubation in 75% RH D_2O chamber, T_0 and T_{24}, respectively. Error bars are the range of 2 independent measurements.

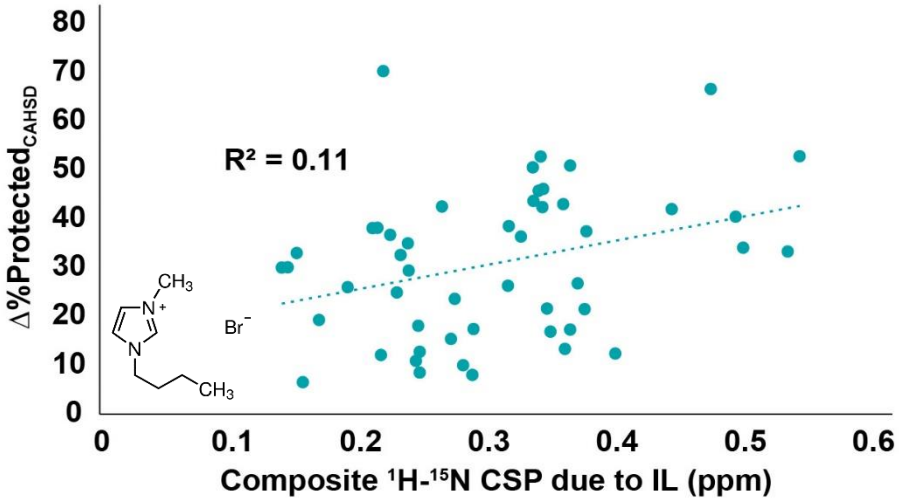


Figure S4.3. Correlation between composite chemical shift perturbations induced by an ionic liquid and Δ%Protected induced by drying with 5 g/L CAHS D. All non-quench-labelled GB1 residues were included in this correlation analysis (n=50). ^{1}H-^{15}N composite chemical shift perturbations (CSPs) induced by the presence of 50% v/v 1-butyl-3-methylimidazolium bromide ([C_4-mim]Br), an ionic liquid (IL), are from Warner et al. [53]. Chemical structure of [C_4-mim]Br shown inside plot. $p < 0.05$.

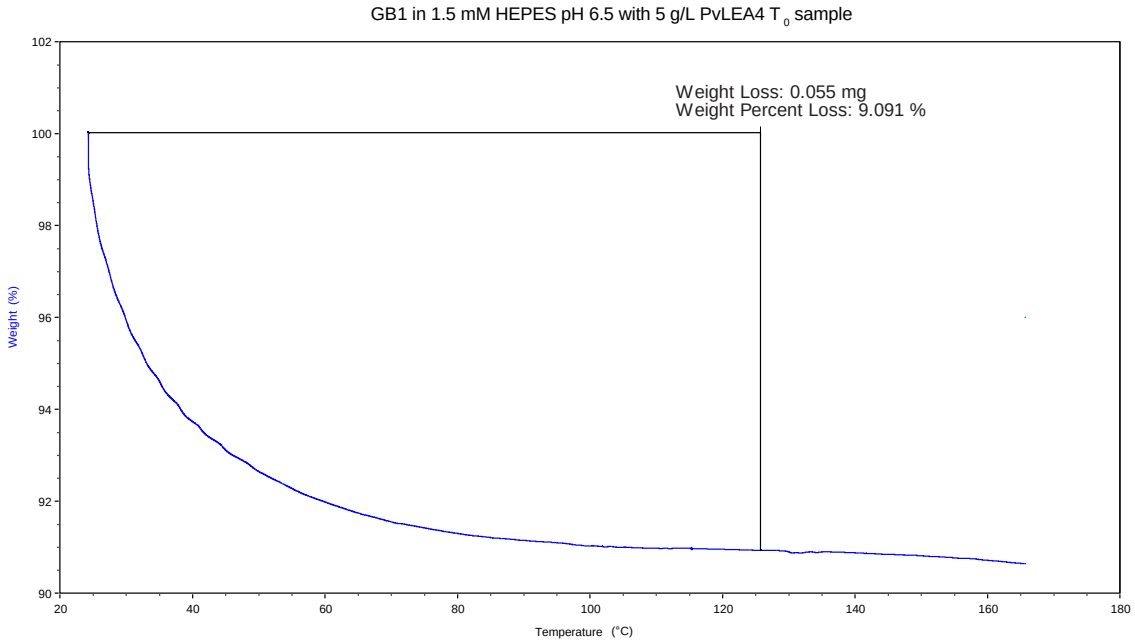


Figure S4.4. A representative thermogram from thermogravimetric analysis. Analysis of the initial change in weight due to water loss is shown.

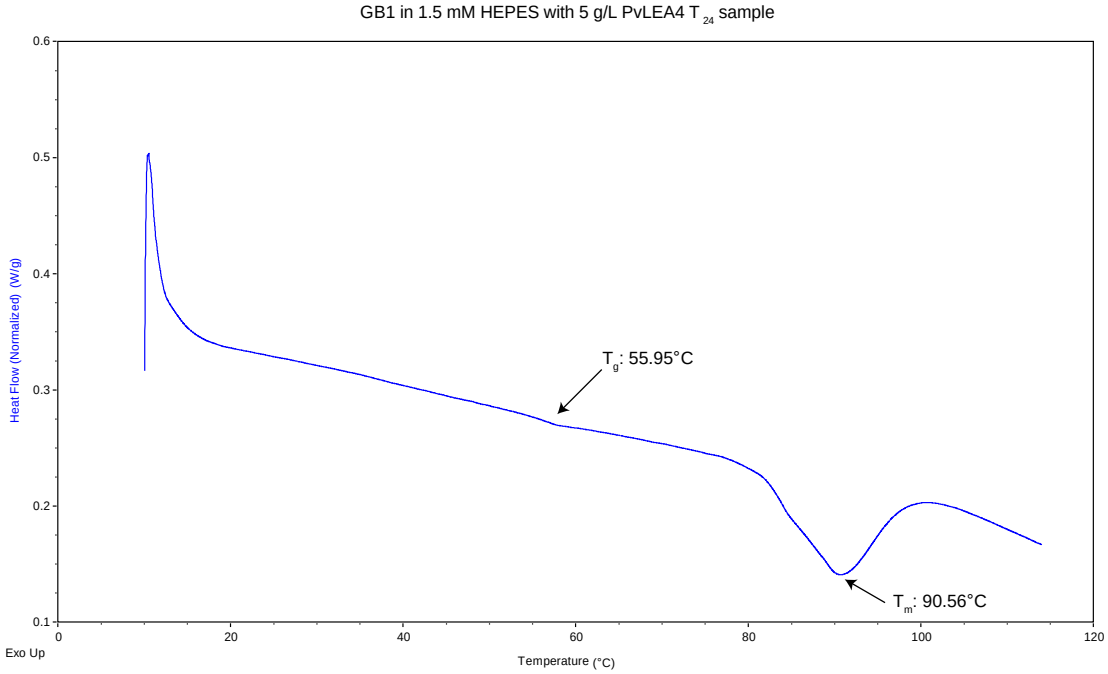


Figure S4.5. A representative thermogram from the second heating scan of differential scanning calorimetry, from which measurements were taken. Glass transition temperature (T_g) and denaturation temperature (T_m) are indicated.

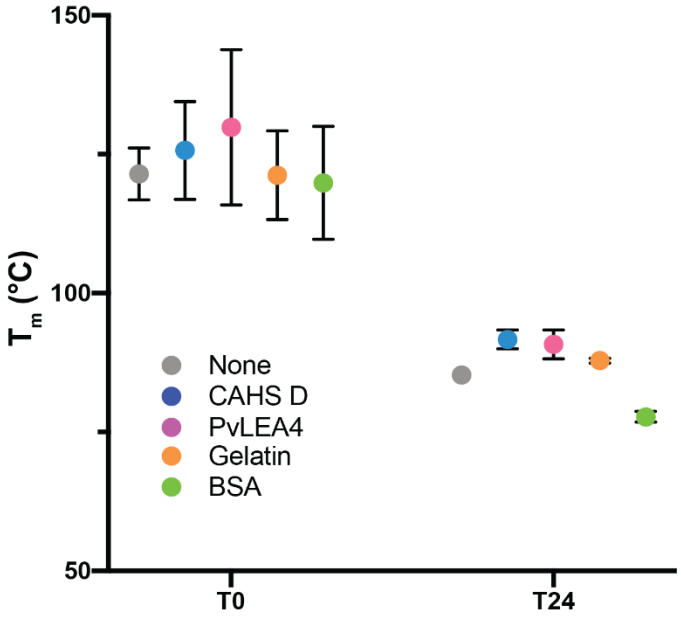


Figure S4.6. Denaturation temperature, T_m, of dehydrated protein mixtures before and after exposure to 75% relative humidity (D_2O). T_m of GB1 lyophilized from 650 μL aliquots of 500 μM GB1 alone or with 5 g/L of indicated protectants before and after incubation in 75% RH D_2O chamber, T_0 and T_{24}, respectively. As expected, as moisture content increases after incubation in the humid chamber, T_m of GB1 decreases.[40, 59] T_m is similar for all protectants, except with BSA as a protectant, T_m is notably low at T_{24}. Error bars are the standard deviation from 3 independent measurements for GB1 with CAHS D and PvLEA4, and represent the range of 2 independent measurements for all other data plotted.

Supplemental tables

Table S4.1. Hydrated fraction of GB1 solvent-accessible surface area (SASA). .

Hydrated fraction of GB1 SASA ± uncertainty		
Protectant	**T_0**	**T_{24}**
None	0.40 ± 0.03	0.70 ± 0.06
CAHS D	0.44 ± 0.03	0.74 ± 0.07
PvLEA4	0.46 ± 0.09	0.6 ± 0.1
Gelatin	0.49 ± 0.03	0.65 ± 0.07
BSA	0.35 ± 0.01	0.77 ± 0.08

For this calculation, we assume that all water in the sample is bound to GB1, that all water in the T_0 sample is H_2O, and that all water in the T_{24} sample is D_2O. These values were calculated by multiplying the molar ratio of water to GB1 (as measured by TGA) by the average amount of protein surface covered by a water molecule (20 $Å^2$),[60] and then dividing that value by the surface area of the native solution structure of GB1 (3727 $Å^2$) as determined by the PyMOL get_area function for PDB 2QMT. Origins of uncertainties are described in the caption of Fig. 2 in the main text.

Table S4.2. Average %Protected values of GB1 dried alone or in the presence of 5 g/L protectant. .

Residue	No protectant	+CAHS D	+PvLEA4	+Gelatin	+BSA
Y3	61 ± 3	90 ± 10	90 ± 1	88 ± 6	69
K4	81 ± 9	108 ± 8	95 ± 1	94 ± 1	76
L5	88 ± 3	102 ± 9	99.2 ± 0.3	96 ± 6	91
I6	92 ± 6	130 ± 10	100.2 ± 0.2	89 ± 3	87
L7	42 ± 3	110 ± 10	109 ± 20	90 ± 10	52
8N	16 ± 1	34 ± 4	40 ± 1	31 ± 3	19
9G	16 ± 1	36 ± 9	25 ± 4	19 ± 2	17
12L	49 ± 8	90 ± 20	82 ± 1	60 ± 20	58
13K	25 ± 1	60 ± 10	51 ± 4	32 ± 2	26
14G	17 ± 1	34 ± 7	37 ± 2	26 ± 2	19
15E	66 ± 3	112 ± 6	90 ± 10	100 ± 10	81
16T	35 ± 2	78 ± 5	81 ± 9	73 ± 7	39
17T	32 ± 2	70 ± 20	48 ± 7	33 ± 4	31
18T	60.7 ± 0.4	110 ± 20	96 ± 1	90 ± 10	64
19E	50 ± 2	80 ± 10	60 ± 10	60 ± 10	51
20A	8 ± 3	16 ± 4	25 ± 1	17 ± 1	6
21V	83 ± 3	117 ± 6	90 ± 20	80 ± 10	64
22D	13.1 ± 0.3	22 ± 3	21 ± 1	16 ± 1	15
23A	45 ± 2	80 ± 10	58 ± 3	54 ± 7	40
24A	13 ± 1	26 ± 4	17 ± 3	16 ± 1	16
25T	15 ± 2	22 ± 1	28.0 ± 0.1	18 ± 4	18
26A	70 ± 10	116 ± 7	107 ± 1	99 ± 7	78
27E	72 ± 6	110 ± 30	96 ± 8	90 ± 10	64
28K	28 ± 3	70 ± 20	78 ± 1	54 ± 2	39
29V	50 ± 5	100 ± 10	92.0 ± 0.4	78 ± 5	53
30F	90 ± 10	116 ± 5	101 ± 1	110 ± 10	87
31K	68 ± 2	98 ± 3	91 ± 8	91 ± 7	60
32Q	13 ± 1	46 ± 8	52 ± 1	30 ± 2	21
33Y	23 ± 8	74 ± 2	85 ± 1	62 ± 5	34
34A	40 ± 10	110 ± 10	99 ± 1	87 ± 5	44
35N	9 ± 4	50 ± 4	52 ± 4	37 ± 3	17
36D	11 ± 1	37 ± 2	32 ± 1	23.3 ± 0.4	14
37N	14 ± 1	52 ± 3	57 ± 9	51 ± 2	22
38G	11 ± 1	28 ± 2	15.2 ± 0.5	14 ± 1	14
39V	16 ± 2	54 ± 7	55 ± 7	56 ± 5	18
41G	35 ± 2	60 ± 20	60 ± 10	45 ± 8	45
42E	12 ± 1	23 ± 3	23 ± 1	20 ± 1	14
43W	14 ± 2	38 ± 9	25 ± 5	16 ± 2	14
44T	73 ± 2	110 ± 7	98.1 ± 0.2	90 ± 4	72
45Y	59 ± 4	112 ± 8	80 ± 10	70 ± 20	63
46D	63 ± 6	90 ± 10	86 ± 2	79 ± 3	62
48A	26 ± 8	60 ± 10	46 ± 9	38 ± 6	35
49T	10 ± 3	20 ± 1	17 ± 1	14 ± 1	15
50K	13 ± 2	26 ± 2	42 ± 1	25 ± 3	19
51T	100 ± 20	120 ± 10	102 ± 2	97 ± 5	86
52F	86 ± 2	120 ± 10	110 ± 20	110 ± 20	87
53T	84 ± 3	117 ± 3	98 ± 1	95 ± 5	83
54V	81 ± 9	100 ± 20	98.2 ± 0.4	81 ± 4	63
55T	27 ± 2	56 ± 6	72 ± 1	61 ± 3	32
56E	8 ± 1	20 ± 5	21 ± 2	19 ± 1	16

Standard deviation from the mean for three independent measurements is reported for all conditions except for GB1 dried in the presence of BSA, which was measured once.

Table S4.3. Average %Protected values of CI2 dried alone or in the presence of 5 g/L protectant. .

Residue	No protectant	+CAHS D	+PvLEA4	+Gelatin	+BSA
4	11.3 ± 0.7	12 ± 1	11.7 ± 0.3	12.1 ± 0.5	12
5	60 ± 4	71 ± 2	65.9 ± 0.4	76 ± 3	68
7	37 ± 3	32 ± 2	40 ± 9	43 ± 2	40
8	68 ± 9	90 ± 8	73.8 ± 0.7	78 ± 3	77
9	38 ± 9	39 ± 1	35 ± 1	37.6 ± 0.9	41
11	86 ± 1	94.6 ± 0.6	92 ± 1	93 ± 5	90
12	11.0 ± 0.1	16 ± 6	11.6 ± 0.5	12.0 ± 0.5	13
13	18 ± 5	16 ± 4	17.0 ± 0.1	18.9 ± 0.8	17
14	13.9 ± 0.5	20 ± 10	16 ± 1	16.7 ± 0.4	16
15	15.2 ± 0.6	18 ± 3	15.2 ± 0.3	15.9 ± 0.7	16
16	90 ± 20	95 ± 3	85.7 ± 0.9	91 ± 4	87
17	16 ± 3	16 ± 1	17 ± 1	17.6 ± 0.9	18
18	24 ± 6	24 ± 2	25.1 ± 0.8	25.0 ± 0.7	25
19	85 ± 5	99 ± 8	90 ± 2	92 ± 5	89
20	94 ± 2	110 ± 10	94 ± 1	103 ± 5	99
21	91 ± 1	105 ± 10	92.5 ± 0.8	100 ± 5	96
22	71 ± 8	86 ± 5	74.4 ± 0.9	82 ± 4	75
24	79 ± 5	94 ± 3	85.0 ± 0.6	91 ± 5	81
26	24 ± 2	30 ± 20	24 ± 2	27.4 ± 0.5	26
27	18 ± 1	22 ± 1	20.9 ± 0.1	24.6 ± 0.9	20
28	19 ± 4	24 ± 3	24 ± 4	27 ± 2	17
29	12.2 ± 0.3	15 ± 5	12.5 ± 0.1	13.1 ± 0.7	13
30	93.5 ± 0.6	110 ± 10	98 ± 2	100 ± 4	106
31	79 ± 3	78 ± 2	82 ± 2	82 ± 5	70
32	96.4 ± 0.5	110 ± 10	101 ± 2	103 ± 5	100
34	13 ± 2	15.1 ± 0.4	17.1 ± 0.7	15.6 ± 0.8	15
35	58 ± 7	75 ± 1	63.4 ± 0.1	70 ± 3	62
37	24 ± 1	22.1 ± 0.4	26.8 ± 3.2	27 ± 1	26
38	14 ± 4	20 ± 9	15.8 ± 0.8	16.2 ± 0.5	15
39	60 ± 20	46 ± 4	60 ± 10	68 ± 4	62
42	40 ± 10	41 ± 4	50 ± 10	54 ± 2	50
43	13 ± 3	16 ± 4	15 ± 9	21 ± 2	15
44	13.3 ± 0.8	19 ± 6	16.5 ± 0.4	13.6 ± 0.6	14
46	89 ± 3	98 ± 2	85 ± 2	92 ± 22	93
47	83.9 ± 0.7	98 ± 9	82.4 ± 0.2	92 ± 5	87
48	50 ± 10	58 ± 1	62.4 ± 0.8	60 ± 5	49
49	98 ± 2	110 ± 10	102 ± 2	104 ± 5	102
50	86 ± 4	98 ± 8	92 ± 1	90 ± 5	88
51	87 ± 2	99 ± 7	92 ± 2	96 ± 6	87
52	85 ± 2	91 ± 4	92 ± 2	92 ± 5	82
54	65 ± 3	70.3 ± 0.2	70.8 ± 0.9	71 ± 3	61
56	94.4 ± 0.4	101 ± 2	98 ± 2	100 ± 5	91
57	86 ± 3	98 ± 8	88 ± 1	95 ± 5	89
58	90 ± 2	98 ± 8	96 ± 3	99 ± 5	91
59	92 ± 3	102 ± 6	97 ± 1	98 ± 5	92
60	80 ± 10	110 ± 10	93 ± 1	100 ± 5	88
62	56 ± 7	64 ± 3	63.4 ± 0.5	61 ± 4	55
63	32 ± 6	40 ± 2	37.3 ± 0.5	41 ± 2	37
64	62 ± 3	69 ± 2	70 ± 2	76 ± 7	55

Standard deviation from the mean for three independent measurements is reported for all conditions except for CI2 dried in the presence of BSA, which was measured once.

REFERENCES

1. Morrow, T.; Felcone, L. H., Defining the difference: what makes biologics unique. *Biotechnol. Healthc.* **2004,** *1*, 24-29.

2. Frokjaer, S.; Otzen, D. E., Protein drug stability: a formulation challenge. *Nat Rev Drug Discov* **2005,** *4* (4), 298-306.

3. Piszkiewicz, S.; Pielak, G. J., Protecting enzymes from stress-induced inactivation. *Biochemistry* **2019,** *58*, 3825-3833.

4. Wang, W., Lyophilization and development of solid protein pharmaceuticals. *Int. J. Pharm.* **2000,** *203*, 1-60.

5. Walters, R. H.; Bhatnagar, B.; Tchessalov, S.; Izutsu, K. I.; Tsumoto, K.; Ohtake, S., Next generation drying technologies for pharmaceutical applications. *J Pharm Sci* **2014,** *103* (9), 2673-2695.

6. Bjelosevic, M.; Zvonar Pobirk, A.; Planinsek, O.; Ahlin Grabnar, P., Excipients in Freeze-dried Biopharmaceuticals: Contributions Toward Formulation Stability and Lyophilisation Cycle Optimisation. *Int. J. Pharm.* **2020,** *576*, 1-12.

7. Ohtake, S.; Kita, Y.; Arakawa, T., Interactions of formulation excipients with proteins in solution and in the dried state. *Adv Drug Deliv Rev* **2011,** *63* (13), 1053-73.

8. Moorthy, B. S.; Iyer, L. K.; Topp, E. M., Characterizing Protein Structure, Dynamics and Conformation in Lyophilized Solids. *Current Pharmaceutical Design* **2015,** *21* (40), 5845-5853.

9. Mensink, M. A.; Frijlink, H. W.; van der Voort Maarschalk, K.; Hinrichs, W. L., How sugars protect proteins in the solid state and during drying (review): Mechanisms of stabilization in relation to stress conditions. *Eur. J. Pharm. Biopharm.* **2017,** *114*, 288-295.

10. Hill, A. B.; Kilgore, C.; McGlynn, M.; Jones, C. H., Improving global vaccine accessibility. *Curr Opin Biotechnol* **2016,** *42*, 67-73.

11. 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.

12. Englander, S. W.; Kallenbach, N. R., Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q Rev Biophys* **1983,** *16* (4), 521-655.

13. Desai, U. R.; Osterhout, J. J.; Klibanov, A. M., Protein Structure in the Lyophilized State: A Hydrogen Isotop exchange/NMR Study with Bovine Pancreatic Trypsin Inhibitor. *Journal of the American Chemical Society* **1994,** *116*, 9420-9422.

14. Percy, A. J.; Rey, M.; Burns, K. M.; Schriemer, D. C., Probing protein interactions with hydrogen/deuterium exchange and mass spectrometry-a review. *Anal Chim Acta* **2012,** *721*, 7-21.

15. Li, S.; Chakraborty, N.; Borcar, A.; Menze, M. A.; Toner, M.; Hand, S. C., Late embryogenesis abundant proteins protect human hepatoma cells during acute desiccation. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109* (51), 20859.

16. Yamaguchi, A.; Tanaka, S.; Yamaguchi, S.; Kuwahara, H.; Takamura, C.; Imajoh-Ohmi, S.; Horikawa, D. D.; Toyoda, A.; Katayama, T.; Arakawa, K.; Fujiyama, A.; Kubo, T.; Kunieda, T., Two novel heat-soluble protein families abundantly expressed in an anhydrobiotic tardigrade. *PLoS ONE* **2012,** *7* (8), e44209-e44209.

17. Rebecchi, L.; Guidetti, R.; Borsari, S.; Altiero, T.; Bertolani, R., Dynamics of long-term anhydrobiotic survival of lichen-dwelling tardigrades. *Hydrobiologia* **2006,** *558* (1), 23-30.

18. Jonsson, K. I.; Rabbow, E.; Schill, R. O.; Harms-Ringdahl, M.; Rettberg, P., Tardigrades survive exposure to space in low Earth orbit. *Curr. Biol.* **2008,** *18* (17), R729-R731.

19. Yamaguchi, A.; Tanaka, S.; Yamaguchi, S.; Kuwahara, H.; Takamura, C.; Imajoh-Ohmi, S.; Horikawa, D. D.; Toyoda, A.; Katayama, T.; Arakawa, K.; Fujiyama, A.; Kubo, T.; Kunieda, T., Two novel heat-soluble protein families abundantly expressed in an anhydrobiotic tardigrade. *PLoS ONE* **2012,** *7* (8), e44209.

20. Boothby, T. C.; Tapia, H.; Brozena, A. H.; Piszkiewicz, S.; Smith, A. E.; Giovannini, I.; Rebecchi, L.; Pielak, G. J.; Koshland, D.; Goldstein, B., Tardigrades use intrinsically disordered proteins to survive desiccation. *Mol. Cell* **2017,** *65* (6), 975-984 e5.

21. 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.

22. Wise, M. J.; Tunnacliffe, A., POPP the question: what do LEA proteins do? *Trends Plant Sci* **2004,** *9* (1), 13-17.

23. Battaglia, M.; Olvera-Carrillo, Y.; Garciarrubio, A.; Campos, F.; Covarrubias, A. A., The enigmatic LEA proteins and other hydrophilins. *Plant Physiol* **2008,** *148* (1), 6-24.

24. Hand, S. C.; Menze, M. A.; Toner, M.; Boswell, L.; Moore, D., LEA proteins during water stress: not just for plants anymore. *Annu Rev Physiol* **2011,** *73*, 115-34.

25. Hatanaka, R.; Hagiwara-Komoda, Y.; Furuki, T.; Kanamori, Y.; Fujita, M.; Cornette, R.; Sakurai, M.; Okuda, T.; Kikawada, T., An abundant LEA protein in the anhydrobiotic midge, PvLEA4, acts as a molecular shield by limiting growth of aggregating protein particles. *Insect Biochem Mol Biol* **2013,** *43* (11), 1055-67.

26. Kamilari, M.; Jorgensen, A.; Schiott, M.; Mobjerg, N., Comparative transcriptomics suggest unique molecular adaptations within tardigrade lineages. *BMC Genomics* **2019,** *20* (1), 607.

27. Das, R. K.; Pappu, R. V., Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110* (33), 13392-13397.

28. Hesgrove, C. S.; Nguyen, K. H.; Biswas, S.; Childs, C. A.; Shraddha, K.; Medina, B. X.; Alvarado, V.; Sukenik, S.; Yu, F.; Malferrari, M.; Francia, F.; Venturoli, G.; Martin, E. W.; Holehouse, A. S.; Boothby, T. C., Molecular Swiss Army Knives: Tardigrade CAHS Proteins Mediate Desiccation Tolerance Through Multiple Mechanisms. *bioRxiv* **2021**, 2021.08.16.456555.

29. Glasoe, P. K.; Long, F. A., Use of Glass Electrodes to Measure Acidities in Deuterium Oxide. *J. Phys. Chem.* **1960,** *64* (1), 188-190.

30. Charlton, L. M.; Barnes, C. O.; Li, C.; Orans, J.; Young, G. B.; Pielak, G. J., Residue-Level Interrogation of Macromolecular Crowding Effects on Protein Stability. *J. Am. Chem. Soc.* **2008,** *130* (21), 6826-6830.

31. Monteith, W. B.; Pielak, G. J., Residue level quantification of protein stability in living cells. *Proc. Natl. Acad. Sci. U.S.A.* **2014,** *111* (31), 11335-11340.

32. Boothby, T. C.; Tapia, H.; Brozena, A. H.; Piszkiewicz, S.; Smith, A. E.; Giovannini, I.; Rebecchi, L.; Pielak, G. J.; Koshland, D.; Goldstein, B., Tardigrades Use Intrinsically Disordered Proteins to Survive Desiccation. *Mol. Cell* **2017,** *65* (6), 975-984.

33. Joan C. May, E. G., Roscoe M. Wheeler, and Jerry Westy, Determination of residual moisture in freeze-dried viral vaccines: Karl Fischer, gravimetric and thermogravimetric methodologies. *Journal of Biological Standardization* **1982,** *10*, 249-259.

34. Joan C. May, R. M. W., and Elizabeth Grim, The Gravimetric Method for the Determination of Residual Moisture in Freeze-Dried Biological Products. *Cryobiology* **1989,** *26*, 277-284.

35. Kalichevsky, M. T.; Jaroszkiewicz, E. M.; Blanshard, J. M. V., Glass transition of gluten. 1: Gluten and gluten-sugar mixtures. *Int J Biol Macromol* **1992,** *14*, 257-266.

36. Aguilera, J. M.; Levi, G.; Karel, M., Effect of water content on the glass transition and caking of fish protein hydrolyzates. *Biotechnol. Prog.* **1993,** *9*, 651-654.

37. Sochava, I. V.; Smirnova, O. I., Heat capacity of hydrated and dehydrated globular proteins. Denaturation increment of heat capacity. *Food Hydrocolloids* **1993,** *6*, 513-524.

38. Rouilly, A.; Orliac, O.; Silvestre, F.; Rigal, L., DSC study on the thermal properties of sunflower proteins according to their water content. *Polymer* **2001,** *42*, 10111-10117.

39. Duddu, S. P.; Monte, P. R. D., Effect of glass transition temperature on the stability of lyophilized formulations containing a chimeric therapeutic monoclonal antibody. *Pharm. Res.* **1997,** *14* (5), 591-595.

40. Bell, L. N.; Hageman, M. J.; Muraoka, L. M., Thermally Induced Denaturation of Lyophilized Bovine Somatotropin and Lysozyme As Impacted by Moisture and Excipients. *J. Pharm. Sci.* **1995,** *84* (6), 707-712.

41. Medina-Vivanco, M.; Sobral, P. J. A.; Sereno, A. M.; Hubinger, M. D., Denaturation and the Glass Transition Temperatures of Myofibrillar Proteins from Osmotically Dehydrated Tilapia: Effect of Sodium Chloride and Sucrose. *Int. J. Food Prop.* **2007,** *10* (4), 791-805.

42. Samouillan, V.; Delaunay, F.; Dandurand, J.; Merbahi, N.; Gardou, J. P.; Yousfi, M.; Gandaglia, A.; Spina, M.; Lacabanne, C., The use of thermal techniques for the characterization and selection of natural biomaterials. *J. Funct. Biomat.* **2011,** *2* (3), 230-48.

43. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995,** *6* (3), 277-93.

44. Johnson, B. A.; Blevins, R. A., NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **1994,** *4* (5), 603-14.

45. Orban, J.; Alexander, P.; Bryan, P.; Khare, D., Assessment of stability differences in the protein G B1 and B2 domains from hydrogen-deuterium exchange: comparison with calorimetric data. *Biochemistry* **1995,** *34* (46), 15291-300.

46. Grasmeijer, N.; Stankovic, M.; de Waard, H.; Frijlink, H. W.; Hinrichs, W. L. J., Unraveling protein stabilization mechanisms: Vitrification and water replacement in a glass transition temperature controlled system. *Biochim. Biophys. Acta, Proteins Proteomics* **2013,** *1834* (4), 763-769.

47. Crowe, J. H.; Clegg, J. S.; Crowe, L. M., Anhydrobiosis: the water replacement hypothesis. In *The Properties of Water in Foods ISOPOW 6*, Reid, D. S., Ed. Springer: Boston, MA, 1998; pp 440-455.

48. Carpenter, J. F.; Crowe, J. H., An Infrared Spectroscopic Study of the Interactions of Carbohydrates with Dried Proteins. *Biochemistry* **1989,** *28*, 3916-3922.

49. Tunnacliffe, A.; Wise, M. J., The continuing conundrum of the LEA proteins. *Naturwissenschaften* **2007,** *94* (10), 791-812.

50. Chakrabortee, S.; Tripathi, R.; Watson, M.; Schierle, G. S.; Kurniawan, D. P.; Kaminski, C. F.; Wise, M. J.; Tunnacliffe, A., Intrinsically disordered proteins as molecular shields. *Mol Biosyst* **2012,** *8* (1), 210-9.

51. 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.

52. Belott, C.; Janis, B.; Menze, M. A., Liquid-liquid phase separation promotes animal desiccation tolerance. *Proc. Natl. Acad. Sci. U.S.A.* **2020,** *117* (44), 27676-27684.

53. Warner, L.; Gjersing, E.; Follett, S. E.; Elliott, K. W.; Dzyuba, S. V.; Varga, K., The effects of high concentrations of ionic liquid on GB1 protein structure and dynamics probed by high-resolution magic-angle-spinning NMR spectroscopy. *Biochem. Biophys. Rep.* **2016,** *8*, 75-80.

54. Graether, S. P.; Boddington, K. F., Disorder and function: a review of the dehydrin protein family. *Front. Plant Sci.* **2014,** *5*, 576.

55. Das, R. K.; Ruff, K. M.; Pappu, R. V., Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2015,** *32*, 102-112.

56. Boothby, T. C.; Pielak, G. J., Intrinsically Disordered Proteins and Desiccation Tolerance: Elucidating Functional and Mechanistic Underpinnings of Anhydrobiosis. *BioEssays* **2017,** *39* (11), 1700119.

57. Eggers, D. K.; Valentine, J. S., Molecular confinement influences protein structure and enhances thermal protein stability. *Protein Sci.* **2001,** *10* (2), 250-261.

58. McGinnis, S.; Madden, T. L., BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* **2004,** *32* (Web Server issue), W20-W25.

59. Tsereteli, G. I.; Belopolskaya, T. B.; Grunina, N. A.; Vaveliouk, O. L., Calorimetric study of the glass transition process in humid proteins and DNA. *J. Therm. Anal. Calorim.* **2000,** *62*, 89-99.

60. John A. Rupley, E. G. a. G. C., Water and globular proteins. *Trends in Biochemical Sciences* **1983,** (January 1983), 18-22.

CHAPTER 5: NEXT STEPS FOR METHOD ESTABLISHMENT

I. INCREASE PRECISION AND ACCURACY

To our knowledge, LOVE NMR is the highest resolution technique currently available for studying protein structure and interactions in the dry state. However, the relatively high level of uncertainty associated with the %Protected values measured by LOVE NMR (~5-10% standard deviation from triplicate analysis) makes it difficult to identify significant, residue-level differences between experimental conditions. The experimental uncertainty arises from several sources, including slight variations in sample preparation, e.g. differences in the total time and strength of vacuum applied during freeze-drying due to addition of other samples to lyophilizer, or differences in the amount of time spent above 4°C due to speed of sample transfer, etc.; slight variations in spectrometer settings after calibration to a new sample, e.g. the 90° pulse length, quality of the shims, and efficacy of water suppression; impact of cosolutes and deuteration on NMR relaxation and, in turn, peak volume; and imperfect phasing and peak integration. Although differences in sample preparation can be minimized by further streamlining the experimental workflow and purchasing better instrumentation, optimizing acquisition parameters to minimize variation in spectrometer sensitivity is more difficult, especially considering that one must minimize that amount of time the sample spends in solution before

signal acquisition. Fortunately, the last two potential sources of variation – sample differences in signal relaxation and imperfect data processing – are related areas that can be explored to improve the accuracy of LOVE NMR.

Account for differences in signal relaxation rates between T_0 and T_{24}.

Like all other spectroscopic techniques, the signal produced by an NMR spectrometer is concentration-dependent and decays with time.[1] The former is how amide-proton hydrogen-deuterium exchange works – as the concentration of amide protons decreases due to exchange with deuterium from the environment, so does the signal from ^{1}H-^{15}N resonances, which allows one to monitor the extent and/or rate of exchange in real-time via NMR. The latter property – the time-dependent decay of the signal – is modulated by the characteristics of the system being analyzed, e.g. a probe's local chemical environment, how quickly the probe is rotating in solution, whether/how rapidly the probe is undergoing chemical exchange with another species, etc.;[2] and, because most modern NMR experiments are done through Fourier-transform- (FT-) NMR,[3] the time-dependent decay of the signal determines the width of the peak (linewidth) (Fig. 5.1).

In 1D FT-NMR experiments, all signals are collected immediately after excitation, and therefore the peak areas of all signals share (effectively) the same proportionality to concentration, though the peak shapes may differ. However, in 2D NMR experiments, coherence must be transferred from the first nucleus to a second – a process that is rarely 100% efficient; in addition, the frequency of the second dimension is determined indirectly *via* time incrementation (Fig. 5.2).

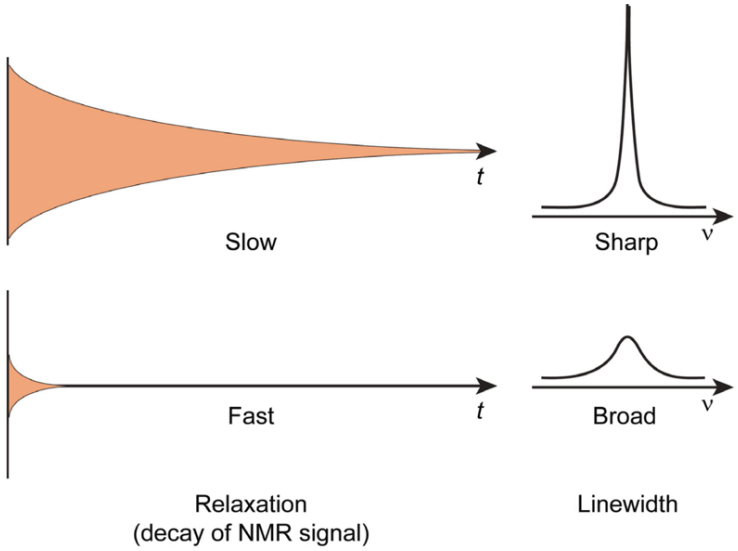


Figure 5.1. Rate of signal relaxation influences peak width. Due to the Heisenberg uncertainty principle ($\Delta E \Delta t > \hbar/2$, where $\hbar$ is Planck's constant), the slower a signal relaxes (i.e. the longer you can observe it) the more certain you can be about the energy (frequency) of that signal, hence the smaller linewidth. Figure adapted from "Key Labeling Technologies to Tackle Sizeable Problems in RNA Structural Biology"[4] with permission from the International Journal of Molecular Sciences.

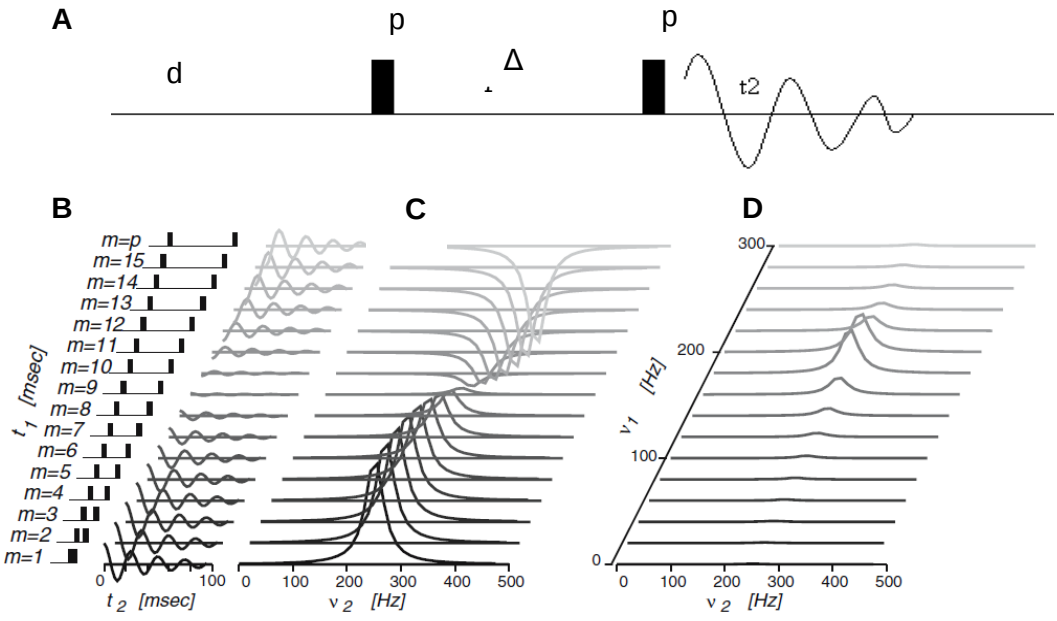


Figure. 5.2. The role of time incrementation in a multidimensional NMR experiment. A) Simplified pulse-sequence of a 2D NMR experiment. After initial excitation with pulse p1, magnetization will precess at the frequency of the indirectly-detected nucleus (e.g. ^{15}N) during the evolution time t1, before being transferred to the directly-detected nucleus (typically ^{1}H) with the mixing pulse p2 and subsequent detection over the period t2. Before beginning another pulse train in

which the evolution time t1 has been incremented, there is a delay d1 to allow magnetization to return to equilibrium. B) At each increment, the time t1 between the excitation and mixing pulses is elongated, which shifts the free-induction decay (FID) by a phase that depends on the frequency of the indirectly-detected nucleus. C) Fourier transformation of the directly-detected FID at each increment. D) Fourier-transformation of the second dimension. Adapted by permission from Springer Nature from Fundamentals of Protein NMR Spectroscopy[5] copyright 2006.

Due to this time-delay in signal acquisition, the signal collected for a given probe in a multi-dimensional NMR experiment is modulated by its unique, time-dependent relaxation, among other dampening factors; this is why the ^{1}H-^{15}N peak volumes of the amide-protons of a protein are not the same, despite being at the same concentration (Table 5.1).

Table 5.1. Representative ^{1}H-^{15}N peak volumes of a fully-protonated GB1 sample.

Residue	Volume
Q2	223
Y3	179
K4	146
L5	161
I6	153
L7	185
N8	226
G9	216
K10	165

In theory, the simplest way to determine the concentration of amide-protons from a 2D ^{1}H-^{15}N HSQC experiment is to compare each peak volume to that of the corresponding peak in a spectrum (acquired using the same acquisition parameters) of a sample of known concentration; this is what we do with LOVE NMR.[6] However, the concentration calculated from this ratiometric comparison is only accurate if the relaxation properties of the two samples are the same. For LOVE NMR, assuming the relaxation properties of the T_0 and T_{24} samples are the same may not always be correct, because deuteration of neighboring amide-protons (due to local unfolding

and exchange with D_2O vapor) could theoretically affect dipolar relaxation.[7] Additionally, deuteration can alter the properties of cosolutes such as protective proteins or peptides,[8-9] which can in turn affect their interactions with the client protein, inducing a change in client protein relaxation.

One potential way to account for differences in relaxation is by using a two-dimensional time-zero extrapolated HSQC pulse sequence such as the one developed by Hu, Westler, and Markley.[10] In this experiment, the entire HSQC pulse sequence between the first ^1H excitation pulse and signal acquisition is repeated an additional 1-2 times before the signal is acquired. The data are then processed individually and plotted against the number of HSQC pulse trains used (1, 2, or 3) to determine the linear relationship and extrapolate back to zero time (0 repetitions, i.e. no relaxation). The downside to this approach is that the experiment takes more time, during which back-exchange may be occurring. Though it may be possible to calculate the rate of back-exchange separately, one should be wary of back-exchange altering the relaxation properties of the nucleus, which could affect the time-zero extrapolation.

Alternatively, one can use a standard (or sensitivity-enhanced) ^1H-^{15}N HSQC experiment and, after proper processing, fit the peak shape using freely-available programs such as TITAN[11] to determine the rate of transverse relaxation (R_2) – the primary source of relaxation in protein NMR. Then, the fitted relaxation rate for each peak can theoretically be used to back-calculate the peak volume before any relaxation occurred.[12] However, quantitative peak fitting to obtain R_2 often requires model specification (e.g. you must state whether your protein is in chemical

exchange, and with how many species), which is not always known. Additionally, peak-fitting is quite sensitive to imperfect spectral phasing - which brings us to another potential way to improve the accuracy and precision of LOVE NMR:

Use better peak phasing and integration methods.

Currently, LOVE NMR data are processed, i.e. zero-filled, filtered, Fourier-transformed, phase-adjusted, and baseline-corrected,[13] in NMRPipe (see processing script in Appendix) and integrated automatically using the ellipse tool in NMRViewJ.[14-15] In NMRPipe, spectral phase-adjustment is performed manually by the user, who inspects each peak of the Fourier-transformed spectrum by eye and adjusts the zero-order (frequency-independent) and first-order (frequency-dependent) phasing parameters to maximize the number of peaks that have an ideal Lorentzian peak shape entirely above the baseline (Fig. 5.3).[16-17]

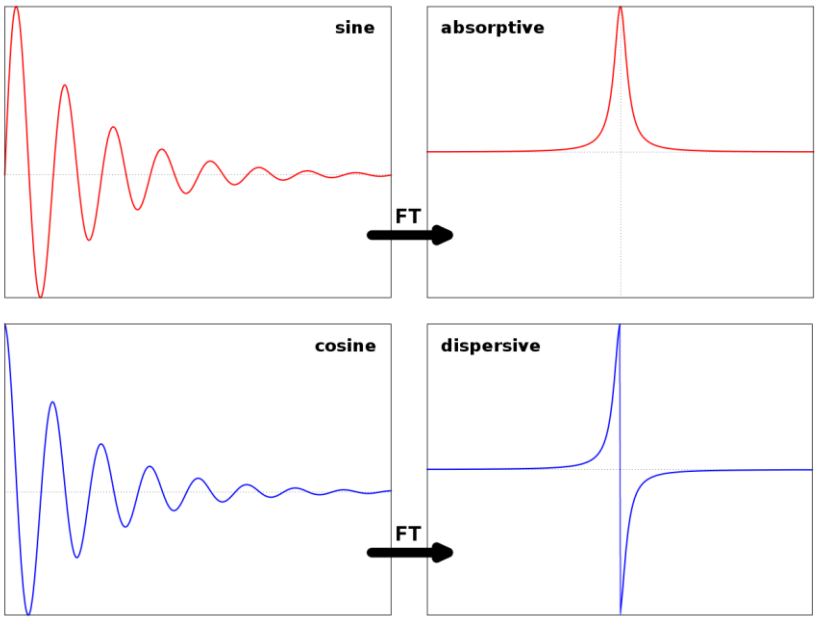


Figure 5.3. How signal phase determines peak shape. *Source: UCSD NMR Facility blog.*

However, given that peak phase is affected by several frequency-dependent factors and can be distorted by the use of non-optimal spectral acquisition parameters (e.g. scan repetition time set to less than $5xT_1$, the longitudinal relaxation time),[18-19] it is extremely challenging to obtain perfectly-phased peaks for all 50+ amide protons of a protein. And unfortunately, extracting peak volumes from improperly phased spectra using programs such as NMRViewJ or TopSpin, which integrate peaks by summing-up positive signal intensity over an automatically- or semi-automatically determined “footprint” area, can lead to volume under-estimations. Moreover, using the aforementioned “footprint” integration method to extract peak volumes of overlapping peaks may lead to volume over-estimation due to “double counting”; this is why severely overlapped peaks are currently excluded from LOVE NMR analyses.

To avoid such problems, one may opt to use automated phasing methods; however, such methods are not yet optimized and do not solve the “double counting” problem that occurs with footprint-based integration of overlapped peaks. To overcome the latter problem and enable more thorough data extraction from crowded HSQC spectra, one can use volume-extraction methods that integrate a fitted peak, such as PINT;[20] but again, peak fits are not immune from poor phasing. Current research is exploring the use of combined phasing-fitting algorithms for 1D NMR spectra,[21] which, if extrapolated to two-dimensions, may be useful for LOVE NMR in the future.

II. EXPLORE CONDITION EFFECTS

LOVE NMR is still in the proof-of-concept stage. To convince the scientific community that LOVE NMR is a valid way to probe dehydrated protein structure and interactions at the residue level, we must gain (and publish) a deeper understanding of the inner-workings of the method. This is why I believe that in addition to implementing measures to improve the accuracy and precision of LOVE NMR, future work should thoroughly explore how modifying experimental conditions such as temperature during vapor exchange, relative humidity, and sample pH before drying impact the results of LOVE NMR experiments.

Fortunately, the developers of solid-state HDX mass spectrometry (ssHDX-MS), a technique similar to LOVE NMR that provides peptide-level information on dehydrated protein structure, have studied how experimental conditions impact their results and have used these data to create quantitative models of the vapor exchange process.[22-25] Given that ssHDX-MS and LOVE NMR share nearly the same experimental procedure, we expect (but should verify) that the impact of changing conditions will be similar to those seen in ssHDX-MS.

CONCLUDING REMARKS

Residue-level information on protein structure and dynamics has provided countless insights into biochemical phenomena, some of which led to paradigm shifts and still others that inspired the design of life-saving medicines. However, before any of these insights could be made, the groundwork had to be laid for the techniques that produced that residue-level information. I believe that improving

LOVE NMR through the approaches outlined in this chapter will make it better poised to produce truly ground-breaking insights into protein-water and protein-excipient interactions.

REFERENCES

1. Bloch, F., Nuclear Induction. *Physical Review* **1946,** *70* (7-8), 460-474.

2. Keeler, J., Understanding NMR Spectroscopy. 2004.

3. Kumar, A.; Welti, D.; Ernst, R. R., NMR Fourier zeugmatography. *J. Magn. Reson.* **1975,** *18* (1), 69-83.

4. Dayie, K. T., Key labeling technologies to tackle sizeable problems in RNA structural biology. *Int. J. Mol. Sci.* **2008,** *9* (7), 1214-1240.

5. Rule, G. S.; Hitchens, T. K., Two Dimensional Homonuclear J-Correlated Spectroscopy. In *Fundamentals of Protein NMR Spectroscopy*, Springer Netherlands: Dordrecht, 2006; pp 169-195.

6. 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.

7. Nietlispach, D.; Clowes, R. T.; Broadhurst, R. W.; Ito, Y.; Keeler, J.; Kelly, M.; Ashurst, J.; Oschkinat, H.; Domaille, P. J.; Laue, E. D., An Approach to the Structure Determination of Larger Proteins Using Triple Resonance NMR Experiments in Conjunction with Random Fractional Deuteration. *J. Am. Chem. Soc.* **1996,** *118* (2), 407-415.

8. Stadmiller, S. S.; Pielak, G. J., Enthalpic stabilization of an SH3 domain by D2O. *Protein Sci.* **2018,** *27* (9), 1710-1716.

9. Canrinus, T. R.; Cerpentier, F. J.; Feringa, B. L.; Browne, W. R., Remarkable solvent isotope dependence on gelation strength in low molecular weight hydro-gelators. *Chem. Commun.* **2017,** *53* (10), 1719-1722.

10. Hu, K.; Westler, W. M.; Markley, J. L., Simultaneous Quantification and Identification of Individual Chemicals in Metabolite Mixtures by Two-Dimensional Extrapolated Time-Zero 1H−13C HSQC (HSQC0). *J. Am. Chem. Soc.* **2011,** *133* (6), 1662-1665.

11. Waudby, C. A.; Ramos, A.; Cabrita, L. D.; Christodoulou, J., Two-Dimensional NMR Lineshape Analysis. *Sci. Rep.* **2016,** *6* (1), 24826.

12. Rai, R. K.; Tripathi, P.; Sinha, N., Quantification of Metabolites from Two-Dimensional Nuclear Magnetic Resonance Spectroscopy: Application to Human Urine Samples. *Anal. Chem.* **2009,** *81* (24), 10232-10238.

13. Rule, G. S.; Hitchens, T. K., Introduction to Signal Processing. In *Fundamentals of Protein NMR Spectroscopy*, Springer Netherlands: Dordrecht, 2006; pp 65-88.

14. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995,** *6* (3), 277-93.

15. Johnson, B. A., Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **2004,** *278*, 313-52.

16. Binczyk, F.; Tarnawski, R.; Polanska, J., Strategies for optimizing the phase correction algorithms in Nuclear Magnetic Resonance spectroscopy. *Biomed. Eng. Online* **2015,** *14 Suppl 2* (Suppl 2), S5-S5.

17. Ernst, R. R., Numerical Hilbert transform and automatic phase correction in magnetic resonance spectroscopy. *J. Magn. Reson.* **1969,** *1* (1), 7-26.

18. Sheberstov, K. F.; Sistaré Guardiola, E.; Jeannerat, D., Everything you wanted to know about phase and reference frequency in one- and two-dimensional NMR spectroscopy. *Magn. Reson. Chem.* **2020,** *58* (5), 376-389.

19. Torres, A. M.; Price, W. S., Common problems and artifacts encountered in solution-state NMR experiments. *Concepts in Magnetic Resonance Part A* **2016,** *45A* (2), e21387.

20. Ahlner, A.; Carlsson, M.; Jonsson, B.-H.; Lundström, P., PINT: a software for integration of peak volumes and extraction of relaxation rates. *J. Biomol. NMR* **2013,** *56* (3), 191-202.

21. Matviychuk, Y.; Steimers, E.; von Harbou, E.; Holland, D. J., Improving the accuracy of model-based quantitative nuclear magnetic resonance. *Magn. Reson.* **2020,** *1* (2), 141-153.

22. Tukra, R.; Gardner, S.; Topp, E. M., Effects of temperature and relative humidity in D2O on solid-state hydrogen deuterium exchange mass spectrometry (ssHDX-MS). *Int. J. Pharm.* **2021,** *596*, 120263.

23. Kammari, R.; Topp, E. M., Solid-State Hydrogen–Deuterium Exchange Mass Spectrometry (ssHDX-MS) of Lyophilized Poly-d,l-Alanine. *Mol. Pharm.* **2019,** *16* (7), 2935-2946.

24. Bersin, L. M.; Patel, S. M.; Topp, E. M., Effect of ‘pH’ on the Rate of Pyroglutamate Formation in Solution and Lyophilized Solids. *Mol. Pharm.* **2021,** *18* (8), 3116-3124.

25. Moussa, E. M.; Wilson, N. E.; Zhou, Q. T.; Singh, S. K.; Nema, S.; Topp, E. M., Effects of Drying Process on an IgG1 Monoclonal Antibody Using Solid-State Hydrogen Deuterium Exchange with Mass Spectrometric Analysis (ssHDX-MS). *Pharm. Res.* **2018,** *35* (1), 12.

APPENDIX

I. BATCH SPECTRAL PROCESSING SCRIPT FOR NMRPIPE

```
#!/bin/csh

foreach i (1 2 3 4 5 6 7 8 9 10)

echo "Converting and processing spectrum $i..."
echo

bruk2pipe -in ./$i/ser \
  -bad 0.0 -ext -aswap -AMX -decim 1472 -dspfvs 20 -grpdly 67.984130859375  \
  -xN              2048  -yN              128  \
  -xT              1024  -yT               64  \
  -xMODE            DQD  -yMODE  Echo-AntiEcho  \
  -xSW        13586.957  -ySW         4308.402  \
  -xOBS         850.284  -yOBS          86.168  \
  -xCAR           4.700  -yCAR         120.000  \
  -xLAB              HN  -yLAB             15N  \
  -ndim               2  -aq2D         Complex  \
  -out ./test.fid -verb -ov

# Basic 2D Phase-Sensitive Processing:
#   Cosine-Bells are used in both dimensions.
#   Use of "ZF -auto" doubles size, then rounds to power of 2.
#   Use of "FT -auto" chooses correct Transform mode.
#   Imaginaries are deleted with "-di" in each dimension.
#   Phase corrections should be inserted by hand.

nmrPipe -in test.fid \
| nmrPipe  -fn SOL \
| nmrPipe  -fn SP -off 0.5 -end 1.00 -pow 2 -c 0.5    \
| nmrPipe  -fn ZF -auto                          \
| nmrPipe  -fn FT -auto                          \
| nmrPipe  -fn PS -p0 0 -p1 10  -di -verb        \
| nmrPipe  -fn EXT -left -sw \
| nmrPipe  -fn TP                              \
| nmrPipe  -fn LP -auto -fb                  \
| nmrPipe  -fn SP -off 0.5 -end 1.00 -pow 2 -c 0.5    \
| nmrPipe  -fn ZF -auto                          \
| nmrPipe  -fn FT -auto                          \
| nmrPipe  -fn PS -p0 -90 -p1 0.00 -di -verb        \
   -ov -out ./test-$i.ft2

rm test.fid

end
```

II. MATLAB SCRIPT FOR FITTING BACK-EXCHANGE CURVES

```
%How to format/label your excel sheet %
% Row 1, Column A - "Peak Number"
% Row 1, Column B - "Residue number
% Row 1, Column C onward - Time values in seconds (not corrected for deadtime)
%Row 1, last column – VT0 (peak volume of non-vapor-exchange sample

% Column A, Row 2 down = Peak # (the order in which NMRviewJ picked the peaks);
% Column B, Row 2 down = Residue number (based on the actual HSQC assignment);
% Column C, Row 2 downward = Initial concentration-corrected peak volume of T24 at time 0, i.e. the
first HSQC
% Column D onward, Row 2 downard = Next concentration-corrected peak volumes;
% Column P - VT0 values for each residue (peak volumes of non-exchanged sample) - if your data
go beyond column P, you can edit line 18 to reflect that

%Values to specify or edit below%

filename = '    '; % Specify excel filename from which you want to read data – make sure it is saved in
the working directory;
sheet = '     '; % Specify sheet name in which data is stored; e.g. 'Sheet 1' of your excel file
deadtime = 1200; % Specify deadtime
finaltime = xlsread(filename,sheet,'M1'); %Where to find the final time value - make sure to edit in line
19 and 30 as well
xnull = xlsread(filename,sheet,'C1:M1'); % Read in time values, where row C corresponds to time 0
and the final letter corresponds to the column for the final time point
x1 = 2*(finaltime)./3; %Determining where to label graph - 2/3 of the way across the x axis
x = xnull + deadtime; %Times with dead time added

%Beginning of for loop - edit i values on line 24%
for i = 1:52 % Last index should be number of resonances for which you wish to fit

%Locating data and defining variables - edit s2 to reflect final time - i+1 accounts for headers%

s1= ['C', num2str(i+1)]; % Specifying where to find initial volume value for the ith value in excel
s2 = ['M',num2str(i+1)]; % Specifying location of final volume value in excel for the ith residue
s3 = ['B', num2str(i+1)]; % Specifying where to find the residue label (e.g. E56)
s4 = ['P',num2str(i+1)]; %Specifying where to find T0 values
title1 = ['B', num2str(i+1)]; % Title of graph (residue label) %title of graph
resnum = xlsread(filename,sheet,[title1]);
res = string(resnum);
t0 = xlsread(filename,sheet,s4);
t24initial = xlsread(filename,sheet,s1);
t24final =xlsread(filename,sheet,s2);
folddifference = t24final./t24initial;
PCTuncorrected = 100.*(t24initial./t0);

y = xlsread(filename,sheet,[s1,':',s2]); % For the ith resonance, read in NMR volume values at each
time point

% Calculating the best fit curve for the given data that exhibits measurable change - skips values that
fall in an infinite well and puts all zeros for those

if (folddifference < 1.05)
```

```
Rates(i,1) = num2cell(i);
Rates(i,2) = num2cell(resnum);
Rates(i,3) = num2cell(folddifference);
Rates(i,4) = num2cell(0);
Rates(i,5) = num2cell(0);
Rates(i,6) = num2cell(PCTuncorrected);
Rates(i,7) = num2cell(PCTuncorrected);

elseif folddifference >=1.05

try

start = xlsread(filename,sheet, [s1]);
[xData, yData] = prepareCurveData( x, y );
ft1 = fittype( 'c+(a*(1-exp(-b*(x))))', 'independent', 'x', 'dependent', 'y' );
opts = fitoptions( 'Method', 'NonlinearLeastSquares' );
opts.Display = 'Off';
opts.Robust = 'Bisquare';

opts.StartPoint = [800 0.001 0]; % Enter your guesses for [ a b c] – where a is the total possible volume change, b is the estimated observed rate constant, and c is the true starting peak volume. If fits are poor, change initial value for first column

[fitresult, gof] = fit( xData, yData, ft1, opts );

  catch ME
    fprintf('Infinite fit for residue', num2str(i));

    Rates(i,1) = num2cell(i);
    Rates(i,2) = num2cell(res);
    Rates(i,3) = num2cell(folddifference);
    Rates(i,4) = num2cell(0);
    Rates(i,5) = num2cell(0);
    Rates(i,6) = num2cell(0);
    Rates(i,7) = num2cell(0);


    continue;  % Jump to next iteration (i) and print zeros for values. These peaks will need to be refit using new initial values on line 33, remember to also change the index just to refit these
end

kobs = fitresult.b; %Observed rate constant
PCTcorrected = 100*(fitresult.c./t0); %Corrected VT24

% Plotting the data and best fit curve
figure( 'Name', res);
plot( fitresult, xData, yData );
xlabel ('Time (seconds)'); %Change if your time is in another unit
ylabel ('Peak Volume');
title (res);
grid off;
hold off;
legend off;

% Set position of printed text in graphs
ylim = get(gca, 'ylim');
```

```
ymax = ylim(2);
y1 = (ylim(2)-ylim(1))/10;
y2 = ylim(1)+y1*8;
y3 = ylim(1)+y1*7;
y4 = ylim(1)+y1*6;


text(x1,y2,sprintf('R^{2} = %.2f',gof.adjrsquare))
text(x1,y3,sprintf('k_{obs} = %.2g s^{-1}',fitresult.b))
text(x1,y4,sprintf('i = %.0f',i))

%saveas(gca,str,'svg'); %Save figure as .svg file - optional, uncomment if desired
%saveas(gca,str,'jpg'); %Save file as .jpg file - optional, uncomment if desired
%close % Delete this line if you want the figure to stay open on the computer after it is saved.

Rates(i,1) = num2cell(i);
Rates(i,2) = num2cell(resnum);
Rates(i,3) = num2cell(folddifference);
Rates(i,4) = num2cell(kobs);
Rates(i,5) = num2cell(gof.adjrsquare);
Rates(i,6) = num2cell(PCTuncorrected);
Rates(i,7) = num2cell(PCTcorrected);


end

end

xlRange = 'A2'; %Where to write data in excel sheet
header = {'i','Residue','Fold Difference','kobs', 'Rsquared', 'PctT24uncorrected','PctT24corrected';
%header labels
xlswrite(filename,header,'Fitvalues'); %Creates new sheet in the excel file your using to paste
fitvalues
xlswrite(filename,Rates,'Fitvaluesretry',xlRange); %Inserts fitvalue matrix into sheet
```

III. UNSUCCESSFUL AND ONGOING APPROACHES TO UNCOVERING THE MECHANISM OF DEHYDRATION PROTECTION BY CAHS D

Table A1. Unsuccessful approaches.

Attempted approach	Reason for failure/concern	Alternative solution
Measuring changes in client protein stability in the presence of high concentrations of CAHS D using solution amide-proton exchange or ^{19}F NMR	Client proteins transiently stick to CAHS D gel network, resulting in extremely fast transverse relaxation that broadens NMR signal into the noise and thus impedes accurate peak integration.	Performing same experiments with magic-angle spinning NMR.
Measuring changes in client protein stability in the presence of high concentrations of CAHS D using quantitative cysteine reactivity.[1]	Sulfhydryl chemistry is not specific enough, so probe will react primarily with CAHS D (or any other protein-based crowder) simply because of its relatively high concentration.	Using bio-orthoganol probes, e.g. click chemistry, with unnatural amino acids.[2]
Performing random-mutagenesis on CAHS D and assessing gel formation capacity with microrheology and/or viscosity-sensitive dyes such as RY3.[3]	-Cell slurries are too inconsistent to obtain reliable results with microrheology. -Dyes may stick to proteins, reducing accuracy of viscosity reported.	N/A (gelation in cells is very difficult to quantitatively assess in a high-throughput manner).
Performing random-mutagenesis on CAHS D and assessing cell-protective ability with high-throughput "Start Growth Time" survival assay.[4]	-Difficult to obtain homogeneous drying in a 96-well plate. -Cell growth rate can depend on cell metabolism and rate of "cell reawakening", which could be altered by the presence of protein mutants. -We have also found that expression of CAHS proteins can artificially increase $OD_{600,}$ which can interfere with this assay	Perhaps viability stains combined with flow-cytometry could work,[5] though dyes sticking to the gelled protein is still a concern. (Idea credit: Octavio Origel)

Table A2. Initiated and/or ongoing approaches.

Measuring client-protein diffusion in the presence of CAHS D with fluorescence recovery after photobleaching (FRAP) and/or fluorescence correlation spectroscopy (FCS).[6]
Dark-state exchange saturation transfer (DEST) NMR spectroscopy to reveal K_d of client-protein – CAHS D gel interaction.[7]
Assignment of CAHS D ^{1}H-^{15}N resonances + temperature coefficient analysis of both CAHS D and client protein as a function of CAHS D concentration.[8-9]

REFERENCES

1. Isom, D. G.; Vardy, E.; Oas, T. G.; Hellinga, H. W., Picomole-scale characterization of protein stability and function by quantitative cysteine reactivity. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107* (11), 4908-13.

2. Lee, K. J.; Kang, D.; Park, H.-S., Site-Specific Labeling of Proteins Using Unnatural Amino Acids. *Mol. Cells* **2019,** *42* (5), 386-396.

3. Peng, X.; Yang, Z.; Wang, J.; Fan, J.; He, Y.; Song, F.; Wang, B.; Sun, S.; Qu, J.; Qi, J.; Yan, M., Fluorescence Ratiometry and Fluorescence Lifetime Imaging: Using a Single Molecular Sensor for Dual Mode Imaging of Cellular Viscosity. *J. Am. Chem. Soc.* **2011,** *133* (17), 6626-6635.

4. Hazan, R.; Que, Y.-A.; Maura, D.; Rahme, L. G., A method for high throughput determination of viable bacteria cell counts in 96-well plates. *BMC Microbiology* **2012,** *12* (1), 259.

5. Buysschaert, B.; Byloos, B.; Leys, N.; Van Houdt, R.; Boon, N., Reevaluating multicolor flow cytometry to assess microbial viability. *Appl. Microbiol. Biotechnol.* **2016,** *100* (21), 9037-9051.

6. Chenette, E. J., FCS and FRAP: illuminating cellular processes. *Nat. Cell Biol.* **2009,** *11* (1), S13-S14.

7. Fawzi, N. L.; Ying, J.; Torchia, D. A.; Clore, G. M., Probing exchange kinetics and atomic resolution dynamics in high-molecular-weight complexes using dark-state exchange saturation transfer NMR spectroscopy. *Nat. Protocols* **2012,** *7* (8), 1523-1533.

8. Bah, A.; Vernon, R. M.; Siddiqui, Z.; Krzeminski, M.; Muhandiram, R.; Zhao, C.; Sonenberg, N.; Kay, L. E.; Forman-Kay, J. D., Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **2015,** *519* (7541), 106-109.

9. Trainor, K.; Palumbo, J. A.; MacKenzie, D. W. S.; Meiering, E. M., Temperature dependence of NMR chemical shifts: Tracking and statistical analysis. *Protein Sci.* **2020,** *29* (1), 306-314.