12-2021

Computational and biochemical characterizations of anhydrobiosis-related intrinsically disordered proteins.

Brett R. Janis
University of Louisville

COMPUTATIONAL AND BIOCHEMICAL CHARACTERIZATIONS OF ANHYDROBIOSIS-RELATED INTRINSICALLY DISORDERED PROTEINS

By
Brett R. Janis
B.A., Cornell College 2014

A Dissertation
Submitted to the Faculty of the
College of Arts and Sciences of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
in Biology

Department of Biology
University of Louisville
Louisville, Kentucky

December 2021

COMPUTATIONAL AND BIOCHEMICAL CHARACTERIZATIONS OF ANHYDROBIOSIS-RELATED INTRINSICALLY DISORDERED PROTEINS

By
Brett R. Janis
B.A., Cornell College 2014

A Dissertation Approved on

November 12, 2021

By the following Dissertation Committee:

_____
Dissertation Director
Michael A. Menze

_____
David J. Schultz

_____
Mark P. Running

_____
Vladimir N. Uversky

_____
John H. Crowe

DEDICATION

This dissertation is dedicated to my parents, Janet Louis Janis, Gregory Charles Janis, and Charles Richard Tuten, who nurtured my sense of curiosity and encouraged my unique manner of thinking. Their love, guidance, and unwavering support were the pillars upon which I could grow and adapt to life's many challenges. Thank you for everything that you have done and continue to do for me.

This dissertation is also dedicated to three people who inspired me to become a scientist. These role models taught me what it means to love the process of discovery, and the sense of freedom that comes with knowing that every answer is a series of new questions.

My grandmother, Phyllis Janis (Robinson) – Who showed me that it is so much more fun to learn than it is to know. Rather than being satisfied with knowing the answer to a question, you demonstrated to me that there are always gaps in what we know, and that it is a beautiful thing.

My undergraduate adviser, Craig Tepper, PhD – who taught me to love molecular biology, and that the process of scientific discovery is fun and rewarding. You advocated for me to continue my education, and you taught me the love for scientific literature.

My PhD mentor, Michael A. Menze, PhD – who took a chance on an underperforming student who waited outside of his office for 2 hours. Thank you for guiding me and encouraging me to start my own projects. It is no overstatement that you made me the scientist that I am today.

ACKNOWLEDGEMENTS

Tyler Boyd, Ryan Napier, Benjamin Cadle, Benjamin Xie, Charles Elder, and Jensen Smith assisted in protein expression, conduction of long-term experiments, and presented research at international conferences.

Jonathan Kopechek, PhD, George Pantalos, PhD, Mariah Priddy, Connor Centner, Emily Murphy, Sienna Shacklette, and Clara Jones contributed to projects that, while not reported in this dissertation, inspired novel methods of experimentation that affected the experiments within.

Scott Janis, Ozgur Yavuzcetin, PhD, and Nicholas Tippery, PhD, at the University of Wisconsin Whitewater assisted in early experiments using scanning electron microscopy and atomic force microscopy that, while not published, informed later experiments found in this dissertation.

Close collaboration with Clinton Belott dramatically improved the collection of data and the determination of critical experiments found in this dissertation, which is represented in the co-authorship of several publications. His assistance in establishing the complete animal LEA database used in chapter II made the statistical analyses possible on the short timeline provided to complete the study.

My research was funded through several sources, including Eastern Illinois University, the University of Louisville, the National Science Foundation, the National Institutes of Health, the National Aeronautics and Space Administration, and the Department of Defense.

D.O.D. grant ‘Development of Dried Blood for Prolonged Field Care in Austere Environments’ (10/2020-09/2023, Jonathan Kopecheck, PI, Michael A. Menze, Co-PI, Brian Harbrecht, Co-I).

NASA grant ‘Evaluation of Preserved Blood for Transfusion Therapy in Reduced Gravity’ (01/2019-12/2020, Michael A. Menze, PI; Jonathan Kopecheck, Co-PI; George Pantalos Co-I).

NSF-PIF grant ‘A High-Precision Sonoporation System for Cell Transfection and Preservation. (08/2018-07/2021, Paula Bates, PI, Michael A. Menze, Co-PI, Jonathan Kopecheck, Co-PI).

NIH grant ‘ExCITE’: An Integrated Microfluidic Device for Long-Term Preservation of Dried Red Blood Cells at Ambient Temperature’ (08/2017 - 01/2019 Jonathan Kopecheck, PI; Brett Janis, PI; Michael A. Menze, PI, *U01HL12751*).

University of Louisville Arts and Sciences Graduate Student Research Grant. 2017

Eastern Illinois University Research and Creative Activity Grant Spring. 2016

NSF grant “Collaborative Research: Mechanisms of Tolerance to Severe Water Stress in Animals “(05/2015 - 04/2020; Michael A. Menze; PI, IOS-1659970, Steven Hand, PI, IOS-1457061).

ABSTRACT

COMPUTATIONAL AND BIOCHEMICAL CHARACTERIZATIONS OF ANHYDROBIOSIS-RELATED INTRINSICALLY DISORDERED PROTEINS

Brett R. Janis

November 12, 2021

Anhydrobiosis is the remarkable phenomenon of “life without water”. It is a common technique found in plant seeds, and a rare technique utilized by some animals to temporarily stop the clock of life and enter a stasis for up to several millennia by removing all of their cellular water. If this phenomenon can be replicated, then biological and medical materials could be stored at ambient temperatures for centuries, which would address research challenges as well as enhance the availability of medicine in areas of the world where refrigeration, freezing, and cold-chain infrastructure are not developed or infeasible. Furthermore, modifying crop tissues could make them resistant to droughts, addressing one of the greatest threats to food stability around the world. This work utilizes a combination of computational techniques and novel approaches to performing biochemistry without water to elucidate the mechanisms of function of specialized proteins that are responsible for anhydrobiosis in animals, particularly the anhydrobiotic cysts of the brine shrimp *Artemia franciscana*. A detailed evaluation of the chemical properties of anhydrobiosis-related, intrinsically disordered proteins indicates that there are multiple protein-based

strategies to achieve anhydrobiosis, but that late embryogenesis abundant (LEA) proteins are the most well understood. However, the mechanisms of LEA protein function have never been demonstrated, resulting in a wide variety of hypotheses regarding their ability to confer desiccation tolerance. This work demonstrates that a group 1 LEA protein, *Af*LEA1.1, and a group 6 LEA protein, *Afr*LEA6, undergo liquid-liquid phase separations during desiccation and thereby transiently form novel protective membraneless organelles which partition specific proteins and nucleic acids. These desiccation-induced cellular compartments are a novel mechanism to explain how LEA proteins confer desiccation tolerance, and the drivers of this behavior have been linked to the consensus sequences that define these LEA proteins. Therefore, the separation of aqueous proteins into a specialize compartment during drying is unlikely to only be a function of *Af*LEA1.1 and *Afr*LEA6, but actually the mechanism by which group 1 and group 3 LEA proteins function in plant seeds and anhydrobiotic animals. These results indicate that when water is unavailable, anhydrobiotic organisms substitute it with their own solvents.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

All known life on Earth uses water as its primary solvent, in part because it is abundant, and in part because it has unique chemical properties that make it suitable for a broad range of chemical reactions. Some animals have evolved remarkable strategies to withstand virtually complete water loss for prolonged periods of time, despite the cellular damage associated with desiccation, such as membrane destabilization, protein and nucleic acid denaturation, oxidative stress, and metabolic dysregulation (*1-5*). This transient state of life has been an enigma since 1702 when Van Leeuwenhoek first noted anhydrobiosis in rotifers or 'wheel animals' (*6, 7*). Since then, desiccation tolerance has been confirmed to occur in several other animal phyla including Arthropoda, Tardigrada, and Nematoda (*8, 9*). Remarkably, many desiccation-tolerant species can survive in an anhydrobiotic state for years, or even decades, with limited impacts on viability (*10*). Understanding the constrains that govern desiccation tolerance has obvious biotechnological applications, particularly in crop-drought resistance and stabilization of clinically relevant cells and tissues at ambient temperatures (*11*). To translate insights from anhydrobiotic animals into clinical applications, it is imperative to compare and contrast the molecular principles among these organisms to distinguish between fundamental and unique strategies.

These remarkable organisms survive complete desiccation by utilizing a variety of behaviors, protective osmolytes, proteins, or a combination of any or all of these methods.

The brine shrimp, *Artemia franciscana*, survives the freezing temperatures of winter by encapsulating their embryos into cysts which rapidly undergo diapaus(*12, 13*). These cysts are enriched with protective osmolytes such as trehalose, a non-reducing disaccharide with multiple protective properties (*14*), and glycerol (*15*). In addition to osmolyte-based strategies, plants and animals utilize specialized proteins. In most anhydrobiotic plant and animal cells, late embryogenesis abundant (LEA) proteins are utilized (*8*). The mechanisms of function for LEA proteins are much less understood than the roles of osmolytes in part because the proteins are harder to generate and because there appears to be a very large variety of proteins with specialized functions and unique physicochemical properties (*16*). To add to this complexity, the definition of a LEA protein has come into question, with disagreements over whether or not tardigrades rely on LEA proteins or on their own unique anhydrobiosis related intrinsically disordered proteins (*17-19*). This uncertainty has produced several mechanistic hypotheses for the functions of protective proteins during anhydrobiosis. While some of these mechanisms may occur simultaneously or may be represented in different families of LEA proteins, some others may not make a substantial impact during desiccation or rehydration.

Disorder: regulatory element or essential property for hydration-level specific function(s)?

The intrinsic disorder of LEA proteins has been hypothesized to serve a role in several functions such as 1) molecular shields that block protein aggregation (*20*), 2) regulating desiccation rates as hydration buffers (*21*), 3) binding divalent metal ions (*22, 23*), and 4) reinforcing trehalose glasses (*24*). Recently discovered TDPs from tardigrades have been demonstrated to form protective hydrogels and glasses that reinforce structural integrity of

the animal during desiccation (*25*). An additional hypothesis for LEA proteins states that they undergo *in vivo* conformational transitions from randomly-coiled hydrated chains to semi-folded, activated proteins at lower hydration levels (*26*). Furthermore, LEA proteins in plants have been found to confer membrane protection and freeze tolerance even in the hydrated, disordered state (*27*). These properties can be mechanistically explained by binding-partner induced conformational transitions *via* specific molecular recognition features (MoRFs) (*28*). Protein-protein interactions may be modulated by conformational transitions as water is reduced during freezing or drying (*29*). Multiple intrinsically disordered regions found on a single protein may fold upon recognition of distinct binding partners, thus allowing one-to-many targeting or protein moonlighting (*30*). Some of these regions may only fold under specific crowding and hydration conditions such that one protein may have several functions and targets during desiccation. Nevertheless, the specific role of disorder in dehydration-related proteins warrants further investigations and unresolved questions remain: 1) Is protein disorder an intrinsic mode of function or a regulatory element of functional properties? 2) Is the impact of the hydration state on protein function(s) similar among proteins? Although many of the above stated hypotheses were initially developed on LEA proteins, their relevance to other IDPs will also be discussed.

Molecular shielding

The molecular shielding hypothesis predicts that protection is conferred by entropic chains that act as steric and/or electrostatic barriers against protein aggregation during water stress (*20*). The protective IDP has low target specificity which is limited to fuzzy and small

MoRF-induced structural interactions with specific surfaces on target proteins. The target protein may partly unfold upon binding to the IDP but refolds by entropic energy transfer upon release from the protein (*31*). Molecular shielding-like anti-aggregation has been observed in LEA proteins expressed by the sleeping chironomid *P. vanderplanki* (*32*). Most of the IDP's structure exists as a highly plastic ensemble of conformations that encompasses a large hydrodynamic radius. This hypothesis is consistent with the finding that IDPs, compared to globular proteins, in general have a greatly enlarged hydrodynamic radius (*33*), but it does not fully address the functional relevance of the observed increase in secondary structure of LEA proteins during desiccation.

One challenge to this hypothesis might be that a high ratio of protective proteins to targets is required for shielding. To protect each target would likely require several molecular shields to insulate them from multiple angles of interaction. Target selection might be complicated by increasing crowding during desiccation and recruitment of additional molecular shields might be hindered by already interacting IDPs. Furthermore, LEA proteins can gain as much as an additional 40% of defined secondary structure and reduce their hydrodynamic radii substantially as cellular water depletes (*29*). The reduced hydrodynamic radii in turn would likely reduce the molecular shielding efficiency due to the decrease in area of steric and electrostatic repulsion. Therefore, the observed gain in secondary structure of some LEA proteins during desiccation would be counterintuitive for this model. Molecular shielding does not offer a mechanistic explanation for the membrane stabilizing effect observed for several LEA proteins. Consequently, molecular shielding by LEA proteins may only be one of several possible functions.

Hydration buffers

The intrinsic disorder of anhydrobiosis-related proteins gives rise to unique solution properties when compared to globular proteins, such as large hydration shells and conformational plasticity. These properties have been hypothesized to play a role in desiccation tolerance by modulating the solution properties and desiccation rates of anhydrobiotic organisms (*21*). The large protein hydration shell may act as a water reservoir that is released upon assuming secondary structures during desiccation. Solid state NMR studies have demonstrated enlarged hydration shells of IDPs, and also suggest that those of dehydrins from plants are particularly extensive (*34*). Although loss of water upon disorder-to-order transitions is likely to occur, a functional role of this released water would likely be small. While there may be some impact on the overall drying kinetics of the animal due to water release, the relatively small fraction of water that could be released, even assuming relatively high levels of LEA expression, is likely insignificant in the context of variable drying conditions found in nature. Furthermore, the lack of free water, once desiccation is complete, is the major source of cellular protection because unregulated chemical reactions are inhibited in the vitrified state. A similar principle is noted in cryobiology, wherein the relatively mobile water found at temperatures above the glass-transition temperature of water (~136 °K) contributes to degradation in frozen samples (*35*).

Ion sequestration

During water removal, dissolved molecules and ions will concentrate and, eventually, precipitate from the solution depending on their physicochemical properties. Precipitation,

if not carefully regulated, often produces damaging aggregation events that permanently inactivate proteins and other biomolecules. The ion-sequestration hypothesis states that, during water loss, ions are buffered by binding to LEA proteins (*22, 23*). Considering the large amounts of inorganic ions relative to protein in the cell, this effect is likely limited to metal cations that are found in low concentrations. However, highly charged LEA proteins might serve as nucleation sites for salt precipitation during desiccation. Promoting precipitation of salts would only be possible for LEA proteins with relatively high hydrophilicity and frequency of charged residues (FCR).

Reinforcing sugar glasses and protein vitrification

During the past decade, group 3 LEA proteins have been associated with the vitrification of the intracellular space by reinforcing sugar-based glasses (*24*) and, more recently, proteins and LEA model peptides have been demonstrated to form glasses themselves (*36*). Protein and sugar glasses are a non-crystalline, physical state that is characterized by a high viscosity above 10^8 Pa-s which greatly impedes molecular movement and prevents chemical reactions (*37*). Sugar vitrification is a well-established mechanism in anhydrobiosis for both plants and animals, which is generally associated with membrane and protein stabilization (*38*). Trehalose vitrifies at low water contents, but the capacity for protection is impacted by the glass transition temperature (Tg), which dictates the temperature and degree of hydration where the sugar will form or maintain a glassy state (*35, 39*). Although the mechanism by which IDPs can reinforce and/or stabilize sugar glasses is not well understood, some LEA proteins and peptides have been shown to increase the Tg of sugar glasses (*24, 36*). In the case of tardigrades, vitrification occurs

rapidly during desiccation provided that the rate of drying is sufficiently slow for the organism to accumulate sufficient amounts of TDPs. Protein-based vitrification was further demonstrated by ectopic expression of TDPs in both yeast and bacteria (*25*).

Disorder-to-order transitions and membrane stabilization

LEA proteins have been shown to undergo conformational changes in response to various solutes and crowding effects (*40*) and two group 3 LEA proteins from *A. franciscana* increased their amount of α-helical structures *in vitro* when exposed to sodium dodecyl sulfate, tetrafluoroethylene, or when desiccated (*29*). Proteins from *P. vanderplanki* and *A. franciscana* undergo conformational transitions when dried and protect enzymatic activity of lactate dehydrogenase during desiccation. Additionally, both LEA proteins were shown to prevent aggregation of casein better than bovine serum albumin (*29, 32*). Furthermore, two LEA proteins from *Arabidopsis thaliana* were recently shown to undergo conformational transitions when crowded by glycerol and when localized near a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylcholine liposome (*41*). The conformational state was maintained in the desiccated state, and the protein inserted into the phospholipid bilayer as measured using FTIR. These findings support the hypothesis of direct membrane interactions and insertion of some LEA proteins, although the exact thermodynamics are not fully understood. This evidence also suggests desiccation induces conformational transitions, rather than MoRF-induced membrane interactions alone, because the protein needed to be folded by glycerol crowding before associating with the liposome membrane (*41*).

Membrane interactions of some folded LEA proteins are hypothesized to occur through bundled, amphipathic α-helices. Two hydrated group 3 LEA proteins from *A. franciscana* contain regions of high α-helical propensity that, if folded, may form amphipathic helices with stripes of positive and negative residues separating the hydrophobic face from the hydrophilic portion of the protein (*42*). These proteins protect the membranes of liposomes during desiccation and rehydration, potentially via interactions between the phospholipids and the hydrophobic face of the amphipathic α-helices. Stripes of positive and negative charge distributed may not be a prerequisite for all LEA proteins to stabilize phospholipid bilayers and some LEA proteins may actually insert into the membrane during desiccation (*41*).

Phase separation and the formation of desiccation induced 'membraneless' organelles

With the discovery of the glass and gel propensities of TDPs, the question remains if or which desiccation tolerance proteins may form fibrils, glasses, or gels during desiccation. To date, no stable super-molecular LEA assembles have been observed. However, the question if LEA proteins form higher order oligomers has been discussed in the literature, and some LEA proteins form multimers in solution (*43, 44*). A more recently described physicochemical property of some IDPs is the separation of the protein from the solvent as another liquid phase (*45-47*). Protein liquid-liquid phase separations, comprised of loose associations of LEA proteins, could serve different functions including molecular shielding and hydration buffering. Protein liquid-phase separations have become a rapidly expanding topic in biology that has explained the behavior of the disordered tail of DEAD-box helicase 4 (Ddx4) and several other previously uncharacterized IDPs (*48-50*). These liquid

protein droplets, termed “membraneless organelles,” are found under osmotic and oxidative stress in eukaryotes and are predicted for some LEA proteins in *A. franciscana* (*51-53*).

The observed desiccation-induced folding of LEA proteins suggests that intramolecular interactions occur as surface water is depleted. Protein interactions between similar LEA proteins or partner molecules may result in a protein droplet with specific physicochemical properties as governed by the amino acid sequence of the nucleating protein. Similar to other proteins that undergo liquid-liquid phase separations, LEA proteins are highly repetitive, have low complexity, and tend to have high overall charges and charge separation (*54, 55*). Furthermore, the amyotrophic lateral sclerosis (ALS) related fused in sarcoma (FUS) protein still forms liquid protein droplets when its multivalent interaction sites are substituted with LEA motifs (*56*). Unlike the molecular shielding hypothesis, where the induced folding of several LEA proteins in response to interactions with each target protein is necessary for protection, membraneless organelles can nucleate off a core material and incorporate a variety of targets (*57*). The interior of the droplet should then function like a molecular shield, but each added target further expands the droplet radius and increases the odds of collision and fusion with other LEA proteins and/or target molecules. If these droplets form early during drying, or in preparation for drying, then they may require a lower protein content for protection compared to molecular shields. Although not all desiccation tolerance proteins should be expected to form anhydrobiosis associated membraneless organelles, the sequence characteristics of some LEA proteins and TDPs suggest that they are candidates for this super-molecular structure (e.g., MH351624).

CHAPTER II

ANHYDROBIOSIS PROTEINS OF THE ANIMAL KINGDOM – A SEQUENCE ANALYSIS AND COMPARISON

SUMMARY

Intrinsically disordered proteins (IDPs) play a central role in desiccation tolerance in animals belonging to a broad taxonomic range. Late embryogenesis abundant (LEA) proteins are found in most anhydrobiotic organisms and have recently been distinguished from tardigrade proteins (TDPs) that appear to only be found in the phylum Tardigrada. Despite their similar functions, a surprisingly wide range in protein sequence characteristics, including hydropathy and the frequency and distribution of charges, was discovered between these two groups of desiccation tolerance conferring proteins. However, two distinct clusters of similar proteins were found for LEA proteins that potentially correlate with distinct functions for this group of polypeptides. Further analysis indicates two broad groups of LEA proteins, one that may only collapse into functional conformations during desiccation and a second group that potentially displays functions in the fully hydrated state. A broad range of physiochemical properties suggest that protein folding in both groups may be induced by factors such as hydration level, molecular crowding, and interactions with binding partners. This plasticity in folding behavior may be required to fine tune protein functions at different hydration levels during desiccation.

Janis et al. 2018. Proteomics

Furthermore, the sequence properties of some LEA proteins share qualities with IDPs known to undergo liquid-liquid phase separations during environmental challenges.

INTRODUCTION

Plants and animals that survive desiccation do so by utilizing specialized metabolites (*58*), proteins (*8, 59*), or a combination of these two molecular strategies. These protective osmolytes tend to prevent protein aggregation (*60*), impose entropic challenges to protein unfolding (*61, 62*), or increase cytoplasmic viscosity to the point of a glass transition during extreme desiccation (*18, 38*). The proteins most readily associated with desiccation tolerance are LEA proteins and, more recently, tardigrade proteins (TDPs) (*18, 60, 63, 64*). LEA proteins were discovered by Dure et al. in the late embryogenic stage of cotton (*Gossypium hirsutum*) seeds and have more recently be linked to anhydrobiosis in animals (*65, 66*). Plant LEA proteins were initially grouped based on the presence of specific sequence motifs (*67*). Since then, several nomenclatures have been proposed for LEA proteins (*54, 68*). This work follows the classification scheme by Tunnacliffe and Wise, which proposes 6 distinct groups of LEA proteins based on the primary amino acid sequence (*16*).

Group 1 LEA proteins contain one or more repeats of a hydrophilic 20 amino acid motif, while group 2 LEA proteins, termed 'dehydrins', contain two or more specific motifs denoted as Y, S, and K. Group 3 contains the largest number of LEA proteins and are characterized by a specific 11 amino acid motif (*16, 59*). While most LEA proteins in plants seem to fall into groups 1-3, other minor groups have been described. Group 4 LEA proteins lack any consensus sequence, and group 5 LEA proteins are characterized by an

unusually high content of hydrophobic residues (*69*). Finally, group 6 LEA proteins are characterized by the presence of at least one seed maturation protein motif and have recently been associated with the long-term stability of seeds in the desiccated state (*70*). Interestingly, only group 3 LEA proteins have been identified in anhydrobiotic animals with the exception of *Artemia,* which expresses LEA proteins from groups 1, 3, and 6 in their desiccation tolerant embryos. TDPs are still a less-characterized group of proteins, and a limited library of TDPs has been isolated and reported from the anhydrobiotic tardigrade *Hypsibius durardina* (*18*). Other tardigrades are capable of anhydrobiosis, such as, but not limited to, *Richtersius coronifer* (*71*), *Milnesium tardigradum* (*72*), *Paramacrobiotus richters* (*17*), *Ramazzottius oberhaeuseri (73).*

The reason(s) why different groups of anhydrobiotic animals rely on different types of proteins to achieve desiccation tolerance are unresolved. One hypothesis may be that the lack of trehalose or presence of other protective compounds encourages different adaptive trajectories for proteins involved in desiccation tolerance. Data presented on three *Triops* species demonstrated that the cysts undergo vitrification in absence of trehalose (*74*). These data are strikingly like those presented for tardigrades (*18*). Blasting LEA protein sequence data from *Artemia* against EST libraries for *Triops* yielded no significant hits, while searches against EST libraries derived from tardigrades yielded low-identity hits (*data not shown*). This may support the hypothesis that the absence of trehalose requires proteins with a different set of physicochemical properties compared to animals that accumulate substantial levels of this sugar (e.g. <0.5% dry weight in *Triops longicaudatus, Triops cancriformis, and Triops australiensis vs.* 13-18% dry weight in *A. franciscana* and *P. vanderplanki* (*74*)).

METHODS

Datasets

Amino acid sequences for animal LEA proteins were retrieved using "LEA" or "Late Embryogenesis Abundant Protein" as search terms in the protein and nucleotide sequence databases at National Center for Biotechnology Information (NCBI). The retrieved sequences were individually cross-referenced using BLAST-P and rejected if they did not share sequence similarity within at least $E < 1*10^{-3}$ to a confirmed LEA protein (Supplemental File 1). 101 LEA protein sequences failed to be rejected and comprise the LEA protein dataset for the following analysis (Table 1). Published tardigrade protein sequences were used (*18*) and no additional proteins were found using BLAST algorithms limiting the total dataset to only 14 sequences (Table 2). The globular protein dataset was retrieved from RCSB PDB selecting for proteins between 10 and 80 kDa with structures verified by X-ray crystallography to avoid intrinsically disordered regions (Supplemental File 2).

Sequence Analysis

All sequences were analyzed in bulk using localCIDER, a freely accessible program designed by Holehouse et al. (*75*), and flavor predictions were made using values generated from the VL2 predictor with a window length of 21 amino acids (*76*). Individual protein sequences were tested using standard sliding window averages with a range of 5 amino acids. Statistics were performed using one-way ANOVA tests with a Tukey post-hoc test

via the program R-Studio. No additional plugins were necessary to perform these tests. SigmaPlot 12.5 (Systat Software Inc., San Jose, CA) was used to generate the graphs.

RESULTS AND DISCUSSION

Mean net charge and hydropathy analysis

The ratio of mean net hydropathy and the absolute value of the mean net charge indicates the likelihood for disorder based on an arbitrary boundary established by statistical analysis of known proteins (*77*). As shown in Fig. 1, the distribution of LEA proteins overlaps with the diverse range of known IDPs originally used to generate the disorder-order boundary and both groups are clustering separately from known globular proteins as expected (Fig. 1A, B). Although TDPs have been confirmed as IDPs, they distribute evenly between the areas of the plot where most globular proteins or IDPs are located (Fig. 1A, C). Some LEA proteins are also rather hydrophobic and overlap with the distribution pattern of known globular proteins (Fig. 1A, D). In case desiccation induces conformational transitions to 'activate' proteins, then more hydrophobic proteins may begin folding earlier during water stress than more hydrophilic polypeptides. A large range of hydropathies could indicate a temporal regulation of the LEA-proteome response during drying. Alternatively, more hydrophobic desiccation tolerance proteins may separate from solution more readily into gels or liquid droplets than their more hydrophilic counterparts.

Phase diagrams – charge ratios and distributions

The reduced range of amino acid expressed in IDPs results in a lower overall complexity than observed in globular proteins (*78*). This low complexity increases the impact of overrepresented amino acids in a protein sequence. IDP properties are often governed by the frequency of charged residues (FCR), the frequency of order-promoting residues (e.g., alanine and phenylalanine), and the spatial distributions of both (*79*). The frequency of positive and negative charges, for example, offers insights into a protein's capacity to maintain intramolecular interactions and has been used to define three compositional categories: polar tracts, polyelectrolytes, and polyampholytes (*79*). These categories were further expanded into five; weak polyampholytes and polyelectrolytes (R1), Janus sequences (R2), strong polyampholytes (R3), negatively charged polyampholytes (R4), and positively charges polyampholytes (R5) (*80*). R1 proteins tends to be globular or represent IDPs with a globular domain. R2 proteins tend to be more plastic, with conformations that depend on salt concentration, ligand binding, or other factors. R3 proteins are highly charged but have a low mean net charge due to an equal balance of charges. R3 proteins with regularly distributed charges have strong self-repulsion and exhibit more coiled structures, whereas proteins with more localized charges increase their likelihood to form intramolecular interactions such as hairpins or chimeras. R4 and R5 proteins form expansive coils due to polyanionic or polycationic repulsion (*79*). Employing this analysis predicts that LEA proteins have a wide range of potential behaviors, which is represented by proteins distributing across regions R1, R2, and R3 of the phase diagram (Fig. 2A). However, LEA proteins cluster into two major groups across R2 and R3 with only few outliers in R1 (Fig. 2B). Conversely, all known TDPs are found in region R2, and the group of globular protein used for comparison is mainly represented

in regions R1 and R2 (Fig. 2C, D). Although the tight clustering of TDPs may be due to some general properties for this protein family, the small sample size likely accounts for some lack in variance (Table 1).

The separation of LEA proteins into two clusters at the border of R1/R2 and R2/R3 may indicate differences in function. Proteins in R1/R2 are likely more environmentally regulated than those in R2/R3 but may undergo a limited amount of charge-mediated folding due to their relatively low FCR. This suggests that hydropathy and MoRF regions should have a higher impact than charge distribution on induced-fit or desiccation-induced conformational transition for LEA proteins falling at the R1/R2 border. LEA proteins in the R2/R3 cluster have a higher FCR and should be more influenced by their charge distribution, pH fluctuations, and ion concentrations. For proteins plotting in R2, which predicts environmentally modulated structures, conformational transitions during desiccation, such as induced folding at phospholipid bilayers or in response to ligand binding, are supported. The localization of LEA proteins in R1 and R3also suggest that some of these proteins may not need to gain additional structure during desiccation and, therefore, function at high water contents before any environmentally induced folding.

Flavor categorization based on amino acid composition

Like compositional categories, flavors of disorder are an established criterion that, while mainly used in disorder algorithms, correlates broadly to predicted functions and partner-binding behaviors. The flavors described by Vucetic et al include “V”, “C”, and “S”-flavored proteins (VIDPs, CIDPs, and SIDPs, respectively) (*76*). VIDPs are enriched in structure-promoting and hydrophobic amino acids residues, most notable cysteine,

phenylalanine, isoleucine, and tyrosine, relative to CIDPs and SIDPs and are often associated with ribosomal proteins. CIDPs are enriched in alanine, histidine, and methionine and are generally associated with DNA and RNA binding. Finally, SIDPs are relatively depleted in histidine and are generally associated with protein-binding behavior. The proteins examined showed a clear bias towards SIDPs (Fig. 3A) and each of the four TDPs that were categorized as CIDPs are secreted and not cytoplasmic localized proteins. LEA proteins were mostly classified as SIDPs but some VIDPs and CIDPs were also discovered. Interestingly, no tardigrade VIDPs were identified, and all rotifer LEA proteins were CIDPs. Given the propensity of TDPs to form protein-glasses, the SIDP flavor that suggests protein-protein interactions is not surprising. However, flavor categories cannot predict the potential membrane interactions associated with some LEA proteins.

A more in-depth analysis using composition profiling reveals that LEA proteins and TDPs both are each enriched in alanine, acidic residues, and lysine (Fig. 3 B, C). The hydrophilic nature of desiccation tolerance proteins probably allows them to maintain disorder in the hydrated state despite their remarkable depletion in the major structure-breaking amino acids, glycine and proline. A possible explanation for the enrichment in alanine observed for LEA proteins, and to a lesser extent TDPs, might be the propensity to form α-helices in the desiccated state. Furthermore, relative to ordered proteins, TDPs are enriched in all positively charge amino acids whereas LEA proteins are enriched in lysine but depleted in arginine (Fig. 3 D). This bias towards lysine over arginine may allow for a greater variety of post-translational modifications and lower surface-charge interactions that should increases the relative solubility of LEA proteins (*81*).

Differences in protein properties among animal genera

Proteins were separated into 5 groups to investigate genera-specific properties. Not surprisingly, some genera-specific clustering was observed in the charge-hydropathy plot and phase diagram (Fig. 4A, B). Rotifers express less hydrophilic LEA proteins, which may interact more readily with membranes than other animal LEA proteins. LEA proteins in animals that utilize trehalose are mainly localized in the disordered region of the plot, although *Artemia* and *Polypedilum* each express at least one rather hydrophobic protein (Table 1). TDPs have a lower absolute mean net charge and higher mean net hydropathy on average than LEA proteins. LEA proteins are generally longer than TDPs, perhaps allowing for larger hydrodynamic radii even after folding (Fig. 4C). LEA proteins found in *Caenorhabditis* were on average substantially longer than observed in the other groups. However, a larger variation in the frequency of charged residues was found for *Artemia* and *Polypedilum* compared to the other groups (Fig. 4D).

An overabundance of like-charges promotes self-repulsion, whereas distinct regions of different charges may facilitate intermolecular and intramolecular electrostatic interactions. This distribution of charges can be represented by a scale from 0 to 1, denoted by the variable κ, where 0 represents a completely uniform charge distribution and 1 describes complete separation of charges (*79*). LEA proteins with larger size such as observed in *Caenorhabditis* and *Polypedilum* should have increased odds of intramolecular interactions forming tertiary structures compared to shorter proteins. Both, the FCR and protein length will modify the impact of κ on the probability of electrostatic interactions. For example, longer proteins will in general have a higher probability to form favorable intramolecular interactions which in turn reduces the impact of κ. However, as the FCR

increases, the impact of κ also increases. Overall, larger proteins appear to exhibit larger FCRs and smaller κ values which may increases electrostatic repulsion within the polypeptide chain (Fig. 3E). The small sample sizes of protein sequences available for *Adineta* and *Hypsibius/Paramacrabiotus* (Table 1) may contribute to the smaller range in metrics compared to proteins from the other genera. However, based on our analysis, rotifer LEA proteins and TDPs in tardigrades are both characterized by high FCRs and moderate to high κ values. Interestingly, anhydrobiosis does not depend on trehalose in the species from either genus (*17*, *18*).

CONCLUSIONS

Returning to the initial question of intrinsic disorder as a regulatory element or functional property of desiccation tolerance proteins, we conclude that it may be either, depending on the specific LEA protein or TDP in question. For any given protein, the degree of desiccation-mediated regulation could be governed by its hydropathy and the quantity and distribution of charged residues. Furthermore, our analyses demonstrate that the impact of the hydration state on protein function may vary substantially among proteins. Anhydrobiotic animals may require a physiochemically diverse LEA or TDP proteome to elicit a temporal progression of responses during desiccation. A temporal distribution of responses would better accommodate variable rates of desiccation and cellular structures may require targeted protection from different sources of deterioration depending on the hydration state of the animal.

Tables

Table 1**:** Selected properties of LEA-related proteins in animal species. Protein sequences were deduced from full-length nucleotide sequences* as indexed by the US National Center for Biotechnology Information (NCBI).

Organism	Accession Nr.	Length (aa)	Frac. Dis. Prom.	Frac. Char. Res.	GRAVY	References
Adineta ricciae	ABU62808	376	0.75	0.35	-0.457	(*82*)
Adineta ricciae	ABU62809	421	0.74	0.35	-0.460	(*82*)
Adineta ricciae	ABU62810	376	0.75	0.35	-0.465	(*82*)
Adineta ricciae	ABU62811	420	0.74	0.35	-0.461	(*82*)
Adineta vaga	ADD91471	354	0.75	0.35	-0.627	
Adineta vaga	ADD91479	354	0.76	0.35	-0.626	
Ancylostoma ceylanicum	EPB73657	387	0.80	0.28	-0.466	
Ancylostoma duodenale	KIH53544	314	0.83	0.37	-0.887	
Ancylostoma duodenale	KIH57747	359	0.78	0.26	-0.350	
Aphelenchus avenae	Q95V77	143	0.85	0.39	-1.585	(*83*)
Aphelenchus avenae	AAL18843	143	0.85	0.39	-1.585	(*83*)
Aphelenchus avenae	ABQ23232	102	0.82	0.44	-1.376	(*84*)
Aphelenchus avenae	ABQ23233	85	0.84	0.52	-1.832	(*84*)
Artemia franciscana	ABR67402	182	0.83	0.27	1.365	
Artemia franciscana	ABX89317	182	0.84	0.27	-1.410	(*85*)
Artemia franciscana	ACM16586	307	0.79	0.37	-1.295	(*86*)
Artemia franciscana	ACX81197	97	0.74	0.27	-1.158	
Artemia franciscana	ACX81198	217	0.79	0.27	-1.257	
Artemia franciscana	ADE45145	142	0.84	0.27	-1.418	
Artemia franciscana	ADE45146	122	0.83	0.27	-1.312	
Artemia franciscana	ADE45147	62	0.82	0.27	-1.234	
Artemia franciscana	MH351624	257	0.70	0.21	-0.418	
Artemia franciscana	ACA47267	357	0.81	0.27	-1.027	(*87*)
Artemia franciscana	ACA47268	364	0.80	0.27	-0.884	(*87*)
Artemia parthenogenetica	AEM72698	85	0.78	0.27	-1.235	

Artemia parthenogenetica	AEM72699	182	0.83	0.26	-1.396	
Artemia persimilis	AEM72697	85	0.78	0.30	-1.235	
Artemia sinica	AMQ80946	182	0.83	0.30	-1.412	(*88*)
Artemia sinica	AOV81545	364	0.79	0.30	-0.885	(*88*)
Bemisia tabaci	XP_018915417	136	0.75	0.37	-0.839	
Brachionus plicatilis	ADE05593	613	0.84	0.38	-1.248	(*89*)
Brachionus plicatilis	ADE05594	248	0.85	0.33	-1.219	(*89*)
Caenorhabditis brenneri	EGT57645	935	0.86	0.39	-1.255	
Caenorhabditis brenneri	EGT57648	789	0.84	0.39	-1.020	
Caenorhabditis brenneri	EGT59057	917	0.85	0.39	-1.244	
Caenorhabditis brenneri	EGT59115	379	0.84	0.37	-0.870	
Caenorhabditis brenneri	EGT59117	724	0.85	0.40	-1.032	
Caenorhabditis briggsae	CAP25432	324	0.82	0.36	-0.833	
Caenorhabditis briggsae	CAP25462	379	0.79	0.38	-1.125	
Caenorhabditis briggsae	CAP25449	925	0.82	0.39	-1.252	
Caenorhabditis elegans	AAB69446	733	0.83	0.38	-1.126	
Caenorhabditis elegans	CCF23420	821	0.81	0.36	-1.104	
Caenorhabditis elegans	CCF23421	1166	0.82	0.36	-1.067	
Caenorhabditis elegans	CCF23422	1214	0.83	0.38	-1.066	
Caenorhabditis elegans	CCF23423	1381	0.82	0.37	-1.065	
Caenorhabditis elegans	CCF23424	1198	0.83	0.36	-1.066	
Caenorhabditis elegans	CCF23425	1349	0.82	0.37	-1.065	
Caenorhabditis elegans	CCF23426	1397	0.82	0.37	-1.066	
Caenorhabditis elegans	NP_001256160	1397	0.82	0.36	-1.065	
Caenorhabditis elegans	NP_001256161	1214	0.83	0.37	-1.066	
Caenorhabditis elegans	NP_001256162	821	0.81	0.37	-1.104	
Caenorhabditis elegans	CAB05543	733	0.83	0.36	-1.126	(*90*)
Caenorhabditis elegans	NP_001256163	1381	0.82	0.37	-1.065	
Caenorhabditis elegans	NP_001256164	1198	0.83	0.38	-1.066	
Caenorhabditis elegans	NP_001256165	805	0.81	0.40	-1.105	
Caenorhabditis elegans	NP_001256166	1349	0.82	0.37	-1.066	
Caenorhabditis elegans	NP_001256167	1166	0.82	0.36	-1.067	

Caenorhabditis elegans	NP_001256168	773	0.81	0.37	-1.108	
Caenorhabditis elegans	NP_001256169	1309	0.83	0.36	-1.075	
Caenorhabditis elegans	NP_001256170	1126	0.83	0.37	-1.078	
Caenorhabditis elegans	NP_001256171	733	0.83	0.40	-1.126	
Caenorhabditis elegans	NP_001256172	409	0.86	0.37	-1.226	
Caenorhabditis elegans	CAB05548	497	0.84	0.36	-1.054	
Caenorhabditis elegans	NP_001256173	556	0.82	0.37	-0.997	
Caenorhabditis elegans	NP_001256174	497	0.84	0.37	-1.054	
Caenorhabditis elegans	CAI46598	556	0.82	0.36	-0.997	
Caenorhabditis elegans	CBZ01819	1126	0.83	0.37	-1.078	(*90*)
Caenorhabditis elegans	CCA65580	409	0.86	0.36	-1.226	
Caenorhabditis elegans	CCA65581	1309	0.83	0.37	-1.075	(*90*)
Caenorhabditis elegans	CCF23418	805	0.81	0.36	-1.105	
Caenorhabditis elegans	CCF23419	773	0.81	0.36	-1.108	
Caenorhabditis remanei	EFO95235	821	0.83	0.38	-1.184	
Caenorhabditis remanei	EFO95236	843	0.83	0.38	-1.189	
Caenorhabditis remanei	EFO95291	1172	0.82	0.42	-1.369	
Caenorhabditis remanei	XP_003116339	821	0.83	0.38	-1.184	
Caenorhabditis remanei	XP_003116340	843	0.83	0.38	-1.189	
Caenorhabditis remanei	XP_003116395	1172	0.82	0.42	-1.369	
Cherax quadricarinatus	ALC79587	169	0.73	0.23	-0.057	
Dictyocaulus viviparus	KJH51853	535	0.71	0.25	-0.595	
Drosophila hydei	XP_023160045	233	0.77	0.26	-0.829	
Limulus polyphemus	XP_013783717	198	0.66	0.32	-0.196	
Oesophagostomum dentatum	KHJ93211	740	0.84	0.37	-0.921	
Oesophagostomum dentatum	KHJ93212	453	0.79	0.30	-0.510	
Polypedilum vanderplanki	BAE92617	180	0.74	0.45	-1.263	(*64*)
Polypedilum vanderplanki	BAE92618	484	0.58	0.27	-0.340	(*64*)
Polypedilum vanderplanki	BAN67644	143	0.78	0.44	-1.487	(*91*)
Polypedilum vanderplanki	BAN67645	709	0.76	0.44	-1.082	(*91*)
Polypedilum vanderplanki	BAE92616	742	0.67	0.35	-0.643	(*64*)
Ramazzottius varieornatus	BAQ94586	293	0.83	0.25	-0.942	(*92, 93*)

Ramazzottius varieornatus	A0A0E4AVP3	293	0.83	0.25	-0.942	(*92, 93*)
Saccoglossus kowalevskii	XP_006818499	118	0.52	0.26	-0.566	
Steinernema carpocapsae	ABQ23230	87	0.79	0.40	-1.211	
Steinernema carpocapsae	ABQ23231	95	0.76	0.41	-1.194	(*84, 90*)
Steinernema carpocapsae	ABQ23240	70	0.81	0.43	-1.139	
Teladorsagia circumcincta	PIO62605	594	0.81	0.34	-0.874	
Teladorsagia circumcincta	PIO73643	595	0.74	0.39	-0.905	
Teladorsagia circumcincta	PIO73975	634	0.74	0.41	-0.996	
Teladorsagia circumcincta	PIO74047	1580	0.83	0.34	-0.815	
Toxocara canis	KHN83840	600	0.82	0.27	-0.504	
Trichinella papuae	KRZ64074	66	0.74	0.17	0.170	

*ESTs were excluded from the analysis.

Table 2: Selected properties of tardigrade protein sequences retrieved from the supplemental materials of Boothby et al (*18*).

Organism	Accession Nr.	Length (aa)	Frac. Dis. Prom.	Frac. Char. Res.	GRAVY
Hypsibius dujardini	P0CU39.1	224	0.73	0.35	-0.314
Hypsibius dujardini	P0CU40.1	224	0.73	0.35	-0.529
Hypsibius dujardini	P0CU41.1	237	0.77	0.34	-0.635
Hypsibius dujardini	P0CU42.1	414	0.74	0.27	-0.41
Hypsibius dujardini	P0CU43.1	227	0.69	0.34	-1.175
Hypsibius dujardini	P0CU44.1	229	0.69	0.34	-1.167
Hypsibius dujardini	P0CU45.1	238	0.79	0.31	-0.984
Hypsibius dujardini	P0CU46.1	227	0.76	0.35	-0.878
Hypsibius dujardini	P0CU47.1	227	0.76	0.35	-0.825
Hypsibius dujardini	P0CU48.1	172	0.67	0.27	-0.985
Hypsibius dujardini	P0CU48.1	168	0.63	0.26	-0.985
Hypsibius dujardini	P0CU49.1	163	0.67	0.26	-1.019
P. richtersi	P0CU52.1	174	0.68	0.28	-1.074
P. richtersi	P0CU51.1	227	0.76	0.35	-0.894

Figure Captions

Figure 1: Mean net charge versus mean net hydropathy plots. (A) Intrinsically disordered proteins (white circles) and globular proteins (black circles) separate, which allows for deducing an arbitrary border (black line) for order-disorder prediction (*77*). (B) LEA proteins share distribution patterns with other IDPs. (C) TDPs distribute similarly to either side of the arbitrary border and (D) the comparison group of globular proteins mirrors the previously established boundary.

Figure 2: Phase diagrams of anhydrobiosis-related and known globular proteins. (A) Summary of phase-diagram regions for LEA proteins (black), TDPs (light gray), and globular proteins (dark gray). Individual phase diagrams for LEA proteins (B), TDPs (C), and globular proteins (D) are shown.

Figure 3: VL2 flavor and Sequence Compiler Analysis. (A) LEA proteins (black) and TDPs (gray) show different proportions in flavor of disorder. Frequencies of amino acid that compose LEA proteins relative to ordered proteins (B), TDPs relative to ordered proteins (C), and LEA proteins relative to TDPs (D). Bootstrap significance values are shown (<0.05 (*), <0.01 (**), <0.001 (***)).

Figure 4: Analysis of anhydrobiosis-related proteins from selected genera. (A) Mean net charge versus mean net hydropathy plot and phase diagram (B) for IDPs from *Adineta* (red), *Artemia* (dark green), *Caenorhabditis* (blue), *Polypedilum* (cyan), and *Hypsibius/Paramacrabiotus* (magenta). The comparison group of globular proteins is

shown in gray. Protein length (C), FCR (D), separation of charges I, and isoelectric point, for proteins from *Adineta* (Adi.), *Artemia* (Art.), *Caenorhabditis* (Cae.), *Polypedilum* (Pol.), and *Hypsibius/Paramacrabiotus* (Tar.).

Figures

Figure 1

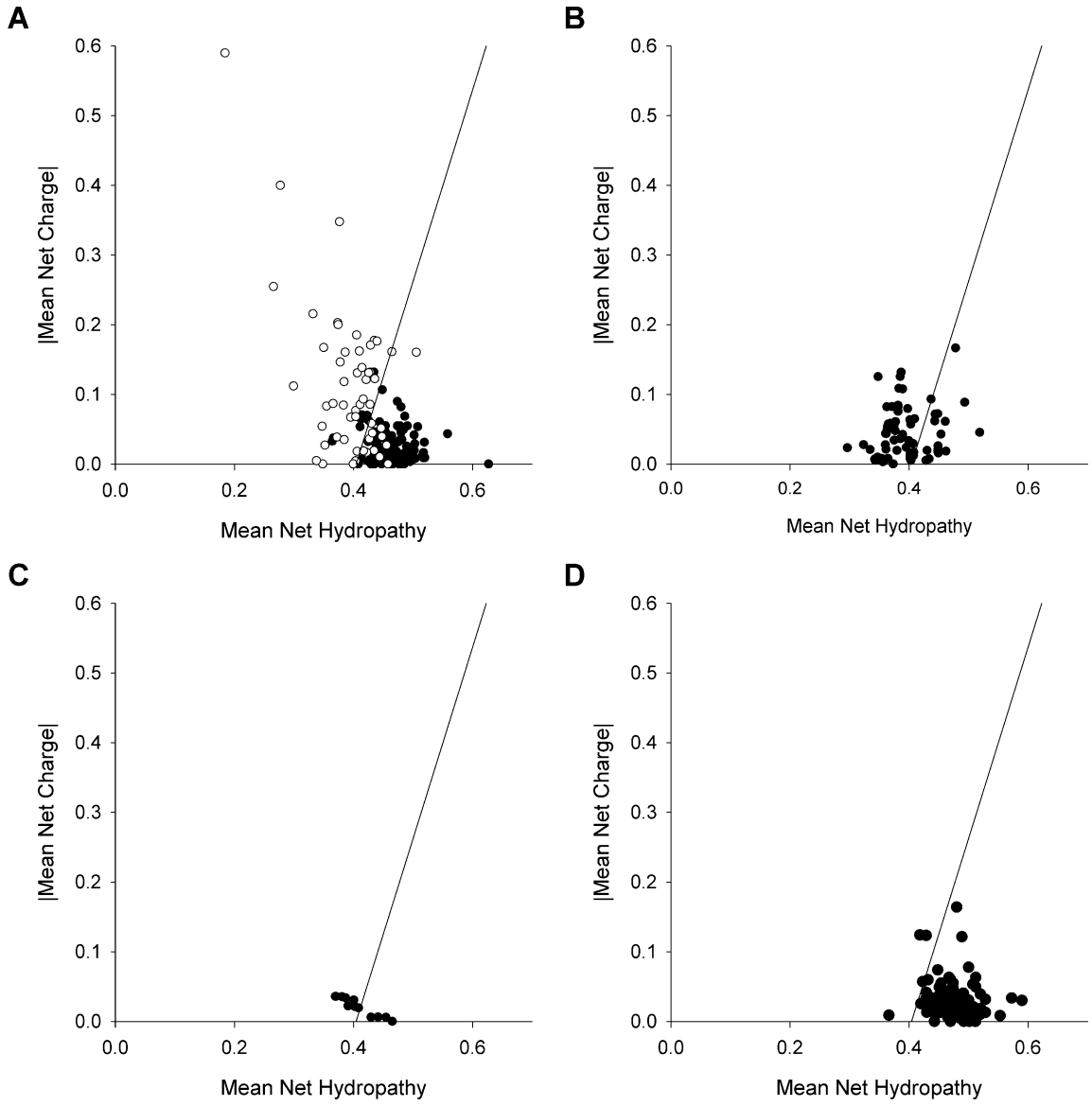


Figure 2

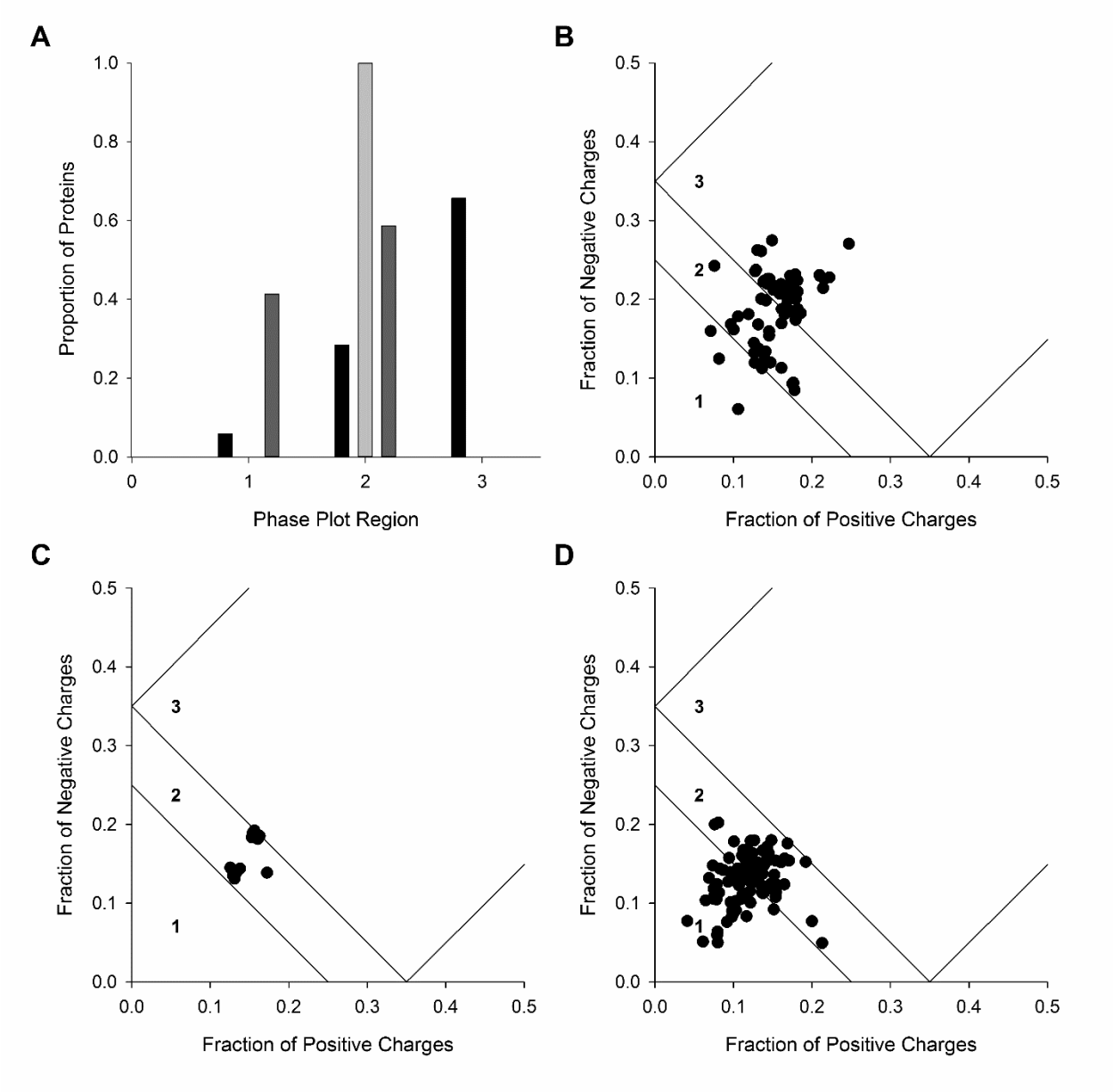

A
Proportion of Proteins
Phase Plot Region
B
Fraction of Negative Charges
Fraction of Positive Charges
C
Fraction of Negative Charges
Fraction of Positive Charges
D
Fraction of Negative Charges
Fraction of Positive Charges

Figure 3

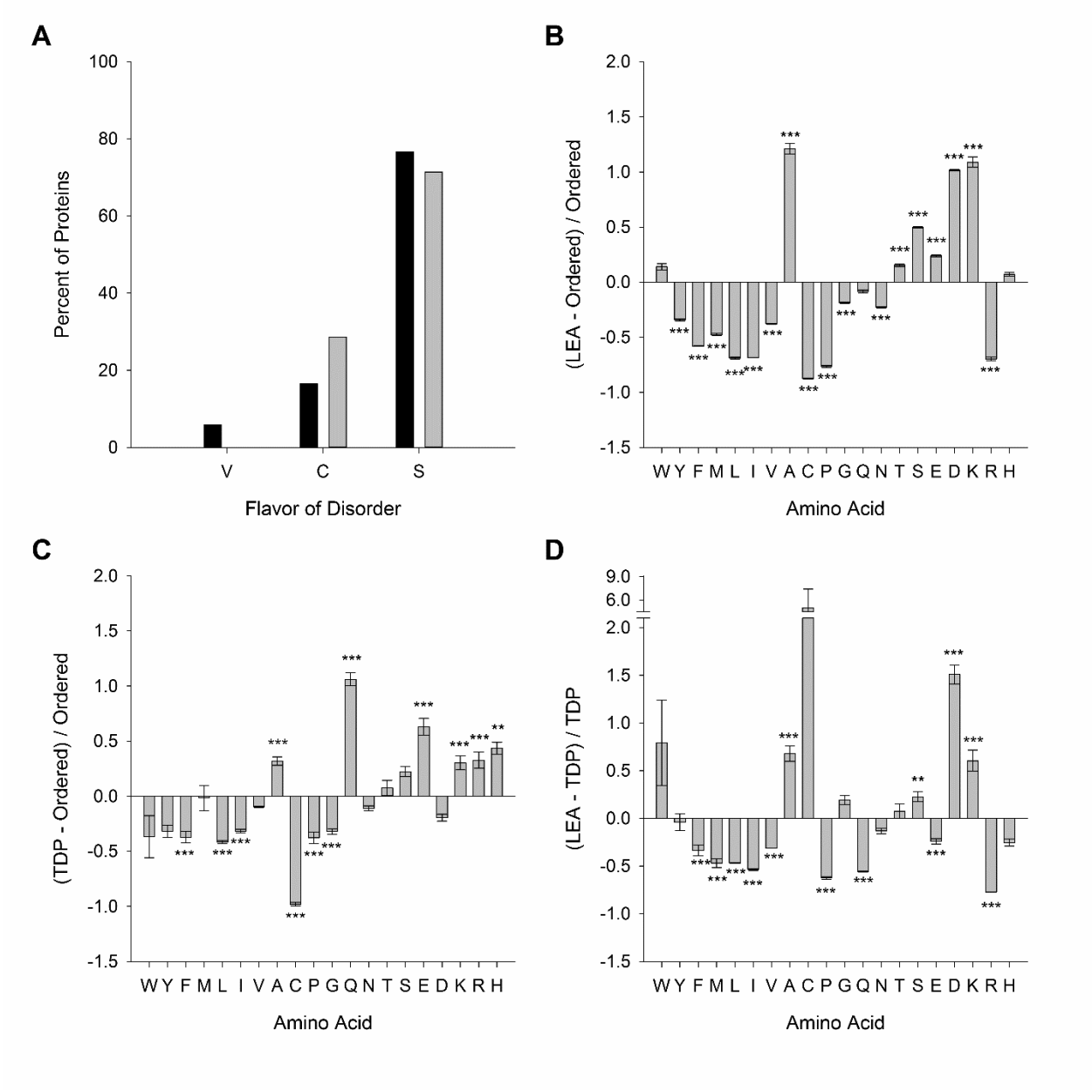
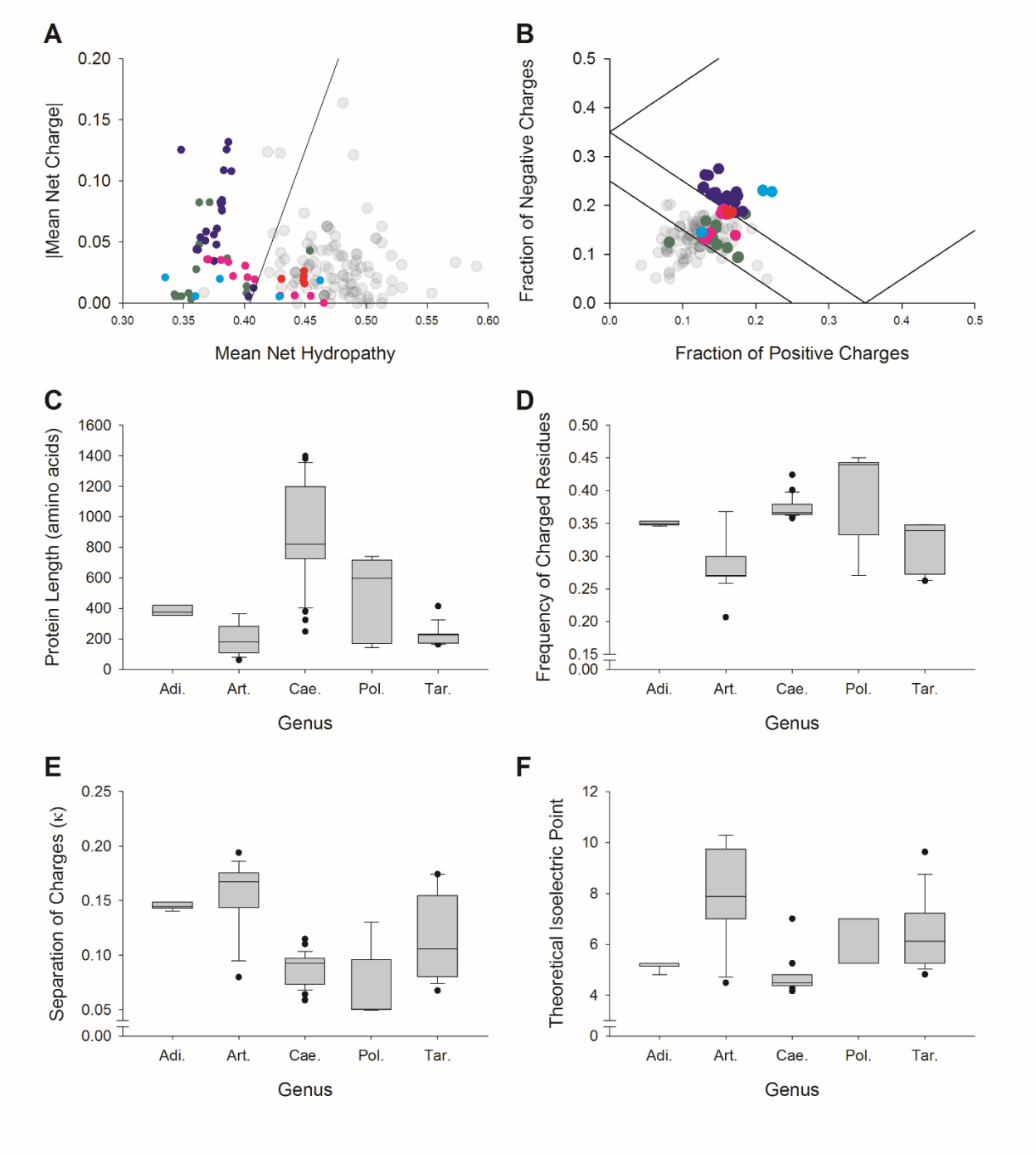

A
Percent of Proteins
100
80
60
40
20
0
V
C
S
Flavor of Disorder
B
(LEA - Ordered) / Ordered
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
W Y F M L I V A C P G Q N T S E D K R H
Amino Acid
C
(TDP - Ordered) / Ordered
Amino Acid
D
(LEA - TDP) / TDP
9.0
6.0
2.0
Amino Acid

Figure 4


A
|Mean Net Charge|
Mean Net Hydropathy
B
Fraction of Negative Charges
Fraction of Positive Charges
C
Protein Length (amino acids)
Adi.
Art.
Cae.
Pol.
Tar.
Genus
D
Frequency of Charged Residues
Genus
E
Separation of Charges (κ)
Genus
F
Theoretical Isoelectric Point
Genus

CHAPTER III

POTENTIAL FUNCTIONS OF LEA PROTEINS FROM THE BRINE SHRIMP ARTEMIA FRANCISCANA – ANHYDROBIOSIS MEETS BIOINFORMATICS

SUMMARY

Late embryogenesis abundant (LEA) proteins are a large group of anhydrobiosis-associated intrinsically disordered proteins (IDP), which are commonly found in plants and some animals. The brine shrimp *Artemia franciscana* is the only known animal that expresses LEA proteins from three, and not only one, different groups in its anhydrobiotic life stage. The reason for the higher complexity in the *A. franciscana* LEA proteome (LEAome), compared with other anhydrobiotic animals, remains unknown. To address this issue, a suite of bioinformatics tools was employed to evaluate the disorder status of the *Artemia* LEAome and to analyze the roles of intrinsic disorder in functioning of brine shrimp LEA proteins. *A. franciscana* LEA proteins from different groups are more like each other than one originally expected, while functional differences among members of group 3 are possibly larger than commonly anticipated. Our data show that although these proteins are characterized by a large variety of forms and possible functions, as a general strategy, *A. franciscana* utilizes glassy matrix forming LEAs concurrently with proteins that more readily interact with binding partners. It is likely that the function(s) of both types, the matrix-forming and partner-binding LEA proteins, are regulated by changing water availability during desiccation.

Janis et al. 2017. J. Biomol. Struct. Dyn.

INTRODUCTION

Late embryogenesis abundant (LEA) proteins constitute a large group of intrinsically disordered proteins (IDP) associated with anhydrobiosis, or 'life without water' (*94-96*). LEA proteins have been shown to improve desiccation tolerance in anhydrobiotic organisms (*90, 97*) and in desiccation sensitive cell lines that ectopically express them (*98*). Given the nature of anhydrobiosis, proteins that improve desiccation tolerance are difficult to characterize, because they likely remain mostly inactive in the fully hydrated state. Elucidation of LEA function(s) in the dried state represents another challenge since it excludes a variety of biochemical techniques commonly used to study proteins in solution. As a result, the functional structure of LEA proteins and their mechanisms of conferring desiccation tolerance have proven difficult to understand (*8, 16, 68, 69, 99*). Due to the challenges in directly or indirectly observing LEA protein structure and function at distinct water levels, several hypotheses for their mechanism(s) of functionality have been presented, including molecular shielding (*16, 20*), membrane stabilization (*41, 100-104*), sequestration of divalent ions (*105*), increasing the glass transition temperature of sugar glasses (*36, 106*), protection of proteins by prevention of protein aggregation (*60, 107, 108*), and acting as hydration buffers (*21*). Furthermore, functions of a given LEA protein may change with changes in hydration levels.

LEA proteins were first discovered in cotton seeds (*65, 109, 110*) and later were also found in seeds and vegetative tissues of several other plants (for review see (*4, 69, 70, 111, 112*)) and, more recently, in some anhydrobiotic animals, such as nematodes (*66, 83*), rotifers (*113, 114*), tardigrades (*115*), springtails (*116*), the sleeping chironomid *Polypedilum vanderplanki* (*64*), and the brine shrimp *Artemia franciscana* (*87*). LEA

proteins have been proven to be difficult to conceptually organize, resulting in several different classification schemes that propose 6 to 12 different protein families (for overview see: (*54, 117*)) (Table 3). Despite ongoing efforts to categorize LEA proteins into different functional groups, no classification method has been universally accepted. This lack of consensus may further illustrate the complex nature of these proteins and may resembles challenges associated with characterizing and classifying IDPs in general.

Depending on the amount of residual structure found in them, intrinsically disordered proteins (IDPs), at the whole molecule/domain level, can be organized into three distinct classes, such as native coils, native pre-molten globules, and native molten globules (*118, 119*). This categorization of whole molecule IDPs is based on their structural similarity to unfolded and different partially folded conformations detected for several globular proteins under various denaturing conditions (*33, 120-123*). Therefore, it seems that structurally, functional proteins can be classified as intrinsically disordered (coils, pre-molten globules, molten globules), and ordered (globular). In reality, this picture is more complex, since different parts of a protein can be differently disordered, thereby forming a protein structure continuum (*124, 125*). Of these structural subtypes, only globular proteins are considered as ordered in the classic sense, typically serving as illustrations of the standard 'lock and key' model of the protein structure-function paradigm. Note that transmembrane and structural, e.g., fibrillar, proteins are intentionally excluded from this consideration.

Coil-like polypeptide chains, or almost entirely disordered proteins, can have a hydrodynamic radius dramatically exceeding that of 'classic' globular proteins (*33, 119, 126*). The large hydrodynamic volume and the highly accessible structure of extended IDPs

makes them especially susceptible to degradation. Importantly, highly disordered polypeptides are frequently found as spacers and linkers between functional domains in globular proteins might have additional functions. For example, ligand-binding elements linked together by random coils increase binding affinity through the chelate effect (*127*). Extended IDPs and IDP regions (IDPRs) are highly susceptible to post-translational modifications, for example, containing up to 10 times as many phosphorylation sites as globular proteins (*128*). Pre-molten globular proteins (both IDPs or partially folded intermediates of globular proteins) contain of significant levels of secondary structure, but exhibit no globular tertiary structure and occupy about twice the volume of the molten globular proteins. Molten globular proteins are characterized by compact, globular conformations that contain high levels of defined secondary structure, but display limited tertiary features (*118, 119*).

IDPs/IDPRs display a variety of functions and functional mechanisms. They can show activity in their disordered state, often acting as chaperones, entropic chains, and recognition regions for interactions with a variety of partner molecules (*129*). Alternatively, IDPs/IDPRs can undergo disorder-to-order transitions, when their environment changes, such as during desiccation (*40, 130*), or in response to recognition of binding partners (*131*). In comparison with ordered proteins and domains, IDPs/IDPRs hold a variety of functional benefits (*132*), such as conformational plasticity (*133*), one-to-many and many-to-one signaling mechanisms (*134*), binding-site plasticity (*135*), thermodynamic regulation (*136*), and reduced cellular lifespan for transient expression patterns (*133*). In the case of environmental conditions causing a conformational transition in the polypeptide chain, random coil-like regions tend to undergo disorder-to-order

transitions more readily than pre-molten globular regions (*137*). These state transitions can occur due to target binding, changes in the chemical environment, or activation by post-translational modification, and several useful bioinformatics tools were developed to investigate potential biological functions of polypeptide chains with low structural complexity (*138-141*). In the study presented here, a variety of open access bioinformatics tools were employed to gain insights into the intrinsic disorder and potential function(s) of LEA proteins from the brine shrimp *Artemia franciscana*.

While several biochemical methods can be applied to characterize IDPs (for reviews see Methods in Molecular Biology volumes 895 (*142*) and 896 (*143*)), great strides have been made in developing bioinformatics tools to explain and/or predict potential structural and functional elements of IDPs/IDPRs, to guide future research, and to assist in data interpretation (*144-146*). Many of these programs have a high accuracy in predicting IDPs and the localization of IDPRs. In general, IDPs have an amino acid composition biased towards residues that promote disorder such as alanine (Ala), glycine (Gly), aspartic acid (Asp), methionine (Met), lysine (Lys), arginine (Arg), serine (Ser), glutamic acid (Glu), and proline (Pro) (*78*). Additionally, certain motifs of physicochemical characteristics are common in amino acids sequences of IDPs, such as Positive (pos)-Pos-X-Pos, Negative (Neg)-Neg-Neg, Glu-Glu-Glu, Lys-X-X-Lys-X-Lys, and Pro-X-Pro-X-Pro (*147, 148*). The amino acid composition may also be associated with different "flavors" of disorder (*76, 147*) that have weak but statistically significant associations with protein function. Given the relatively low complexity of IDP structures, amino acid sequence data has been used in bioinformatics programs in order to predict IDP function with some success (*149*).

The brine shrimp *Artemia franciscana* is the only known animal expressing three different groups of LEA proteins (1, 3, and 6; for alternative classifications please refer to Table 3) in the anhydrobiotic life stage (*2, 63, 150, 151*). The reason for the higher complexity in the *Artemia* LEA proteome compared with other anhydrobiotic animals that only express group 3 LEA proteins is unknown. We hypothesized that distinct functional differences among the three LEA groups may exist and offer additive or synergistically advantages to the anhydrobiotic stage in brine shrimp. To test this hypothesis, we employed a wide spectrum of bioinformatics tools.

METHODS

Predictor of Naturally Disordered Regions (PONDR)

The Predictor of Naturally Disordered Regions (PONDR) is a web server for intrinsic disorder prediction based on the input amino acid sequence of a query protein. PONDR server utilizes a combination of computational tools including several feedforward neural networks (PONDR® VLXT (*152, 153*), PONDR® VSL2 (*154*), and PONDR® VL3 (*155, 156*)), and two binary disorder predictors that evaluate the probability of a query protein to be disordered as whole, Charge-Hydropathy (CH-plot) analysis (*77*), and a Cumulative Distribution Function (CDF) analysis (*157, 158*). CDF analysis is based on one of the outputs of PONDR® VLXT and summarizes the per-residue predictions by plotting PONDR scores against their cumulative frequency, which allows ordered and disordered proteins to be distinguished based on the distribution of prediction scores (*157, 158*). PONDR is freely available at http://www.pondr.com.

Metaserver of Disorder (MeDor)

Metaserver of Disorder (MeDor) is a freely available platform that predicts the structure of a protein based on the input amino acid sequence, provides a hydrophobic cluster analysis (HCA) plot that projects the protein in α-helical orientation, submits the sequence to several protein disorder and localization prediction servers (such as MeDor submits the amino acid sequence to IUPRED (*159*), PreLINK (*160*), RONN (*161*), FoldUnfold (*162*), DisEMBL1.5 (REM 465, loops, and hotloops) (*163*), FoldIndex (*164*), Globplot2.3 (*165*), PONDR® VL3 (*155*), PONDR® VL3H (*155*), PONDR® VSL2B (*166*), and Phobius (*167*)), and juxtaposes the results of each for ease of analysis (*168*). MeDor offers secondary structure prediction using the Secondary Structure Predictor (SSP) Pred2ary (*169*). The HCA plot is a useful tool for visual detection of disordered and potential binding regions by highlighting hydrophobic clusters and representing the characteristics of secondary structures by coloring residues based on their chemical properties (*170*). MeDor is no longer available at this time. However, every program aside from Pred2ary is still available at their dedicated websites.

IUPred and ANCHOR

IUPred is a disorder prediction server that uses pairwise energies of potential interactions between amino acid to predict the likelihood of disorder (*159, 171*). IUPred predicts disorder based on two reading lengths, long regions of 30 or more amino acids and short regions of 25 or fewer amino acids. The updated versions of IUPred are freely available at https://iupred2a.elte.hu/ (IUPred 2) and http://iupred.enzim.hu/https://iupred.elte.hu/ (IUPred3).

ANCHOR is a Molecular Recognition Feature (MoRF) prediction server that uses similar pairwise energies as IUPred employs, but combines them with characteristics of known MoRF regions (*28, 137*). ANCHOR, while not a trained algorithm, was tested on various data sets and predicted protein binding MoRF sites with 70% accuracy and a false-positive rate of <5% in globular protein datasets (*28*). ANCHOR specifically identifies protein-binding MoRF regions. ANCHOR is freely available as an incorporated component of IUPred 2 and 3.

DisEMBL1.5

DisEMBL is a disorder prediction server that utilized three artificial neural networks for structural analysis, Loops/Coils, Hot Loops, and Remark-465 (*163, 172*). The Loops/Coils predictor is based on proteins from the Dictionary of Secondary Structure of Proteins (DSSP) and contains ~57% of disordered residues (*163, 173*). It accurately predicts only ~50% of ordered sequences, but regions known to be disordered are extremely rarely predicted to be ordered. The Hot Loops predictor utilizes B-factors from the X-ray crystallography structures (*163*). It also was trained on DSSP proteins with disordered residues and includes proteins representing members of each protein family listed in the database. The Remark465 neural network is trained on the stretches of amino acids with missing electron density in X-ray crystallography structures (*163*). Remark465 has a false positive rate of ~16%, likely because missing electron density is only partly due to protein disorder. DisEMBL1.5 is freely available at http://dis.embl.de/.

GlobPlot2.3

GlobPlot is a propensity-based server for prediction of structural disorder and globular domains (*165*). GlobPlot utilized the Remark465 propensities, and its output may be adjusted. The default output is a sloped graph, in which negative slopes represent propensity for ordered domains and positive slopes indicate disorder predictions. Using the SMART server, coiled-coil regions and low complexity regions are highlighted as striped boxes or empty boxes, respectively. Along the bottom of the graph, GlobPlot gives a color-coded predictor of regional structure, with no color indicating uncertainty or structural flexibility. GlobPlot2.3 and all propensity sets are freely available at http://globplot.embl.de/.

Heliquest

Heliquest projects amino acid sequences as α-helices, calculates the physicochemical properties of these α-helices, and plots two superimposed graphs of hydropathy and hydrophobic moment at each amino acid position (*174*). Corresponding projections and graphs are derived from a sliding window, which the user can select to range from 11 to 54 residues. For each projection, an accompanying table includes the number of charged, polar, and uncharged residues, as well as special residues such as proline and cysteine. The table also includes standard hydropathy (*175*), hydrophobic moment (*175*), and net charge assuming a pH of 7.4. Heliquest is freely available at http://heliquest.ipmc.cnrs.fr/.

DISPHOS 1.3

DISPHOS 1.3 is an online phosphorylation prediction server specialized in

identifying phosphorylation sites in the context of protein disorder (*176*). To assess potential phosphorylation sites, DISPHOS 1.3 predicts the surface exposure, electrostatic charge, hydropathy, and flexibility of amino acids that neighbor serine, threonine, and tyrosine. DISPHOS 1.3 is trained on specific data sets in the SWISS-PROT database, such as Eukaryotes, or specific model organisms to reduce mischaracterizations (e.g. *Caenorhabditis elegans*, *Drosophila melanogaster*, *Homo sapiens*, etc.). For the purposes of the analysis presented here, we used the predictor trained on proteins from *D. melanogaster* since both *A. franciscana* and *D. melanogaster* are arthropods. DISPHOS 1.3 is freely available at: http://www.dabi.temple.edu/disphos/.

CIDER/localCIDER

CIDER is a server that returns sequence-specific parameters such as the length, distribution of opposite charges (κ), the Frequency of Charged Residues (FCR), the Net Charge Per Residue (NCPR), hydropathy according to the Kyte & Doolittle scale (*177*), the proportion of disorder promoting residues, and plots the protein on a diagram of states for a prediction of the structural qualities of a query protein (*75*). The distribution of opposite charges, represented as κ, is scaled between 0 and 1, where 0 represents a perfectly even distribution of charges across the protein and 1 indicates complete separation of charges. This measure is useful for identifying self-repulsion or attraction, especially in the desiccated state for LEA proteins. LocalCIDER is a high-performance software package that offers a more advanced analysis of protein sequences, including plotting parameters, such as NCPR for example, with a defined window size. Several other parameters may also be calculated or modified, such as calculating poly-proline helix propensity and changing

the hydropathy or complexity. CIDER and localCIDER are freely available at http://pappulab.wustl.edu/CIDER/analysis/.

Sequences used for Analysis

>AfLEA1.1
MELSSSKLNRSIFKRRSKMSEQGKLSRQEAGQRGGQARAEQLGHEGYVEMGRK
GGQARAEQLGHEGYQEMGQKGGQARAEQLGTEGYQEMGQKGGQKRAEQLGH
EGYQEIGQKGGQTRAEQLGTEGYQEMGQKGGQTRAEQLGHEGYVQMGKMGG
EARKQQMSPEDYAAMGQKGGLARQK

>AfrLEAI
MAEPEEPPGIYEKVKSAFVSAPDRAQEAYNQAYESARSVFDDAVRSARKMKNT
AAEQAQGAYEGLKESPENLQRVTRDIYHQAQDTGKGAYETVAGSADDAYRRA
QETAQAAQEQSKGFLNRVKDTLTAPFSSSSDQAKETYDRTKDEAQYRAQQAAD
AGQGFFGKVKDTITAPFTSGYDQTQEGYERARRSAEEAAQQAADQGQTLFERA
KDTITSPFSSGSEQAQESFERAKRAAEEQVEQSKGMFQNIKGTITSPFNSAADTAK
EAGQRAKKQAEEAADQSQGFMQKVKDTVASPFLSAGEESQEAIERTKREAEEAR
HQGEGFLHRVADTIMHPFQSSSEQVGEAADRIKRGA

>AfrLEA2
MPKAAAKGIGETVKADADVVEGMASTGYEKLKSAFGIASNKTKDAAENVAESA
RATKDYTVDSAKSAYDKTVDSTKSAYDKTTDSAKSVHDSTADTAKSAYNKATE
TLGSAYDKTKDTAQSTYDQVTGAAHSAYDKTAEATKSAYDKTADAAHSVYNK
TGDAGKQAYDSTKEAARSTGKSISDAAYFTGKGAERQGDQVKSELPSYSPSSSG
EKLAQHLVKSEKEGKKLTEEALKDRDLSQVPGFRSVKKAHEPDAKEDISAVDFA
SASPSQRKVADTEGVWSSPVDRQESRFFSDLAGKIGDMLGGGKINAIQTPEEMD
HERLIHKSSQSQVAGNVPGRAKTAWTPEDRIILHQERFPKENPE

>AfrLEA3m
MLSKRLIKSLSCVSRTELRAFSGTTSCCLQQKDLDKNKGDTPPPSREHEEQEGVF
KRAMEKAKGEYDPEYPLSSSMKATKDVAKDVAEGAKEKVKSAYESIKESVSSTS
SEAQNRGESMYGKTKETVSDTANKAKEKAESMYDTAKETAKSGADKLSWEDT
KETYKEKAGEIKERIQDTAESMKERMGETGHNMKEKMQHTGQSMKEGMKESW
ESLKDTAKQTKEGAHDQWNTAKDKTKEVKDAASEKMSNSVDKTLKRGEKVSE
RVTEMYSGTKGDSKGGSGFNQITPEQTENMKGQQSASGAHER

>AfrLEA6

MSENIGHININANLQNVDRRDAAAIQSVERKLLGYNPPGGLASEAQSAAALNEGI
GQPMNRGISTDIPAPADIDVDRGTASKDFGHVRFDVDLNQVRPEEAAALQAAES
KIEGLAPSITVGGIGSAAQSMAAFNEREQSETGPFHPGIKATEPLPGPTYYQGVEL
SPSALPTYAPDVSVFPPSLSTNTSNVGAVPPSITTYSPDAGANDWERVYRKTTKT
TQRIAIPGGIEDIVDEGKLGEAPRTNIRS

RESULTS AND DISCUSSION

Within the last 14 **y**ears, LEA proteins have been found to accumulate in some desiccation tolerant animals including the brine shrimp *A. franciscana* (for review see (*2, 63*)). However, *A. franciscana* expresses multiple LEA proteins from three classification groups (group 1, 3, and 6) in the desiccation tolerant embryo, making it unique among anhydrobiotic animals (*2, 85, 151*). The reason(s) for the presence of a larger variety of LEA groups in *A. franciscana*, compared to other anhydrobiotic animals, is unknown. It seems reasonable to assume that proteins from different LEA groups may offer distinct or additive benefits to the animal if group specific differences in protein functions exist. However, even in the absence of group-specific functional differences, a large variety of LEA proteins might be necessary to confer desiccation tolerance in anhydrobiotic animals. The reasons for concurrent expression of multiple LEA proteins may include targeting different types of macromolecules (lipids, nucleic acids, proteins) or different members of the same macromolecular type, to serve different molecular functions (ion chaperones, molecular shields, structural reinforcement), and/or are to localize to different subcellular compartments.

Group 1

AfLEA1.1

A large number of highly similar group 1 LEA proteins has been described in *A. franciscana* (*85*) and two of them, *Af*LEA1.1 and *Af*LEA1.3, are almost identical, except that *Af*LEA1.3 contains an N-terminal signal sequence and localized to the mitochondria, whereas *Af*LEA1.1 lacks a signal sequence and is retained in the cytoplasm (*98, 178-180*). Mitochondrial signal sequences are usually cleaved after incorporation of the protein into the mitochondrial matrix (*181, 182*). Therefore, the cytoplasmic protein *Af*LEA1.1 will be analyzed below as an illustrative representative for the other *A. franciscana* group 1 LEA proteins that basically differ only in the numbers of a repeat of a 20-amino acid long sequence motif (*179*).

The first group 1 LEA protein in *A. franciscana* was described by Sharon and colleagues as a heat stable and highly hydrophilic 21-kDa protein (*85*). This protein contains a characteristic 20-amino acid motif (GGQTRREQLGEEGYSQMGRK), and several protein variants including 2 to 8 repeats of this motif have been discovered (*85, 179, 180*). The mean net charge and low hydropathy shown in the CH-plot (Fig. 5A) place *Af*LEA1.1 in the category of proteins with extended disorder, which is not surprising given the particularly high percentage of charged and polar residues (52.8%) in this protein. This is also in agreement with the output of CDF analysis (see Fig. 5B) which further supports the notion of a highly-disordered nature of *Af*LEA1.1. In fact, it was established earlier that seven boundary points located in the 12^{th} through 18^{th} bin provided the optimal separation of the ordered and disordered protein sets in the CDF plots and that classification of a query protein as wholly ordered or wholly disordered is based on whether a corresponding CDF

curve was above or below a majority of boundary points, respectively (*158*). According to these criteria, *Af*LEA1.1 is expected to be disordered as a whole.

In the CH-plot, *Af*LEA1.1 is closest to the group 3 LEA protein *Afr*LEA3m (*86*) whose secondary structure, along with that of *Afr*LEA2, has been characterized using circular dichroism (*29*). *Af*LEA1.1 most closely resembles *Afr*LEA2 in terms of its proportion of charged residues, but *Af*LEA1.1 has greater separation of its charged residues (Table 4), although both *Af*LEA1.1 and *Afr*LEA2 are being classified as Janus sequences by CIDER (Fig. 5C). In the desiccated state, electrostatic interactions likely hold greater impact on folding dynamics than in the hydrated state.

Therefore, lower absolute mean net charges combined with higher κ values may become particularly influential in predicting secondary and tertiary structure motives in the dry state. The distribution of positive and negative charges alternates repetitively due to the 20-amino acid sequence motif, creating several points for favorable electrostatic interactions within this center region of the protein (Fig. 6). This separation of charges along the sequence likely causes *Af*LEA1.1 to adopt electrostatically driven structures in the dry state that could be influenced by the presence or absence of ions.

We combined disorder predictions derived from applying several different algorithms to understand potential structural features in the hydrated and desiccated states of *Af*LEA1.1. DisEMBLE predicts *Af*LEA1.1 to be overall disordered (63.3%), with different likelihoods for ordered or disordered states at distinct regions within the polypeptide chain (see Supplementary Materials, Fig. 7). Stretches of the protein where ordered structure is predicted by MeDOR-based DisEMBL (Fig. 8A), show a strong tendency for β-strands in the hydrated state, a structure not as commonly found in group 1

LEA proteins as α-helices (*99*). However, experimental analysis is needed to confirm this prediction. Our knowledge of secondary structure of group 1 LEA proteins is limited to plants, where most group 1 LEA proteins have been shown to be highly disordered, or to contain up to 47% of α-helices (for review see: (*69*)). Surprisingly, most of the predicted α-helices in *Af*LEA1.1 fall into regions that are likely to be disordered, suggesting that α-helices can only be formed in response to interactions with a binding partner or during desiccation. The β-strands, however, appear to more likely occur in the hydrated protein, with the potential for increased folding in less polar solvents or during desiccation.

Fig. 9A shows that *Af*LEA1.1 is predicted to have several regions that possess an ambiguous propensity for ordered and disordered structure that coincide with the positions of MoRF regions predicted by ANCHOR (Fig. 9A). Although ANCHOR predicts MoRF regions spaced relatively evenly across the protein, four MoRF regions with a likelihood greater than 80% are localized in pairs at the protein termini (amino acid positions 1-20 and 39-53 in the N-terminal region, and regions 140-154 and 163-180 and the C-tail). These terminal MoRF regions have distinct amino acid sequences not found in other regions. Sudden dips in the Globplot slope (see Supplementary Materials, Fig. 10) are predicted to form β-strands at positions 39-53 and 140-154 that separate the terminal MoRF pairs from the internal regions. These predicted β-strands are characterized by a specific clustering of hydrophobic residues, high glycine content, and complementary charges of basic and acidic amino acids. The finding that the primary amino acid sequence of these two 15 amino acids long MoRF regions are distinct from the other regions, while sharing an almost identical 13 amino acid overlap, suggests that they are either separating functional segments or are involved in orienting them. The two terminal MoRF regions

(residues 1-20, 163-180) have pronounced structural differences, which is shown by ANCHOR as well as by sudden drops in PONDR® VLXT profile (Fig. 9A). Furthermore, PONDR® VL3 predictor weakly indicates that the N- and C-terminal MoRF sites might exhibit unique structural elements.

DISPHOS predicts the N-terminal region of the *Af*LEA1.1 to be heavily phosphorylated if translated in *D. melanogaster* (see Supplementary Materials, Fig. 11). The N-terminal region is highly enriched in positively charged residues (30%), serine residues (30%) predicted by DisPHOS to be phosphorylated, and contains a cluster of hydrophobic residues (25%). All eight serine residues within the N-terminus are predicted to be phosphorylated, with seven phosphorylation sites being located within the first 26 amino acids of the protein. Residues 4, 5, and 6 are consecutive serine residues resembling an α-helix cap, which may promote α-helical stability during desiccation (*183*). This high concentration of likely phosphorylation sites further distinguishes the two N-terminal MoRF regions from the other regions of the protein.

Considering that α-helices may be important to LEA protein structure and function, Heliquest algorithm was used to evaluate the properties of any α-helices that might be formed in the hydrated and/or desiccated states (Fig. 12A). Interestingly, an α-helix within the N-terminal MoRF region would have a high hydrophobic moment, due to a small but concentrated hydrophobic face. An α-helix within the MoRF region at the C-terminus, on the other hand, would have a very low hydrophobic moment due to a relatively even distribution of hydrophobic residues. This means that, if *Af*LEA1.1 would interact with phospholipid membranes, this may occur at the N-terminus, but not at the C-terminus. The penultimate MoRF regions both exhibit a hydrophobic face composed of the six amino

acids sequence “AMGGY”, although the hydrophobic face of the MoRF in the 140-154 region is extended to “LMGAMGGY”. This suggests that, if α-helices were to form in these two regions**,** then the formed structure would be an amphipathic α-helix with substantial flexibility due to the 2 or 3 glycine residues in this structure. However, given the helix-breaking propensity of glycine residues, the odds of these structures forming are low.

The Heliquest-based predictions of an α-helical region with a continuous hydrophobic stripe can be visualized on the HCA (Fig. 8A). This band becomes most pronounced at the predicted internal MoRF regions and less pronounced at the termini of the protein, again suggesting different functional behaviors for the termini compared to the internal regions of the protein. Another noteworthy observation is that the SMART server describes *Af*LEA1.1 as a protein containing quadruple repeat of LEA_5 (PF00477) domains, which are found in hydrophilic plant seed proteins. Furthermore, close homologies of the PFAM LEA5 domain to EM-like proteins were found using NCBI tblastp (e.g., e-value of $8e^{-60}$ to GEA1 from *Camelina sativa* XP_010503885).

Group 3

AfrLEAI

*Afr*LEAI maintains a ratio of hydropathy to mean net charge of 0.094 which is similar to the group 6 LEA protein *Afr*LEA6, but higher than those of the other Artemia LEA proteins (Fig. 5A, Table 4). The overall charge of *Afr*LEAI is negative, and the CDF analysis predicts the protein to be pre-molten globular or to contain a mixture of coils and globular structures (Fig. 5B). CIDER predicts *Afr*LEAI to be a Janus sequence, like

*Af*LEA1.1, and to undergo environmental conditions-dependent conformational transitions (Fig. 5C). *Afr*LEAI is predicted by the SMART server to be the most repetitive of the group 3 LEA proteins identified in *A. franciscana* with two distinct sets of repeating motifs. The first set of repeats spans amino acid position 5-47 and 56-98 (Fig. 8B). Further inspection of the sequence suggests that the physicochemical properties of the repeats are conserved for positions 5-58 and 60-118. These repeats are highly enriched in aromatic residues, which is a unique feature among the LEA proteins in *A franciscana*. Furthermore, both repeats are enriched in alanine, which is well-established as an α-helix forming amino acid (*184*).

The second set of repeats spans amino acid positions 116-221 and 244-331. Each of these repeats consists of three highly conserved motifs composed of a hydrophobic cluster containing a proline-phenylalanine pair, which is followed by regions enriched in alanine, aromatic residues, and clusters of negative amino acids which are separated by three arginine residues (Fig. 8B). These repeats are predicted to contain coiled-coil regions at positions 98-125, 186-252, and 304-327. The hydrophobic cluster regions are, once again, enriched in aromatic residues, such as phenylalanine and tyrosine. Being enriched in alanine and complementary charges, these regions are predicted by MeDOR-based Pred2ary algorithm to readily form α-helices. Additionally, any α-helices in this region would have a hydrophobic face due to the linear alignment of hydrophobic residues on the helix surface. This face would be flanked on one side by an alternating negative-positive-negative stripe and a thin polar stripe, similar to *Af*LEA1.1 (Fig. 8B). The N-terminal domain has a consistently oscillating hydropathy, correlating to the charged, alanine rich regions, and aromatic hydrophobic clusters. Combining these amphipathic α-helical

tendencies with the coiled-coil behavior predicted by the SMART server suggests that *Afr*LEAI may form a bundle of amphipathic α-helices capable of interacting with phospholipid bilayers and monolayers. An NCBI BLAST of *Afr*LEAI supports this interpretation considering the homologies found to perilipin proteins that are known to interact with phospholipid monolayers (e-value of 3e-12c to perilipin-4, XP_013194305).

ANCHOR and PONDR® both predict several different MoRF regions in *Afr*LEAI (Fig. 9B). The close agreement between these two programs suggests that *Afr*LEAI may undergo extensive conformational transitions either through the loss of water interactions or by contact with target molecules. Given the degree of shift in the PONDR® VLXT score, it may be that *Afr*LEAI binds to proteins or lipids under conditions of minimal water reduction, or even in the hydrated state, but considering the negative charge of the protein, it is unlikely that *Afr*LEAI will interact with nucleic acids. PONDR® VL3 profile suggests that the C-terminal repeats can be structurally segregated into separate domains, as well as the first two repeats of the N-terminal region (Fig. 9B). The second pair of repeats are combined in one domain, which correlates to a coiled-coil prediction by the SMART server (Fig. 8B). PONDR® VL3 predicts that the final two repeats of the N-terminus fall into one single structural domain, which is distinct from the first two domains. Both programs predict high MoRF potential in the hydrophobic, aromatic half of each repeat, which is separated by a proline residue from the more hydrophilic half. The conservation of aromatic residues in this region offers insights into potential binding partners, or points to aromatic stabilization of the structure (*185*). The highly charged, alanine-rich halves of the N-terminal repeats are predicted to be disordered in the hydrated state by both IUPred and PONDR, but any conformational shifts during desiccation would favor α-helical

conformations with a high capacity for tertiary structure due to alternating charges represented by a κ value of 0.145 (Table 4).

AfrLEA2

Compared to *Afr*LEAI the protein *Afr*LEA2 has a substantially lower mean net charge over hydropathy ratio and is the second most hydrophobic LEA aside from *Afr*LEA6 (Fig. 5A, Table 4). *Afr*LEA2 has been shown to have protective effects on lipid vesicles (*42*) and cytoplasmic and mitochondrial enzymes during desiccation, although the protection was not dramatically better than that conferred by bovine serum albumin (*29*). Many group 3 LEA proteins are characterized by repeating amino acid motifs that may fold into amphipathic α-helices during desiccation (*16*), however *Afr*LEA2 does not contain a repeating sequence. Additionally, secondary structure data for *Afr*LEA2 using circular dichroism values at $[\theta]_{200}$ (-10205.4) and $[\theta]_{222}$ (-1509.88) suggest that the protein is most likely pre-molten globular in the hydrated state, with a net ensemble of ~19% β-sheets, ~4% α-helices, ~15% turns, and ~62% random coils. When desiccated, *Afr*LEA2 exhibited only ~5% β-pleated sheet structure, but the α-helical content increased from ~4% to ~50%, while turns remained at 15% (*29*), which agrees with CIDER prediction of Janus sequence-like structural plasticity (Fig. 5C). While this data sheds light on the degree of secondary structure adoption that *Afr*LEA2 undergoes during desiccated, the actual structure of any given polypeptide strand in the sample may vary substantially within the conformational ensemble, or may shift from one conformation to another in the hydrated state (*122*). Furthermore, some LEA proteins have been observed to undergo different conformational transitions depending on the presence of monovalent or divalent ions (*186*).

However, even with structural plasticity and ion-interactions considered, the shift in the prevalence of ordered secondary structure during desiccation suggests a transition from a native pre-molten globular structure to a potentially active molten globule. This prediction is further supported by the CDF analysis, which places *Afr*LEA2 both above and slightly below the boundary for molten globular and globular proteins (Fig. 5B). Therefore, experimental evidence regarding structural uniformity or localization of structural motifs in the *Afr*LEA2 polypeptide is needed to gain further insight into the specific mechanisms by which this protein may increases desiccation tolerance in *A. franciscana*.

DisEMBL predicts an overall degree of disorder of approximately 70.3% (see Supplementary Material, Fig. 13), which is close to the circular dichroism data and the IUPred and GlobPlot outputs according to which *Afr*LEA2 is expected to have 78.6% and 78.5% disorder, respectively. The agreement among the predicted and experimental data is encouraging for our approach of combining the localized structural predictions with the circular dichroism data of *Afr*LEA2 to elucidate local structural propensities in the polypeptide chain. Given that these programs are trained to distinguish IDPs and IDPRs from globular proteins and domains, and they accurately predict the degrees of order in *Afr*LEA2, then the positions of these ordered regions might be reliable. Furthermore, the IUPred accuracy in determining *Afr*LEA2 structure is inspiring for the application of ANCHOR, which uses similar techniques (*28*). Based on this analysis, *Afr*LEA2 in the hydrated state is likely composed of a β-sheet in the first 81 amino acids of the structure and a highly disordered, C-terminal tail with some α-helical tendency at amino acid position 280-300. Perhaps most notably, the Remark-465 predictions were the most accurate from GlobPlot and DisEMBLE, which suggests that the *Afr*LEA2 curve on the

CDF suggests a combination of ordered structures and disordered regions rather than a cohesive molten globule in the hydrated state. It should be noted that the Pred2ary predictions from MeDOR significantly deviated from the experimental data, which suggests that these ordered regions are small and may interact with turns (Fig. 8C).

For *Afr*LEA2 to follow molten globular and globular folding patterns, it would need to be structurally distinct from the other group 3 LEA proteins in *A. franciscana*. This hypothesis is supported by the difference in both structure and conformational changes during drying observed for *Afr*LEA2 when compared to *Afr*LEA3m (*29*). From a bioinformatics perspective, the amino acid sequence of *Afr*LEA2 is indeed distinct from all other *A. franciscana* LEA proteins. As aforementioned, the net mean charge of *Afr*LEA2 is low, due to the positively and negatively charge residues being well balanced (58 negative and 53 positive residues), which make up approximately 30% of the protein. Furthermore, *Afr*LEA2 shows no signs of repeat sequences, whereas all other LEA protein contain several repeating sequences, sometimes making up almost the entire protein. The lack of repeating sequences is particularly surprising because *Afr*LEA2 is the largest known LEA protein in *A. franciscana*. This finding becomes even more noteworthy in the context of LEA proteins in general, which are characterized by the presence of specific repeating motifs that are typically used for the classification of LEA proteins (*16*).

Several other unique features are observed in *Afr*LEA2. The protein shows an uneven distribution of proline and arginine residues throughout the polypeptide chain. Of the 12 proline residues in its sequence, 11 are observed after position 200 and six of them fall between amino acids positions 200 and 290 (Fig. 8C). Similarly, of the 12 arginine residues in of the protein, nine are observed after position 235, whereas the other charged

residues appear to be relatively equally distributed throughout the protein. This suggests that in the region from amino acid 200 to 364, any secondary structure elements that may form under any condition would be interrupted by proline or glycine residues every 10-40 amino acids. DISPHOS predicts 18 phosphorylated serine residues in *Afr*LEA2 (see Supplementary Material, Fig. 14) and 13 fall between amino acid positions 200 and 290. These predicted phosphate groups may help to overcome electrostatic repulsion in the protein.

Also, contrasting to the other group 3 LEA proteins, *Afr*LEA2 does not show an even distribution of its predicted MoRF regions (see Fig. 9C). Aside from small MoRF regions with relatively low probability at positions 30-37 and 151-156, ANCHOR predicts the MoRFs to mainly occur downstream of a high probability MoRF at position 180 – 189. Following this 10 amino acid MoRF are three MoRF regions of nine amino acid that are spaced evenly every 15 amino acids apart from each other. Unlike other LEA proteins, these MoRFs are not highly similar in sequence. After this region of small MoRFs follows a region containing three larger MoRFs ranging from 15 to 23 amino acids. These three MoRFs are quite different from each other except for a reoccurring small region of 3 hydrophobic amino acids flanked by charged and polar residues on either side.

PONDR® VLXT predicts a particularly ordered N-terminus, which suggests that its structure is mainly regulated by hydrophobic interactions and may explain the increase in α-helices observed by CD (*29*) (Fig. 9C). The stretch of amino acids 29-98, which is associated with high α-helical propensity and a high hydrophobic moment, has previously been predicted to form amphipathic α-helices (*42*) (Fig. 12C). The N-terminal region of the protein up to amino acid position 180 correlates to the observed ~40% of α-helices in

the desiccated state. This suggests that the C-terminus functions as either a functional domain that utilizes intrinsic disorder or functions as a targeting domain that undergoes a conformational transition when in contact with a binding partner rather than due to environmental factors.

The C-terminal domain is separated into two sub-domains by PONDR® VL3 (Fig. 9C). The first sub-domain spans from amino acid position 180 – 290 and contains a 10 residue-long MoRF and a cluster of three 9 residue-long MoRFs, which are simultaneously predicted by both PONDR® VLXT and ANCHOR. This region is enriched in serine residues which are likely to be phosphorylated and leucine, valine, phenylalanine, and lysine residues (Fig. 9C). The second C-terminal domain is composed mainly of three large MoRFs, enriched in isoleucine, methionine, leucine, and arginine residues.

Given the unique feature of the C-terminal region ranging from approximately amino acid position 180-364, it may be predicted that this the region is subjected to desiccation-induced folding. Expectedly, it appears that the length of this region directly correlates with the degree of secondary structure detected by circular dichroism in the dry as state (*29*). This is of particular importance considering the content of proline and glycine in the region that would break apart any α-helices that might be forming in this region. Furthermore, this region has an amino acid composition that is not conducive to form amphipathic α-helices. Heliquest predicts that possible α-helices in this region would have a lower hydrophobic moment than at any other position in the protein, except for a region spanning from about amino acid position 275 to about 300 (Fig. 12C). While it is unlikely that the CD detected secondary structure is exclusively located within this C-terminal region, it is reasonable to suggest that the degree of secondary structure in this region is

higher than in the remainder of the polypeptide. This information can be highly useful for experiments regarding the function of *Afr*LEA2, such as ectopic expression of the C-terminal region and comparing effects of this region and full length *Afr*LEA2 on physiological properties of model cells under water stress or using site-directed mutagenesis to remove the prolines separating the MoRF regions and observing the shift in secondary structure during desiccation of the protein via CD spectroscopy.

AfrLEA3m

*Afr*LEA3m has been shown to localize in the mitochondria (*29, 86*), which explains the peculiar cysteine residues near the N-terminus, which is most likely being cleaved off after the protein is incorporated into the mitochondrial matrix (*86*). Therefore, the first 31 amino acids, which was predicted to serve as the signal sequence, are excluded from the bioinformatics analyses conducted in this study. *Afr*LEA3m is the least hydrophobic LEA protein known to occur in *A. franciscana* that belongs to group 3, falling very close to the group 1 protein *Af*LEA1.1 (Fig. 5A). Compared to the other group 3 members, this LEA protein contains the largest fraction of charged residues, making up approximately 38.8% of the sequence, but the distribution of charges is the most even observed for LEA proteins from *A. franciscana*, with a κ value of 0.072 (Tab. 4). CIDER predicts *Afr*LEA3m to be a strong polyampholyte, which, having such a low κ value, should be self-repulsive unless the charges are aligned via the adoption of secondary structure (Fig. 5C).

The protein is predicted by CDF analysis to be mainly intrinsically disordered, making it the only group 3 LEA protein to fall below the boundary of the CDF (Fig. 5B). This further suggests a somewhat structured protein with high self-repulsion in the

hydrated state. Its proximity to the group 1 protein *Af*LEA1.1 on the CH-plot is of particular interest given its sequence length and its classification as a member of group 3 LEA proteins. DisEMBL disorder prediction for *Afr*LEA3m suggest that this protein is that about 89.5% disordered in the hydrated state, although this percentage drops to 69.1% if Remark-465 is not being considered (see Supplementary Material, Fig. 15). The observed degree of disorder for *Afr*LEA3m by CD spectroscopy (*29*) is approximately 74% in the hydrated state and reduces to approximately 60% during desiccation. The predictions by DisEMBL, after removing the consideration of missing electron density in structures of globular protein domains, falls closely between the hydrated and dry states measured experimentally. This is particularly interesting because it implies that *Afr*LEA3m may fulfill some functions in the hydrated state, that only a few key regions are regulated by desiccation, or perhaps that its secondary structure is not as important to its function as previously hypothesized. This is not to say that tertiary structure, such as the predicted coiled-coil region, may not be regulated by desiccation and be crucial for function, but the methods currently employed do not adequately address these possibilities.

The Smart Server predicts two 46-48 residue-long repeats at positions 116-163 and positions 191-236 separated by a coiled-coil region spanning amino acids 157-185. These repeats and the coiled-coil region each coincide with α-helices predicted by MEDOR (Fig. 8D) and GlobPlot (see Supplementary Material, Fig. 16). Furthermore, each of the α-helices is predicted to be amphipathic in nature by Heliquest, implying helical interactions among the three regions (Fig. 12D). These higher-order folding patterns may be relevant to interactions with lipids and/or membranes. In this way, *Afr*LEA3m resembles *Afr*LEAI, although the former protein is potentially less ordered in the desiccated state. This may

offer support to the hypothesis that tertiary structure is relevant to *Afr*LEA3m function in the desiccated state. Combined with the potential relevance of *Afr*LEAI tertiary structure to its function, it may be that group 3 LEA proteins adopt more tertiary structure during desiccation compared to members from groups 1 and 6.

The mature protein most likely spans from amino acid position 31-307 based on the indications from IUPred, GlobPlot, and a review of signal peptides from *D. melanogaster*. ANCHOR predicts a similar MoRF region pattern as observed for the cytoplasmic *Af*LEA1.1 and *Afr*LEAI proteins in that two distinct and different MoRF regions are found around the protein termini, and the appearance of internal MoRF regions correlates with repeating amino acid patterns (Fig. 9D). The PONDR® VLXT plot shows several peaks and troughs with extreme slopes spanning the entirety of the protein, suggesting that most of the folding should be regulated by some binding partner (Fig. 9D). The PONDR® VL3 predictor also shows three distinct domains, which correlates with the arrangement of MoRF sites predicted by ANCHOR. It appears that *Afr*LEA3m, like *Afr*LEAI, may be associated with membranes or other lipids due to the amphipathic coiled-coil region predicted to occur roughly in the middle of the protein. Perhaps a unique role for *Afr*LEA3m might be to undergo a conformational shift exclusively in the presence of a membrane to orientate its hydrophobic face. The distribution of charges may also allow *Afr*LEA3m to interact in some way with others of itself, forming a loosely associating matrix with nanogel like properties, even the proteins will only interact among each other via non-covalent bonding.

Group 6

AfrLEA6

*Afr*LEA6 is unique due to its position on the CH-plot being well within the region where most globular proteins fall (Fig. 5A). While it is flanked by two well-characterized IDPs, α-synuclein and γ-synuclein, its location is right on the edge of where such exceptions are observed. The mean net charge to hydropathy ratio of 0.094 is comparable to the one observed for *Afr*LEAI. *Afr*LEA6 is classified as a group 6 LEA protein, which is the most recently defined group that shows, compared to other LEA groups, unusual characteristics and hydropathy is not considered a major characteristic of this LEA group. CIDER predicts *Afr*LEA6 to be a weak polyampholyte, potentially forming a tadpole or globular structure (Fig. 5C). This globular structure would seem to agree with the CH-plot. Despite being predicted to be globular by the CH-plot, the overall DisEMBL prediction of disorder for *Afr*LEA6 is 80.9% (see Supplementary Material, Fig. 17) and CDF analysis places *Afr*LEA6 well below the boundary (Fig. 5B). This may be indicative that it is another exceptional IDP, but the predictions from each program, and even within the same program, may offer additional insight into the structure and behavior of this protein.

Algorithms using missing electron density from x-ray crystallography data tend to suggest that *Afr*LEA6 is a globular protein with less than 40% disorder, including DisEMBL Remark-465 prediction. Algorithms that predict disorder using secondary structure propensity such as Pred2ary from MEDOR, Loops/Coils from DisEMBL, and IUPred predict that the degree of disorder for *Afr*LEA6 ranges from about 50%-75%. The Hot-Loops predictor, which is based on the B-factor, predicts that only 32.3% of the polypeptide is disordered, and therefore agrees with the results of GlobPlot analysis. While

the programs appear to disagree on whether or not the disorder propensity breaches an appropriate threshold, they are quite consistent in showing the locations of possible disordered regions and domains. Each predictor suggests that there are regions with a high likelihood of order juxtaposed to regions with a high propensity for disorder. Programs that smooth the data appear to favor an ordered interpretation, whereas programs with smaller windows or less smoothing tend to favor disorder, implying that there are small, defined regions of order and disorder scattered throughout the protein.

The SMART server predicts that *Afr*LEA6 has two Pfam-SMP domains, one at position 9-55 and the other at position 90-137 (Fig. 8E). Pfam-SMP, or seed maturation proteins, are associated with desiccation tolerance in seeds, but have not been characterized in animals, with *Artemia* being the only animal known to express a protein containing Pfam-SMP domains. In contrast to the second SMP domain, the first domain is recognized by NCBI BLAST, although the second region has a very high sequence similarity to the first domain. Both domains appear to be parts of a larger repeat, spanning from amino acid positions 2-70 and 81-155. At position 140 to 184, appears to be a large region with a very high concentration of proline residues. Half of the proline residues in the entire protein are concentrated into this relatively short region, spanning approximately 11% of the sequence. Given the nature of proline as an α-helix and β-sheet disruptor, it is unlikely that defined secondary structures fall within this region. The prolines are also spaced in such a way as to make a poly-proline helix unlikely, which suggests that this region remains disordered at any hydration level. In addition to the high content of prolines, this region contains several hydrophobic residues, making it exceptionally hydrophobic for a disordered region. Aromatic residues such as tyrosine and phenylalanine are disproportionately included in

this region as well. This may also explain the problems that missing electron density programs have for predicting secondary structure features in this location. Following the proline-rich region is again a region with similarity to the Pfam-SMP domain, although it is somewhat more degenerated from the two aforementioned domains. Given the length of each repeat, it appears that the protein is composed of 3 repeats, with one region of the last repeat being less conserved and enriched in proline residues than the other two regions. This may indicate that the third region has evolved from an SMP domain into a distinct domain with unknown functions. The C-terminal region exhibits a unique staggering of positive and negative charges separated by proline and glycine residues, potentially allowing folding in the desiccated state (Fig. 8E).

ANCHOR predicts two conserved MoRF regions within the N-terminal SMP domains, which fall in a region of relatively low disorder-propensity (Fig. 9E). The second half of the second SMP domain has a large MoRF region ranging from amino acid 105 to 134, which is not shared with the first SMP domain. PONDR® VL-XT predicts weak potential binding capacity shared between the last 20 amino acids of the second SMP domain, the proline-rich region, and the first half of the C-terminal region (Fig. 9E). An N-terminal disordered region correlates with the disorder prediction of IUPred (Fig. 9E), and the C-terminus has a disordered region with limited binding capacity that coincides with the MoRF region predicted by ANCHOR. PONDR® VL3 predicts three distinct domains, separated as an N-terminal domain at the point where the SMP domains meet, a large domain spanning the combined MoRF regions described above including the proline-rich region and the neighboring regions, and a C-terminal domain downstream of the MoRF region. Due to the occurrence of charges in the internal region that may be complementary

to charges at N-terminal and C-terminal regions the desiccated protein likely forms a structure resembling a bio-glass.

CONCLUSIONS

We have utilized a broad suite of open-source bioinformatics tools to gain insights into the dynamic structures of LEA proteins from the brine shrimp *A. franciscana*. Results of our analysis were used to refine current hypotheses regarding the function of LEA proteins in animals. Our analysis indicates that LEA proteins from different groups are more similar than we originally hypothesized, while functional differences among members of group 3 are possibly larger than commonly anticipated. Each of the LEA proteins analyzed, except for *Afr*LEAI, had three distinct domains; one at each terminus with potential binding sites connected by an intermediary domain. We predict that *Afr*LEA1.1 is a highly disordered protein with coil-like structure that appears to have two distinct MoRF domains on either side of a repeating internal spacer domain and is predicted to be a Janus sequence that exists as a mostly random coil in the hydrated state. The internal domain may undergo a conformational transition during water loss, pulling the terminal MoRF sites, and potentially attached binding partners, closer together during desiccation.

The group 3 LEA proteins all showed domains with amphipathic α-helix propensities, but otherwise showed substantial differences among each other. *Afr*LEAI, as previously noted, is the only LEA protein with just two distinct domains, an N-terminal domain with more even distribution of hydrophobic and charged residues, and a C-terminal domain with six repeats of a coiled region that may form amphipathic, potentially self-interactive, α-helices, which could form a perilipin-like bundle. *Afr*LEAI also appears to

be the most readily protein-binding LEA protein found in *A. franciscana*, potentially interacting with multiple partners, and is one of the two LEA proteins that appears to be molten globular in the hydrated state. *Afr*LEAI is predicted to function as a Janus sequence which should undergo conformational changes during desiccation.

*Afr*LEA2 is more hydrophobic than the other group 3 LEA proteins and has no detectable internal repeats in its sequence. It has a uniquely stable intermediary domain that likely includes the observed α-helical MoRFs found in CD spectra (*29*). This increase in orderly structure supports our prediction that *Afr*LEA2 functions as a Janus sequence, and bolsters our confidence in similar results for the other the proteins not yet characterized by CD spectroscopy. The relatively small N- and C-terminal domains likely interact with binding partners and *Afr*LEA2 appears to be natively either molten globular or to contain globular regions in the hydrated state. *Afr*LEA3m uniquely categorizes as a strong polyampholyte of low mean net charge with a low κ value, which suggests that it should maintain a relatively high degree of disorder despite desiccation. The termini appear to have MoRFs, which are separated by an intermediate spacer region (Table 4). The distribution of charges may be overcome by folding into an α-helical conformation in this region, but not at the termini. Staggering of two or more of this protein might also facilitate favorable protein interactions, rather than gaining substantial structure on its own. Most certainly, *Afr*LEA3m will need a compatible binding partner before it undergoes a conformational transition, instead of being regulated more readily by desiccation as the other group 3 LEA proteins appear to be.

*Afr*LEA6 is the most distinct LEA protein compared to the other LEAs in *A. franciscana*. It is by far the most hydrophobic and the protein contains two SMP domains,

which appear to function only when they interact with another sequence. *Afr*LEA6 has a tremendously proline-enriched intermediate domain that may either function as a highly flexible spacer or as a unique binding site. The N-terminal domain is composed of a proline- and isoleucine-rich region flanked by two SMP domains, which begin with low PONDR score and transition suddenly to a high score. This slope does not strictly indicate a binding site but may points to the potential of self-interaction between the SMP domains. The juxtaposition of SMP domains upstream of proline regions indicates that this pattern might be important for its function, which has yet to be elucidated. The C-terminus has a distinct separation of charges that makes it very susceptible to binding other proteins in the desiccated state, contributing to the model of a weak polyampholyte tadpole. In such a model, the N-terminus might act as a globular "head" whereas the C-terminus would act as a sticky "tail" which coil to form a glassy or gel-like matrix.

Overall, our investigation indicates a variety of differences in form and potential function(s) of LEA proteins expressed in *A. franciscana* during anhydrobiosis but indicates that as a general strategy the animal utilizes glassy matrix forming LEAs concurrently with proteins that more likely interact with more specific binding partners. Nevertheless, the function(s) of both types, the matrix-forming and partner-binding LEA proteins, are likely regulated by changing water availability during desiccation.

Tables

Table 3: Classifications of LEA proteins found in the brine shrimp *Artemia franciscana**.

Protein	Tunnacliffe & Wise(*68*)	Dure et al. (*67*)	Hundertmark & Hincha (*187*)	LEApb (*117*)	PFAM
***Af*LEA1.1**	Group 1	D19, D132	LEA_5	Class 5	PF00477
***Afr*LEAI**	Group 3	D7	LEA_4	Class 6	PF02987
***Afr*LEA2**	Group 3	D7	LEA_4	Class 6	PF02987
***Afr*LEA3m**	Group 3	D7	LEA_4	Class 6	PF02987
***Afr*LEA6**	Group 6	D34	SMP	Class 11	PF04927

*In this dissertation, classification scheme proposed by Tunnacliffe and Wise is used.

Table 4: CIDER and PONDR Parameters* of LEA Protein Sequences from *A. franciscana*.

Protein	κ	FCR	κ/FCR	|MNC|	MNH	|MNC|/MNH
***Af*LEA1.1**	0.194264	0.283333	1.458491	0.0278	0.3490	0.0797
***Afr*LEAI**	0.145081	0.29972	2.065885	0.0364	0.3858	0.0943
***Afr*LEA2**	0.079765	0.304945	3.098031	0.0137	0.4017	0.0341
***Afr*LEA3m**	0.072713	0.387681	0.187558	0.0109	0.3388	0.0322
***Afr*LEA6**	0.142528	0.206226	1.446918	0.0428	0.4536	0.0945

*The κ, FCR, and the fraction of both values. As κ increases, the likelihood of self-interaction increases, whereas if κ decreases, then the protein becomes self-repelling. Men net charge (MNC) and mean hydrophathy (MNH) were calculated based on PONDR. For more information please refer to text.

Figure Captions

Figure 5. Global analysis of intrinsic disorder predispositions of LEA proteins from *A. franciscana*. A. CH-plot including LEA proteins from *A. franciscana* (diamonds) that are plotted together with a set of known IDPs (red circles), and globular proteins (blue squares). B. CDF analysis of LEA proteins from *A. franciscana*. The order-disorder boundary is shown by bold black line. C. CIDER state predictions of each LEA proteins based on their FCRs, separated into positively and negatively charged residues. *Afr*LEA6 and *Afr*LEA3m are the only two LEA proteins that fall into their own distinct regions of the plot as weak and strong polyampholytes, respectively. *Afr*LEA1.1, *Afr*LEAI, and *Afr*LEA2 are predicted to be Janus sequences with independent conformational transitions.

Figure 6. NCPR distribution in *Af*LEA1.1 with a window size of five. The protein displays a distinct separation of charges based on the region of the protein. The N-terminus has a strongly positively charges region, whereas the C-terminus has two adjacent positive and negative regions.

Figure 7: DisEMBL disorder predictions for *Af*LEA1.1 by loops/coil (blue), Remark465 (Green), and HotLoops (red) predictors, with dotted line thresholds for disorder with the correlating colors.

Figure 8. MeDor-based analysis of LEA proteins from *A. franciscana*. For *Af*LEA1.1 (A), Pred2ary predicts β-sheets separating the termini from the central protein domain, which are shown within the boxes. The HCA shows series of small hydrophobic clusters

embedded inside the regions enriched in charged and polar residues. B. In *Afr*LEAI, the two N-terminal internal repeats (red boxes), contain several hydrophobic clusters enriched in tyrosine, followed by a proline. The six C-terminal repeats (blue boxes) are composed of a hydrophobic cluster enriched in phenylalanine and is interrupted by a proline as well as a stretch of alternating charges enriched in a hydrophobic face of alanine residues. SMART server predicts coiled coil regions throughout the protein (black bar). C. The *Afr*LEA2 protein has three distinct domains (black boxes). The N-terminal domain has a likely amphipathic α-helix propensity due to the arrangement of polar and nonpolar residues and enrichment in alanine. The second domain is enriched in leucine and valine residues, with little likely structure due to enrichment of regularly spaced proline residues. The third domain begins with a hydrophobic cluster enriched in glycine. The domain is enriched in isoleucine and methionine. D. The *Afr*LEA3m protein has two internal repeats from positions 116 – 236 (black boxes) which are separated by a coiled-coil region predicted by SMART server (black bar). E. The *Afr*LEA6 protein has two internal SMP domains towards the N-terminus (black boxes) and a proline-rich intermediary domain (red box) connecting a C-terminal domain (blue box).

Figure 9. Analysis of LEA proteins from A. *franciscana* (*Afr*LEA1.1 (A), *Afr*LEAI (B), *Afr*LEA2 (C), and *Afr*LEA3m (D), *Afr*LEA6 I) by a set of per-residue disorder predictors, such as PONDR® VL3 (red), PONDR® VLXT (black), PONDR® VSL2 (green), PONDR® FIT (pink), IUPred_short (yellow), and IUPred-long (blue). Bold dashed cyan lines show the mean disorder propensity calculated by averaging disorder profiles of individual predictors. Light pink shadow around the PONDR® FIT shows error distribution. In these

analyses, the predicted intrinsic disorder scores above 0.5 are considered to correspond to the disordered residues/regions, whereas regions with the disorder scores between 0.2 and 0.5 are considered flexible. The plots also include the results of functional analysis of these proteins by ANCHOR to evaluate the MoRF probability (dark pink).

Figure 10: GlobPlot disorder prediction for AfLEA1.1 using the Remark465 propensity set. Positive slopes denote propensity towards disorder and a blue bar at the bottom of the figure denotes structural disorder prediction.

Figure 11: DISPHOS 1.3 phosphorylation prediction of *Af*LEA1.1 based on phosphorylation patterns in D. melanogaster. The phosphorylation propensity of serine residues (red triangles) and tyrosine residues (green squares) are shown for all residues above a 50% threshold. AfLEA1.1 has 100% serine phosphorylation, 14.3% tyrosine phosphorylation, and 0% threonine phosphorylation.

Figure 12. Heliquest output of local hydropathy (red) and hydrophobic moment (blue) for *Afr*LEA1.1 (A), *Afr*LEAI (B), *Afr*LEA2 (C), and *Afr*LEA3m (D), *Afr*LEA6 I.

Figure 13: DisEMBL disorder predictions for *Afr*LEA2 by loops/coil (blue), Remark465 (Green), and HotLoops (red) predictors, with dotted line thresholds for disorder with the correlating colors.

Figure 14: DISPHOS 1.3 phosphorylation prediction of *Afr*LEA2 based on phosphorylation patterns in *D. melanogaster*. The phosphorylation propensity of serine

residues (red triangles) and tyrosine residues (green squares) are shown for all residues above a 50% threshold. AfrLEA2 has 43.9% serine phosphorylation, 0% tyrosine phosphorylation, and 0% threonine phosphorylation.

Figure 15: DisEMBL disorder predictions for *Afr*LEA3m by loops/coil (blue), Remark465 (Green), and HotLoops (red) predictors, with dotted line thresholds for disorder with the correlating colors.

Figure 16: GlobPlot disorder prediction for *Afr*LEA3m using the Remark465 propensity set. Positive slopes denote propensity towards disorder and a blue bar at the bottom of the figure denotes structural disorder prediction. The yellow bar at the top depicts a low-complexity region and the striped bar indicates a coiled-coil region.

Figure 17: DisEMBL disorder predictions for *Afr*LEA6 by loops/coil (blue), Remark465 (Green), and HotLoops (red) predictors, with dotted line thresholds for disorder with the correlating colors.

Figure 5

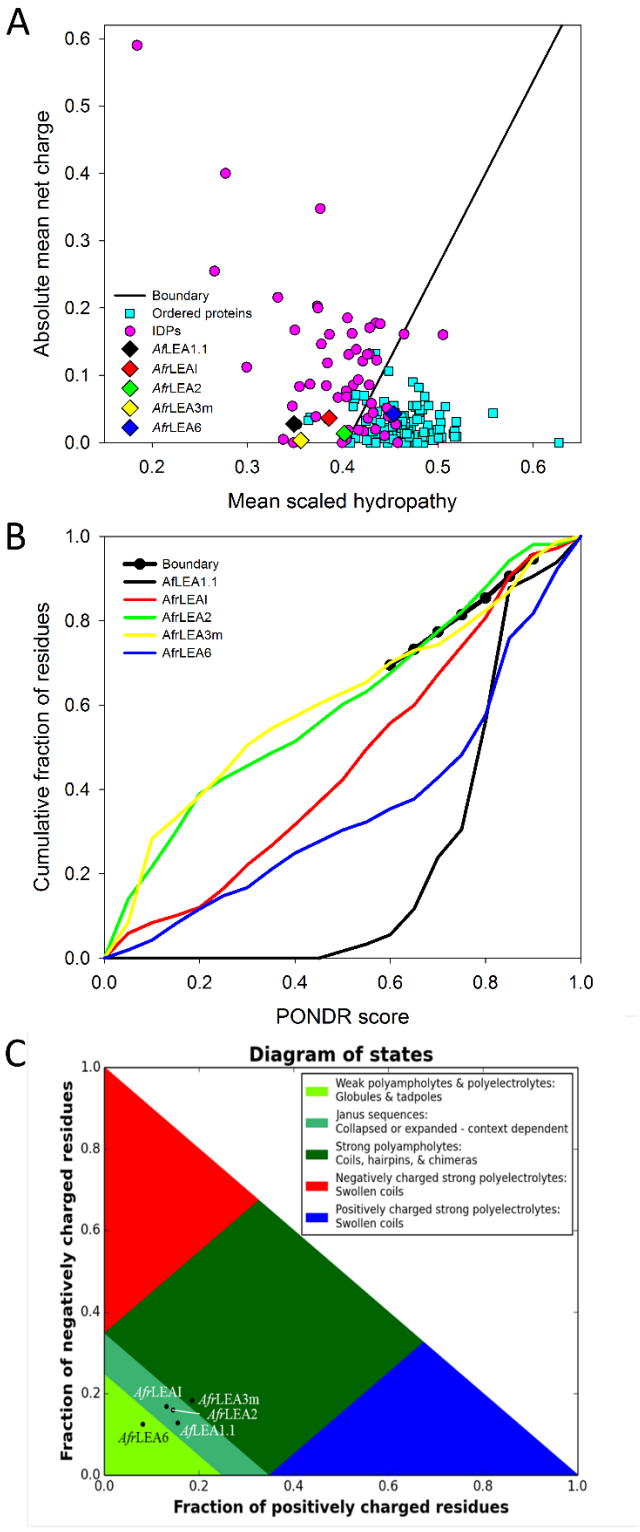

A
Absolute mean net charge
Mean scaled hydropathy
Boundary
Ordered proteins
IDPs
AfLEA1.1
AfrLEAI
AfrLEA2
AfrLEA3m
AfrLEA6
B
Cumulative fraction of residues
PONDR score
Boundary
AfLEA1.1
AfrLEAI
AfrLEA2
AfrLEA3m
AfrLEA6
C
Diagram of states
Fraction of negatively charged residues
Fraction of positively charged residues
Weak polyampholytes & polyelectrolytes: Globules & tadpoles
Janus sequences: Collapsed or expanded - context dependent
Strong polyampholytes: Coils, hairpins, & chimeras
Negatively charged strong polyelectrolytes: Swollen coils
Positively charged strong polyelectrolytes: Swollen coils
AfrLEAI
AfrLEA3m
AfrLEA2
AfLEA1.1
AfrLEA6

Figure 6

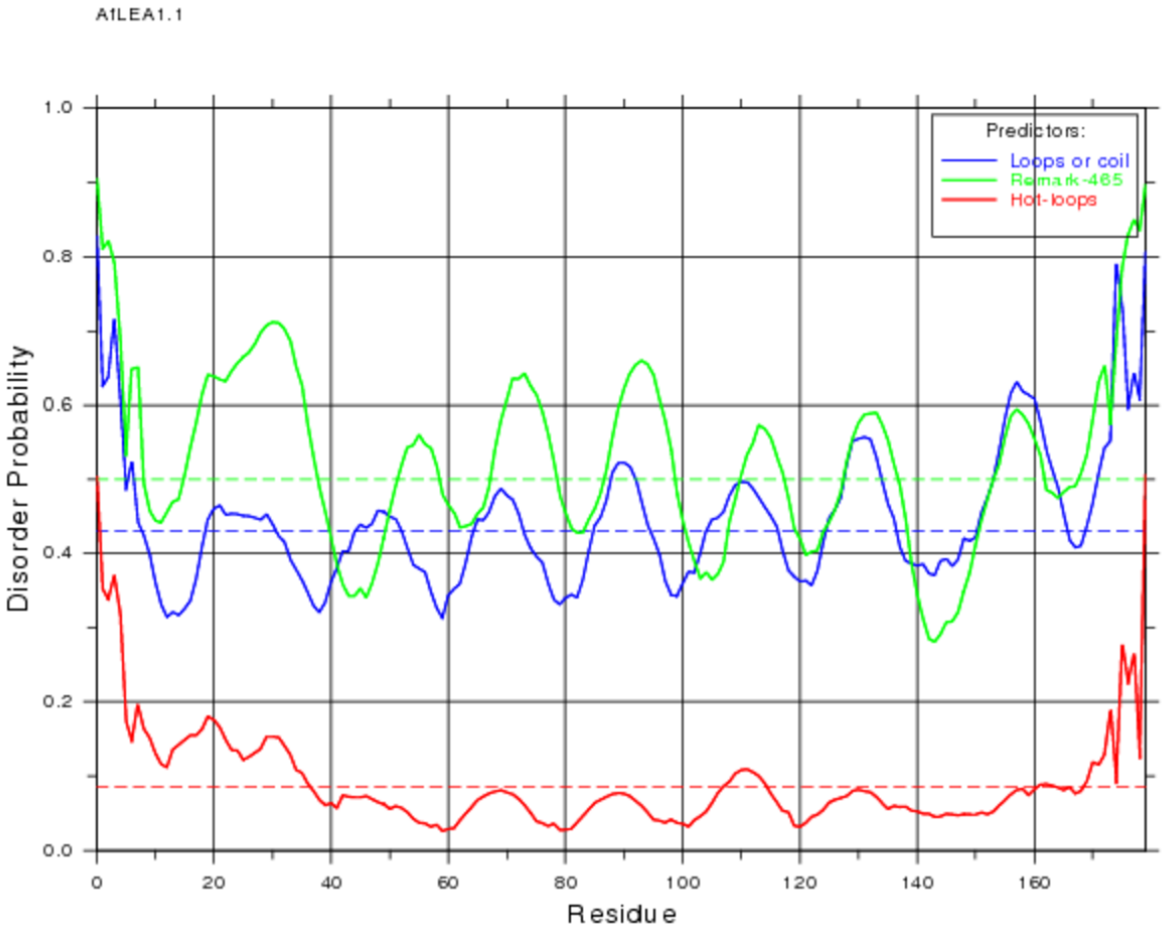
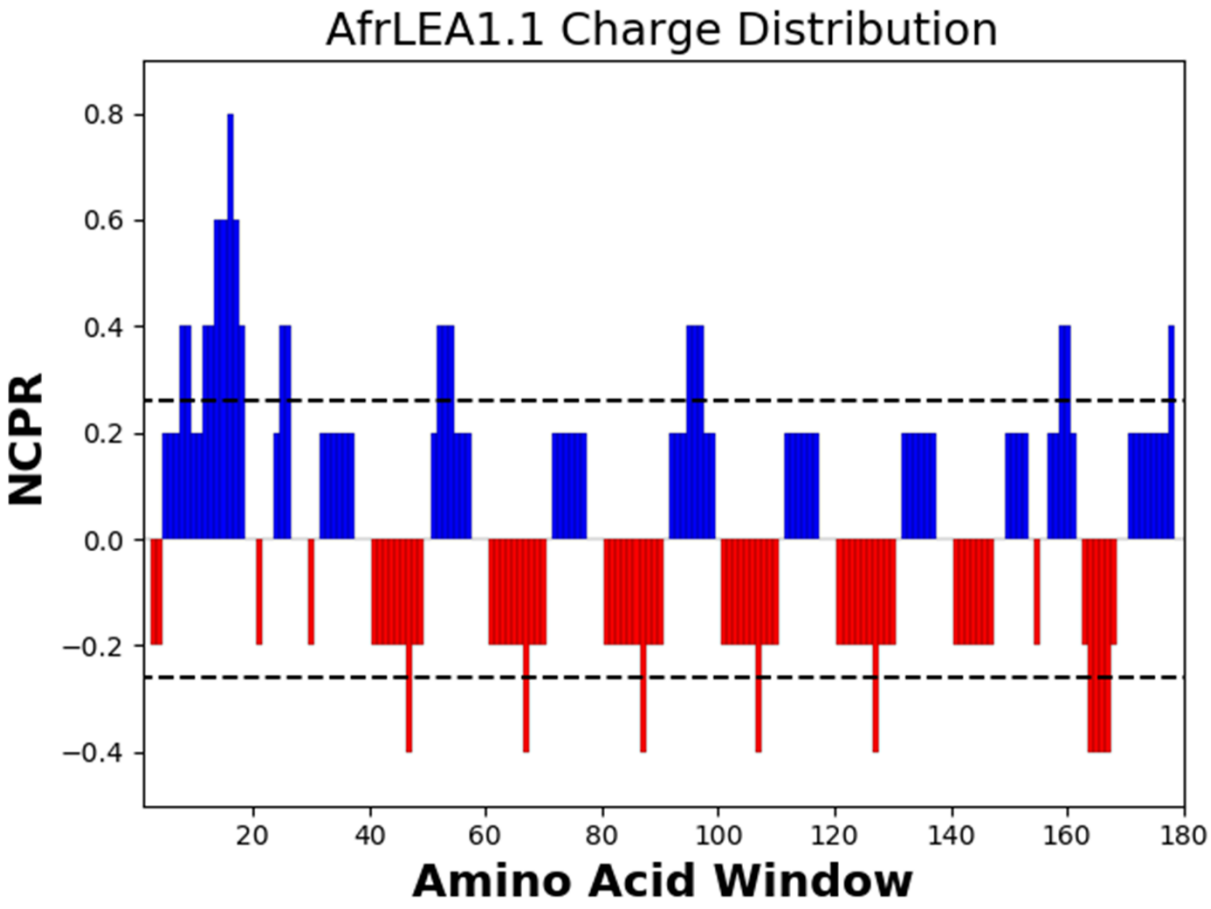

A1LEA1.1
Predictors:
Loops or coil
Remark-465
Hot-loops
Disorder Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
120
140
160
Residue

Figure 7


AfrLEA1.1 Charge Distribution
NCPR
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
20
40
60
80
100
120
140
160
180
Amino Acid Window

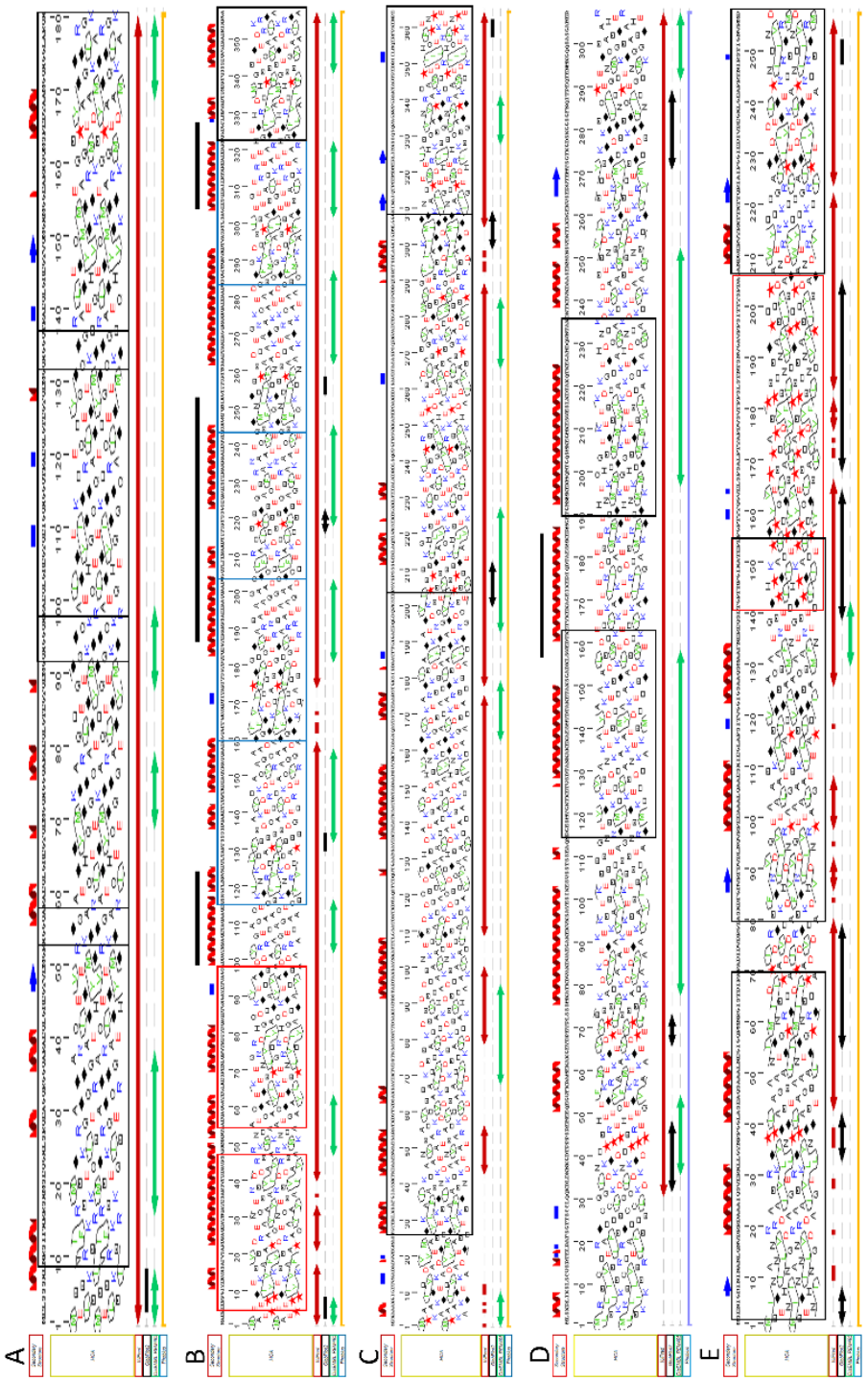


Figure 8

Figure 9

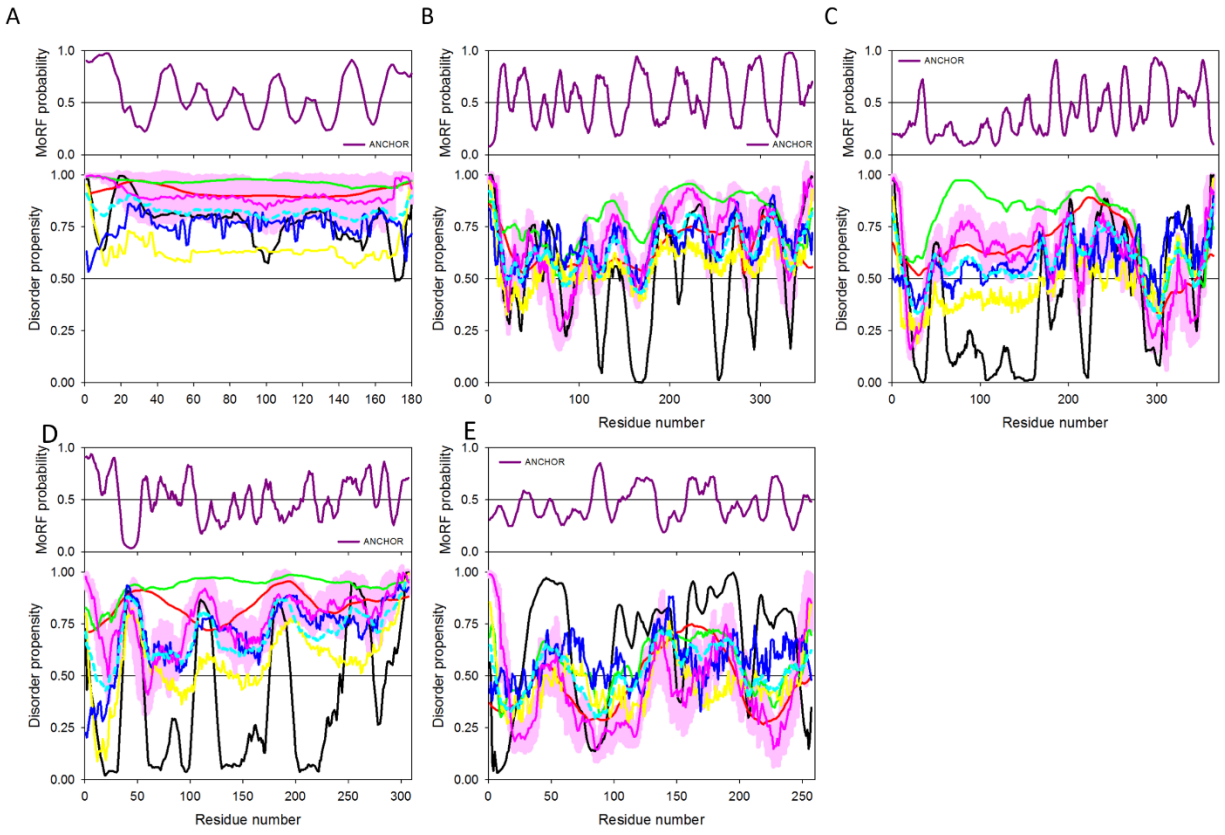
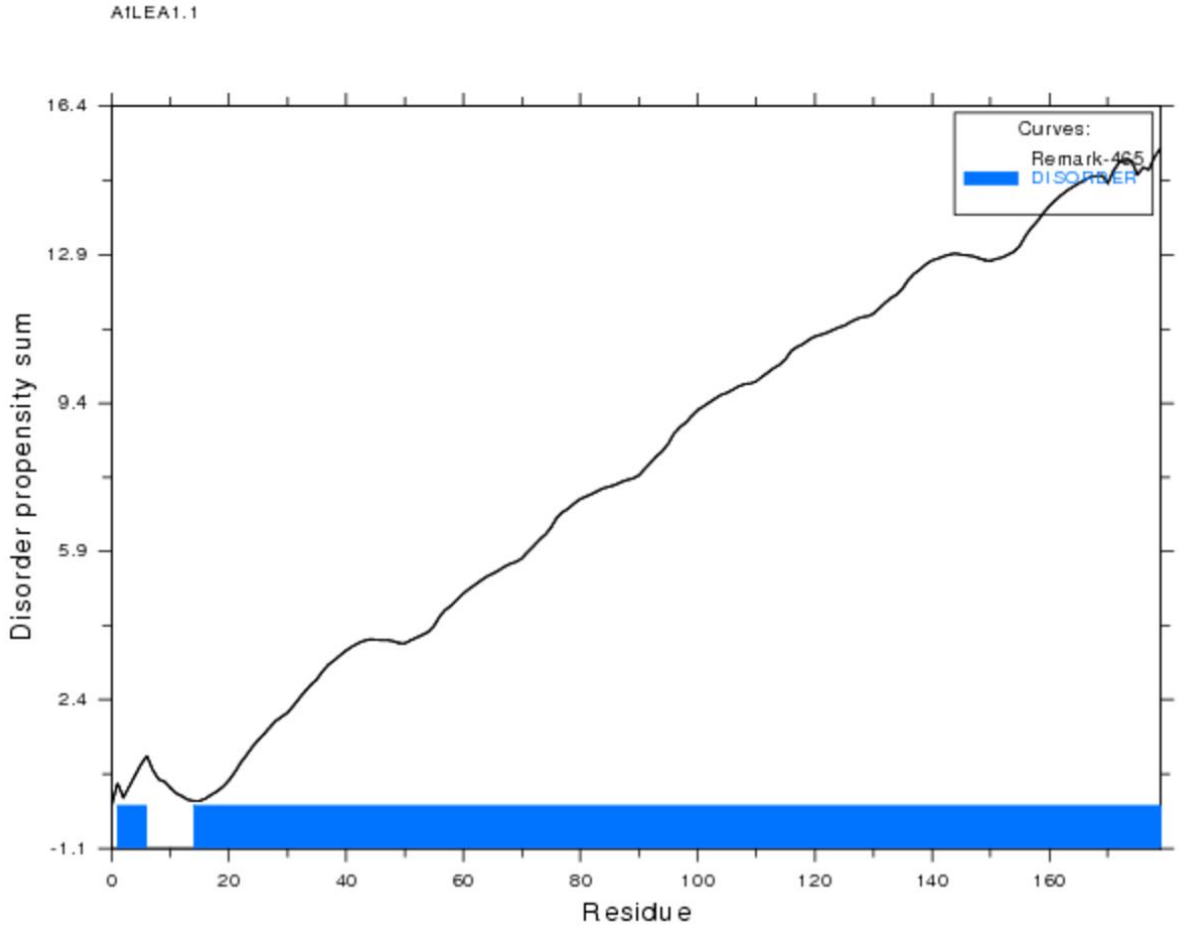
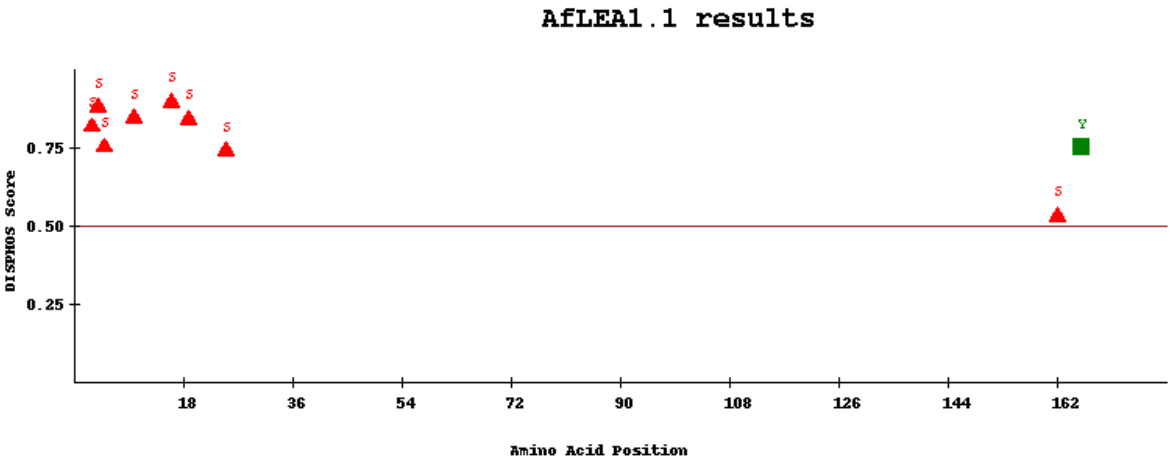

A
B
C
D
E
MoRF probability
Disorder propensity
ANCHOR
Residue number

Figure 10



Janis et al. 2017. J. Biomol. Struct. Dyn.

Figure 11



Janis et al. 2017. J. Biomol. Struct. Dyn.

Figure 12

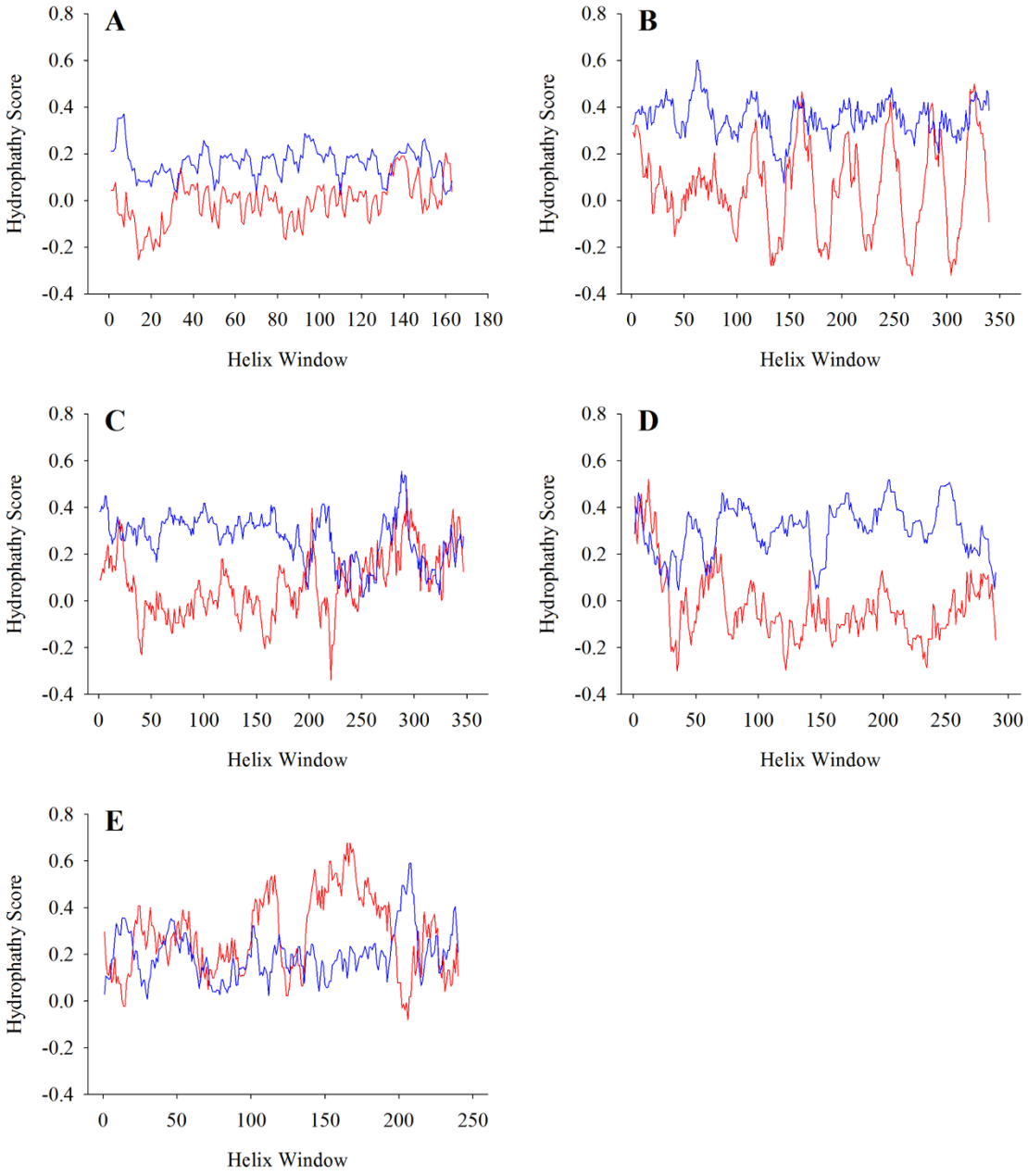

A
B
C
D
E
Hydrophathy Score
Helix Window
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
0 20 40 60 80 100 120 140 160 180
0 50 100 150 200 250 300 350
0 50 100 150 200 250 300
0 50 100 150 200 250

Figure 13

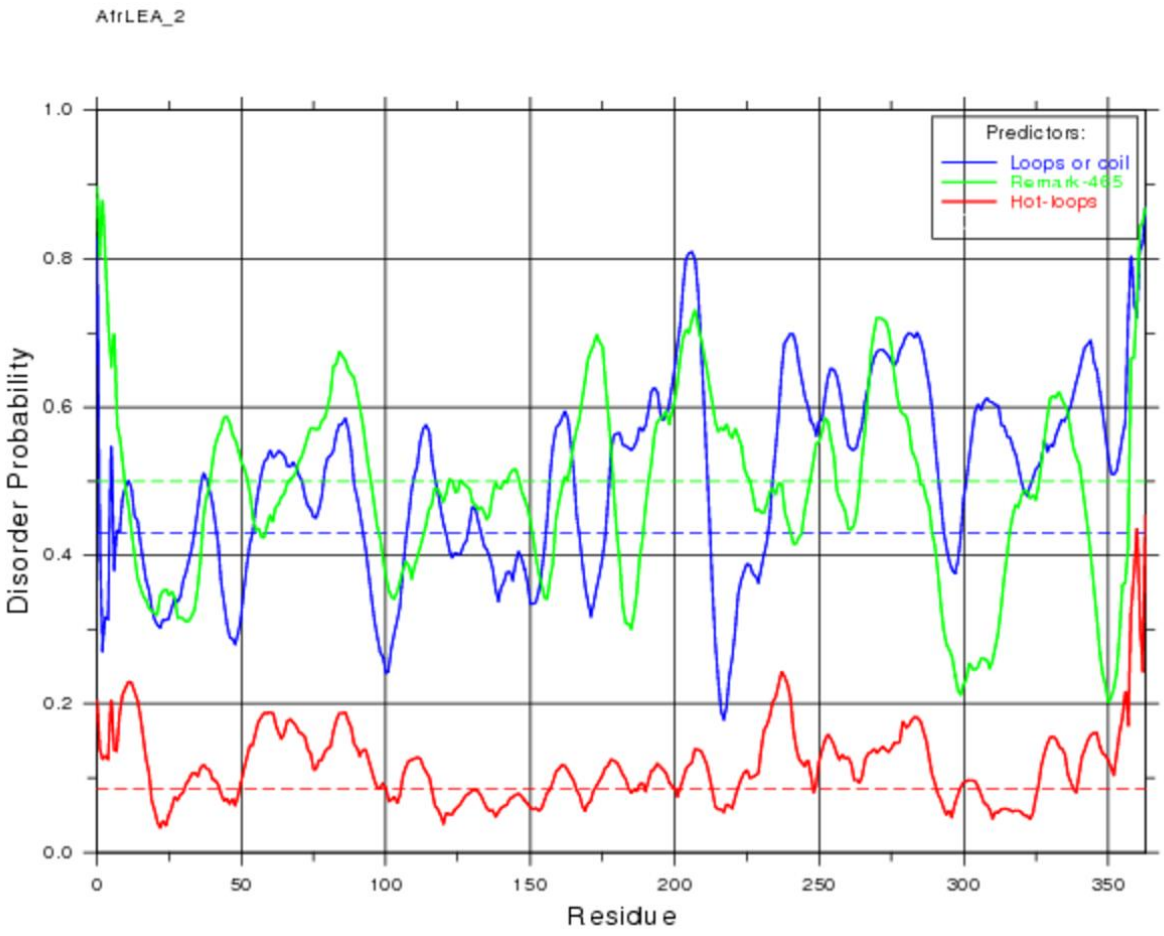
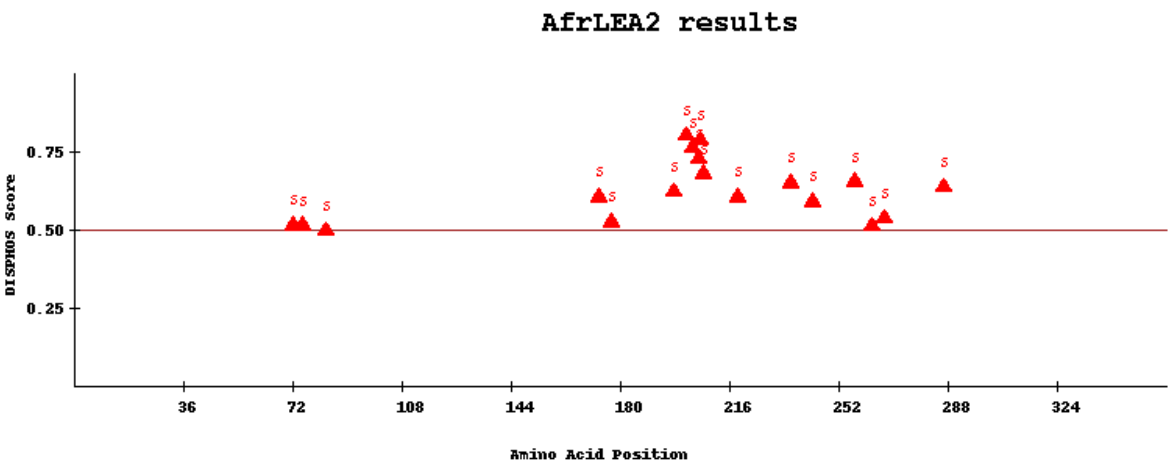

AfrLEA_2
Predictors:
Loops or coil
Remark-465
Hot-loops
Disorder Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
300
350
Residue

Figure 14



Janis et al. 2017. J. Biomol. Struct. Dyn.

Figure 15

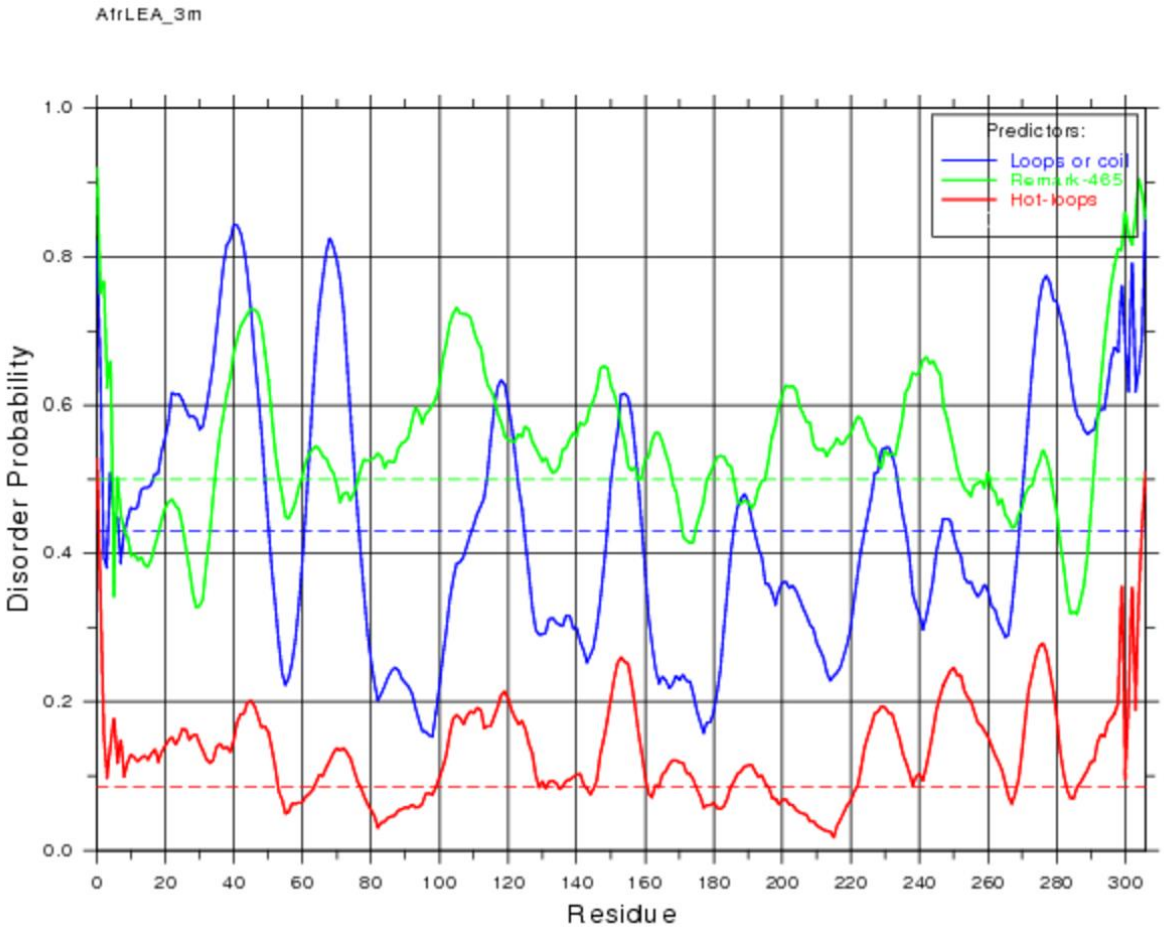

AfrLEA_3m
Predictors:
Loops or coil
Remark-465
Hot-loops
Disorder Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
Residue

Figure 16

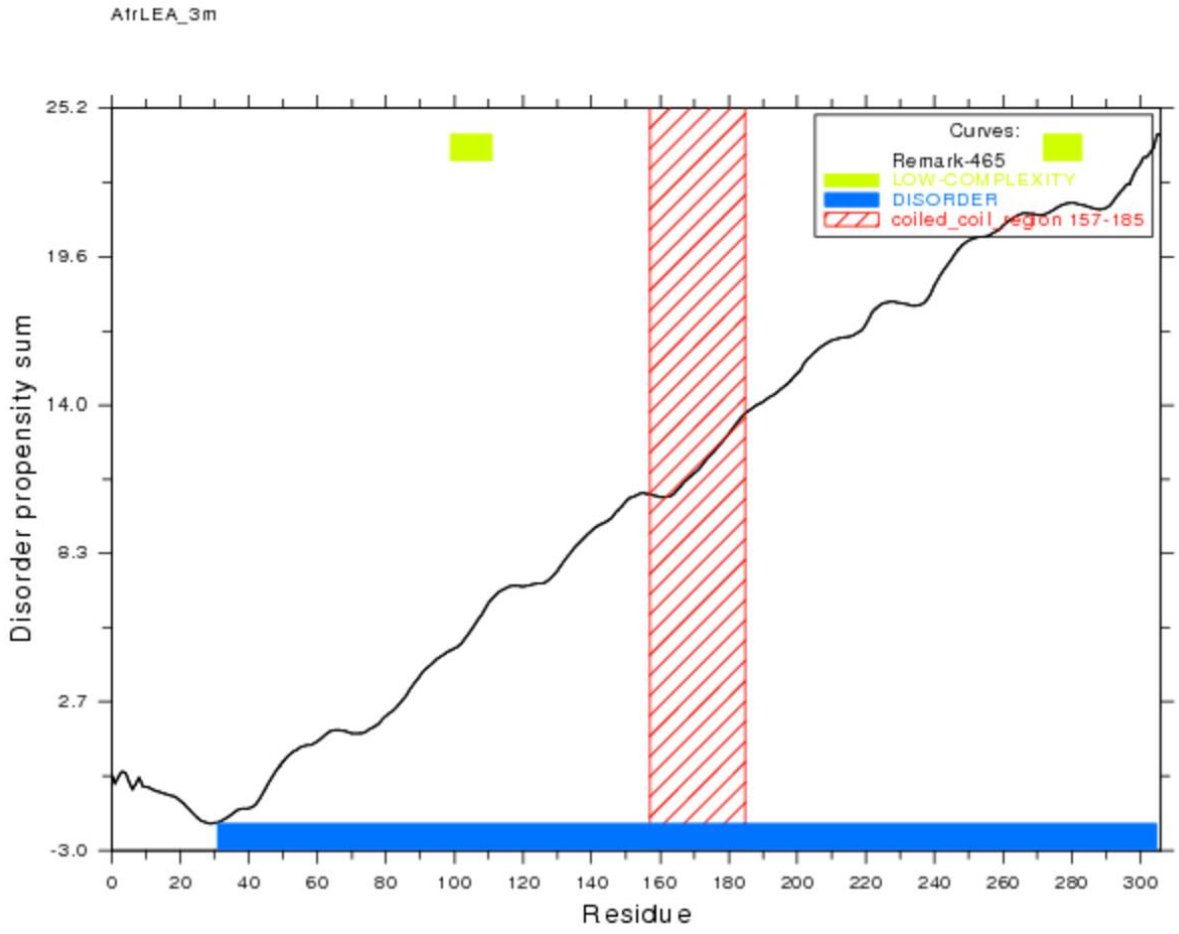
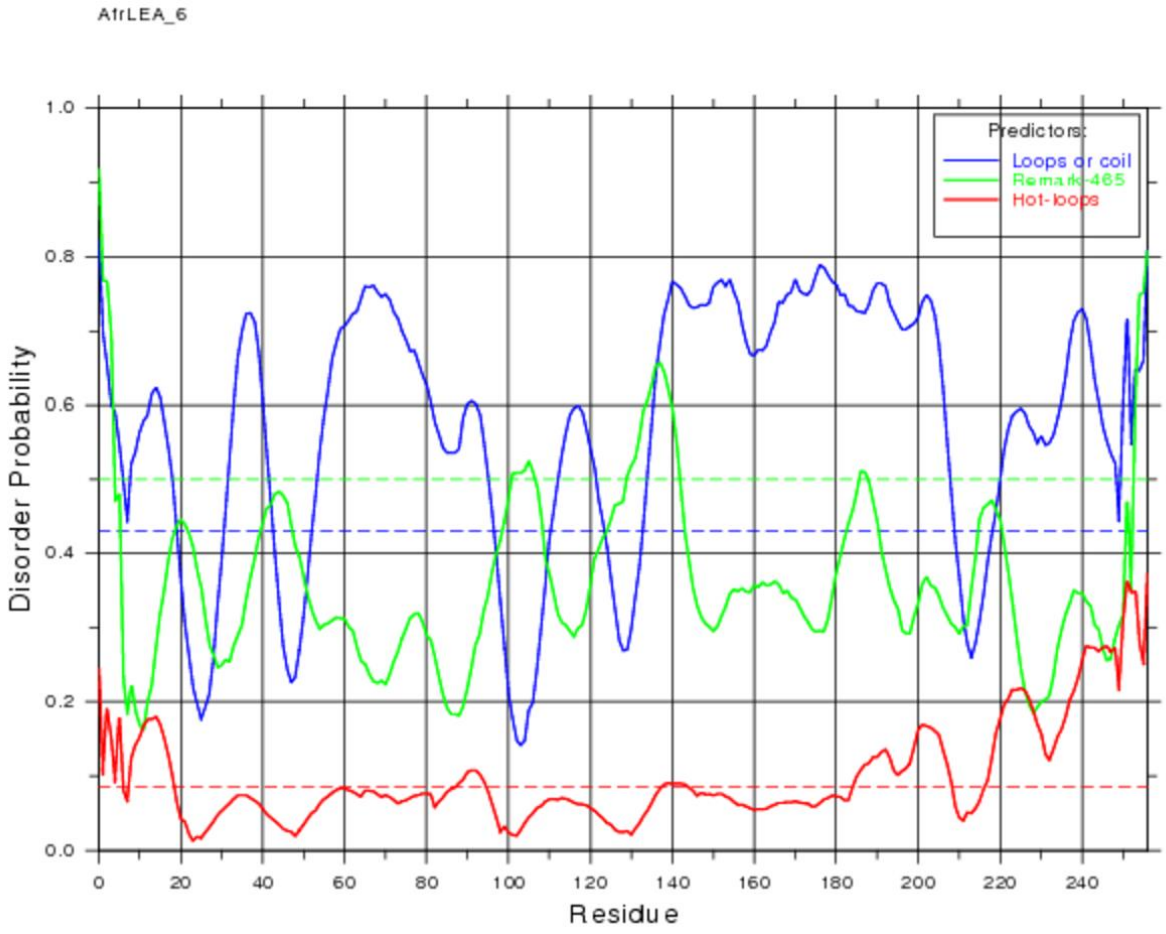

AfrLEA_3m
Curves:
Remark-465
LOW-COMPLEXITY
DISORDER
coiled_coil_region 157-185
Disorder propensity sum
25.2
19.6
14.0
8.3
2.7
-3.0
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
Residue

Figure 17


AfrLEA_6
Predictors:
Loops or coil
Remark-465
Hot-loops
Disorder Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
120
140
160
180
200
220
240
Residue

CHAPTER IV

CHARACTERIZATION OF THE MULTI-PHASE BEHAVIOR OF A GROUP 6 LEA PROTEIN FROM ARTEMIA FRANCISCANA – PROTEINS AS 'SOLVENTS' DURING ANHYDROBIOSIS

In the absence of water the role of hydrophobicity in protein folding and behavior decreases, leaving electrostatic and steric interactions as the prime drivers of structure and function. The protein *Afr*LEA6 is found in the desiccation-tolerant life stage of the animal extremophile *Artemia franciscana*, and the protein engages in a series of charge-mediated phase changes that may confer protection to the animal during water loss. As water loss induces molecular crowding and an increase in ionic strength, *Afr*LEA6 precipitates as a liquid phase (biomolecular condensate) that selectively incorporates positively charged proteins and excludes neutral or negatively charged ones. *Afr*LEA6 further transitions into a gel phase and finally a bio-glass as water loss persists, encapsulating any partitioned proteins within. These partitions may protect the function of critical, positively charged proteins such as transcription factors, ribosomal subunits, and heat-shock proteins, allowing for immediate cellular repair upon rehydration. The liquid-liquid phase separation of *Afr*LEA6 is governed by its seed maturation protein domains, which define it as a group 6 LEA protein. Therefore, it is plausible that all group 6 LEA proteins undergo

liquid-liquid phase separation as an essential characteristic of their mechanism for conferring desiccation tolerance.

INTRODUCTION

It is a biological truism that all known life uses water as a solvent (*188-190*). Some organisms, termed extremophiles, have developed the ability to survive at exceptionally low water levels, or to survive sporadic decreases in available water (*191, 192*). However, these organisms ultimately still require some water. A completely separate category of organisms has evolved to survive even complete desiccation almost indefinitely, which is a phenomenon referred to as anhydrobiosis (for review see: (*193*)). Unlike extremophiles, which can live and even thrive in harsh environments, anhydrobiotic organisms can go dormant in a desiccated state and wait for more favorable conditions. Anhydrobiosis is relatively common phenomenon in plants, which utilize it as a means of prolonging the lifespan of seeds prior to germination (*112*). As a part of their life cycle, many plant seeds completely dry out before they germinate. In this dried state, the rate of chemical reactions which typically occur using water as a solvent is dramatically reduced (*4, 194*). Several structures within plant cells require water to maintain their structure such as the cell membrane and many critical proteins (*195, 196*). To protect these cellular components from desiccation-induced damage, desiccation-tolerant organisms employ a variety of protective osmolytes and proteins. Some of these protectants, such as the non-reducing disaccharides sucrose and trehalose, are simply repurposed energy storage molecules (*197*). However, both metabolites appear to be insufficient to fully protect cells, and all known anhydrobiotic organisms also employ protective proteins, such as heat shock

proteins or other proteins that are specialized for anhydrobiosis (*198-201*). Of these specialized proteins, late embryogenesis abundance (LEA) proteins are among the most prevalent and well-studied group (*94*). LEA proteins derive their name from their prevalence during late embryogenesis (*65, 200*), and they are found in abundance in desiccation tolerant plant seeds (*194*), frost-tolerant and desiccation tolerant plant tissues (*202, 203*), and more recently in anhydrobiotic animals (*8, 193*). Since their discovery in plants, LEA proteins have been identified in Rotifera, Nematoda, and Arthropoda (*6, 64, 66, 82, 83, 87, 204*). Despite the occurrence of a wide variety of LEA protein families, or groups, in plants almost all anhydrobiotic animals only express group 3 LEA proteins according to the nomenclature of Wise and Tunnacliffe (*68*). The primitive brine shrimp, *Artemia franciscana*, is the only known animal to also express group 1 LEA proteins (*85, 180*) and at least one group 6 LEA protein (*151*). Similar to plant seeds, some of these proteins are only present during the embryonic stage of *A. franciscana*'s life cycle, particularly in the encysted embryos that must survive overwintering (*63*). This unique quality of *A. franciscana* makes it a particularly exciting model organism for LEA protein function.

Decades of research has revealed that there are multiple different strategies for achieving anhydrobiosis, that many of these mechanisms may be exclusive to specific taxa, and a clear understanding of the anhydrobiotic cell is still lacking (*6, 8, 58, 107, 205-207*). One major challenge to the understanding of anhydrobiosis is that it is incompatible with standard water-based biochemical methods. Studying anhydrobiosis requires one to perform biochemistry without water, which is the most common media for biochemical experiments. Furthermore, fundamental chemical characteristics are pushed to extremes

during desiccation. Protein and ion concentrations approach infinite as water is depleted, and the general conception of pH as the ratio of H_3O^+, OH^-, and H_2O is replaced with the much more complicated ratio of protonated and unprotonated hydroxyl and amine groups that remain after the water has evaporated (*208*). Certain protein behaviors that are generally considered as anomalous under physiological conditions such as salting out (*209*), vitrification (*18*), and spontaneous conformational transitions are prone to occur during the rapidly changing physicochemical environment (*40*). A protein's behavior during desiccation will be driven by its own physicochemical properties and its propensity for intramolecular and intermolecular interactions. In the hydrated state, the collapse of hydrophobic cores and the exposure of hydrophilic surface residues are a major driving force for protein structure (*210-212*). However, the lack of available water permits these hydrophobic residues to unfold at low water contents, which generally results in denaturation and aggregation of globular proteins (*213*). In contrast, electrostatic interactions and steric repulsion likely maintain their influences over protein structure during desiccation.

*Afr*LEA6 (AWM11684) is the only group 6 LEA protein known to be expressed in an animal and is therefore an exciting protein for investigation. Group 6 LEA proteins are defined primarily by the presence of seed maturation protein (SMP) domains according to the Wise and Tunnacliffe nomenclature (*68*). Proteins containing SMP domains are common in plants and have been linked to the longevity of orthodox seeds in the desiccated state, but the specific function(s) of the SMP domains remain unclear (*1, 206, 214*). Previous characterization of *Afr*LEA6 using bioinformatics suggests that it is composed of two SMP domains close to the n-terminus, a spacer, and a c-terminal binding region with

some resemblance to an SMP domain (*53*). These same analyses suggest that *Afr*LEA6 is uncharacteristically hydrophobic compared to other LEA proteins from *A. franciscana*, but that its enrichment in disorder-promoting residues such as proline and xxx and the high mean net charge would promote large sections of structural disorder. However, the presence of low-complexity, alanine-rich regions within the SMP domains indicated a propensity for disorder-to-order transitions into α-helices, and the alternating charge distribution at the c-terminal protein-binding region may promote β-sheet formation.

Circular dichroism data revealed that these predictions were true for *Afr*LEA6 during desiccation (*215*). Dissolved in pure water, *Afr*LEA6 exists in a 89% randomly coiled conformation, but folds into an ~52% ordered conformation comprised of ~47% α-helices and ~5% β-sheets when the sample is desiccated (*215*). Simulated molecular crowding with 2% SDS produced slightly less folding: ~45% order comprised of ~34% α-helices and ~11% β-sheets (*215*). This degree of conformational transition in the slightly crowded water solution indicates that *Afr*LEA6 changes its conformation early during osmotic stress, rather than during severe desiccation. Therefore, its role in enhancing desiccation tolerance likely occurs well before the cell is desiccated. One possible explanation, considering *Afr*LEA6's hydrophobicity, the increase in ionic strength of the cell. An increase in small osmolytes decreases the available water to maintain the solution of *Afr*LEA6 and may force it separate from solution. If this behavior is physiologically relevant, then *Afr*LEA6 will separate at some critical concentration below its cellular concentration of 0.173 ± 0.016 mg/ml (*215*). This hypothesis is further strengthened by recent data collected using dielectrophoresis to characterize the electrical properties of KC167 cells from *Drosophila melanogaster* expressing *Afr*LEA6 (*216*). When

osmotically stressed with sucrose, which is impermeant to KC167 cells, the cytoplasmic conductivity of cells expressing *Afr*LEA6 increased by ~51%, whereas the value for non-transfected control cells only increased by ~26%. These results suggest that the stokes radius of *Afr*LEA6 decreased, thereby allowing more rapid movement of K^+ and Cl^- ions through the cytoplasm. This secondary indication of *Afr*LEA6's early response to water stress supports the hypothesis that its function occurs well before complete desiccation and, considering its prevalence of low-complexity regions and multivalent interaction sites, that it may undergo a liquid-liquid phase separation (LLPS) like some other osmotic stress proteins (*53, 200, 217, 218*).

The LLPS of proteins within the cell to for biomolecular condensates has given rise to an exciting frontier in molecular biology in the form of "membraneless organelles" (MLOs). Although the name may be self-contradicting, MLOs are quaternary superstructures composed of proteins undergoing weak multivalent interactions with each other (for review, see (*219, 220*)). Sometime these interactions are so weak that they are functionally transient, allowing proteins to move freely across each other but still attracted to each other. In this way, these MLOs can behave as a true liquid, similar to a drop of oil in water. They derive their name from their ability to partition specific biomolecules while excluding others based on the unique physicochemical properties of the proteins that comprise it despite not being bound by a membrane (*48, 221-224*). The differences in molecular inclusion and exclusion as well as the influence of the separating proteins' properties create a distinct environment that can regulate enzymatic pathways (*225*), modulate gene expression (*226-228*), and modify cellular signaling (*229*). Furthermore, MLOs are adaptive to the cellular state. Some MLOs are stable during cellular homeostasis,

such as the nucleolus. These MLOs would be considered "persistent," however, this is only because these MLOs are generally observed in the hydrated state (*230*). Similarly, some MLOs are considered "transient" because they form in the specific environment of a cell outside of homeostasis. Stress granules, for example, form when the cell is exposed to some form of insult (e.g., osmotic stress (*57*)). It is critical in the context of desiccation tolerance to understand that MLOs may still contain some water, but it persists in a different liquid state than the cytoplasmic free water (*231*). Similar behaviors are noted for non-protein polymers such as glycerol, which undergoes a first-order phase transition resulting in a water phase and a glycerol-water phase (*232*).

The formation of MLOs can be driven by a number of stimuli, but not all proteins will react similarly to every stimulus. Most physiological relevant MLO-forming proteins, and some proteins that form LLPS under non-physiological conditions, will undergo a phase transition when excessive amounts of salt outcompete the protein for water, resulting in their precipitation from solution. This process of "salting out" is a common during protein crystallization experiments and has been used for protein purification for decades (*209*). Some proteins aggregate when attempts to salt them out are performed without the correct salts or binding partners. These proteins, such as bacterial single-stranded DNA binding protein, often use nucleotide polymers as their binding partner (*233*). This phase separation specifically partitions DNA repair proteins to enhance the repair of damaged genetic material using single-stranded DNA as a nucleation site. Other MLO-forming proteins are pH sensitive or temperature sensitive (*234*). Shifts in these conditions can be observed during the natural life cycle in the anhydrobiotic cysts of *A. franciscana*. The cysts undergo cooling, which promotes LLPS in some proteins, and they are osmotically

stressed, which may induce salting out. Furthermore, the anhydrobiotic cysts of *A. franciscana* undergo one of the most extreme pH changes in a eukaryotic cell (*235*). The pH of the hydrate diapause cyst begins at ~7.8 but drops to as low at 6.7 in one hour, and even further to 6.3 after several hours. Each of these conditions is then pushed to the extreme when some of the cysts wash up on shore as desiccate (*236*). Due to the various conditions that may promote the formation of a novel anhydrobiosis-related MLO, a series of i*n vitro* experiments was performed. *Afr*LEA6 was considered the prime candidate for LLPS due to its relatively high hydrophobicity, the evidence of its change in behavior at a relatively high content of water, and to the presence of multiple SMP domains that are predicted to undergo multivalent interactions (*53*).

METHODS

Bioinformatics

Several bioinformatics tools were utilized to gain insights into the structural propensities of *Afr*LEA6 and to develop testable hypothesis for wet-bench experiments. SMART EMBL, available at http://smart.embl-heidelberg.de/, was used to identify homologous regions and internal repeats of amino acids in the *Afr*LEA6 protein sequence (*237*, *238*). The net charge per residue (NCPR) was calculated at a pH of 7.2 with a sliding window of 5 amino acids using the program LocalCIDER, available at https://pappulab.github.io/localCIDER/ (*239*). The I-Tasser server was used to model potentially secondary and tertiary structure motifs of *Afr*LEA6 by comparison to homologous structures in proteins with known crystal structures (*240-242*). To determine possible functions of the group 6 LEA proteins, the SMP subgroup domain (IPR007011) was searched on the InterPro protein family classification database. The domain

architectures for all SMP-containing proteins are reported along with any verified domains attached to the proteins. Descriptions of domain function were adapted from the descriptions presented in the database, and suggested references for specific information were included.

Protein Cloning, Expression, and Purification

DNA encoding *Afr*LEA6 (GenBank: MH351624) and green fluorescent protein (Addgene, Watertown, MA, #51562) were cloned into the Ptxb1 (NEB Biolabs Ipswich, MA) vector using standard techniques and expressed as a fusion-protein composed of *Afr*LEA6 or Sgfp (net surface charge of -7) in frame with a chitin-binding protein (CBP) and self-splicing intein protein spacer. The resulting constructs were used to transform the chemically competent *Escherichia coli* strain BL21 Star (Thermofisher, Waltham, MA) and cells were grown on Luria-Bertani (LB) medium-based agarose plates containing 100 µg/ml ampicillin. Antibiotic resistant colonies were selected at random and grown to an optical density of ~0.6 at 595 nm in liquid culture on an orbital shaker at 225 rpm and 37˚C in LB containing 100 µg/ml ampicillin. Protein expression was induced by adding isopropyl-β-1-thiogalactopyranoside (IPTG) to a final concentration of 0.4 mM and the bacteria were harvested after 2 h via centrifugation at 5,000 x g for 30 min at 4˚C. The bacterial pellets were then resuspended in buffer A (500 mM NaCl, 50 mM Tris-HCl, pH 8.5) containing 1 mM phenylmethylsulphonyl fluoride (PMSF) to inhibit serine protease activity. For protein purification, cells were lysed by sonication (Q500, Qsonica, Newtown, CT) and bacterial debris was removed by centrifugation for 30 min at 5,000 x g at 4˚C. The supernatant was loaded via gravity flow onto a 15 ml chitin resin (NEB Biolabs Ipswich, MA) containing

column. The column was washed with 20 column volumes of buffer A. Proteins were eluted from the column after incubation with 50 mM DTT dissolved in buffer A at 4˚C for 48 h then dialyzed into 50 mM phosphate buffer at a pH of 7.0 and concentrated using centrifugal filter units (Amicon Ultra 10 kDa, Millipore Sigma, St. Louis, MO). The purity of the protein was confirmed by SDS-PAGE and averaged at least 95%. Purified protein aliquots were snap frozen in liquid nitrogen and stored at -80˚C until used in experiments. Two additional supercharged GFP constructs both containing a 6XHIS tag for purification purposes, Pgfp (surface net charge: +36; Addgene, Watertown, MA #62937) and Ngfp (surface net charge: -30; Addgene, Watertown, MA #62936), were also expressed in *E. coli*, but the induction with 0.4 mM IPTG was performed at ambient temperature for 16 h (*243*). Bacterial lysates were prepared in binding buffer (20 mM imidazole, 500 mM NaCl, pH 7.4) by sonication and bacterial debris was removed by centrifugation for 30 min at 5,000 x g at 4˚C. The cleared lysates were then applied to 1 ml HisTrap FF columns (Thermofisher, Waltham, MA) using a 20 ml syringe. Columns were washed with 20 ml of binding buffer and the proteins were eluted by raising the imidazole concentration to 500 mM. The GFP containing elution fraction was dialyzed against 50 mM sodium phosphate buffer at a pH of 7.0 and the proteins were concentrated using centrifugal filter units with a 10 kDa molecular weight cutoff (Amicon Ultra 10 kDa, Millipore Sigma, St. Louis, MO). The purity of the protein was confirmed by SDS-PAGE electrophoresis and averaged above 95%. Purified protein aliquots were snap frozen in liquid N_2 and stored at -80˚C until used in experiments.

AfrLEA6 Liquid-Liquid Phase Separation

Stocks of *Afr*LEA6 were dialyzed against 20 mM sodium phosphate buffer (Ph 6.5), or a solution with a composition designed to resemble the crowded conditions in the cytoplasm and concentrations of major osmolytes measured in *A. franciscana* cysts ('osmosome solution': 32 mM NaCl, 98 mM KCl, 11 mM K_2PO_4, 5 mM $CaCl_2$, 340 mM trehalose, 2.9 % w/v glycerol, and 25% Ficoll 400, pH 6.5) (*244*). A positive LLPS control solution of 50 mM DTT and 200 mM NaCl was used to verify LLPS behavior under identical conditions. Potential LLPS events were observed by light microscopy after pipetting osmosome solution containing *Afr*LEA6 (0.17 mg/ml) onto glass plates and allowing water to evaporate from the samples at ambient conditions. To investigate interactions between *Afr*LEA6 droplets and GFP constructs, the fluorescent proteins (35 mM) were added to 1 mg/ml *Afr*LEA6 (37 mM) in sodium phosphate buffer and the solutions were allowed to desiccate by convective drying as described above. Samples were imaged every five minutes using a Nikon A1R confocal microscope (Nikon Instruments Inc., Melville, NY) until no liquid water was observed to ensure that interactions between *Afr*LEA6 and GFP constructs were consistent throughout the drying process.

SEM and AFM Imaging

*Afr*LEA6 superstructures in the desiccated state were observed after drying the protein (1 mg/ml) in sodium phosphate buffer (Ph 6.5) on aluminum stages. The air-dried samples were further dried in a sealed desiccation chamber over anhydrous calcium sulfate (Drierite, W A Hammond, Xenia, OH) for one week. The dried samples were then sputter coated with gold and examined using SE2 scanning electron microscopy using a Zeiss Supra 35 instrument.

RESULTS AND DISCUSSION

Bioinformatics

The *A. franciscana* protein *Afr*LEA6 is composed of three structurally distinct regions that exhibit unique sequence profiles. According to SMART-EMBEL, the n-terminus is dominated by two SMP domains with e-values of $2.6E^{-6}$ and $1.6E^{-4}$ compared to the PF04927 (IPR007011) consensus sequence. Each of these SMP domains overlaps with a larger internal repeat which is enriched in a low-complexity alanine region. I-Tasser predicts that the two n-terminal SMP domains fold into α-helices (Fig. 18A, B) which are overall negatively charged with a mean net charge value of -0.2 (Fig. 18C). The n-terminal α-helices are separated from the c-terminal region by a 40-60 amino acid long, intrinsically disordered spacer region starting at amino acid position 140, that is enriched in proline, glycine, and aromatic or hydrophobic residues. At the c-terminus, starting around amino acid position 200, the net charge distribution alternates between highly positive (+0.5) and highly negative (-0.5), which is indicative of a protein binding site. This region is predicted by I-Tasser to bind transcription-associated proteins such as beta-catenin (1g3Ja, C-score: 0.05). Overall, proteins with structural homology to *Afr*LEA6 and TM scores larger 0.7 are associated with nuclear pores (e.g. PDB: 4knh; 4kf7; 5ijn) and DNA repair mechanisms (e.g. PDB: 5yz0; 5dlq; 5loi).

The overlapping behavior of these SMP domains shows a propensity for self-interactions. In globular proteins, which as significantly less dynamic than the structural predictions of *Afr*LEA6 would suggest, this would result in the formation of rigid, conserved intramolecular tertiary structures. However, in the disordered state, these regions are also likely to engage in intermolecular binding events (*55*). Considering the

relatively hydrophobic properties of *Afr*LEA6 and the rapidly depleting amount of available water (Fig. 18C), then the potential for *Afr*LEA6 to precipitate from solution during desiccation is very high. Their intrinsic disorder and self-repulsion due to their high mean net charge should limit these interactions to transient SMP-SMP interactions, thereby preventing the irreversible formation of solid aggregate or fibrils. Therefore, this domain architecture is likely to force *Afr*LEA6 to separate from solution in a liquid phase rather than a solid phase (*55*).

Behavior of AfrLEA6 During Simulated Water Stress Conditions

*Afr*LEA6 dissolves readily at a physiological concentration of ~0.17 mg/ml (*215*) in a buffer mimicking the ionic strength and concentrations of the most dominant solutes present in the cytoplasm of hydrated diapausing cysts of *A. franciscana* (osmosome solution). During evaporative water loss from an osmosome solution containing *Afr*LEA6 at a starting concentration of 0.17 mg/ml, the protein separates readily from the remaining solvent as a liquid condensate when reaching a concentration of approximately 0.25 mg/ml even in the absence of other proteins or nucleic acids (Fig. 19A). The formed droplets fuse on contact with each other and maintain a spherical structure, both during and after any fusion events. In a solution of pure water containing *Afr*LEA6 at a starting concentration of ~0.2 mg/ml, droplets do not form during drying, and a thin layer of *Afr*LEA6 that resembles a glassy state lines the bottom of the sample after complete desiccation, but no LLPS was detected.

The LLPS of *Afr*LEA6 in osmosome solution is likely to occur in the cytoplasm of desiccating brine shrimp embryos, but the cyst shell does not allow for direct imaging of

this process and cell fixatives used in preparation for electron microscopy may disrupt the detection of these structures. However, the LLPS of *Afr*LEA6 *in vitro* does not fully demonstrate that *Afr*LEA6 undergoes LLPS *in vivo*. There are several potential binding partners within the cytoplasm that could change *Afr*LEA6's behavior, and *in vivo* experimentation is required for true verification. In collaboration with Clinton Belott at the University of Louisville, *Afr*LEA6 was shown to undergo LLPS in the cellular space under mild osmotic stress (*245*). These experiments also verified that cell expressing *Afr*LEA6 tolerated osmotic stress and desiccation better than controls, and that the SMP domains were essential for LLPS. However, surprisingly, the c-terminal binding region facilitated the formation of larger, more organized structures. This indicates that the c-terminal region may function as a partner-binding domain.

AfrLEA6 Partitioning Based on Protein Surface Charge

Given the high overall negative charge of *Afr*LEA6, it appeared unlikely that negatively charges molecules such as nucleic acids would be partitioned inside. To investigate if protein surface charge affects intermolecular interactions between *Afr*LEA6 and other proteins, a 35 mM solution of *Afr*LEA6 was desiccated in a buffer containing 20 mM sodium phosphate (Ph of 6.5) to induce droplet formation in the presence of supercharged GFP proteins added at an equimolar ratio. The utilized GFP proteins have a calculated net surface charge of -7 (Sgfp-7), +36 (Pgfp+36), and -30 (Ngfp-30) (*246*). When a solution of *Afr*LEA6 plus Sgfp-7 was desiccated, LLPS formed droplets of similar size and shape compared to droplets formed in a solution of *Afr*LEA6 only. Interestingly, Sgfp-7 is clearly excluded from the interior space of the formed protein droplets (Fig. 20A). In contrast,

when *Afr*LEA6 is desiccated in the presence of Pgfp+36 the protein droplets form more readily and are smaller compared to those that form in absence of other proteins and Pgfp+36 is incorporated into the droplet with such high efficiency that the surrounding solution becomes depleted of the fluorescent protein (Fig. 20B). As desiccation proceeds, these protein condensates expand and maintain a seemingly constant ratio of Pgfp+36 and *Afr*LEA6. *Afr*LEA6 condensates formed in presence of Ngfp-30 also exclude this negatively charged fluorescent protein (Fig. 20C).

The partitioning of positively charged GFP+36 is unsurprising given the highly negative electrochemical environment of the *Afr*LEA6 MLO. However, what is quite surprising is the apparent enrichment of positively charged GFP+36 (Fig 3B). Rather than simply permitting entry into the proteinaceous phase, it appears to actively draw in the target protein. Also surprising was the extent to which a quite modestly negative protein, Sgfp-7 is excluded (Fig. 20A). This ability to highly selectively incorporate target proteins strongly indicates that structure should offer some protect to key proteins during desiccation.

The internal space of the *Afr*LEA6 MLO will likely confer similar protections against misfolding as has been proposed in the molecular shielding hypothesis (*20*). The disordered regions of *Afr*LEA6 that attract target proteins will also act as physical barriers in between them during water loss, thus preventing aggregation of globular proteins. Additionally, any proteins that maintain their structure under the physicochemical conditions of the *Afr*LEA6 MLO are stabilized by the interactions between themselves and *Afr*LEA6. Even during extreme desiccation, these protein-protein interactions should

maintain the collapse of hydrophobic clusters and attract the polar and charged residues to the surface of the protein, not unlike water or trehalose (*62*).

AfrLEA6 during extreme desiccation

To characterize the behavior of *Afr*LEA6 during transition from a low-water environment to a fully desiccated state, samples of *Afr*LEA6 were observed using confocal microscopy during desiccation in 20 mM sodium phosphate buffer (Ph 6.5) (Fig. 20A). As evaporation concentrates the proteins, the *Afr*LEA6 MLOs expand in size and lose some of their liquid properties. These larger structures do not completely return to a spherical shape but encapsulate spherical droplets of media containing Sgfp-7, thus behaving more similarly to a hydrogel than a true liquid (*247*) (Fig. 20B). The hydrogel structure appears to be preserved in the desiccated state as shown by scanning electron microscopy (SEM) and atomic force microscopy (Fig. 20C, D), although it loses all flexibility. In ultrapure water, *Afr*LEA6 does not undergo LLPS, and remains in solution until it is amorphously deposited onto the surface that the solution was dried on. When completely dry, *Afr*LEA6 appears to vitrify into a bio-glass (Fig. 21). It becomes hard and brittle, but allows light to pass through it similar to the sugar glasses that stabilize proteins and membranes during desiccation in *A. franciscana* and several other anhydrobiotic plant and animal cells (*38*). It even shatters when exposed to mechanical stress, leaving small shards that maintain their rigid shapes so long as they remain desiccated (Fig. 22). This final transition into a glassy state during complete desiccation does not change the partitioning properties of *Afr*LEA6, despite the loss of dynamic motion. Even when encased in NaCl crystals, *Afr*LEA6 MLOs maintain their shape and continue to exclude Sgfp from within the proteinaceous phase (Fig. 23).

Trends in SMP Domain Architecture

To better evaluate the significance of the role of the SMP domain in the LLPS of *Afr*LEA6, a broad search for all proteins containing SMP domains was performed using Interpro (Table 5). Interpro is a database that uses sequence characteristics and user annotations to label structural domains within proteins. This search revealed several noteworthy characteristics of SMP-containing proteins, although not all of them may be considered group 6 LEA proteins. The three most common protein architectures represented 97.2% of all reported SMP-containing proteins. These three architectures were proteins containing one, two, or three SMP domains and, generally, some other uncharacterized amino acids. An additional 1% of the reported architectures were comprised entirely of four, five, or six SMP domains without any other verified protein domains, meaning that over 98% of SMP-containing proteins are not reported as being associated with any other formal protein domain. This suggests that the functions of group 6 LEA proteins, and SMP-containing proteins in general, operate using additional protein domains which have not been characterizes, such as *Afr*LEA6's c-terminal binding domain. Furthermore, the variations in the number of SMP repeats in these proteins indicates that the phase separation induced by the SMP domain is one of its primary functions. Group 6 LEA proteins containing between one and three SMP repeats are highly prevalent and often present together within the same species. However, the larger sized SMP repeats are found independently in certain species. This appears similar to the variety of group 1 motif repeats reported for group 1 LEA proteins, such as those found in *A. franciscana*.

The remaining 2% of proteins containing SMPs may offer some insights into the general functions of group 6 LEA proteins. Although most proteins with additional domains besides SMP were rare, generally only one example of each protein was reported, this domain reporting may be due to the extrapolation of domains from sequence characteristics rather than due to the actual presence of the domains. Therefore, the functions assigned to these domains, while not entirely reliable, offer initial indications of their functions. Transcription regulation and histone binding domains account for over 23% of the domains predicted in these proteins, whereas another 33% of the domains were involved in RNA processing, reverse transcription, and translation. Overall, 52% of proteins containing an SMP domain and any other protein domain were associated with regulating the production, processing, or binding of RNA. This association with nucleotide-interacting proteins is not unreasonable. As previously stated, nucleic acid polymers are among the most common binding partners during protein LLPS (*248*). Nucleotide-binding proteins also have a tendency to be positively charged, which increases their affinity for the negatively-charged phosphate backbone of nucleic acids (*249*). Therefore, the similarities of these protein domains and the general selection of positively charged proteins into the *Afr*LEA6 MLO indicate that its function is likely related to the regulation, storage, and protection of DNA and RNA-binding proteins such as transcription factors, ribosomes, RNA-processing proteins, and gene expression regulatory proteins.

CONCLUSIONS

*Afr*LEA6 is a unique protein among LEA proteins because it is the only known SMP-containing protein expressed in animals, despite there being several anhydrobiotic animals

from different phyla (*8*). The SMP domain found in all group 6 LEA proteins is now understood to drive the precipitation of these LEA proteins from solution as a separate liquid phase that selectively partition biomolecules and confer osmotic stress tolerance and desiccation tolerance (*245*). This novel, anhydrobiosis-related MLO is an elegant solution to several hypotheses for LEA protein function that are seemingly in conflict. The ability to transiently form a protective structure that amplifies its target within it allows for these targets to be well-protected even if *Afr*LEA6 represents only a small fraction of the proteome. The incorporated proteins also contribute to the structure of this MLO, thereby converting would-be desiccation-vulnerable biomolecules into participants in a protective cellular compartment. The transition from a liquid state to a glassy state seamlessly connects the understood mechanism of anhydrobiosis and protection in the desiccated state with a contemporary understanding of stress granule formation and function (*38*, *51*). Even the well-measured but seemingly inconsequential contribution of hydration buffering appears to participate in this multi-phase mechanism as the removal of the *Afr*LEA6 hydration layer is what drives the transition from aqueous to liquid to hydrogel and eventually to glass (*21*). These results strongly indicate that the LLPS of group 6 LEA proteins is a conserved mechanism of function, rather than a unique property of *Afr*LEA6. The variety in the number of SMP domains further suggests that this LLPS behavior may be tunable to the specific vulnerabilities of the particular cell, to offer protection during different rates of desiccation, or to undergo LLPS as appropriate in different cellular compartments with different physicochemical conditions and responses to osmotic stress. Further research is needed to investigate the behavior of other group 6 LEA proteins from plants, which can be performed observing the desiccation of purified proteins in a droplet

of salt water while under an inverted microscope. It may be that when anhydrobiotic organisms run out of solvent, they simply make their own.

Tables

Table 5: A breakdown of the domain architectures of proteins that contain SMP domains as reported by InterPro.

Domain Architecture	Other Domains	Domain Functions	# Of Hits
SMP			525
SMP-SMP			489
SMP-SMP-SMP			611
SMP-SMP-SMP-SMP			6
SMP-SMP-SMP-SMP-SMP			7
SMP-SMP-SMP-SMP-SMP-SMP			5
SMP-IPR009311	IPR009311: Interferon alpha-inducible protein IFI6/IFI27-like	IPR009311: Pro-apoptotic pathway regulator (*250, 251*)	3
IPR009311-**SMP**	IPR009311: Interferon alpha-inducible protein IFI6/IFI27-like	IPR009311: Pro-apoptotic activity regulator (*250, 251*)	3
IPR014030-**SMP**-IPR014031	IPR014030: Beta-ketoacyl synthase, N-terminal IPR014031: Beta-ketoacyl synthase, C-terminal	IPR014030: Involved in substrate binding for the catalysis of fatty acid chains (*252*) IPR014031: Involved in substrate binding for the catalysis of fatty acid chains (*252*)	3
SMP-IPR000182	IPR000182: GNAT domain	IPR000182: N-Acetyltransferases often associated with antibiotic tolerance (*253*) and histone acetyltransferases (HATS) (*254, 255*)	2
IPR041430-IPR000330-IPR001650-**SMP**-**SMP**	IPR041430: ATRX, ADD domain	IPR041430: Regulates heterochromatin formation via DNA binding (*256*) and patterns of histone methylation (*257*)	2

IPR041430-IPR000330-IPR001650-**SMP**-**SMP** (continued)	IPR000330: SNF2, N-terminal IPR000330: SNF2, N-terminal (continued)	IPR000330: Generally associated with transcription regulation, DNA repair, DNA recombination, and chromatin unwinding (*258-260*). Can hydrolyze ATMP to disrupt DNA-histone interactions to increase transcription factor interactions (*261*)	
	IPR001650: SNF2, N-terminal	IPR001650: Part of superfamilies 1 and 2. Generally associated with eukaryotic translation initiation (*262*)	
IPR003691- IPR003691-**SMP**	IPR003691: Putative fluoride ion transporter CrcB	IPR003691: Transports toxic fluoride ions out of the cell (*263*)	1
SMP-IPR001680-IPR001680	IPR001680: WD40 repeat	IPR001680: Common 7-8 bladed propeller structure associated with signal transduction, cell cycle regulation, transcription regulation, and apoptosis. Repeated domains induce protein-protein or protein-DNA interactions (*264-266*)	1
SMP-IPR008586	IPR008586: Protein of unknown function DUF868, plant	IPR008586: Function unknown	1
IPR02558-**SMP**	IPR02558: Domain of unknown function DUF4283	IPR02558: Function unknown, but it is paired with a wide variety of other domains	1
IPR032691-IPR015403-IPR032817- IPR032817-**SMP**	IPR032691: Guanine nucleotide exchange factor, N-terminal	IPR032691: Associated with guanine nucleotide exchange factors involved in Golgi transport, but it is not the binding component (*267-269*)	1
	IPR015403: Sec7, C-terminal	IPR015403: Involved in scaffolding the COPII-COPI protein switch for VTC maturation and Golgi compartment biogenesis (*269*)	
	IPR032817: Mon2, C-terminal	IPR032817: Essential domain for scaffolding proteins for	

		endomembrane trafficking (*270*, *271*)	
SMP-**SMP**-IPR026960	IPR026960: Reverse transcriptase zinc-binding domain	IPR026960: Zinc binding during transcription of DNA using an RNA template	1
IPR025558-IPR000477-IPR026960-**SMP**	IPR025558: Domain of unknown function DUF4283	IPR025558: Function unknown, but it is paired with a wide variety of other domains	1
IPR025558-IPR000477-IPR026960-**SMP**	I PR025558: Domain of unknown function DUF4283	IPR025558: Function unknown, but it is paired with a wide variety of other domains	
IPR025558-IPR000477-IPR026960-**SMP**	IPR000477: Reverse transcriptase domain	IPR000477: Transcribes DNA using an RNA template (*272*)	
	IPR026960: Reverse transcriptase zinc-binding domain	IPR026960: Zinc binding during transcription of DNA using an RNA template	
IPR004360-**SMP**-**SMP**-**SMP**	IPR004360: Glyoxalase/104osfomycin resistance/dioxygenase domain	IPR004360: Catalyzes the reaction of lactoylglutathione into lactic acid (*273*)	1
IPR016140-IPR001680-IPR001680- IPR001680-**SMP**-**SMP**	IPR016140: Bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain	IPR016140: Structural domain found in seed lipid storage (*274*) and transfer proteins (*275*, *276*) and proteases/ α -amylases inhibitors (*277*, *278*)	1
	IPR001680: WD40 repeat	IPR001680: Common 7-8 bladed propeller structure associated with signal transduction, cell cycle regulation, transcription regulation, and apoptosis. Repeated domains induce protein-protein or protein-DNA interactions (*264-266*)	
IPR003737-**SMP**	IPR003737: N-acetylglucosaminyl phosphatidylinositol deacetylase-related	IPR003737: Catalyzes the second step in glycosylphosphatidylinositol (GPI) biosynthesis (*279*, *280*)	1
IPR001752-**SMP**-**SMP**-**SMP**	IPR001752: Kinesin motor domain	IPR001752: Microtubule-associated protein that may	1

		transport organelles or participate spindle elongation (*281*) and in chromosome localization during nuclear fusion (*282*), mitosis (*281*, *283*), and neuronal differentiation (*284*)	
IPR000999-**SMP**	IPR000999: Ribonuclease III domain	IPR000999: Double-stranded RNA-specific endonuclease involved in post-transcriptional modification of mRNA, in ribosomal RNA precursor processing, Trna and tRNA precursor processing, processing of small nucleolar RNAs and snRNAs, and is involved in RNAi and miRNA gene silencing (*285-289*)	1
IPR016140-**SMP**-**SMP**	IPR016140: Bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain	IPR016140: Structural domain found in seed lipid storage (*274*) and transfer proteins (*275*, *276*) and proteases/ α -amylases inhibitors (*277*, *278*)	1
SMP-IPR041366-IPR001950	IPR041366: Pre-PUA domain	IPR041366: A domain commonly found before the PUA domain, which is associated eukaryotic translation, tRNA and rRNA post-transcriptional modifications, and ribosomal biogenesis (*290-292*)	1
	IPR001950: SUI1 domain	IPR001950: Directs the ribosome to the start codon in conjunction with Eif-2 and tRNA-Met (*293*)	
SMP-IPR000477-IPR041577-**SMP**-**SMP**	IPR000477: Reverse transcriptase domain	IPR000477: Transcribes DNA using an RNA template (*272*)	1
	IPR041577: Reverse transcriptase/retrotransposon-derived protein, RNAse H-like domain	IPR041577: Transcribes single-stranded RNA into double-stranded DNA (*294-296*)	

Figure Captions

Figure 18: *Afr*LEA6 is predicted to have three distinct regions. A) The n-terminal SMP repeats (purple) exhibit alpha-helical propensity. The c-terminal domain (green) exhibits fuzzy self-interactions. These two domains are linked by an intrinsically disordered spacer enriched in proline, glycine, and aromatic residues (blue). I-Tasser structural prediction of *Afr*LEA6 is based on hierarchical stability of known crystal structures, thus associating this structure with a possible conformation in the dried state. B) SmartEMBLE identifies two n-terminal SMP domains in *Afr*LEA6 (purple). C) The protein is overall negatively charged, with alternating charges (green) at the c-terminus promoting interactions with other proteins or itself.

Figure 19: *Afr*LEA6 undergoes an LLPS that sequesters *in vitro*. A) *Afr*LEA6 (0.17 mg/ml) separates from solution into a liquid phase in a buffer mimicking the intracellular milieu of *A. franciscana*. B) DTT in 200 mM NaCl undergoes LLPS with similar behaviors to *Afr*LEA6.

Figure20: *Afr*LEA6 undergoes an LLPS that sequesters GFP based on surface charge. A) Standard GFP (stGFP-7) is partitioned outside of the *Afr*LEA6 droplet. B) Positive GFP (Pgfp+36) is selectively partitioned and enriched within the droplet and C) highly negatively charged GFP is also excluded (Ngfp-30).

Figure 21: *Afr*LEA6 undergoes phase transitions during desiccation. A) Confocal microscopy shows that *in vitro Afr*LEA6 condensates are spherical and heterogenous at low to moderate dehydration. B) *Afr*LEA6 condensates *in vitro* increase in viscosity and

form a gel-like matrix at moderate to severe desiccation. C) SEM imaging shows that some *Afr*LEA6 condensates maintain a spherical structure in the desiccated state. D) AFM imaging reveals a series of mobile parallel proteins, aligning into a hydrogel structure.

Figure 22: *Afr*LEA6 forms a glassy layer when dried on an aluminum stage at 0% relative humidty. This protenecous glass is brittle and cracks when the aluminum stage is dropped from a 1 cm height as shown in the SEM image.

Figure 23: *Afr*LEA6 MLOs maintain their structure in the completely desiccated state. A) *Afr*LEA6 MLOs vitrify in their spherical shape and maintain their structure even when encased in NaCl crystals. B) *Afr*LEA6 MLOs maintain their exclusionary behavior against Sgfp even in the completely desiccated state.

Figures

Figure 18

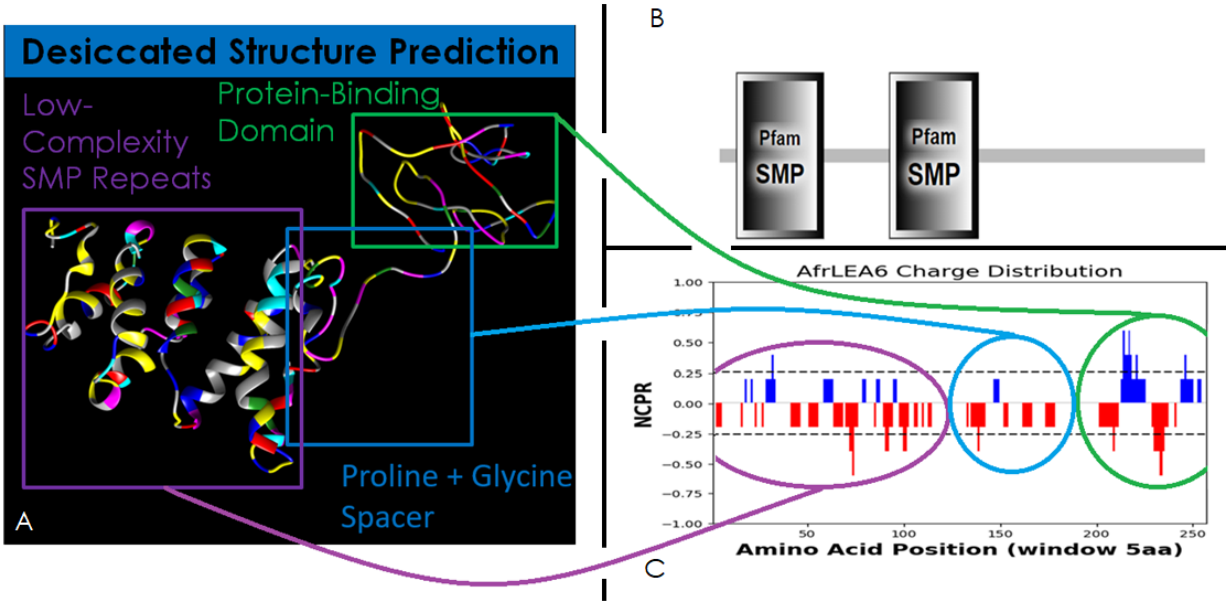


Figure 19

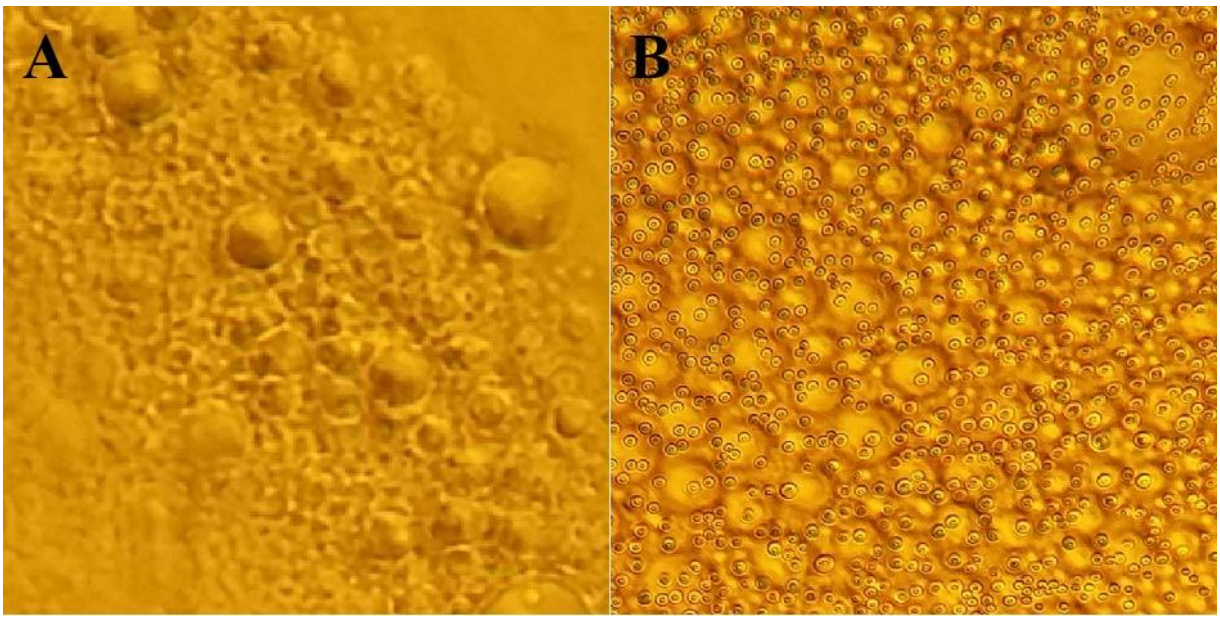


Figure 20

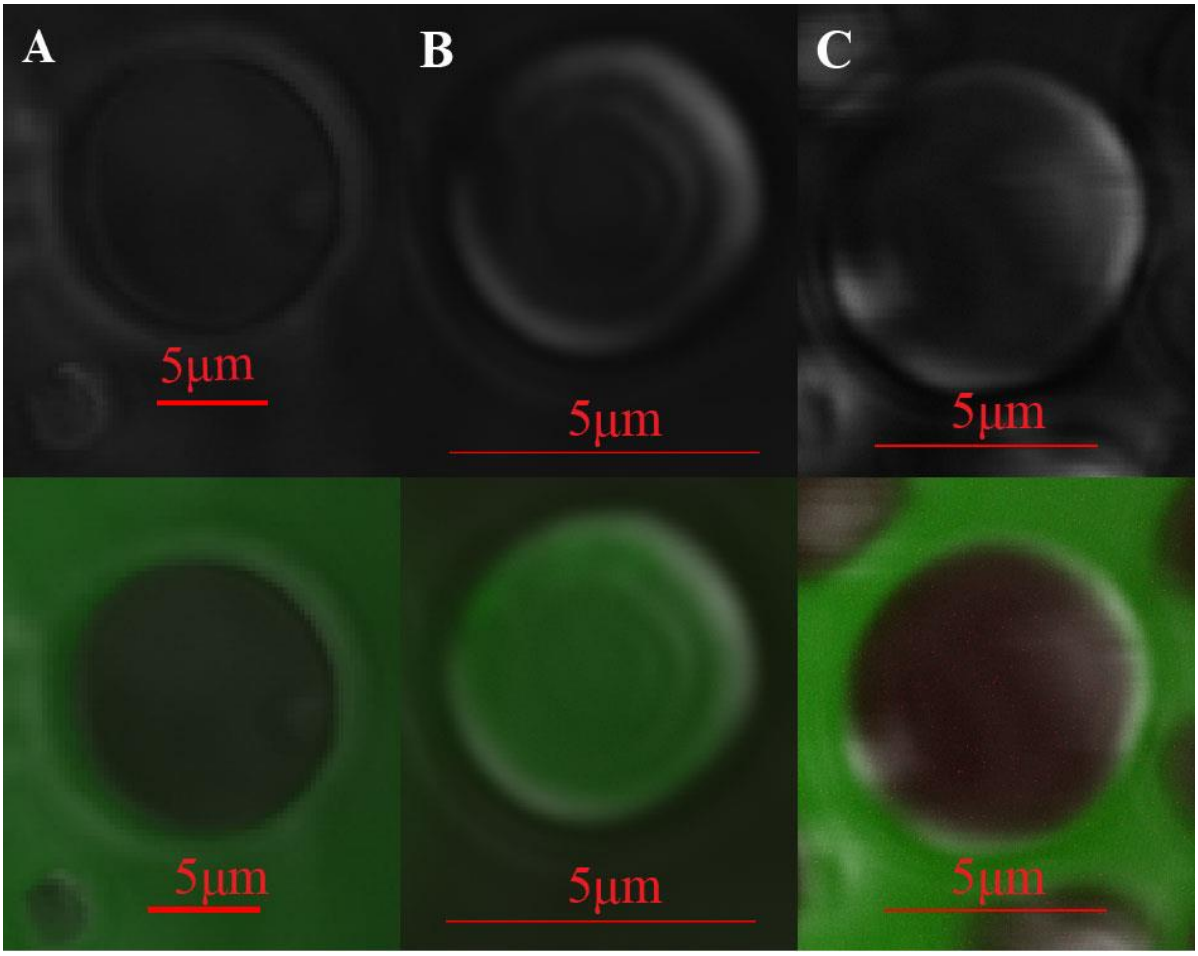


Figure 21

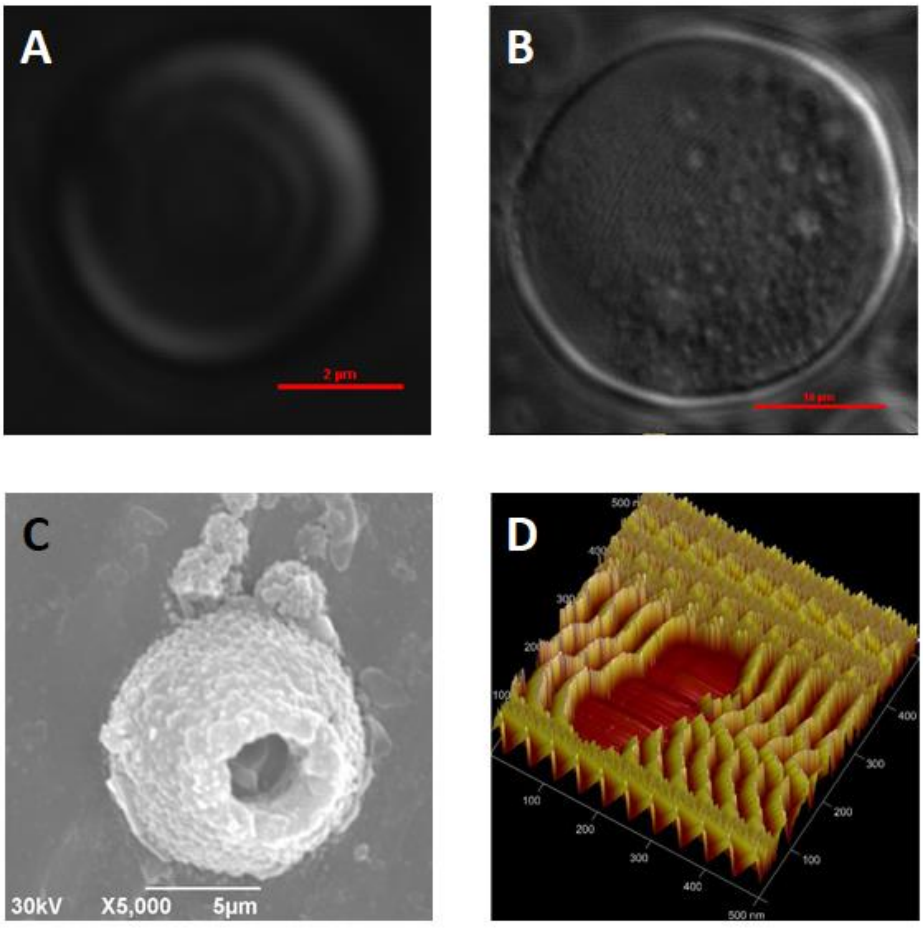

A
2 µm
B
C
30kV X5,000 5µm
D
500 nm
400
300
200
100
100
200
300
400
500 nm

Figure 22

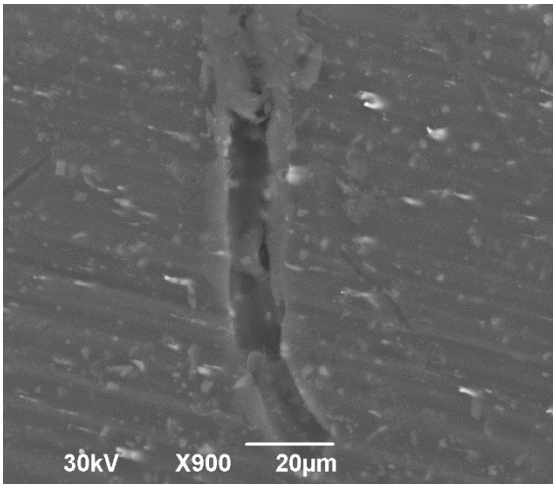
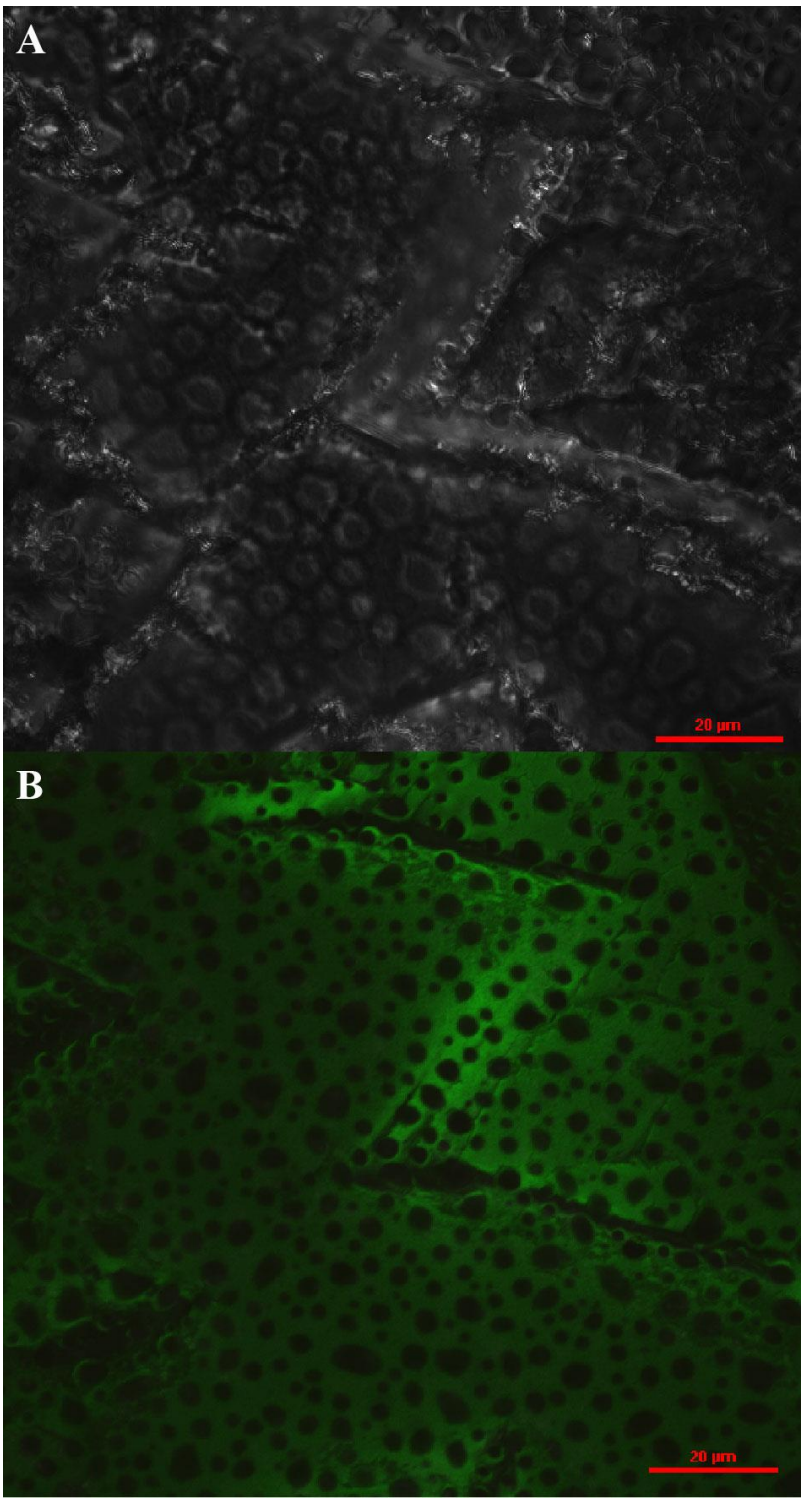


Figure 23



CHAPTER V

STRUCTURAL CHARACTERIZATION OF A GROUP 1 LEA PROTEIN FROM ARTEMIA FRANCISCANA – PROMISCUOUS BINDING BEHAVIOR AND CONFORMATIONAL TRANSITIONS DURING DESICCATION

SUMMARY

*Af*LEA1.1 is a group 1 LEA protein expressed by the encysted embryos of *Artemia franciscana*, which the only group of LEA proteins that has been demonstrated to be essential to their desiccation tolerance. *Af*LEA1.1 is surprisingly ordered in the hydrated state (40% ordered), especially compared to other LEA proteins from *A. franciscana*. During purification and during desiccation, *Af*LEA1.1 engages in conformational transitions and intermolecular interactions with other *Af*LEA1.1 molecules. However, *Af*LEA1.1 was only observed as a monomer when its mass was quantified using size-exclusion chromatography. *Af*LEA1.1 was observed undergoing the most extreme disorder-to-order conformational transition ever recorded for a LEA protein, transitioning from 35% β-sheets and 5% α-helices in the hydrated state to 5% β-sheets and 85% α-helices with no detectable disorder in the completely desiccated state. Simulating molecular crowding with 2% SDS induced a conversion of β-sheets to α-helices, but no conversion of random coils to any other type of structure. *Af*LEA1.1 likely folds into an armadillo-repeat protein-like structure and localizes stripes of positive and negative amino acids along the protein's surface to stabilize this conformation. Strong indication exists that the

internal region of the dried *Af*LEA1.1 protein binds mRNA and promotes the liquid-liquid phase separation of *Af*LEA1.1, and this behavior may be universal to group 1 LEA proteins, but its role in conferring desiccation tolerance

is unknown. *Af*LEA1.1 protects LDH during desiccation, but does not repair LDH after desiccation-induced damage.

INTRODUCTION

Artemia franciscana is the only known animal to express more than one family of LEA proteins (*87*). It is also the only known animal to express group 1 LEA proteins, which are typically found in plants, bacteria, and archaea (*8, 297*). Only the encysted embryos of *A. franciscana* express LEA proteins and can survive desiccation (*12*). *A. franciscana* expresses several LEA proteins from group 1, group 3, and at least one LEA protein from group 6, based on the nomenclature developed by Wise et al. (*68*). Despite its uniquely wide range LEA proteins, the group 1 LEA proteins are the only LEA protein to be experimentally shown to influence desiccation tolerance (*178*). When all group 1 LEA proteins were knocked down in *Artemia* cysts, their ability to survive freezing was reduced by over 50% and their desiccation tolerance was reduced by over 90% (*178*). Therefore, understanding the mechanism by which *Af*LEA1.1 confers desiccation tolerance onto the cyst may lead to the development of methods to desiccate and rehydrate other living cells and tissues such as research cell lines and medical stem cells.

Two of the group 1 LEA proteins of *A. franciscana* are a cytoplasmic protein named *Af*LEA1.1 and a mitochondrial group 1 LEA protein called *Af*LEA1.3 (*87, 180*). *Af*LEA1.3 is an almost identical protein to *Af*LEA1.1 aside from an n-terminal mitochondrial leader sequence. Therefore, experiments meant to elucidate the function of *Af*LEA1.3 may also

be performed using *Af*LEA1.1 with the understanding that the mitochondrial space has different chemical properties from the cytoplasm. The majority of the group 1 LEA proteins expressed in the cysts of *A. franciscana* are extremely similar, with different numbers of repeating LEA group 1 motifs. The requirement of *Af*LEA1.1 to be present in both the cytoplasm and the mitochondria offers some insight into its role in desiccation tolerance. *Af*LEA1.1 is unlikely to specifically target proteins associated with oxidative phosphorylation, for example. Within these compartments, there are several variations of *Af*LEA1.1, each with different numbers of LEA group 1 consensus sequence motifs (*178, 180*). The existence of similar proteins with varying lengths may indicate that the number of group 1 motifs is not critical to the function of *Af*LEA1.1. Alternatively, the number of repeats may regulate the behavior of *Af*LEA1.1 during desiccation. More group 1 repeats may increase the odds of folding, for example, or increase odds of binding a desiccation-sensitive target. Shorter proteins, however, may be more mobile in the increasingly viscous drying cell, which is filled with protective osmolytes such as trehalose and glycerol which may hinder the movement of large polymers (*14, 15*).

Despite over a decade of study focused on *Af*LEA1.1, little is known about its mechanism of function. Its sequence features have been thoroughly investigated using bioinformatics, and several *in vivo* experiments have been performed by transfecting bacteria with *Af*LEA1.1 variants to improve their desiccation tolerance (*53, 179*). Determining the mechanism by which *Af*LEA1.1 confers desiccation tolerance and what its desiccation-sensitive targets is more straightforward using *in vitro* techniques. However, the unique challenge of performing biochemistry without water has led to limited progress on this front. Nonetheless, techniques have been developed to deduce the secondary

structure of LEA proteins in the desiccated state using bioinformatics and circular dichroism (*298, 299*). Other techniques may also be adapted to study the behavior of LEA proteins in the unique chemical environment that they function under. The liquid-liquid phase separation behavior of the group 6 LEA protein from *A. franciscana*, *Afr*LEA6, was verified using a combination of light microscopy techniques, scanning electron microscopy, atomic force microscopy, and simple drying procedures in the presence of osmolytes found in the anhydrobiotic cyst (*245*). By using some of these same techniques and guided by computation tools, it is possible to also gain insight into the behavior of *Af*LEA1.1 during desiccation. These insights can be applied to the various hypotheses regarding LEA protein function, such as molecular shielding (*20*), hydration buffering (*21*), or protein glass reinforcement (*36, 37, 106*).

METHODS

Protein Cloning, Expression, and Purification

DNA encoding *Af*LEA1.1 was cloned into the Ptxb1 (NEB Biolabs Ipswich, MA) vector using standard techniques yielding a fusion protein comprised of *Af*LEA1.1 in frame with a chitin-binding protein (CBP) and a self-splicing intein protein spacer. The resulting construct was used to transform the chemically competent Escherichia coli strain BL21 Star (Thermofisher, Waltham, MA) and cells were grown on Luria-Bertani (LB) medium-based agarose plates containing 100 µg/ml ampicillin. Antibiotic resistant colonies were selected at random and grown to an optical density of ~0.6 at 595 nm in liquid culture on an orbital shaker at 225 rpm and 37˚C in LB containing 100 µg/ml ampicillin. Protein expression was induced by adding isopropyl-β-1-thiogalactopyranoside (IPTG) to a final

concentration of 0.4 mM and the bacteria were harvested after 2 h via centrifugation at 5,000 x g for 30 min at 4˚C. The bacterial pellets were then resuspended in buffer A (500 mM NaCl, 50 mM Tris-HCl, pH 8.5) containing 1 mM phenylmethylsulphonyl fluoride (PMSF) to inhibit serine protease activity. For protein purification, cells were lysed by sonication (Q500, Qsonica, Newtown, CT) and bacterial debris was removed by centrifugation for 30 min at 5,000 x g at 4˚C. The supernatant was loaded via gravity flow onto a 15 ml chitin resin (NEB Biolabs Ipswich, MA) containing column. The column was washed with 20 column volumes of buffer A. Proteins were eluted from the column after incubation with 50 mM DTT dissolved in buffer A at 4˚C for 48 h then dialyzed into 20 mM tris buffer at a pH of 8.0 and further purified by high-performance liquid chromatography (HPLC) (AKTA, Cytiva Life Sciences, Marlborough, MA) using a 1 ml Resource Q anion exchange column (Cytiva Life Sciences, Marlborough, MA). *Af*LEA1.1 was eluted from the column using a slow gradient of 20 mM tris at pH 8.0 and NaCl from 0-125 mM. The purity of the protein was confirmed by SDS-PAGE and averaged at least 95%. The purified protein was then dialyzed into 50Mm phosphate solution at a pH of 7.0 and aliquots were snap frozen at -80 °C in advance of experiments.

Size Exclusion Chromatography

Size exclusion chromatography was performed using a HPLC (AKTA, Cytiva Life Sciences, Marlborough, MA) with a Superdex 200 10/300 column (Cytiva Life Sciences, Marlborough, MA). *Af*LEA1.1 fractions in 50 mM phosphate buffer at pH 7.0 were injected at a volume of 100 µl at a flow rate of 0.1 ml/min. Low molecular weight standards proteins column (Cytiva Life Sciences, Marlborough, MA) including aprotinin (6,500 da),

ribonuclease A (13,700 da), carbonic anhydrase (29,000 da), ovalbumin (44,000 da), conalbumin (75,000 da), and a blue dextran 2000 tracking polymer were injected in 100 µl volumes at a flow rate of 0.1Ml/min. To control for concentration-dependent behaviors such as liquid-liquid phase transitions, *Af*LEA1.1 fractions were loaded onto the Superdex 200 10/300 column at the same protein concentration as they were eluted from the Resource Q column.

Circular Dichroism

*Af*LEA1.1 at 100 µg/ml in 20 mM phosphate buffer at a pH of 7.0 was measured using a J-1500 circular dichroism spectrophotometer (Jasco, Easton, MD). Hydrated *Af*LEA1.1 was plated into a sealed quartz cuvette with a path length 0.1 cm (Starna Scientific, Atascadero, CA) and measured using a wavelength range from 280 nm to 185 nm. To reduce the light-scattering associated with a dried sample, *Af*LEA1.1 was dialyzed into ultrapure water three times at a ratio of 1:1,000 protein solution to water. The *Af*LEA1.1 at 1mg/ml in water was repeatedly plated on an open 0.01cm path-length quartz cuvette (Starna Scientific, Atascadero, CA), each layer was rapidly dried by incubating them for one hour at 0% relative humidity induced by anhydrous calcium sulfate. The rapid drying produced an amorphous protein glass with limited light scattering, and each subsequent layer of protein precipitated onto the last in a seamless column of *Af*LEA1.1. Once the protein column spanned the 0.01 cm path length, 1 µl of ultrapure water was used to dissolve the top layer of *Af*LEA1.1, allowing the cuvette to be assembled and nearly sealed. A final incubation for 24 h at 25 °C at 0% relative humidity produced a perfect 0.01cm column of *Af*LEA1.1 without any apparent light scattering or concave, coffee ring shape

that would warp the beam of the spectrophotometer. Data were averaged over 5 measurements and measurements were taken in 1nm intervals. Secondary structure predictions were performed using the CONTIN and SELCON 3 predictors from DichroWeb, using data sets 4 and 7 as references.

Bioinformatics Structural Predictions

I-Tasser predicts the structure of a query sequence by comparing portions of crystal structures from the PDB with similar sequences as portions of the query protein and combining them into models using a hierarchical ranking system (*240-242*). To visualize the hydrophobic face of *Af*LEA1.1's α-helices, the protein modeling software Swiss PDBViewer (DeepView), which initiates all protein models as 100% continues α-helices (*300*). To investigate the possibility of liquid-liquid phase separation, the amino acid sequence of *Af*LEA1.1 was evaluated using catGranule (*301*).

Light Microscopy Sample Preparation

*Af*LEA1.1 was dialyzed into ultrapure water three times at a ratio of 1:1,000 protein solution volume to water at 4 °C to remove salts. Next, 15 µl of *Af*LEA1.1 was plated onto microscope slides directly adjacent to a 15 µL droplet of ultrapure water. The droplets were connected to allow diffusion of the protein and produce a protein gradient over 10 minutes, then incubated at ambient relative humidity at 25˚C. The samples were then incubated again at 0% relative humidity induced by anhydrous calcium sulfate at 25˚C to remove any residual water. Glass coverslips were placed above the samples and sealed with nail polish

after desiccation to preserve the samples. The samples were stored in a sealed container at 0% relative humidity induced by anhydrous calcium sulfate until they were viewed using a specimen microscope.

Electron Microscopy Sample Preparation

*Af*LEA1.1 was dialyzed into ultrapure water three times at a ratio of 1:1,000. 15 µl of *Af*LEA1.1 was plated onto aluminum scanning electron microscope (SEM) stages directly adjacent to a 15µL droplet of ultrapure water. The droplets were connected to allow diffusion of the protein and produce a protein gradient over 10 minutes, then incubated at ambient relative humidity at 25˚C. The samples were then incubated again at 0% relative humidity induced by anhydrous calcium sulfate at 25˚C to remove any residual water. Once completely dried, the samples were sputter coated with a 10nm layer of gold to prevent sample rehydration and to prevent charging artefacts. The samples were stored in a sealed container at 0% relative humidity induced by anhydrous calcium sulfate until they were viewed using an electron microscope (Jeol, Peabody, MA).

AfLEA1.1 preservation of LDH activity during desiccation in cell lysate

To investigate the ability of *Af*LEA1.1 to protect proteins in the crowded cellular milieu during desiccation, cell lysates were obtained by sonication of Kc167 cells from *Drosophila melanogaster*. Lysates were diluted to a total protein concentration of 2 mg/ml with 100 mM phosphate buffer at a pH of 6.4. This concentration was quantified via Bradford assay. *Af*LEA1.1 or BSA were then added to the lysate to yield a total concentration of 400 µg/ml of *Af*LEA1.1 or BSA and the initial LDH activity was recorded. As a negative control, additional phosphate buffer was added to a set of lysate samples to match the final lysate protein concentrations of the experimental groups. 50 µl aliquots

were desiccated for 7 days in a sealed container at 25 °C and at 0% RH induced by anhydrous calcium sulfate. Samples were rehydrated with 100 µl phosphate buffer and activity was measured using UV-Vis spectrophotometry via a Shimadzu UV-1800 (Shimadzu, Tokyo, Japan).

AfLEA1.1 preservation or repair of purified LDH during desiccation

To distinguish between the protection of LDH and the repair of LDH by *Af*LEA1.1, pure LDH was desiccated in the presence or absence of *Af*LEA1.1 or BSA in 100 mM sodium phosphate buffer at a pH of 6.4. Pure LDH (EC 1.1.1.27) at 0.2 mg/ml was dialyzed against phosphate buffer (100 mM NaPO4, pH 6.5). To test for protection against desiccation-induced damage, *Af*LEA1.1 or BSA were added to the LDH sample at a concentration of 400 µg/ml prior to desiccation. Initial LDH activity was determined and 25 µl aliquots were placed in microtubes and desiccated for 7 days at 0% RH, then samples were rehydrated with 50 µl of phosphate buffer. To distinguish between protection and repair mechanisms, additional samples of purified LDH were desiccated in absence of either *Af*LEA1.1 or BSA, and desiccated for 7 days at 0% RH, then samples were rehydrated on ice with 50 µl of phosphate buffer containing *Af*LEA1.1 or BSA to produce a final added-protein concentration of 400 µg/ml. LDH activity was measured using UV-Vis spectrophotometry via a Shimadzu UV-1800 (Shimadzu, Tokyo, Japan).

Measuring LDH Activity

LDH activity was measured by diluting 50 µl of the sample into 2.9 ml of 50 mM phosphate buffer and adding an additional 50 µl of 12 mM NADH, producing a final volume of 3 ml.

The samples were then measured at a wavelength of 340 nm in kinetics mode while the sample was stirred at 500 rpm. When the NADH was fully expended, 30 µl of 100 mM sodium pyruvate was added to the samples. The sample temperature was actively maintained at 25 °C during enzyme activity measurements by water-based temperature control attachment.

Screening AfLEA1.1 for RNA-induced liquid-liquid phase separation

To determine whether *Af*LEA1.1 undergoes LLPS during desiccation, a solution resembling the cytoplasm of the diapause cysts of *A. franciscana* (32 mM NaCl, 98 mM KCl, 11 mM K_2PO_4, 5 mM $CaCl_2$, 340 mM trehalose, 2.9 % w/v glycerol, and 25% Ficoll 400, pH of 6.5) was prepared with an initial concentration of 150 µg/ml of *Af*LEA1.1 measured via Bradford assay. The protein was then plated onto microscope slides as previously done to detect protein crystals, then the solutions were allowed to desiccated at ambient humidity (83% RH) at ambient temperature (27 °C) under constant surveillance using an inverted microscope. To determine if RNA induces the LLPS of *Af*LEA1.1, 2 µl of mRNA from *A. franciscana* was added directly to the plated droplet, resulting in a total volume of 17 µl.

RESULTS AND DISCUSSION

Protein Cloning, Expression, and Purification

*Af*LEA1.1 was readily expressed in *E. coli* and no complications were encountered utilizing the IMPACT system to purify protein. When further purified using anion exchange chromatography, *Af*LEA1.1 elutes in two distinct fractions (Fig. 24A). The overall purity

of each elution fraction is over 99%, so this purification method was deemed satisfactory (Fig. 24B). However, the presence of two *Af*LEA1.1 elution peaks could indicate that there are two distinct protein modifications, higher order structures, or conformational differences between these proteins.

Previously published bioinformatics reported that *Af*LEA1.1 should be highly conformationally plastic, which supports the hypothesis that it exists in more than one conformation on the Resource Q anion exchange column. However, the binding behavior to the column is also peculiar. The concentration required for *Af*LEA1.1 to elute from the column for both fractions is negatively correlated with the quantity of *Af*LEA1.1 bound to the column (Fig. 25). Despite this shift, the same concentration of NaCl is required to fully elute each fraction, but this elution behavior indicates that *Af*LEA1.1 is not only present in multiple conformations, but that there are interactions among *Af*LEA1.1 molecules on the column that change its affinity for the matrix.

To verify that there are no higher-order structures or conformational transitions that are undetectable using SDS PAGE, the effective molecular weight of *Af*LEA1.1 from each elution fraction during anion chromatography was measured using size exclusion chromatography (Fig. 26.) Both fractions of *Af*LEA1.1 elute from the column between conalbumin (43,000 Da) and carbonic anhydrase (29,000 Da), and the area under the curve suggests that *Af*LEA1.1 has an effective molecular weight of 33kd, which is ~75% larger than a globular protein of the same sequence length. However, intrinsically disordered proteins can have apparent molecular sizes up to 12-times as large as a similarly sized globular protein (*33*). Therefore, *Af*LEA1.1 is similar to a molten globular protein, which characteristically have some secondary and tertiary structure, but no rigid tertiary structure,

and can have an effective molecular weight of up to twice that of a globular protein of similar sequence length. These results further suggest that *Af*LEA1.1 undergoes folding in the hydrated state, but it does not explain the distinct elution fractions nor does it explain the change in elution behavior with protein concentration. The concentration of *Af*LEA1.1 did not affect its effect molecular weight so the concentration of *Af*LEA1.1 is not inducing an appreciable change in its conformational ensemble in solution.

AfLEA1.1 Structural Analysis

To investigate the possibility that *Af*LEA1.1 undergoes a conformational transition from an inactive, partially disordered protein in the hydrated state into a more ordered, functional state during desiccation, the secondary structure of *Af*LEA1.1 was measured under various conditions using circular dichroism (Fig 27). In the hydrated state, *Af*LEA1.1 is composed of 5% α-helices, 35% β-sheets, 18% turns, and 42% random coils, which is approximately 40% ordered. While this means that *Af*LEA1.1 is definitively classified as an intrinsically disordered protein like other LEA proteins, it is also surprisingly ordered in the hydrated state. *Afr*LEA2 and *Afr*LEA3m, two group 3 LEA proteins also found in *A. franciscana*, are only 21% and 25% ordered, respectively (*29*). In the desiccated state, AfLEA1.1 undergoes a dramatic conformational transition (Fig. 27) and the protein is 100% ordered and composed of 85% α-helices, 5% β-sheets, 10% turns. This conformational transition is the most dramatic shift from disorder to order that has been reported for a LEA protein from *A. franciscana*, despite being the most ordered in the hydrated state. The transition from the hydrated conformation to the desiccated conformation is partly characterized by the molecular crowding effect of the 2% SDS (Fig. 27). With this molecular crowding,

*Af*LEA1.1 appears to become slightly less ordered (38%), and its β-sheets appear to transition into α-helices before the random coils eventually fold. Generally, the secondary structure measurements and predictions produced by circular dichroism represent the average conformation of an ensemble of different conformational state that were in the path of the beam. However, the structure of *Af*LEA1.1 is skewed towards α-helices and there is little room for different conformations that could produce. Therefore, it appears that *Af*LEA1.1 is transitioning into a consistent conformation, which is a strong indication of disorder-to-order regulation of its protective functions during desiccation.

The sequence of *Af*LEA1.1 is highly repetitive, and is mainly composed of the eight repeats of the group 1 consensus sequence 'GGOTRREQLGEEGYSQMGRK' (*85*) (Fig. 28A). *Af*LEA1.1's propensity for α-helices can be predicted by the regular appearance of alanine-arginine-alanine "helix cap" motifs found at the ends of the repeating group 1 LEA motif (*183, 302, 303*). The primary structure of *Af*LEA1.1 is extremely repetitive and even the occasional substituted amino acid is generally replaced with one of similar chemical properties. Given the amount of predicted α-helical structure, an α-helical protein projecting is informative (Fig. 28B). Unlike *Afr*LEA2 and *Afr*LEA3m, which have stripes of positive-negative-positive residues stabilizing regions that appear to transition into α-helices during desiccation. These stripes have strong indications of structural function in these proteins because the alternation of charged residues can stabilize α-helices in the absence of water, which is the main force driving the compaction of the protein into this conformation (*304, 305*) *Af*LEA1.1 has relatively little organization of its charges in three-dimensional space (*42*), but does have a thin hydrophobic face, which can also stabilize α-helical formation in the absence of water, but the hydrophobic moment of the face is too

low to offer sufficient stabilization to explain *Af*LEA1.1's α-helix content measured in the desiccated state (Fig. 27) (*53, 306*).

Computational Interpretation of the Structure of AfLEA1.1 in the Desiccated State

Due to I-Tasser's use of x-ray crystallography data to predict protein structure, it is surprisingly effective at predicting the ordered structures of LEA proteins that undergo conformational transitions during desiccation (*240*). I-Tasser predictions are within 2% of each secondary structure content measured by xxxx, which is well within the range of the conformational ensemble of *Af*LEA1.1 (Figure. 28A). The helices predicted by I-Tasser appear to be organized into a large helical bundle composed of four helix-turn-helix structures, each helix being composed of a single group 1 LEA consensus sequence and a short spacer containing a histidine residue. The α-helices propagate from the alanine-arginine-alanine caps, which is in line with previous expectations, but the internal space of the helical bundle is highly enriched in positively charged residues and aromatic residues (Fig. 28B).

Although aromatic residues can interact with positively-charged residues due to their decentralized pi orbitals, the small number of aromatic residues does not appear to be sufficient to prevent the electrostatic repulsion of the α-helices (*307*). Despite this source of instability this structure is similar to a *de novo* designed protein built to emulate armadillo-repeat proteins (RCSB PDB ID: 5CWH), which has a remarkably similar primary, secondary, and tertiary structures (*308*) to *Af*LEA1.1. Armadillo repeat proteins are remarkably stable and are commonly involve in cell signaling and misfolded protein degradation due to their promiscuous binding behaviors (*309, 310*). The structure of

*Af*LEA1.1 appears to be stabilized by similar charged stripes to those found on *Afr*LEA2 and *Afr*LEA3m's helices, but these stripes are generated through the tertiary orientation of the helices with each other. (Fig. 30). If I-Tasser has accurately, or at least closely, predicted the tertiary structure of *Af*LEA1.1, then it would explain the lack of this formal charge motif in *Af*LEA1.1 which was described in previous work and could explain why *Af*LEA1.1 converts its β-sheets into α-helices during molecular crowding (*53*). The self-interactions shown in this structure offer an explanation for the elution behavior from the anion exchange column. If group 1 LEA motifs readily interact with other group 1 LEA motifs, then the α-helices of *Af*LEA1.1 may interface with the α-helices of other *Af*LEA1.1 molecules instead of engaging in the intramolecular interactions that localize many positive residues into the core of the protein. This self-interacting behavior, and presence of compatible charges and helix-forming residues that can interact with each other is similar to the seed maturation protein domain found in *Afr*LEA6, which also formed α-helical bundles and promoted liquid-liquid phase separation during desiccation (*245*).

Behavior of AfLEA1.1 during Desiccation

Although *Af*LEA1.1 appears to evenly distribute itself as an amorphous, glassy deposition during extremely fast desiccation, it undergoes crystallization when dried at over the course of hours, rather than minutes (Fig. 31A). This is not surprising considering that the process for generating crystals generally requires more time than it takes for the protein solution to dry at 0% RH. The formation of crystals does indicate that *Af*LEA1.1 assumes a consistent structure in the desiccated state, as previously hypothesized based on the circular dichroism data and structural predictions by I-Tasser. Scanning electron microscopy reveals,

however, that the crystals formed are of insufficient quality to employ x-ray crystallography to verify the structure of *Af*LEA1.1 in the desiccated state (Fig. 31B). However, it also reveals additional, spherical structures not unlike those observed in dried *Afr*LEA6 samples (*245*).

The program catGranule predicts that *Af*LEA1.1 has exceptionally high propensity to undergo liquid-liquid phase separation, particularly when interacting with RNA (Fig. 32). To determine whether or not *Af*LEA1.1 undergoes liquid-liquid phase separation under a physiologically relevant context, desiccation experiments were conducted in a buffer system that resembles the cellular environment of the encysted embryos of *A. franciscana* termed "osmosome" buffer. At extreme levels of desiccation, *Af*LEA1.1 does appear to undergo LLPS, but this is at nearly complete desiccation and well below 20% g water/g dry weight that is associated with rapid cell death (Fig. 33A). When mRNA from *A. franciscana* is added to *Af*LEA1 samples in osmosome buffer before drying, the protein undergoes a liquid-liquid phase separation after only minor water removal by evaporative drying (Fig. 33B). These results indicate that *Af*LEA1.1 may form an anhydrobiosis-related membraneless organelle in a similar fashion to *Afr*LEA6 (*245*). However, more experiments are required to verify that mRNA is actually selected for incorporation into the *Af*LEA1.1 liquid phase and that the protein offers any kind of protection against mRNA degradation.

AfLEA1.1 Protection of LDH Activity during Desiccation and Rehydration

Although there are indications that *Af*LEA1.1 is involved in RNA stabilization, one of the main hypotheses for LEA protein function is the protection of enzyme function in the dried

state. To test *Af*LEA1.1's ability to prevent enzyme degradation during desiccation or to repair enzyme damage upon rehydration, LDH was desiccated in the presence of *Af*LEA1.1 or BSA (Figure 34). BSA is often used as a negative control for protection against desiccation-induced damage in other LEA protein studies because its promiscuous binding behavior and general hardiness tends to confer protection against protein denaturation under wide variety of conditions. Kc167 cell lysates was used instead of pure protein for this experiment because LDH can denature on contact with the plastic of the microcuvette during desiccation, rather than denaturing due to the lack of water. *Af*LEA1.1 maintained LDH activity after one week of desiccation followed by rehydration, whereas BSA did not (Fig. 34). This indicates that *Af*LEA1.1 does offer some specialized protection against desiccation-induced protein denaturation. However, this preservation of enzyme activity could be due to chaperone-like refolding behavior.

To test the mechanism of enzyme protection, purified LDH was desiccated in the presence or absence of *Af*LEA1.1 or BSA. Samples that were not dried with *Af*LEA1.1 or BSA were instead either rehydrated with phosphate buffer containing either of the proteins, or were just rehydrated with buffer as a negative control. *Af*LEA1.1 preserved nearly 100% of LDH enzyme activity when desiccated together with LDH, but did not significantly affect LDH activity when it was only included in the rehydration solution (Fig. 35). BSA offered lesser protection against desiccation-induced loss of LDH activity when desiccated together with LDH but, again, rehydration with BSA did not significantly preserve LDH activity after rehydration. This indicates that *Af*LEA1.1 is conferring desiccation tolerance through interactions either during drying or in the desiccated state, rather than acting as a

protein-refolding chaperone, and that this protection is greater than what it conferred by general protein-protein interactions.

CONCLUSIONS

*Af*LEA1.1 is the only LEA protein that demonstrably drives desiccation tolerance in the anhydrobiotic cysts of *A. franciscana*. As a group 1 LEA protein, *Af*LEA1.1's repetitive structure appears to result in very consistent folding behaviors during desiccation. Furthermore, these group 1 motifs appear to act as multivalence sites for protein-protein and protein-RNA interactions that induce liquid-liquid phase separation. The α-helical propensity of these motifs also appears to drive a massive conformational transition during desiccation that it initiated by mild molecular crowding with SDS. The cellular space of a cell undergoing desiccation becomes intensely crowded with biomolecules such as proteins, nucleic acids, and lipids. The concentration of monovalent and divalent cation also increase, resulting in a very different physicochemical environment from the hydrated cell. Behaviors such as "salting out" which can cause proteins to separate from solution are likely common-place, and protein-protein interactions become unavoidable (*209*). The discovery of another LEA protein motif potentially driving LLPS during water loss is an exciting prospect. However, associating this behavior with a mechanism of desiccation tolerance is premature.

Further research into the partitioning abilities of *Af*LEA1.1 are needed, and *in vitro* observation of the *Af*LEA1.1 undergoing liquid-liquid phase separtion is critical to establishing this mechanism with its ability to protect biomolecules during desiccation. The indications that *Af*LEA1.1 interacts with mRNA are strong due to the more rapid phase

separation of *Af*LEA1.1 and Mrna together than *Af*LEA1.1 alone. However, *Af*LEA1.1 clearly also confers desiccation tolerance to proteins such as LDH, which have very different structural characteristics. It may be that *Af*LEA1.1 offers mutiple modes of protection against desiccation-induced damage at different cellular hydration levels.

Figure Captions

Figure 24: Purification of *Af*LEA1.1. A) AfLEA1.1 binds quaternary ammonium groups at a pH of 8.0 and elutes at two distinct with low concentrations of NaCl (black) in two distinct fractions (blue). B) SDS Page shows that breakdown products left over from the chitin column are removed in the eluted fractions. Both fractions contain *Af*LEA1.1 at the expected molecular weight of 18 k d.

Figure 25: Concentration-dependent elution of *Af*LEA1.1. The concentration of *Af*LEA1.1 negatively correlates with the NaCl concentration required to initiate elution. The end of the elution fraction is not concentration-dependent.

Figure 26: Size exclusion chromatography of *Af*LEA1.1 fractions. Both fractions of *Af*LEA1.1 elute from a size exclusion column between standards (2) and (3), conalbumin (43,000 Da) and carbonic anhydrase (29,000 Da), respectively. A calibration curve predicts that the molecular weight *Af*LEA1.1 to be 33,400 Da, or approximately 75% larger than the protein**.**

Figure 27: Circular dichroism measurements of secondary structure and conformational shifts of *Af*LEA1.1 during desiccation. In the hydrated state, the secondary structure of *Af*LEA1.1 was an average of 5% α-helices, 35% β-sheets, 18% turns, and 42% random coils. In the completely desiccated state, the secondary structure of *Af*LEA1.1 was an

average of 85% α-helices, 5% β-sheets, 10% turns, and 0% random coils. In the presence of 2% SDS, the secondary structure of *Af*LEA1.1 was an average of 25% α-helices, 13% β-sheets, 16% turns, and 46% random coils.

Figure 28: Sequence Features of *Af*LEA1.1. A) The amino acid sequence of *Af*LEA1.1 is highly repetitive and consists of four group 1 LEA domains, represented as two rows of the repetitive sequence motifs. B) Projected as an α-helix, *Af*LEA1.1 has a distinct hydrophobic face, but does not have distinctly charged amino acid stripes. Amino acid properties are labeled by color representing polar (yellow), nonpolar (grey), positive (red), and negative (blue) amino side chains.

Figure 29: I-Tasser prediction of *Af*LEA1.1 structure in the desiccated state with amino acid properties labeled by color as polar (yellow), nonpolar (gray), aromatic (green), positive (red), and negative (blue). A) *Af*LEA1.1 is predicted to fold into 84% α-helix, 5% β-sheet, and 11% turns. The tertiary structure of *Af*LEA1.1 is similar to a synthetic helical repeat protein composed of helix-turn-helix structures, where each helix-turn-helix is a single group 1 LEA motif. B) The exterior of the *Af*LEA1.1 is stabilized by the distribution of charged amino acids, but the interior of the protein is saturated with positive residues and aromatic residues.

Figure 30: Three-dimensional charge distribution of *Af*LEA1.1. Charge is represented by color as positive (red) and negative (blue). A) Predicted α-helical regions of *Afr*LEA2 were

shown to have characteristic positive-negative-positive residue stripes (*42*). B) Predicted α-helical regions of *Afr*LEA3m were also shown to have characteristic positive-negative-positive residue stripes (*42*). C) *Af*LEA1.1 does not present this pattern in its secondary structure, but its tertiary structure presents adjacent stripes of alternating formal charge (positive-negative-positive-negative) on the protein surface.

Figure 31: *Af*LEA1.1 crystallized readily when completely desiccated in ultrapure water at 80% RH. A) Light microscopy shows that *Af*LEA1.1 forms branching crystals. B) Scanning electron microscopy reveals that these crystals are of insufficient quality for x-ray crystallography and anomalous spherical structures are present in the sample.

Figure 32: *Af*LEA1.1 has a very high propensity towards liquid-liquid phase separation in the presence of RNA. A) A cumulative distribution fraction analysis of the amino acids of *Af*LEA1.1 produces a propensity score of 3.05. A score of 1 or greater is a predictor of LLPS behavior. B) The residue-level propensity of *Af*LEA1.1 to undergo LLPS, where values above 0 indicate increasing likelihood of undergoing LLPS in the presence of RNA.

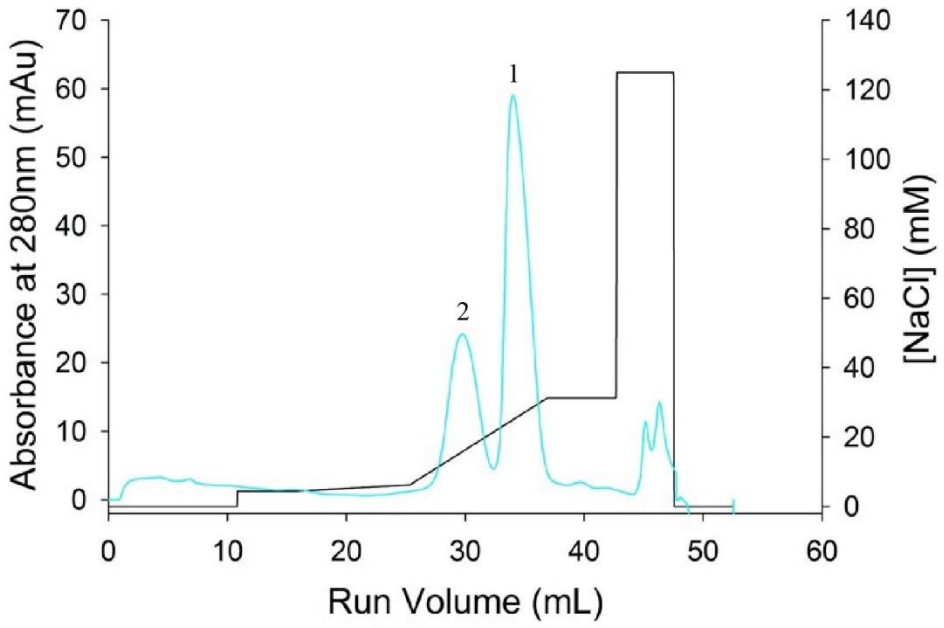
Figure 33: *Af*LEA1.1 undergoes LLPS during desiccation *in vitro*. A) When desiccated in simulated anhydrobiotic *A. franciscana* embryonic cytoplasm in the absence of RNA, *Af*LEA1.1 undergoes LLPS at the maximum concentration of NaCl. B) When desiccated in simulated anhydrobiotic *A. franciscana* embryonic cytoplasm in the presence of mRNA from *A. franciscana*, *Af*LEA1.1 rapidly undergoes LLPS at high water contents.

Figure 34: *Af*LEA1.1 partially preserves LDH activity after desiccation and rehydration in Kc167 cell lysate. (n = 3-4; ±SD, $p<0.05$; *different from control). BSA did not significantly preserve LDH activity.

Figure 35: *Af*LEA1.1 completely preserves purified LDH activity and BSA (dark grey) partially preserves LDH activity after desiccation and rehydration in 100Mm sodium phosphate buffer at pH 6.5 (light grey). Neither protein significantly preserved LDH activity when only added as part of the rehydration solution n = 3-4; ±SD, $p<0.05$; letter groups for significant difference).

Figure 24

A


Absorbance at 280nm (mAu)
[NaCl] (mM)
Run Volume (mL)
1
2
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140

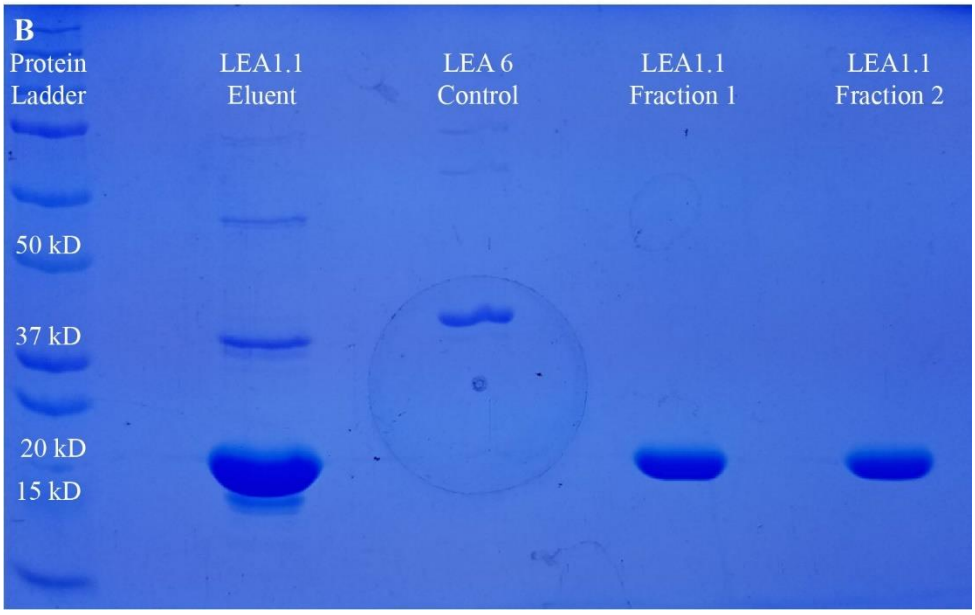

B
Protein Ladder
LEA1.1 Eluent
LEA 6 Control
LEA1.1 Fraction 1
LEA1.1 Fraction 2
50 kD
37 kD
20 kD
15 kD

Figure 25

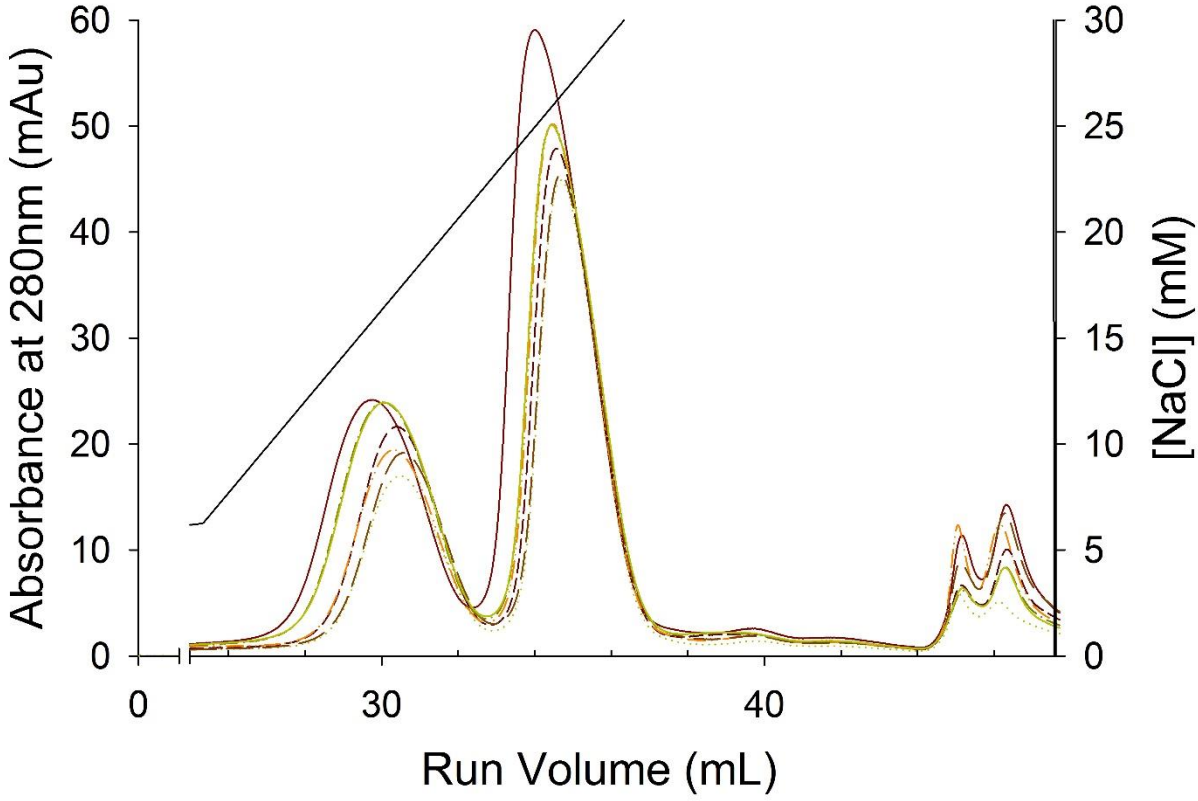

Absorbance at 280nm (mAu)
60
50
40
30
20
10
0
[NaCl] (mM)
30
25
20
15
10
5
0
0
30
40
Run Volume (mL)

Figure 26

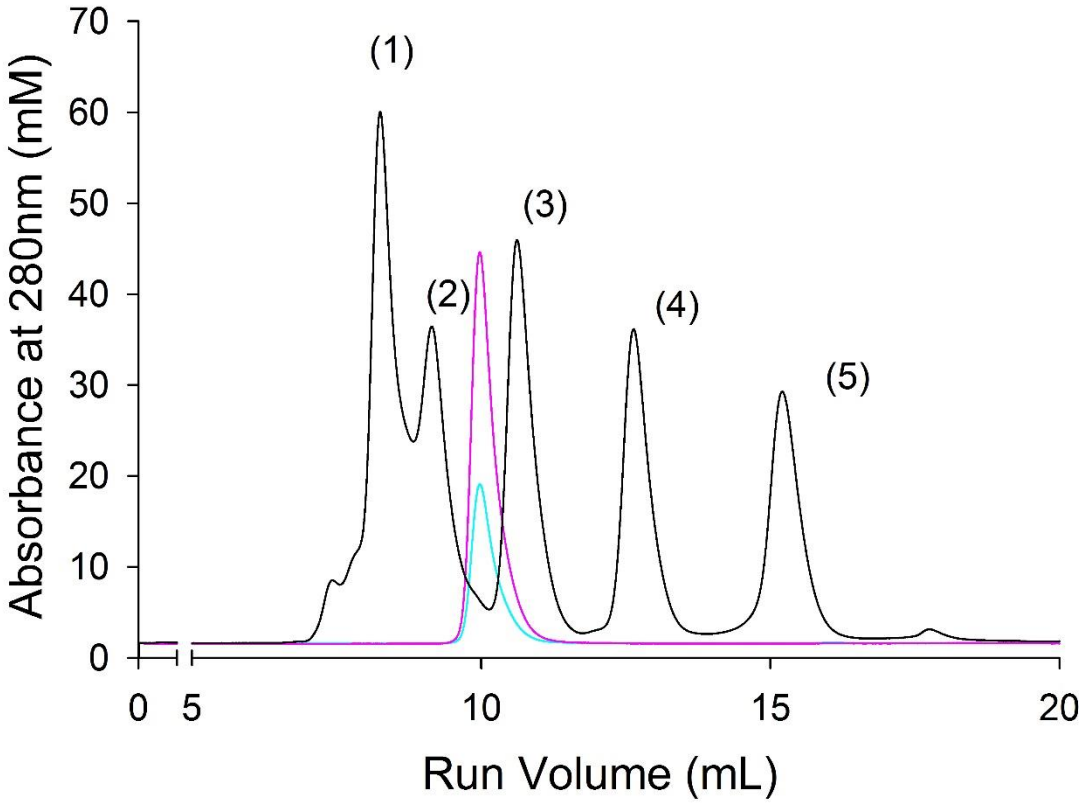

70
60
50
40
30
20
10
0
Absorbance at 280nm (mM)
(1)
(2)
(3)
(4)
(5)
0
5
10
15
20
Run Volume (mL)

Figure 27

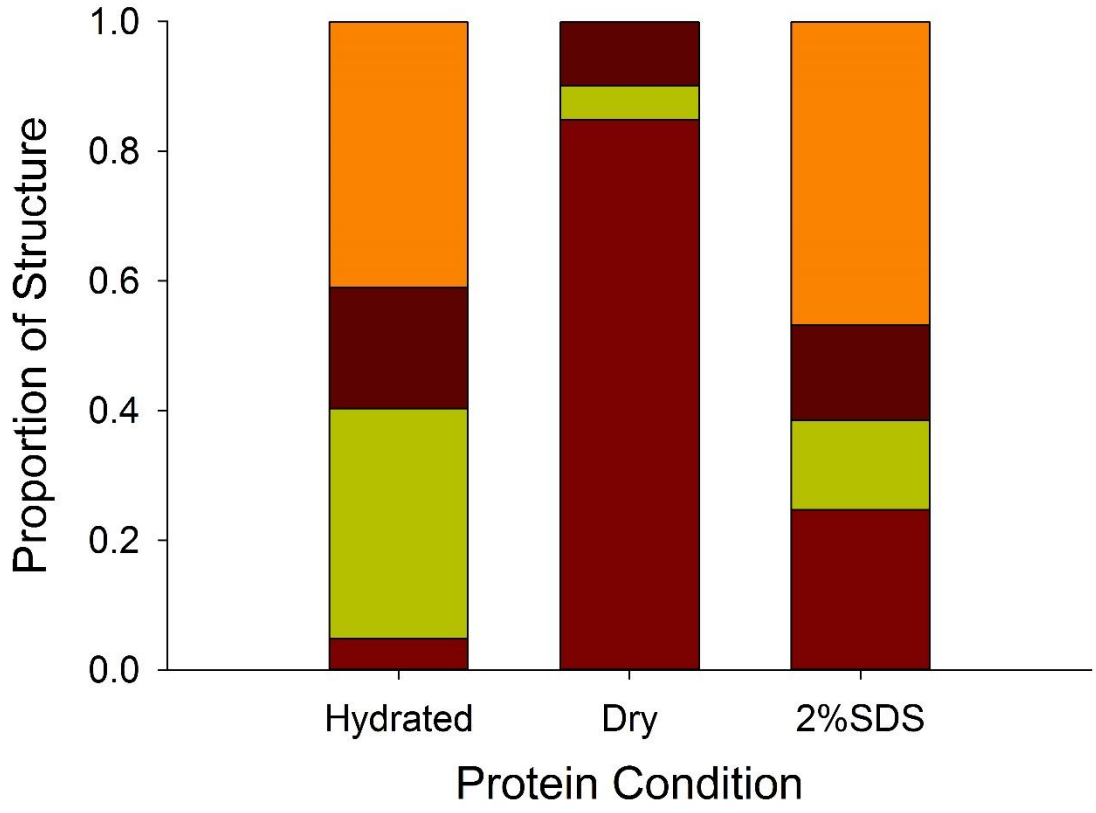

1.0
0.8
0.6
0.4
0.2
0.0
Proportion of Structure
Hydrated
Dry
2%SDS
Protein Condition

Figure 28

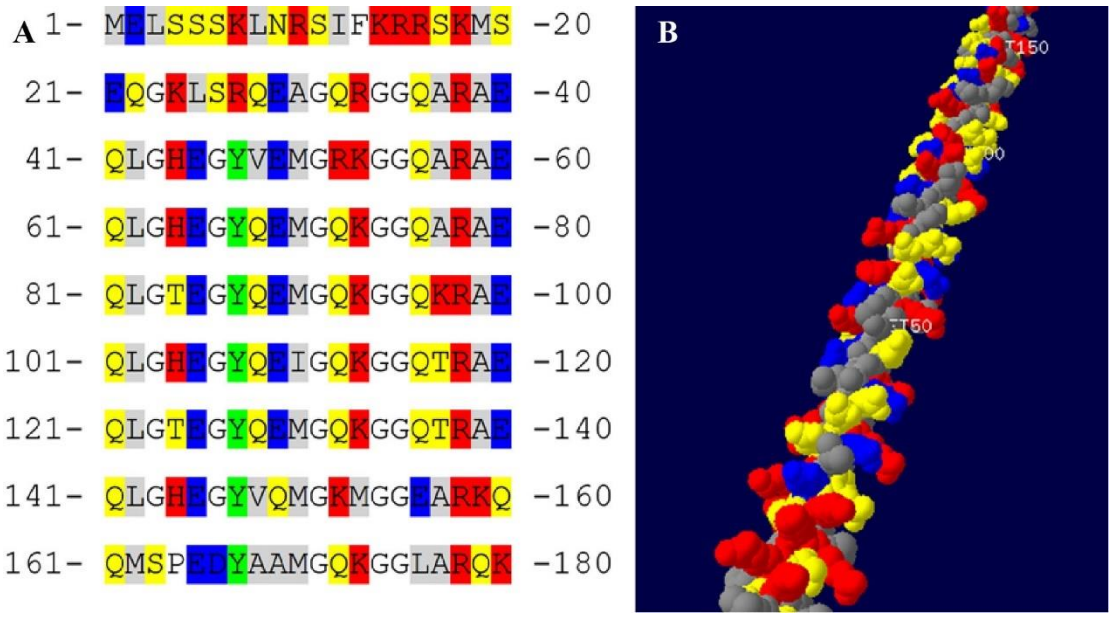

A
1- MELSSSKLNRSIFKRRSKMS -20
21- EQGKLSRQEAGQRGGQARAE -40
41- QLGHEGYVEMGRKGGQARAE -60
61- QLGHEGYQEMGQKGGQARAE -80
81- QLGTEGYQEMGQKGGQKRAE -100
101- QLGHEGYQEIGQKGGQTRAE -120
121- QLGTEGYQEMGQKGGQTRAE -140
141- QLGHEGYVQMGKMGGEARKQ -160
161- QMSPEDYAAMGQKGGLARQK -180
B
T150
T50

Figure 29

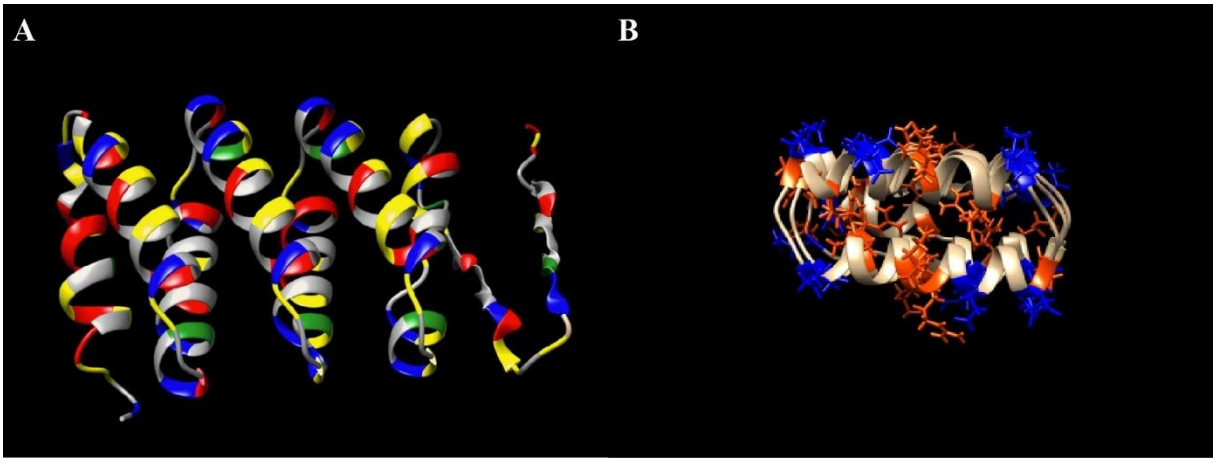


Figure 30

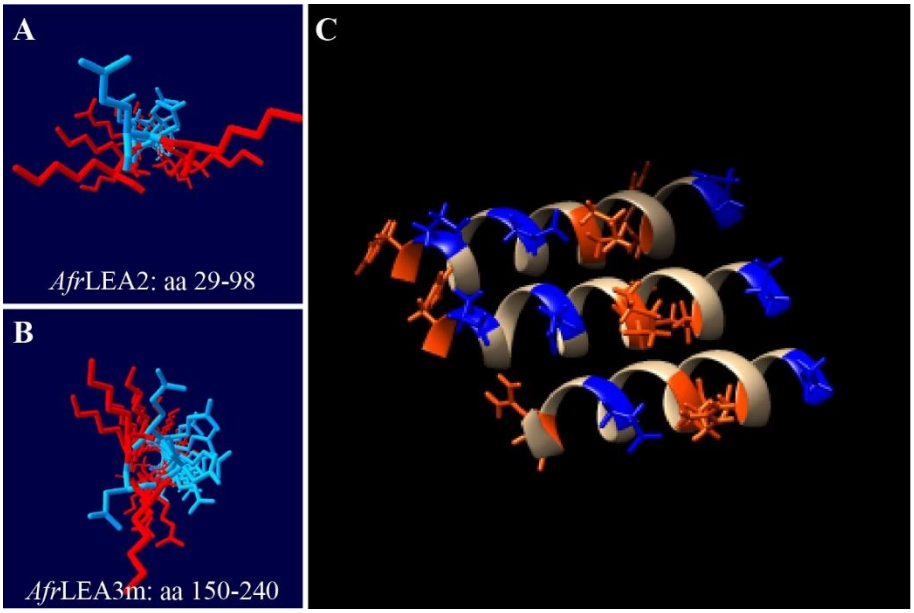


Figure 31

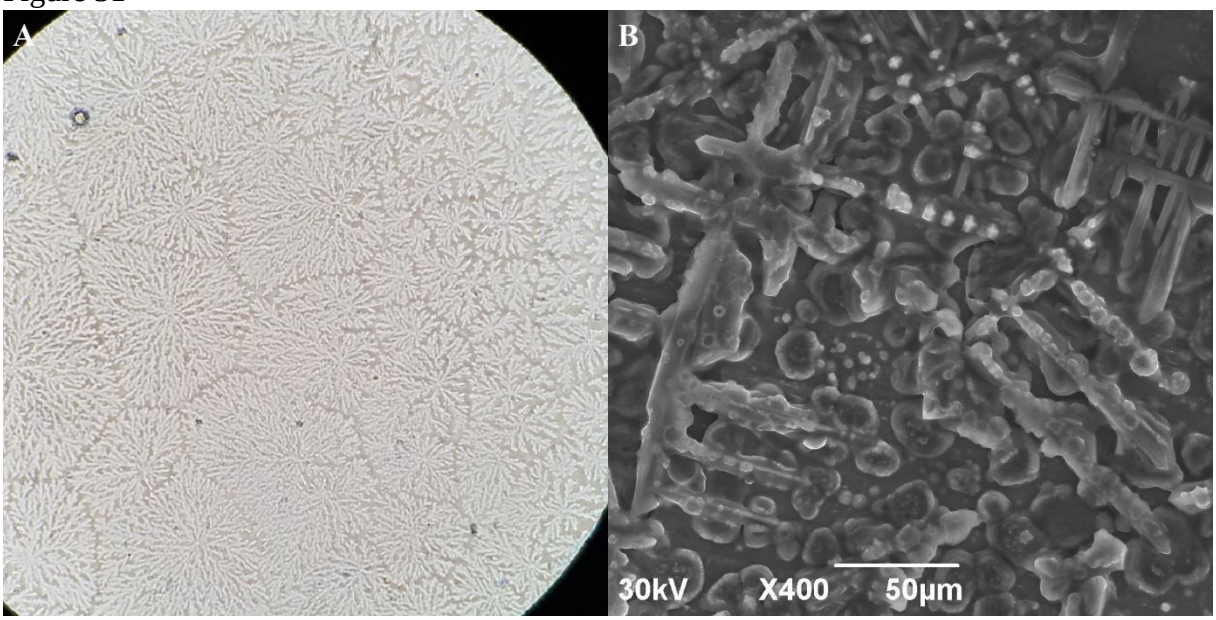


Figure 32

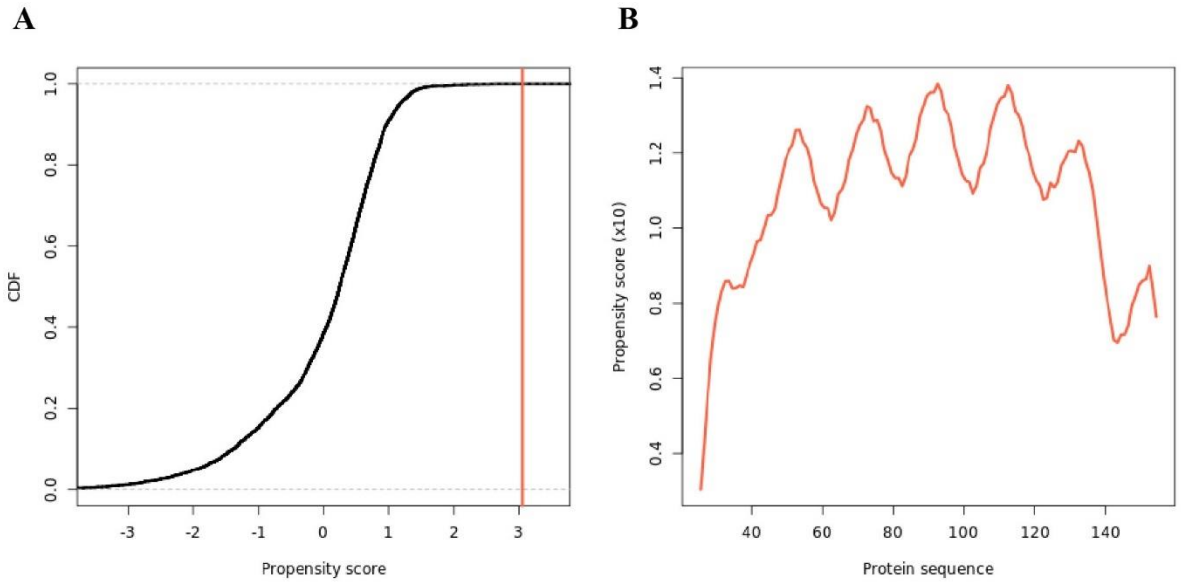

A
B
CDF
0.0
0.2
0.4
0.6
0.8
1.0
-3
-2
-1
0
1
2
3
Propensity score
Propensity score (x10)
0.4
0.6
0.8
1.0
1.2
1.4
40
60
80
100
120
140
Protein sequence

Figure 33

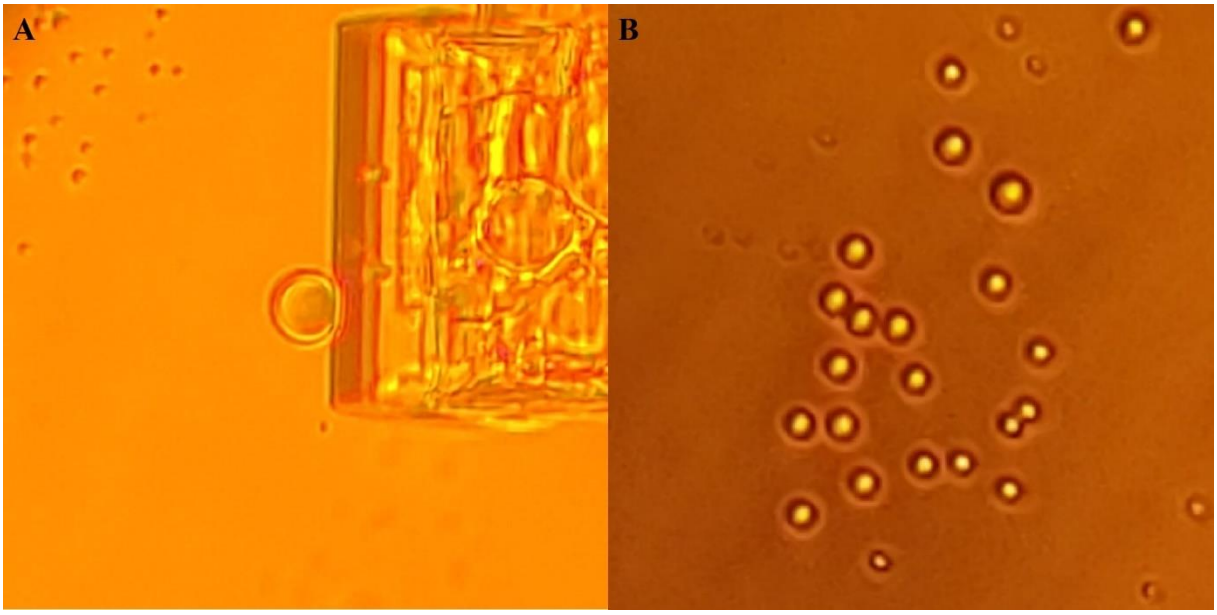


Figure 34

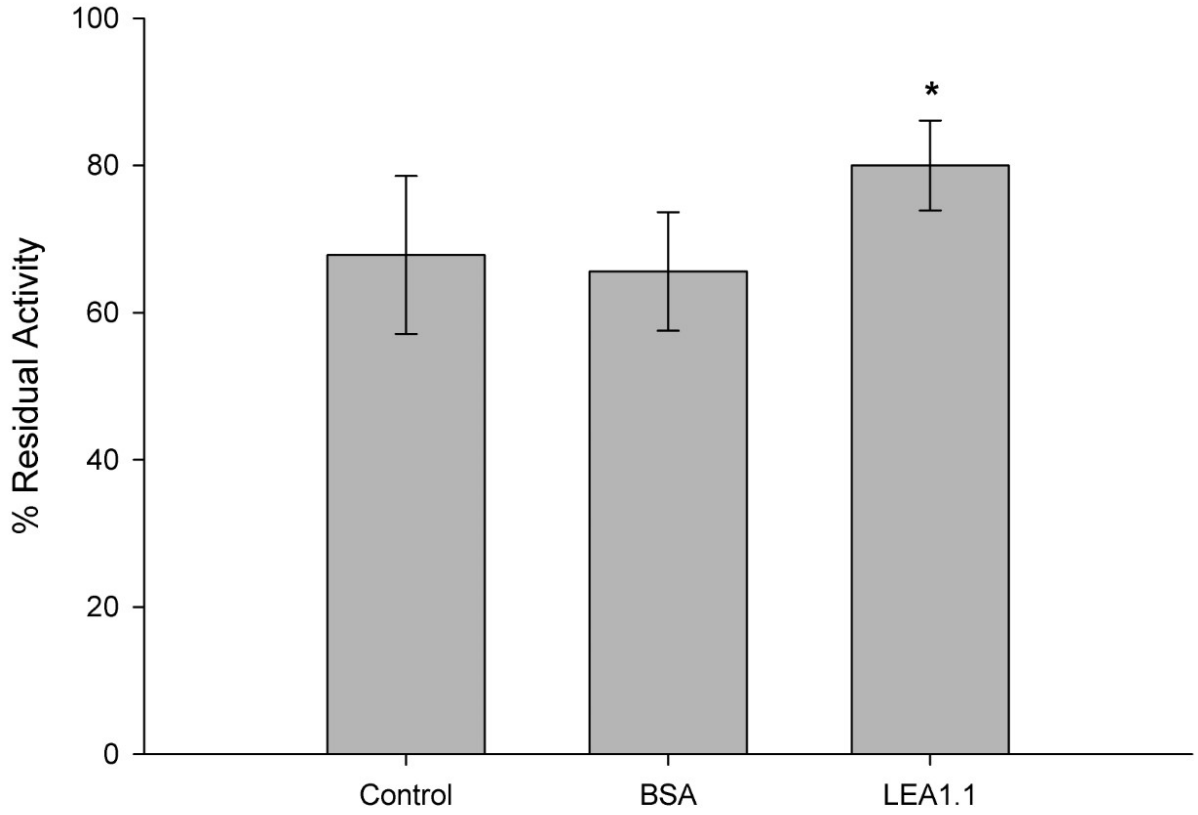
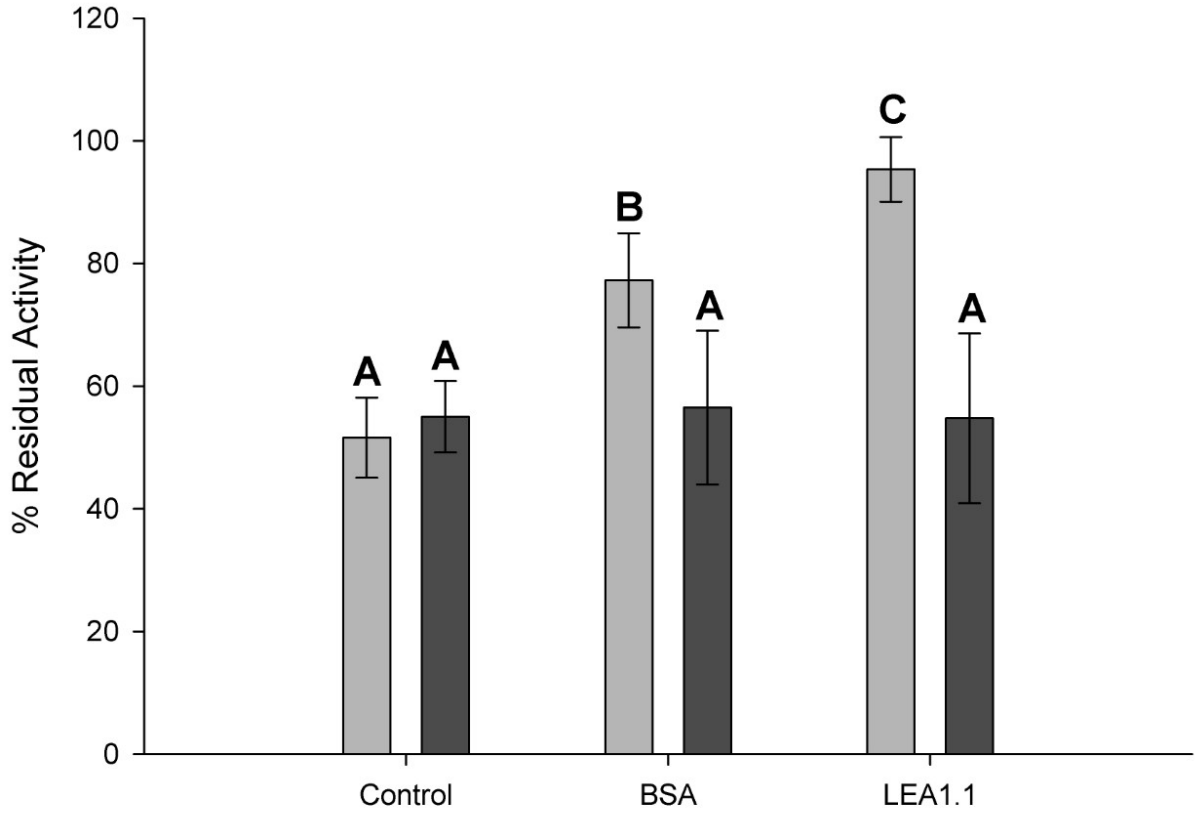


Figure 35


120
100
80
60
40
20
0
% Residual Activity
A
A
B
A
C
A
Control
BSA
LEA1.1

CHAPTER VI

SUMMARY AND FUTURE DIRECTIONS

This dissertation examined the properties and functions of anhydrobiosis related, intrinsically disordered proteins found within the kingdom Animalia. Although there are other families of desiccation tolerance proteins, such as tardigrade proteins, late embryogenesis abundant (LEA) proteins are the most well documented. By utilizing the group 1 and group 6 LEA proteins that are uniquely expressed uniquely among animals by *A. franciscana*, the role of multivalence has been characterized as the driver of phase transitions during desiccation that recontextualize the previously conceived notions of protein function in the dried state.

*Afr*LEA6, for example, combines several seemingly contradictory mechanisms of function. At high water contents, it separates from solution as a liquid phase and partitions positively charged proteins within it. The interior of the liquid *Afr*LEA6 "membraneless organelle" will be made up of a large number of sterically-isolating amino acid polymers that, as described in the molecular shielding hypothesis, should prevent aggregation. The contribution of target protein to the structure and size of the MLO addresses the main concern that no LEA proteins are expressed at high enough concentrations to protect a large number of proteins even if they are paired only 1:1 with their targets. The hydration buffer hypothesis, for example, does not align with the vitrification hypothesis because it suggests the retention of water, thereby reducing the glass transition temperature of trehalose. However, the change of the order of water during desiccation meshes well with the phase

transitions from liquid phase *Afr*LEA6 to gel phase and eventually to final glassy state. In this case, the liquid phase would be composed of a hydration-buffer style re-ordering of water, and the gel transition occurs when this ordered water is removed during extreme desiccation. The liquid phase of *Afr*LEA6 likely incorporates water in some ordered form, however, this will need to be verified using specialized techniques such as confocal Raman spectroscopy. However, when water is forcefully removed from the *Afr*LEA6 MLO, *Afr*LEA6 undergoes a conformation transition that reduces its flexibility and increases it's the adherence of the seed maturation protein (SMP) domains to each other, producing a more rigid hydrogel. In the complete absence of water, *Afr*LEA6 appears to deposit as an amorphous solid similar to a bioglass or bio-ceramic. This material needs to be studied further with specialized techniques such as differential scanning calorimetry to determine if the glass transition temperature of *Afr*LEA6 is sufficient high to protect desiccation-sensitive proteins for prolonged periods of time.

The newly discovered liquid-liquid phase separation of *Af*LEA1.1 is in need of significant additional research. Early experiments may include charged GFP exclusion studies to see if *Af*LEA1.1 is specific to certain surface charges or if it exclusively interacts with RNA. Labeled RNAs, including mRNA, rRNA, and tRNA should be mixed with *Af*LEA1.1 to determine if it is specifically protecting a particular type of nucleotide polymer, such as single-stranded mRNA. Additional attempts to produce high-quality *Af*LEA1.1 crystals or measuring the secondary structure of *Af*LEA1.1 in the desiccated state using Fourier transformed infrared spectroscopy could verify the three-dimensional structure of *Af*LEA1.1 in the desiccated state. Additional gradients of crowding agents,

such as a ficoll gradient or increasing glycerol concentrations, may help identify at what stage of water loss various LEA proteins will activate.

The results of the experiments described in this dissertation support the hypothesis that groups of LEA proteins, despite their differentiation within their protein families, likely share some similar mechanisms of function. Most group 6 LEA proteins, for example, are likely to undergo liquid-liquid phase separation during drying. This should be further investigated using a variety of purified group 6 LEA proteins from plants and animals to determine if this is the functioning mechanism of conferring desiccation tolerance in this family of proteins. If so, then a new focus on identifying the specific targets of protection can begin, and the development of a tunable desiccation-protection MLO can start. This would not only have applications in plant drought resistance and medical protein stabilization technologies, but it could also advance the fledgling field of liquid-phase protein compartments as biochemical catalysts.

REFERENCES

1. O. Leprince, J. Buitink, Introduction to desiccation biology: from old borders to new frontiers. *Planta* **242**, 369-378 (2015).
2. S. C. Hand, M. A. Menze, Molecular approaches for improving desiccation tolerance: insights from the brine shrimp Artemia franciscana. *Planta* **242**, 379-388 (2015).
3. W. Wełnicz, M. A. Grohme, Ł. Kaczmarek, R. O. Schill, M. Frohme, Anhydrobiosis in tardigrades—The last decade. *Journal of Insect Physiology* **57**, 577-583 (2011).
4. F. A. Hoekstra, E. A. Golovina, J. Buitink, Mechanisms of plant desiccation tolerance. *Trends Plant Sci* **6**, 431-438 (2001).
5. J. H. Crowe, F. A. Hoekstra, L. M. Crowe, Anhydrobiosis. *Annu Rev Physiol* **54**, 579-599 (1992).
6. A. Tunnacliffe, J. Lapinski, Resurrecting Van Leeuwenhoek's rotifers: a reappraisal of the role of disaccharides in anhydrobiosis. *Philos T Roy Soc B* **358**, 1755-1771 (2003).
7. A. Van Leeuwenhoek. (dated 9 February 1702).
8. S. C. Hand, M. A. Menze, M. Toner, L. Boswell, D. Moore, LEA proteins during water stress: not just for plants anymore. *Annu Rev Physiol* **73**, 115-134 (2011).
9. P. Alpert, Constraints of tolerance: why are desiccation-tolerant organisms so small or rare? *J Exp Biol* **209**, 1575-1584 (2006).
10. R. Guidetti, K. I. Jonsson, Long-term anhydrobiotic survival in semi-terrestrial micrometazoans. *J Zool* **257**, 181-187 (2002).
11. J. H. Crowe, Anhydrobiosis: An Unsolved Problem with Applications in Human Welfare. *Subcell Biochem* **71**, 263-280 (2015).
12. J. S. Clegg, Metabolic studies of crytobiosis in encysted embryos of Artemia salina. *Comparative Biochemistry and Physiology* **20**, 801-809 (1967).
13. J. Clegg, Embryos of Artemia franciscana survive four years of continuous anoxia: the case for complete metabolic rate depression. *J Exp Biol* **200**, 467-475 (1997).
14. J. S. Clegg, THE ORIGIN OF TREHALOSE AND ITS SIGNIFICANCE DURING THE FORMATION OF ENCYSTED DORMANT EMBRYOS OF ARTEMIA SALINA. *Comp Biochem Physiol* **14**, 135-143 (1965).
15. J. S. CLEGG, FREE GLYCEROL IN DORMANT CYSTS OF THE BRINE SHRIMP ARTEMIA SALINA, AND ITS DISAPPEARANCE DURING DEVELOPMENT. *The Biological Bulletin* **123**, 295-301 (1962).
16. A. Tunnacliffe, M. J. Wise, The continuing conundrum of the LEA proteins. *Naturwissenschaften* **94**, 791-812 (2007).
17. S. Hengherr, A. G. Heyer, H. R. Kohler, R. O. Schill, Trehalose and anhydrobiosis in tardigrades--evidence for divergence in responses to dehydration. *Febs j* **275**, 281-288 (2008).
18. T. C. Boothby *et al.*, Tardigrades Use Intrinsically Disordered Proteins to Survive Desiccation. *Mol Cell* **65**, 975-984 e975 (2017).
19. C. Hesgrove, T. C. Boothby, The biology of tardigrade disordered proteins in extreme stress tolerance. *Cell Communication and Signaling* **18**, 178 (2020).
20. S. Chakrabortee *et al.*, Intrinsically disordered proteins as molecular shields. *Molecular bioSystems* **8**, 210-219 (2012).
21. J. M. Mouillon, P. Gustafsson, P. Harryson, Structural investigation of disordered stress proteins. Comparison of full-length dehydrins with isolated peptides of their conserved segments. *Plant Physiology* **141**, 638-650 (2006).

22. M. Hara, M. Fujinaga, T. Kuboi, Metal binding by citrus dehydrin with histidine-rich domains. *Journal of experimental botany* **56**, 2695-2703 (2005).
23. J. Svensson, E. T. Palva, B. Welin, Purification of recombinant Arabidopsis thaliana dehydrins by metal ion affinity chromatography. *Protein expression and purification* **20**, 169-178 (2000).
24. W. F. Wolkers, S. McCready, W. F. Brandt, G. G. Lindsey, F. A. Hoekstra, Isolation and characterization of a D-7 LEA protein from pollen that stabilizes glasses in vitro. *Biochimica et biophysica acta* **1544**, 196-206 (2001).
25. T. C. Boothby *et al.*, Tardigrades Use Intrinsically Disordered Proteins to Survive Desiccation. *Molecular Cell* **65**, 975-984.e975 (2017).
26. D. Li, X. He, Desiccation induced structural alterations in a 66-amino acid fragment of an anhydrobiotic nematode late embryogenesis abundant (LEA) protein. *Biomacromolecules* **10**, 1469-1477 (2009).
27. K. Sasaki, N. K. Christov, S. Tsuda, R. Imai, Identification of a novel LEA protein involved in freezing tolerance in wheat. *Plant & cell physiology* **55**, 136-147 (2014).
28. Z. Dosztányi, B. Mészáros, I. Simon, ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **25**, 2745-2746 (2009).
29. L. C. Boswell, M. A. Menze, S. C. Hand, Group 3 late embryogenesis abundant proteins from embryos of Artemia franciscana: structural properties and protective abilities during desiccation. *Physiol Biochem Zool* **87**, 640-651 (2014).
30. C. J. Jeffery, Moonlighting proteins: old proteins learning new tricks. *Trends in genetics : TIG* **19**, 415-417 (2003).
31. P. Tompa, P. Csermely, The role of structural disorder in the function of RNA and protein chaperones. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **18**, 1169-1175 (2004).
32. R. Hatanaka *et al.*, An abundant LEA protein in the anhydrobiotic midge, PvLEA4, acts as a molecular shield by limiting growth of aggregating protein particles. *Insect biochemistry and molecular biology* **43**, 1055-1067 (2013).
33. V. N. Uverskii, [How many molten globules states exist?]. *Biofizika* **43**, 416-421 (1998).
34. M. Bokor *et al.*, NMR relaxation studies on the hydrate layer of intrinsically unstructured proteins. *Biophysical journal* **88**, 2030-2037 (2005).
35. T. Chen, A. Fowler, M. Toner, Literature review: supplemented phase diagram of the trehalose-water binary mixture. *Cryobiology* **40**, 277-282 (2000).
36. T. Shimizu *et al.*, Desiccation-induced structuralization and glass formation of group 3 late embryogenesis abundant protein model peptides. *Biochemistry* **49**, 1093-1104 (2010).
37. J. Buitink, O. Leprince, Intracellular glasses and seed survival in the dry state. *Comptes rendus biologies* **331**, 788-795 (2008).
38. J. H. Crowe, J. F. Carpenter, L. M. Crowe, The role of vitrification in anhydrobiosis. *Annu Rev Physiol* **60**, 73-103 (1998).
39. H. Z. Cummins *et al.*, The liquid–glass transition in sugars: Relaxation dynamics in trehalose. *Journal of Non-Crystalline Solids* **352**, 4464-4474 (2006).
40. K. Goyal *et al.*, Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein. *The Journal of biological chemistry* **278**, 12977-12984 (2003).
41. A. Bremer, M. Wolff, A. Thalhammer, D. K. Hincha, Folding of intrinsically disordered plant LEA proteins is driven by glycerol-induced crowding and the presence of membranes. *The FEBS Journal* **284**, 919-936 (2017).

42. D. S. Moore, R. Hansen, S. C. Hand, Liposomes with diverse compositions are protected during desiccation by LEA proteins from Artemia franciscana and trehalose. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1858**, 104-115 (2016).
43. K. Nakayama *et al.*, Arabidopsis Cor15am is a chloroplast stromal protein that has cryoprotective activity and forms oligomers. *Plant Physiology* **144**, 513-523 (2007).
44. L. Y. Rivera-Najera *et al.*, A group 6 late embryogenesis abundant protein from common bean is a disordered protein with extended helical structure and oligomer-forming properties. *The Journal of biological chemistry* **289**, 31995-32009 (2014).
45. D. M. Mitrea, R. W. Kriwacki, Phase separation in biology; functional organization of a higher order. *Cell Communication and Signaling : CCS* **14**, 1 (2016).
46. C. P. Brangwynne *et al.*, Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science* **324**, 1729-1732 (2009).
47. H.-Y. Lee, K.-Y. Cheng, J.-C. Chao, J.-Y. Leu, Differentiated cytoplasmic granule formation in quiescent and non-quiescent cells upon chronological aging. *Microbial Cell* **3**, 109-119 (2016).
48. T. J. Nott *et al.*, Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol Cell* **57**, 936-947 (2015).
49. S. Wegmann *et al.*, Tau protein liquid–liquid phase separation can initiate tau aggregation. *The EMBO Journal* **37**, e98049 (2018).
50. J. A. Riback *et al.*, Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* **168**, 1028-1040.e1019 (2017).
51. O. Bounedjah *et al.*, Macromolecular crowding regulates assembly of mRNA stress granules after osmotic stress: new role for compatible osmolytes. *The Journal of biological chemistry* **287**, 2446-2458 (2012).
52. C. Rabouille, S. Alberti, Cell adaptation upon stress: the emerging role of membrane-less compartments. *Current opinion in cell biology* **47**, 34-42 (2017).
53. B. Janis, V. N. Uversky, M. A. Menze, Potential functions of LEA proteins from the brine shrimp Artemia franciscana - anhydrobiosis meets bioinformatics. *Journal of biomolecular structure & dynamics*, 1-19 (2017).
54. E. Jaspard, D. Macherel, G. Hunault, Computational and Statistical Analyses of Amino Acid Usage and Physico-Chemical Properties of the Twelve Late Embryogenesis Abundant Protein Classes. *PLOS ONE* **7**, e36968 (2012).
55. T. S. Harmon, A. S. Holehouse, M. K. Rosen, R. V. Pappu, Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **6**, (2017).
56. D. S. W. Protter *et al.*, Intrinsically Disordered Regions Can Contribute Promiscuous Interactions to RNP Granule Assembly. *Cell Reports* **22**, 1401-1412 (2018).
57. J. R. Wheeler, T. Matheny, S. Jain, R. Abrisch, R. Parker, Distinct stages in stress granule assembly and disassembly. *eLife* **5**, e18413 (2016).
58. J. S. Clegg, P. Seitz, W. Seitz, C. F. Hazlewood, Cellular responses to extreme water loss: the water-replacement hypothesis. *Cryobiology* **19**, 306-316 (1982).
59. A. C. Cuming, in *Seed Proteins,* P. R. Shewry, R. Casey, Eds. (Springer Netherlands, Dordrecht, 1999), pp. 753-780.
60. K. Goyal, L. J. Walton, A. Tunnacliffe, LEA proteins prevent protein aggregation due to water stress. *Biochem J* **388**, 151-157 (2005).
61. G. Xie, S. N. Timasheff, The thermodynamic mechanism of protein stabilization by trehalose. *Biophysical Chemistry* **64**, 25-43 (1997).

62. S. N. Timasheff, Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components. *Proceedings of the National Academy of Sciences* **99**, 9721-9726 (2002).
63. T. H. MacRae, Stress tolerance during diapause and quiescence of the brine shrimp, Artemia. *Cell Stress Chaperon* **21**, 9-18 (2016).
64. T. Kikawada *et al.*, Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid. *Biochemical and biophysical research communications* **348**, 56-61 (2006).
65. L. Dure, C. Chlan, Developmental Biochemistry of Cottonseed Embryogenesis and Germination : XII. PURIFICATION AND PROPERTIES OF PRINCIPAL STORAGE PROTEINS. *Plant Physiology* **68**, 180-186 (1981).
66. A. Solomon, R. Salomon, I. Paperna, I. Glazer, Desiccation stress of entomopathogenic nematodes induces the accumulation of a novel heat-stable protein. *Parasitology* **121 (Pt 4)**, 409-416 (2000).
67. L. Dure, 3rd *et al.*, Common amino acid sequence domains among the LEA proteins of higher plants. *Plant Mol Biol* **12**, 475-486 (1989).
68. M. J. Wise, A. Tunnacliffe, POPP the question: what do LEA proteins do? *Trends in plant science* **9**, 13-17 (2004).
69. M.-D. Shih, F. A. Hoekstra, Y.-I. C. Hsing, in *Advances in Botanical Research,* K. Jean-Claude, D. Michel, Eds. (Academic Press, 2008), vol. Volume 48, pp. 211-255.
70. M. Battaglia, A. Covarrubias, Late Embryogenesis Abundant (LEA) proteins in legumes. *Frontiers in Plant Science* **4**, (2013).
71. K. I. Jönsson, The Nature of Selection on Anhydrobiotic Capacityin Tardigrades. *Zoologischer Anzeiger - A Journal of Comparative Zoology* **240**, 409-417 (2001).
72. D. D. Horikawa, S. Higashi, Desiccation Tolerance of the Tardigrade Milnesium tardigradum Collected in Sapporo, Japan, and Bogor, Indonesia. *Zoological Science* **21**, 813-816, 814 (2004).
73. L. Rebecchi, R. Guidetti, S. Borsari, T. Altiero, R. Bertolani, Dynamics of Long-term Anhydrobiotic Survival of Lichen-dwelling Tardigrades. *Hydrobiologia* **558**, 23-30 (2006).
74. S. Hengherr, A. G. Heyer, F. Brummer, R. O. Schill, Trehalose and vitreous states: desiccation tolerance of dormant stages of the crustaceans Triops and Daphnia. *Physiol Biochem Zool* **84**, 147-153 (2011).
75. A. S. Holehouse, J. Ahad, R. K. Das, R. V. Pappu, CIDER: Classification of Intrinsically Disordered Ensemble Regions. *Biophysical journal* **108**, 228a (2015).
76. S. Vucetic, C. Brown, K. Dunker, Z. Obradovic, Flavors of protein disorder. *Proteins* **52**, (2003).
77. V. N. Uversky, J. R. Gillespie, A. L. Fink, Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins: Structure, Function, and Bioinformatics* **41**, 415-427 (2000).
78. A. K. Dunker *et al.*, The unfoldomics decade: an update on intrinsically disordered proteins. *BMC genomics* **9 Suppl 2**, S1 (2008).
79. R. K. Das, K. M. Ruff, R. V. Pappu, Relating sequence encoded information to form and function of intrinsically disordered proteins. *Current opinion in structural biology* **32**, 102-112 (2015).
80. A. H. Mao, N. Lyle, R. V. Pappu, Describing sequence-ensemble relationships for intrinsically disordered proteins. *Biochem J* **449**, 307-318 (2013).
81. J. Warwicker, S. Charonis, R. A. Curtis, Lysine and Arginine Content of Proteins: Computational Analysis Suggests a New Tool for Solubility Design. *Molecular Pharmaceutics* **11**, 294-303 (2014).

82. N. N. Pouchkina-Stantcheva *et al.*, Functional divergence of former alleles in an ancient asexual invertebrate. *Science* **318**, 268-271 (2007).
83. J. Browne, A. Tunnacliffe, A. Burnell, Anhydrobiosis: plant desiccation gene found in a nematode. *Nature* **416**, 38 (2002).
84. T. Tyson, W. Reardon, J. A. Browne, A. M. Burnell, Gene induction by desiccation stress in the entomopathogenic nematode Steinernema carpocapsae reveals parallels with drought tolerance mechanisms in plants. *International journal for parasitology* **37**, 763-776 (2007).
85. M. A. Sharon, A. Kozarova, J. S. Clegg, P. O. Vacratsis, A. H. Warner, Characterization of a group 1 late embryogenesis abundant protein in encysted embryos of the brine shrimp Artemia franciscana. *Biochem Cell Biol* **87**, 415-430 (2009).
86. M. A. Menze, L. Boswell, M. Toner, S. C. Hand, Occurrence of Mitochondria-targeted Late Embryogenesis Abundant (LEA) Gene in Animals Increases Organelle Resistance to Water Stress. *The Journal of biological chemistry* **284**, 10714-10719 (2009).
87. S. C. Hand, D. Jones, M. A. Menze, T. L. Witt, Life without water: expression of plant LEA genes by an anhydrobiotic arthropod. *Journal of experimental zoology. Part A, Ecological genetics and physiology* **307**, 62-66 (2007).
88. W. Zhao *et al.*, The Potential Roles of the G1LEA and G3LEA Proteins in Early Embryo Development and in Response to Low Temperature and High Salinity in Artemia sinica. *PLoS ONE* **11**, e0162272 (2016).
89. N. Y. Denekamp *et al.*, Discovering genes associated with dormancy in the monogonont rotifer Brachionus plicatilis. *BMC genomics* **10**, 108 (2009).
90. T. Z. Gal, I. Glazer, H. Koltai, An LEA group 3 family member is involved in survival of C. elegans during exposure to stress. *FEBS letters* **577**, 21-26 (2004).
91. O. Gusev *et al.*, Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge. *Nat Commun* **5**, 4784 (2014).
92. S. Tanaka *et al.*, Novel Mitochondria-Targeted Heat-Soluble Proteins Identified in the Anhydrobiotic Tardigrade Improve Osmotic Tolerance of Human Cells. *PLOS ONE* **10**, e0118272 (2015).
93. A. Yamaguchi *et al.*, Two novel heat-soluble protein families abundantly expressed in an anhydrobiotic tardigrade. *PLoS One* **7**, e44209 (2012).
94. D. K. Hincha, A. Thalhammer, LEA proteins: IDPs with versatile functions in cellular dehydration tolerance. *Biochemical Society transactions* **40**, 1000-1003 (2012).
95. D. Kovacs, B. Agoston, P. Tompa, Disordered plant LEA proteins as molecular chaperones. *Plant signaling & behavior* **3**, 710-713 (2008).
96. P. Tompa, D. Kovacs, Intrinsically disordered chaperones in plants and animals. *Biochem Cell Biol* **88**, 167-174 (2010).
97. J. Yu, Y. Lai, X. Wu, G. Wu, C. Guo, Overexpression of OsEm1 encoding a group I LEA protein confers enhanced drought tolerance in rice. *Biochemical and biophysical research communications* **478 2**, 703-709 (2016).
98. M. R. Marunde *et al.*, Improved tolerance to salt and water stress in Drosophila melanogaster cells conferred by late embryogenesis abundant protein. *J Insect Physiol* **59**, 377-386 (2013).
99. M. Battaglia, Y. Olvera-Carrillo, A. Garciarrubio, F. Campos, A. A. Covarrubias, The Enigmatic LEA Proteins and Other Hydrophilins. *Plant Physiology* **148**, 6-24 (2008).
100. P. L. Steponkus, M. Uemura, R. A. Joseph, S. J. Gilmour, M. F. Thomashow, Mode of action of the COR15a gene on the freezing tolerance of Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **95**, 14570-14575 (1998).

101. D. Tolleter *et al.*, Structure and function of a mitochondrial late embryogenesis abundant protein are revealed by desiccation. *The Plant cell* **19**, 1580-1589 (2007).
102. A. Thalhammer, M. Hundertmark, A. V. Popova, R. Seckler, D. K. Hincha, Interaction of two intrinsically disordered plant stress proteins (COR15A and COR15B) with lipid membranes in the dry state. *Biochim Biophys Acta* **1798**, 1812-1820 (2010).
103. D. S. Moore, R. Hansen, S. C. Hand, Liposomes with diverse compositions are protected during desiccation by LEA proteins from Artemia franciscana and trehalose. *Biochim Biophys Acta* **1858**, 104-115 (2016).
104. D. Tolleter, D. K. Hincha, D. Macherel, A mitochondrial late embryogenesis abundant protein stabilizes model membranes in the dry state. *Biochim Biophys Acta* **1798**, 1926-1933 (2010).
105. A. Garay-Arroyo, J. M. Colmenero-Flores, A. Garciarrubio, A. A. Covarrubias, Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *J Biol Chem* **275**, 5668-5674 (2000).
106. D. K. Hincha, E. Zuther, A. V. Popova, Stabilization of Dry Sucrose Glasses by Four LEA_4 Proteins from Arabidopsis thaliana. *Biomolecules* **11**, 615 (2021).
107. J. Grelet *et al.*, Identification in pea seed mitochondria of a late-embryogenesis abundant protein able to protect enzymes from drying. *Plant Physiology* **137**, 157-167 (2005).
108. A. V. Popova, S. Rausch, M. Hundertmark, Y. Gibon, D. K. Hincha, The intrinsically disordered protein LEA7 from Arabidopsis thaliana protects the isolated enzyme lactate dehydrogenase and enzymes in a soluble leaf proteome during freezing and drying. *Biochim Biophys Acta* **1854**, 1517-1525 (2015).
109. L. Dure, G. A. Galau, Developmental Biochemistry of Cottonseed Embryogenesis and Germination : XIII. REGULATION OF BIOSYNTHESIS OF PRINCIPAL STORAGE PROTEINS. *Plant Physiol* **68**, 187-194 (1981).
110. L. Dure, S. C. Greenway, G. A. Galau, Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by in vitro and in vivo protein synthesis. *Biochemistry* **20**, 4162-4168 (1981).
111. S. P. Graether, K. F. Boddington, Disorder and function: a review of the dehydrin protein family. *Frontiers in Plant Science* **5**, (2014).
112. N. B. Pammenter, P, A review of recalcitrant seed physiology in relation to desiccation-tolerance mechanisms. *Seed Science Research* **9**, 13-37 (1999).
113. N. Y. Denekamp, R. Reinhardt, M. Kube, E. Lubzens, Late embryogenesis abundant (LEA) proteins in nondesiccated, encysted, and diapausing embryos of rotifers. *Biology of reproduction* **82**, 714-724 (2010).
114. A. Tunnacliffe, J. Lapinski, B. McGee, A Putative LEA Protein, but no Trehalose, is Present in Anhydrobiotic Bdelloid Rotifers. *Hydrobiologia* **546**, 315-321 (2005).
115. E. Schokraie *et al.*, Proteomic analysis of tardigrades: towards a better understanding of molecular mechanisms by anhydrobiotic organisms. *PLoS One* **5**, e9502 (2010).
116. M. S. Clark *et al.*, Surviving extreme polar winters by desiccation: clues from Arctic springtail (Onychiurus arcticus) EST libraries. *BMC Genomics* **8**, 475 (2007).
117. G. Hunault, E. Jaspard, LEAPdb: a database for the late embryogenesis abundant proteins. *BMC Genomics* **11**, 221 (2010).
118. V. N. Uversky, Natively unfolded proteins: A point where biology waits for physics. *Protein science : a publication of the Protein Society* **11**, 739-756 (2002).
119. V. N. Uversky, A. K. Dunker, Understanding protein non-folding. *Biochimica et biophysica acta* **1804**, 1231-1264 (2010).

120. O. B. Ptitsyn, Molten globule and protein folding. *Advances in protein chemistry* **47**, 83-229 (1995).
121. V. N. Uversky, Diversity of compact forms of denatured globular proteins. *Protein & Peptide Letters* **4**, 355-367 (1997).
122. V. N. Uversky, O. B. Ptitsyn, "Partly folded" state, a new equilibrium state of protein molecules: four-state guanidinium chloride-induced unfolding of beta-lactamase at low temperature. *Biochemistry* **33**, 2782-2791 (1994).
123. V. N. Uversky, O. B. Ptitsyn, Further evidence on the equilibrium "pre-molten globule state": four-state guanidinium chloride-induced unfolding of carbonic anhydrase B at low temperature. *J Mol Biol* **255**, 215-228 (1996).
124. V. N. Uversky, Unusual biophysics of intrinsically disordered proteins. *Biochimica et biophysica acta* **1834**, 932-951 (2013).
125. V. N. Uversky, Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically Disordered Proteins. *J Biol Chem* **291**, 6681-6688 (2016).
126. V. N. Uversky, Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* **11**, 739-756 (2002).
127. W. P. Jencks, On the attribution and additivity of binding energies. *Proc Natl Acad Sci U S A* **78**, 4046-4050 (1981).
128. H. Xie *et al.*, Functional Anthology of Intrinsic Disorder. III. Ligands, Postranslational Modifications and Diseases Associated with Intrinsically Disordered Proteins. *Journal of proteome research* **6**, 1917-1932 (2007).
129. A. K. Dunker, Z. Obradovic, The protein trinity – linking function and disorder. *Nat Biotechnol* **19**, (2001).
130. M. D. Shih, T. Y. Hsieh, T. P. Lin, Y. I. Hsing, F. A. Hoekstra, Characterization of two soybean (Glycine max L.) LEA IV proteins by circular dichroism and Fourier transform infrared spectrometry. *Plant & cell physiology* **51**, 395-407 (2010).
131. A. K. Dunker *et al.*, Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 473-484 (1998).
132. P. Lieutaud *et al.*, How disordered is my protein and what is its disorder for? A guide through the "dark side" of the protein universe. *Intrinsically disordered proteins* **4**, e1259708 (2016).
133. P. E. Wright, H. J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **293**, 321-331 (1999).
134. P. Romero *et al.*, Thousands of proteins likely to have long disordered regions. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 437-448 (1998).
135. W. E. Meador, A. R. Means, F. A. Quiocho, Target enzyme recognition by calmodulin: 2.4 A structure of a calmodulin-peptide complex. *Science* **257**, 1251-1255 (1992).
136. R. S. Spolar, M. T. Record, Jr., Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**, 777-784 (1994).
137. B. Mészáros, I. Simon, Z. Dosztányi, Prediction of Protein Binding Regions in Disordered Proteins. *PLoS Computational Biology* **5**, e1000376 (2009).
138. Y. Cheng *et al.*, Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. *Biochemistry* **46**, 13468-13477 (2007).
139. F. M. Disfani *et al.*, MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics* **28**, i75-83 (2012).

140. C. J. Oldfield *et al.*, Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* **44**, 12454-12470 (2005).
141. Z. Peng, C. Wang, V. N. Uversky, L. Kurgan, Prediction of Disordered RNA, DNA, and Protein Binding Regions Using DisoRDPbind. *Methods in molecular biology (Clifton, N.J.)* **1484**, 187-203 (2017).
142. V. N. Uversky, A. K. Dunker, *Intrinsically disordered protein analysis Volume 1, Volume 1*. (Springer, New York, 2012).
143. V. N. Uversky, A. K. Dunker, *Intrinsically disordered protein analysis Volume 2, Volume 2*. (Springer, New York, 2012).
144. C. Bracken, L. M. Iakoucheva, P. R. Romero, A. K. Dunker, Combining prediction, computation and experiment for the characterization of protein disorder. *Current opinion in structural biology* **14**, 570-576 (2004).
145. P. Radivojac *et al.*, Intrinsic disorder and functional proteomics. *Biophysical journal* **92**, 1439-1456 (2007).
146. V. N. Uversky, P. Radivojac, L. M. Iakoucheva, Z. Obradovic, A. K. Dunker, Prediction of intrinsic disorder and its use in functional proteomics. *Methods in molecular biology (Clifton, N.J.)* **408**, 69-92 (2007).
147. Lemke, in *Structure and Function of Intrinsically Disordered Proteins,* P. Tompa, Ed. (Taylor & Francis Group, Boca Raton, FL, 2011), vol. 1, chap. 10, pp. 121-142.
148. S. Lise, D. T. Jones, Sequence patterns associated with disordered regions in proteins. *Proteins* **58**, 144-150 (2005).
149. A. Lobley, M. B. Swindells, C. A. Orengo, D. T. Jones, Inferring function using patterns of native disorder in proteins. *PLoS computational biology* **3**, e162 (2007).
150. A. H. Warner, S. Chakrabortee, A. Tunnacliffe, J. S. Clegg, Complexity of the heat-soluble LEA proteome in Artemia species. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* **7**, 260-267 (2012).
151. G. Wu *et al.*, Diverse LEA (late embryogenesis abundant) and LEA-like genes and their responses to hypersaline stress in post-diapause embryonic development of Artemia franciscana. *Comp Biochem Physiol B Biochem Mol Biol* **160**, 32-39 (2011).
152. X. Li, P. Romero, M. Rani, A. K. Dunker, Z. Obradovic, Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome informatics. Workshop on Genome Informatics* **10**, 30-40 (1999).
153. P. Romero *et al.*, Sequence complexity of disordered protein. *Proteins* **42**, 38-48 (2001).
154. K. Peng, P. Radivojac, S. Vucetic, A. K. Dunker, Z. Obradovic, Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* **7**, 208 (2006).
155. Z. Obradovic *et al.*, Predicting intrinsic disorder from amino acid sequence. *Proteins* **53 Suppl 6**, 566-572 (2003).
156. P. Radivojac, Z. Obradović, C. J. Brown, A. K. Dunker, Prediction of boundaries between intrinsically ordered and disordered protein regions. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 216-227 (2003).
157. A. K. Dunker, Z. Obradovic, P. Romero, E. C. Garner, C. J. Brown, Intrinsic protein disorder in complete genomes. *Genome informatics. Workshop on Genome Informatics* **11**, 161-171 (2000).
158. C. J. Oldfield *et al.*, Comparing and combining predictors of mostly disordered proteins. *Biochemistry* **44**, 1989-2000 (2005).
159. Z. Dosztányi, V. Csizmok, P. Tompa, I. Simon, IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433-3434 (2005).

160. K. Coeytaux, A. Poupon, Prediction of unfolded segments in a protein sequence based on amino acid composition. *Bioinformatics* **21**, 1891-1900 (2005).
161. Z. R. Yang, R. Thomson, P. McNeil, R. M. Esnouf, RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **21**, 3369-3376 (2005).
162. S. O. Garbuzynskiy, M. Y. Lobanov, O. V. Galzitskaya, To be folded or to be unfolded? *Protein science : a publication of the Protein Society* **13**, 2871-2877 (2004).
163. R. Linding *et al.*, Protein disorder prediction: implications for structural proteomics. *Structure (London, England : 1993)* **11**, 1453-1459 (2003).
164. J. Prilusky *et al.*, FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**, 3435-3438 (2005).
165. R. Linding, R. B. Russell, V. Neduva, T. J. Gibson, GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic acids research* **31**, 3701-3708 (2003).
166. Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac, A. K. Dunker, Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins* **61 Suppl 7**, 176-182 (2005).
167. L. Käll, A. Krogh, E. L. Sonnhammer, A combined transmembrane topology and signal peptide prediction method. *Journal of molecular biology* **338**, 1027-1036 (2004).
168. P. Lieutaud, B. Canard, S. Longhi, MeDor: a metaserver for predicting protein disorder. *BMC genomics* **9 Suppl 2**, S25 (2008).
169. J. M. Chandonia, M. Karplus, New methods for accurate prediction of protein secondary structure. *Proteins* **35**, 293-306 (1999).
170. I. Callebaut *et al.*, Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cellular and molecular life sciences : CMLS* **53**, 621-645 (1997).
171. Z. Dosztányi, V. Csizmók, P. Tompa, I. Simon, The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *Journal of molecular biology* **347**, 827-839 (2005).
172. J. M. Bourhis, B. Canard, S. Longhi, Predicting protein disorder and induced folding: from theoretical principles to practical applications. *Current protein & peptide science* **8**, 135-149 (2007).
173. W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637 (1983).
174. R. Gautier, D. Douguet, B. Antonny, G. Drin, HELIQUEST: a web server to screen sequences with specific alpha-helical properties. *Bioinformatics* **24**, 2101-2102 (2008).
175. J. Fauchere, V. Pliska, Hydrophobic parameters II of amino acid side-chains from the partitioning of N-acetyl-amino acid amides. *Eur. J. Med. Chem.* **18**, (1983).
176. L. M. Iakoucheva *et al.*, The importance of intrinsic disorder for protein phosphorylation. *Nucleic acids research* **32**, 1037-1049 (2004).
177. J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology* **157**, 105-132 (1982).
178. J. Toxopeus, A. H. Warner, T. H. MacRae, Group 1 LEA proteins contribute to the desiccation and freeze tolerance of Artemia franciscana embryos during diapause. *Cell Stress Chaperones* **19**, 939-948 (2014).
179. A. H. Warner, Z. H. Guo, S. Moshi, J. W. Hudson, A. Kozarova, Study of model systems to test the potential function of Artemia group 1 late embryogenesis abundant (LEA) proteins. *Cell Stress Chaperones* **21**, 139-154 (2016).

180. A. H. Warner *et al.*, Evidence for multiple group 1 late embryogenesis abundant proteins in encysted embryos of Artemia and their organelles. *J Biochem* **148**, 581-592 (2010).
181. W. Neupert, J. M. Herrmann, Translocation of proteins into mitochondria. *Annual review of biochemistry* **76**, 723-749 (2007).
182. D. Roise, G. Schatz, Mitochondrial presequences. *The Journal of biological chemistry* **263**, 4509-4511 (1988).
183. R. Aurora, G. D. Rose, Helix capping. *Protein science : a publication of the Protein Society* **7**, 21-38 (1998).
184. C. N. Pace, J. M. Scholtz, A helix propensity scale based on experimental studies of peptides and proteins. *Biophysical journal* **75**, 422-427 (1998).
185. E. Lanzarotti, R. R. Biekofsky, D. A. Estrin, M. A. Marti, A. G. Turjanski, Aromatic-aromatic interactions in proteins: beyond the dimer. *Journal of chemical information and modeling* **51**, 1623-1633 (2011).
186. T. Furuki, T. Shimizu, T. Kikawada, T. Okuda, M. Sakurai, Salt Effects on the Structural and Thermodynamic Properties of a Group 3 LEA Protein Model Peptide. *Biochemistry* **50**, 7093-7103 (2011).
187. M. Hundertmark, D. K. Hincha, LEA (late embryogenesis abundant) proteins and their encoding genes in Arabidopsis thaliana. *BMC genomics* **9**, 118 (2008).
188. P. Ball, Water is an active matrix of life for cell and molecular biology. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 13327-13335 (2017).
189. S. A. Benner, A. Ricardo, M. A. Carrigan, Is there a common chemical model for life in the universe? *Current opinion in chemical biology* **8**, 672-689 (2004).
190. J. L. Finney, Water? What's so special about it? *Philos Trans R Soc Lond B Biol Sci* **359**, 1145-1163; discussion 1163-1145, 1323-1148 (2004).
191. P. H. Yancey, Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *J Exp Biol* **208**, 2819-2830 (2005).
192. K. E. McCluney *et al.*, Shifting species interactions in terrestrial dryland ecosystems under altered water availability and climate change. *Biological reviews of the Cambridge Philosophical Society* **87**, 563-582 (2012).
193. P. Alpert, The constraints of tolerance: Why are desiccation-tolerant organisms small and rare? *Integr Comp Biol* **44**, 515-515 (2004).
194. O. Leprince, G. A. F. Hendry, B. D. McKersie, The mechanisms of desiccation tolerance in developing seeds. *Seed Science Research* **3**, 231-246 (2008).
195. M. J. Oliver *et al.*, Desiccation Tolerance: Avoiding Cellular Damage During Drying and Rehydration. *Annual Review of Plant Biology* **71**, 435-460 (2020).
196. E. A. Golovina, F. A. Hoekstra, M. A. Hemminga, Drying increases intracellular partitioning of amphiphilic substances into the lipid phase. Impact On membrane permeability and significance for desiccation tolerance. *Plant Physiology* **118**, 975-986 (1998).
197. L. M. Crowe, Lessons from nature: the role of sugars in anhydrobiosis. *Comparative biochemistry and physiology. Part A, Molecular & integrative physiology* **131**, 505-513 (2002).
198. J. H. Crowe, A. E. Oliver, F. Tablin, Is There a Single Biochemical Adaptation to Anhydrobiosis?1. *Integr Comp Biol* **42**, 497-503 (2002).
199. H. Kaur, B. P. Petla, M. Majee, in *Heat Shock Proteins and Plants,* A. A. A. Asea, P. Kaur, S. K. Calderwood, Eds. (Springer International Publishing, Cham, 2016), pp. 3-18.
200. B. Janis, C. Belott, M. A. Menze, Role of Intrinsic Disorder in Animal Desiccation Tolerance. *Proteomics* **18**, e1800067 (2018).

201. J. D. Hibshman, J. S. Clegg, B. Goldstein, Mechanisms of Desiccation Tolerance: Themes and Variations in Brine Shrimp, Roundworms, and Tardigrades. *Frontiers in Physiology* **11**, (2020).
202. Y. Liu, J. Liang, L. Sun, X. Yang, D. Li, Group 3 LEA Protein, ZmLEA3, Is Involved in Protection from Low Temperature Stress. *Frontiers in plant science* **7**, 1011-1011 (2016).
203. M. A. S. Artur *et al.*, Structural Plasticity of Intrinsically Disordered LEA Proteins from Xerophyta schlechteri Provides Protection In Vitro and In Vivo. *Frontiers in Plant Science* **10**, (2019).
204. M. S. Clark *et al.*, Surviving extreme polar winters by desiccation: clues from Arctic springtail (Onychiurus arcticus) EST libraries. *BMC genomics* **8**, 475 (2007).
205. J. H. Crowe *et al.*, The trehalose myth revisited: introduction to a symposium on stabilization of cells in the dry state. *Cryobiology* **43**, 89-105 (2001).
206. O. Leprince, A. Pellizzaro, S. Berriri, J. Buitink, Late seed maturation: drying without dying. *Journal of experimental botany* **68**, 827-841 (2017).
207. J. H. CROWE, Anhydrobiosis: an unsolved problem. *Plant, Cell & Environment* **37**, 1491-1493 (2014).
208. Ľ. Vetráková, V. Vykoukal, D. Heger, Comparing the acidities of aqueous, frozen, and freeze-dried phosphate buffers: Is there a "pH memory" effect? *International journal of pharmaceutics* **530**, 316-325 (2017).
209. T. Arakawa, S. N. Timasheff, Mechanism of protein salting in and salting out by divalent cation salts: balance between hydration and salt binding. *Biochemistry* **23**, 5912-5923 (1984).
210. H. J. Dyson, P. E. Wright, H. A. Scheraga, The role of hydrophobic interactions in initiation and propagation of protein folding. *Proceedings of the National Academy of Sciences* **103**, 13057-13061 (2006).
211. C. Camilloni *et al.*, Towards a structural biology of the hydrophobic effect in protein folding. *Scientific Reports* **6**, 28285 (2016).
212. R. L. Baldwin, G. D. Rose, How the hydrophobic factor drives protein folding. *Proceedings of the National Academy of Sciences* **113**, 12462-12466 (2016).
213. S. Chakrabortee *et al.*, Hydrophilic protein associated with desiccation tolerance exhibits broad protein stabilization function. *Proceedings of the National Academy of Sciences* **104**, 18073-18078 (2007).
214. E. Chatelain *et al.*, Temporal profiling of the heat-stable proteome during late maturation of Medicago truncatula seeds identifies a restricted subset of late embryogenesis abundant proteins associated with longevity. *Plant Cell Environ* **35**, 1440-1455 (2012).
215. B. M. LeBlanc, M. T. Le, B. Janis, M. A. Menze, S. C. Hand, Structural properties and cellular expression of AfrLEA6, a group 6 late embryogenesis abundant protein from embryos of Artemia franciscana. *Cell Stress Chaperones*, (2019).
216. M. Z. Rashed, C. J. Belott, B. R. Janis, M. A. Menze, S. J. Williams, New insights into anhydrobiosis using cellular dielectrophoresis-based characterization. *Biomicrofluidics* **13**, 064113 (2019).
217. L. M. A. Dirk *et al.*, Late Embryogenesis Abundant Protein-Client Protein Interactions. *Plants (Basel)* **9**, (2020).
218. C. L. Cuevas-Velazquez, J. R. Dinneny, Organization out of disorder: liquid-liquid phase separation in plants. *Curr Opin Plant Biol* **45**, 68-74 (2018).
219. E. Gomes, J. Shorter, The molecular language of membraneless organelles. *The Journal of biological chemistry* **294**, 7115-7127 (2019).

220. W. van Leeuwen, C. Rabouille, Cellular stress leads to the formation of membraneless stress assemblies in eukaryotic animal cells. *Traffic*, (2019).
221. H. B. Schmidt, D. Gorlich, Transport Selectivity of Nuclear Pores, Phase Separation, and Membraneless Organelles. *Trends Biochem.Sci.* **41**, 46-61 (2016).
222. P. B. Sehgal, J. Westley, K. M. Lerea, S. DiSenso-Browne, J. D. Etlinger, Biomolecular condensates in cell biology and virology: phase-separated membraneless organelles (MLOs). *Analytical biochemistry*, 113691 (2020).
223. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science* **357**, (2017).
224. V. N. Uversky, Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Current opinion in structural biology* **44**, 18-30 (2017).
225. B. G. O'Flynn, T. Mittag, A new phase for enzyme kinetics. *Nature Chemical Biology* **17**, 628-630 (2021).
226. L. Peng, E.-M. Li, L.-Y. Xu, From start to end: Phase separation and transcriptional regulation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1863**, 194641 (2020).
227. S. L. Moon, T. Morisaki, T. J. Stasevich, R. Parker, Coupling of translation quality control and mRNA targeting to stress granules. *Journal of Cell Biology* **219**, (2020).
228. J. R. Buchan, R. Parker, Eukaryotic stress granules: the ins and outs of translation. *Molecular cell* **36**, 932-941 (2009).
229. P. A. Chong, J. D. Forman-Kay, Liquid–liquid phase separation in cellular signaling systems. *Current opinion in structural biology* **41**, 180-186 (2016).
230. D. L. J. Lafontaine, J. A. Riback, R. Bascetin, C. P. Brangwynne, The nucleolus as a multiphase liquid condensate. *Nature reviews. Molecular cell biology* **22**, 165-182 (2021).
231. T. J. Nott, T. D. Craggs, A. J. Baldwin, Membraneless organelles can melt nucleic acid duplexes and act as biomolecular filters. *Nature Chemistry* **8**, 569-575 (2016).
232. I. Popov, A. Greenbaum, A. P. Sokolov, Y. Feldman, The puzzling first-order phase transition in water–glycerol mixtures. *Physical Chemistry Chemical Physics* **17**, 18063-18071 (2015).
233. G. M. Harami *et al.*, Phase separation by ssDNA binding protein controlled via protein−protein and protein−DNA interactions. *Proceedings of the National Academy of Sciences* **117**, 26206-26217 (2020).
234. C. P. Brangwynne, Phase transitions and size scaling of membrane-less organelles. *The Journal of cell biology* **203**, 875-881 (2013).
235. W. B. Busa, J. H. Crowe, G. B. Matson, Intracellular pH and the metabolic status of dormant and developing Artemia embryos. *Arch Biochem Biophys* **216**, 711-718 (1982).
236. T. J. Abatzopoulos, J. Beardmore, J. S. Clegg, P. Sorgeloos, *Artemia: Basic and Applied Biology*. Biology of Aquatic Organisms (Springer Netherlands, ed. 1, 2002), vol. 1, pp. 286.
237. I. Letunic, P. Bork, 20 years of the SMART protein domain annotation resource. *Nucleic acids research* **46**, D493-d496 (2018).
238. I. Letunic, T. Doerks, P. Bork, SMART: recent updates, new developments and status in 2015. *Nucleic acids research* **43**, D257-260 (2015).
239. A. S. Holehouse, R. K. Das, J. N. Ahad, M. O. Richardson, R. V. Pappu, CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophysical journal* **112**, 16-21 (2017).

240. J. Yang *et al.*, The I-TASSER Suite: protein structure and function prediction. *Nature methods* **12**, 7-8 (2015).
241. J. Yang, Y. Zhang, I-TASSER server: new development for protein structure and function predictions. *Nucleic acids research* **43**, W174-181 (2015).
242. A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols* **5**, 725-738 (2010).
243. D. B. Thompson, J. J. Cronican, D. R. Liu, Engineering and identifying supercharged proteins for macromolecule delivery into mammalian cells. *Methods Enzymol* **503**, 293-319 (2012).
244. J. S. Glasheen, S. C. Hand, Metabolic Heat Dissipation and Internal Solute Levels of Artemia Embryos During Changes in Cell-Associated Water. *Journal of Experimental Biology* **145**, 263-282 (1989).
245. C. Belott, B. Janis, M. A. Menze, Liquid-liquid phase separation promotes animal desiccation tolerance. *Proceedings of the National Academy of Sciences* **117**, 27676-27684 (2020).
246. M. S. Lawrence, K. J. Phillips, D. R. Liu, Supercharging proteins can impart unusual resilience. *Journal of the American Chemical Society* **129**, 10110-10112 (2007).
247. A. M. Jonker, D. W. P. M. Löwik, J. C. M. van Hest, Peptide- and Protein-Based Hydrogels. *Chemistry of Materials* **24**, 759-773 (2012).
248. Q. Guo, X. Shi, X. Wang, RNA and liquid-liquid phase separation. *Non-coding RNA Research* **6**, 92-99 (2021).
249. A. G. Cherstvy, Positively charged residues in DNA-binding domains of structural proteins follow sequence-specific positions of DNA phosphate groups. *The journal of physical chemistry. B* **113**, 4242-4247 (2009).
250. N. Parker, A. C. G. Porter, Identification of a novel gene family that includes the interferon-inducible human genes 6-16 and ISG12. *BMC genomics*. 2004 (10.1186/1471-2164-5-8).
251. H. Gytz *et al.*, Apoptotic properties of the type 1 interferon induced family of human mitochondrial membrane ISG12 proteins. *Biol Cell* **109**, 94-112 (2017).
252. S. Kauppinen, M. Siggaard-Andersen, P. von Wettstein-Knowles, beta-Ketoacyl-ACP synthase I of Escherichia coli: nucleotide sequence of the fabB gene and identification of the cerulenin binding residue. *Carlsberg Res Commun* **53**, 357-370 (1988).
253. D. L. Burk, N. Ghuman, L. E. Wybenga-Groot, A. M. Berghuis, X-ray structure of the AAC(6')-Ii antibiotic resistance enzyme at 1.8 A resolution; examination of oligomeric arrangements in GNAT superfamily members. *Protein science : a publication of the Protein Society* **12**, 426-437 (2003).
254. A. F. Neuwald, D. Landsman, GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein. *Trends Biochem.Sci.* **22**, 154-155 (1997).
255. M. W. Vetting *et al.*, Structure and functions of the GNAT superfamily of acetyltransferases. *Arch Biochem Biophys* **433**, 212-226 (2005).
256. A. Argentaro *et al.*, Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 11939-11944 (2007).
257. S. Iwase *et al.*, ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome. *Nat Struct Mol Biol*. 2011 (10.1038/nsmb.2062).

258. J. A. Eisen, K. S. Sweder, P. C. Hanawalt, Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic acids research* **23**, 2715-2723 (1995).
259. B. Linder, R. A. Cabot, T. Schwickert, R. A. W. Rupp, The SNF2 domain protein family in higher vertebrates displays dynamic expression patterns in Xenopus laevis embryos. *Gene* **326**, 59-66 (2004).
260. S. P. Rowbotham *et al.*, Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1. *Molecular cell* **42**, 285-296 (2011).
261. H. Dürr, A. Flaus, T. Owen-Hughes, K.-P. Hopfner, Snf2 family ATPases and DExx box helicases: differences and unifying concepts from high-resolution crystal structures. *Nucleic acids research* **34**, 4160-4167 (2006).
262. J. M. Caruthers, E. R. Johnson, D. B. McKay, Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 13080-13085 (2000).
263. J. L. Baker *et al.*, Widespread genetic switches and toxicity resistance proteins for fluoride. *Science (New York, N.Y.)* **335**, 233-235 (2012).
264. T. F. Smith, C. Gaitatzes, K. Saxena, E. J. Neer, The WD repeat: a common architecture for diverse functions. *Trends Biochem.Sci.* **24**, 181-185 (1999).
265. B. P. Jain, S. Pandey, WD40 Repeat Proteins: Signalling Scaffold with Diverse Functions. *Protein J* **37**, 391-406 (2018).
266. D. Li, R. Roberts, WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. *Cellular and molecular life sciences : CMLS* **58**, 2085-2097 (2001).
267. J. Lowery *et al.*, The Sec7 guanine nucleotide exchange factor GBF1 regulates membrane recruitment of BIG1 and BIG2 guanine nucleotide exchange factors to the trans-Golgi network (TGN). *The Journal of biological chemistry* **288**, 11532-11545 (2013).
268. T. Achstetter, A. Franzusoff, C. Field, R. Schekman, SEC7 encodes an unusual, high molecular weight protein required for membrane traffic from the yeast Golgi apparatus. *The Journal of biological chemistry* **263**, 11711-11717 (1988).
269. S. B. Deitz, A. Rambourg, F. Képès, A. Franzusoff, Sec7p directs the transitions required for yeast Golgi biogenesis. *Traffic (Copenhagen, Denmark)* **1**, 172-183 (2000).
270. A. K. Gillingham, J. R. C. Whyte, B. Panic, S. Munro, Mon2, a relative of large Arf exchange factors, recruits Dop1 to the Golgi apparatus. *The Journal of biological chemistry* **281**, 2273-2280 (2006).
271. J. A. Efe *et al.*, Yeast Mon2p is a highly conserved protein that functions in the cytoplasm-to-vacuole transport pathway and is required for Golgi homeostasis. *Journal of cell science* **118**, 4751-4764 (2005).
272. S. Inouye, M. Inouye, Structure, function, and evolution of bacterial reverse transcriptase. *Virus Genes* **11**, 81-94 (1995).
273. N. S. Kim, Y. Umezawa, S. Ohmura, S. Kato, Human glyoxalase I. cDNA cloning, expression, and sequence similarity to glyoxalase I from Pseudomonas putida. *The Journal of biological chemistry* **268**, 11217-11221 (1993).
274. M. Rico, M. Bruix, C. González, R. I. Monsalve, R. Rodríguez, 1H NMR assignment and global fold of napin BnIb, a representative 2S albumin seed protein. *Biochemistry* **35**, 15672-15682 (1996).
275. D. Charvolin, J. P. Douliez, D. Marion, C. Cohen-Addad, E. Pebay-Peyroula, The crystal structure of a wheat nonspecific lipid transfer protein (ns-LTP1) complexed with two molecules of phospholipid at 2.1 A resolution. *Eur J Biochem* **264**, 562-568 (1999).

276. D. Samuel, Y.-J. Liu, C.-S. Cheng, P.-C. Lyu, Solution structure of plant nonspecific lipid transfer protein-2 from rice (Oryza sativa). *The Journal of biological chemistry* **277**, 35267-35273 (2002).
277. S. Strobl *et al.*, A novel strategy for inhibition of alpha-amylases: yellow meal worm alpha-amylase in complex with the Ragi bifunctional inhibitor at 2.5 A resolution. *Structure (London, England : 1993)* **6**, 911-921 (1998).
278. C. A. Behnke *et al.*, Structural determinants of the bifunctional corn Hageman factor inhibitor: x-ray crystal structure at 1.95 A resolution. *Biochemistry* **37**, 15277-15288 (1998).
279. R. Watanabe, K. Ohishi, Y. Maeda, N. Nakamura, T. Kinoshita, Mammalian PIG-L and its yeast homologue Gpi12p are N-acetylglucosaminylphosphatidylinositol de-N-acetylases essential in glycosylphosphatidylinositol biosynthesis. *The Biochemical journal* **339 (Pt 1)**, 185-192 (1999).
280. J. T. Maynes *et al.*, The crystal structure of 1-D-myo-inosityl 2-acetamido-2-deoxy-alpha-D-glucopyranoside deacetylase (MshB) from Mycobacterium tuberculosis reveals a zinc hydrolase with a lactate dehydrogenase fold. *The Journal of biological chemistry* **278**, 47166-47170 (2003).
281. S. A. Endow, The emerging kinesin family of microtubule motor proteins. *Trends Biochem.Sci.* **16**, 221-225 (1991).
282. P. B. Meluh, M. D. Rose, KAR3, a kinesin-related gene required for yeast nuclear fusion. *Cell* **60**, 1029-1041 (1990).
283. D. J. Sharp *et al.*, Functional coordination of three mitotic motors in Drosophila embryos. *Mol Biol Cell* **11**, 241-253 (2000).
284. J. Kumar *et al.*, The Caenorhabditis elegans Kinesin-3 motor UNC-104/KIF1A is degraded upon loss of specific binding to cargo. *PLoS Genet* **6**, e1001200-e1001200 (2010).
285. J. Blaszczyk *et al.*, Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. *Structure (London, England : 1993)* **9**, 1225-1236 (2001).
286. J. Blaszczyk *et al.*, Noncatalytic assembly of ribonuclease III with double-stranded RNA. *Structure (London, England : 1993)* **12**, 457-466 (2004).
287. M. A. Carmell, G. J. Hannon, RNase III enzymes and the initiation of gene silencing. *Nat Struct Mol Biol* **11**, 214-218 (2004).
288. M. Tijsterman, R. H. A. Plasterk, Dicers at RISC; the mechanism of RNAi. *Cell* **117**, 1-3 (2004).
289. C. Conrad, R. Rauhut, Ribonuclease III: new sense from nuisance. *Int J Biochem Cell Biol* **34**, 116-129 (2002).
290. B. M. Hallberg *et al.*, The structure of the RNA m5C methyltransferase YebU from Escherichia coli reveals a C-terminal RNA-recruiting PUA domain. *Journal of molecular biology* **360**, 774-787 (2006).
291. J. Sabina, D. Söll, The RNA-binding PUA domain of archaeal tRNA-guanine transglycosylase is not required for archaeosine formation. *The Journal of biological chemistry* **281**, 6993-7001 (2006).
292. C. S. Cerrudo, P. D. Ghiringhelli, D. E. Gomez, Protein universe containing a PUA RNA-binding domain. *The FEBS journal* **281**, 74-87 (2014).
293. H. J. Yoon, T. F. Donahue, The suil suppressor locus in Saccharomyces cerevisiae encodes a translation factor that functions during tRNA(iMet) recognition of the start codon. *Mol Cell Biol* **12**, 248-260 (1992).

294. D. Das, M. M. Georgiadis, The crystal structure of the monomeric reverse transcriptase from Moloney murine leukemia virus. *Structure (London, England : 1993)* **12**, 819-829 (2004).
295. E. Nowak *et al.*, Structural analysis of monomeric retroviral reverse transcriptase in complex with an RNA/DNA hybrid. *Nucleic acids research* **41**, 3874-3887 (2013).
296. E. Nowak *et al.*, Ty3 reverse transcriptase complexed with an RNA-DNA hybrid shows structural and functional asymmetry. *Nat Struct Mol Biol* **21**, 389-396 (2014).
297. F. Campos, C. Cuevas-Velazquez, M. A. Fares, J. L. Reyes, A. A. Covarrubias, Group 1 LEA proteins, an ancestral plant protein group, are also present in other eukaryotes, and in the archeae and bacteria domains. *Molecular genetics and genomics : MGG* **288**, 503-517 (2013).
298. P. Knox-Brown *et al.*, Similar Yet Different-Structural and Functional Diversity among Arabidopsis thaliana LEA_4 Proteins. *Int J Mol Sci* **21**, 2794 (2020).
299. C. Navarro-Retamal *et al.*, Molecular dynamics simulations and CD spectroscopy reveal hydration-induced unfolding of the intrinsically disordered LEA proteins COR15A and COR15B from Arabidopsis thaliana. *Physical chemistry chemical physics : PCCP* **18**, 25806-25816 (2016).
300. N. Guex, M. C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714-2723 (1997).
301. B. Bolognesi *et al.*, A Concentration-Dependent Liquid Phase Separation Can Cause Toxicity upon Increased Protein Expression. *Cell reports* **16**, 222-231 (2016).
302. B. Forood, E. J. Feliciano, K. P. Nambiar, Stabilization of alpha-helical structures in short peptides via end capping. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 838-842 (1993).
303. L. G. Presta, G. D. Rose, Helix signals in proteins. *Science* **240**, 1632-1641 (1988).
304. H. Lu *et al.*, Ionic polypeptides with unusual helical stability. *Nature Communications* **2**, 206 (2011).
305. C. Dalgicdir, C. Globisch, C. Peter, M. Sayar, Tipping the Scale from Disorder to Alpha-helix: Folding of Amphiphilic Peptides in the Presence of Macroscopic and Molecular Interfaces. *PLoS computational biology* **11**, e1004328 (2015).
306. N. Lotan, A. Yaron, A. Berger, The stabilization of the α-helix in aqueous solution by hydrophobic side-chain interaction. *Biopolymers* **4**, 365-368 (1966).
307. S. K. Burley, G. A. Petsko, Amino-aromatic interactions in proteins. *FEBS letters* **203**, 139-143 (1986).
308. T. J. Brunette *et al.*, Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580-584 (2015).
309. J. C. Coates, Armadillo repeat proteins: beyond the animal kingdom. *Trends in cell biology* **13**, 463-471 (2003).
310. R. Tewari, E. Bailes, K. A. Bunting, J. C. Coates, Armadillo-repeat protein functions: questions for little creatures. *Trends in cell biology* **20**, 470-481 (2010).

APPENDIX 1

PERMISSION FROM WILEY TO RESUSE ‘ROLE OF INTRINSIC DISORDER IN ANIMAL DESICCATION TOLERANCE’

OHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Oct 04, 2021

This Agreement between University of Louisville -- Brett Janis ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number

5154320901166

License date

Sep 22, 2021

Licensed Content Publisher

John Wiley and Sons

Licensed Content Publication

Proteomics

Licensed Content Title

Role of Intrinsic Disorder in Animal Desiccation Tolerance

Licensed Content Author

Brett Janis, Clinton Belott, Michael A. Menze

Licensed Content Date

Sep 20, 2018

Licensed Content Volume

18

Licensed Content Issue

21-22

Licensed Content Pages

15

Type of use

Dissertation/Thesis

Requestor type

Author of this Wiley article

Format

Electronic

Portion

Full article

Will you be translating?

No

Title

COMPUTATIONAL AND BIOCHEMICAL CHARACTERIZATIONS OF ANHYDROBIOSIS-RELATED INTRINSICALLY DISORDERED PROTEINS

Institution name

University of Louisville

Expected presentation date

Oct 2021

Requestor Location

University of Louisville
11519 MagnoliaView Ct

LOUISVILLE, KY 40299
United States
Attn: University of Louisville

Publisher Tax ID

EU826007151

Total

0.00 USD

Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.
- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order,** is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley

Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner.**For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.

- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.

- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud)

supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.

- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The Creative Commons Attribution Non-Commercial (CC-BY-NC)License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library http://olabout.wiley.com/WileyCDA/Section/id-410895.html

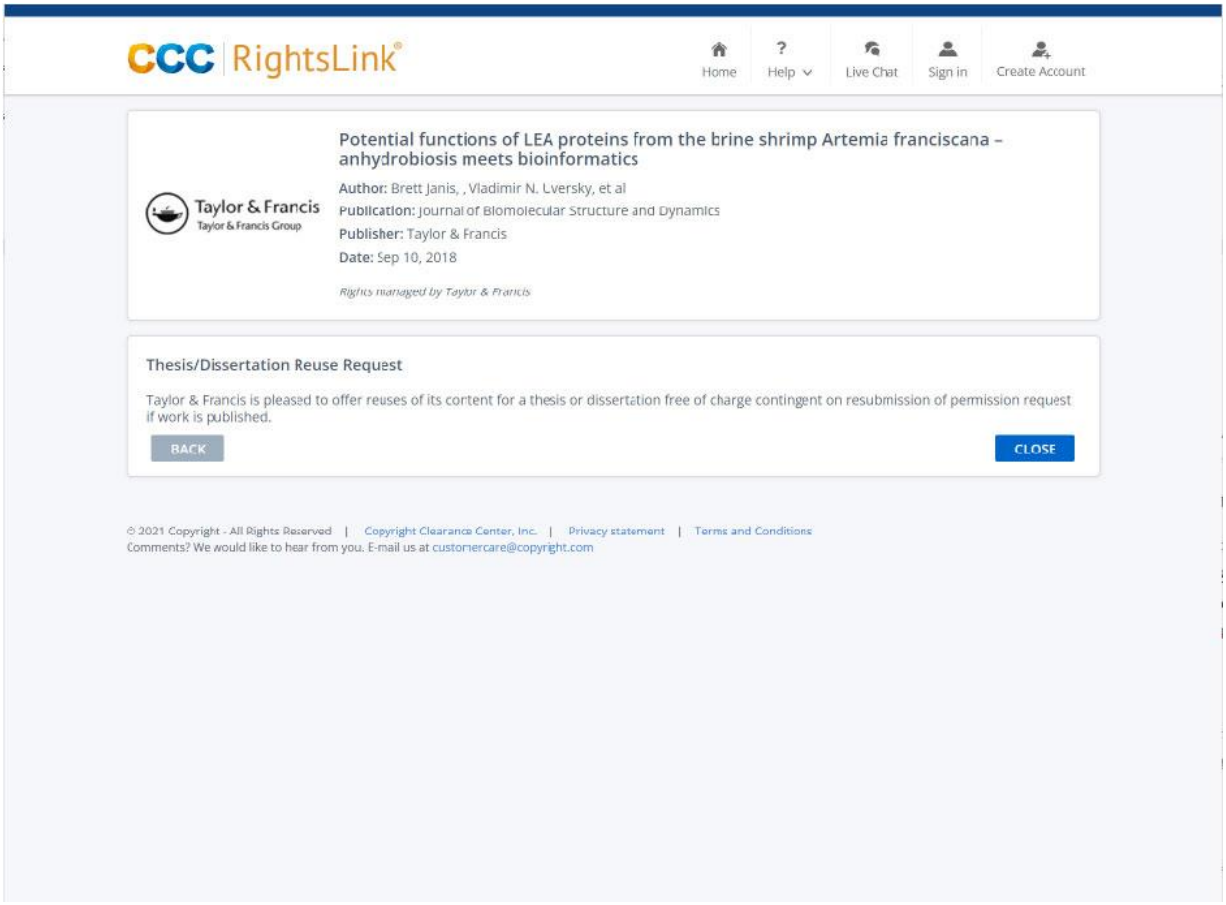
Other Terms and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

APPENDIX 2

PERMISSION FROM TAYLOR & FRANCIS TO REUSE 'POTENTIAL FUNCTIONS OF LEA PROTEINS FROM THE BRINE SHRIMP ARTEMIA FRANCISCANA – ANHYDROBIOSIS MEETS BIOINFORMATICS



APPENDIX 3

PERMISSION FROM THE PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES TO REUSE 'LIQUID-LIQUID PHASE SEPARATION PROMOTES ANIMAL DESICCATION TOLERANCE

Rights and Permissions

← About

PNAS authors need not obtain permission for the following cases:

1. to use their original figures or tables in their future works;
2. to make copies of their articles for their own personal use, including classroom use, or for the personal use of colleagues, provided those copies are not for sale and are not distributed in a systematic way;
3. to include their articles as part of their dissertations; or
4. to use all or part of their articles in printed compilations of their own works.

The full journal reference must be cited and, for articles published in volumes 90–105 (1993–2008), "Copyright (copyright year) National Academy of Sciences" must be included as a copyright note.

CURRICULUM VITAE

Brett Janis

Email: Brett.Janis@Louisville.edu

Education

PhD. in Biology Program, University of Louisville, Louisville, KY.
Anticipated 2021

B.A. in Biochemistry and Molecular Biology, Cornell College, Mt. Vernon, IA.
2015

B.A. in Biology, Cornell College, Mt. Vernon, IA.
2014

Research Experience

Graduate Research Assistantship
2017-2018

Sonoporation of Membrane-Impermeant Compounds into Cells for Cryopreservation.

Developed a novel method of microfluidics-mediated sonoporation to load cell-impermeant preservatives, such as trehalose, into a variety of cell types with minimal cellular damage. This new method improves the recovery of human red blood cells after storage at -80 ˚C and after lyophilization for long-term storage.

Graduate Student Researcher 2016-2021

Characterization of Late Embryogenesis Abundant Proteins from *Artemia franciscana*.

Developed novel scanning electron microscopy, atomic force microscopy, and confocal microscopy techniques to measure late embryogenesis abundant (LEA) protein aggregation behavior *in vitro*. Developed techniques for measuring structural transitions in LEA proteins under simulated desiccation conditions.

Bioinformatics Analysis of Late Embryogenesis Abundant Proteins Found in Animals.

Combined several bioinformatics approaches including machine learning algorithms, hierarchical modeling software, and molecular dynamics simulations to compare late embryogenesis abundant proteins from

animals.

Teaching Experience

Graduate Teaching Assistantship, Department of Biology, University of Louisville
2016-2017

Instructed three sections of a non-major's course in introductory biology focusing on concepts with critical social impact. Led an interactive classroom utilizing hands-on learning approaches to develop functional understandings of information literacy, the scientific method, vaccination, evolution, and climate change. Student interest was cultivated using interactive modeling activities and through direct scientific engagement such as contributing to the US National Phenology Network.

Graduate Teaching Assistantship, Department of Biology, University of Louisville
Fall, 2018

Collaborated with professors to develop a culminating project assignment for students in a non-major biology course: Biology Issues and Applications. Developed a four-stage process of peer collaboration and mentorship to develop individualized projects that incorporated the student's professional interests such as websites, field guides, lesson plans, legal documents, and civil engineering projects.

Other Appointments

DesiCorp Inc.

Cofounder
2017-present

Coinvented the fundamental technology for DesiCorp Inc. and assisted in co-founding the company. Maintains an active, uncompensated role in the company as a consultant and board member outside of his contracted hours with the University of Louisville.

Secretary
2017-present

Updates the DesiCorp board members and maintains records of company matters. Establishes connections with potential industry partners, consultants, and academic collaborators and maintains those relationships.

Chief Scientific Officer
Starting 2021

The obligations of this position include leading the research and

development endeavors at DesiCorp Inc. and securing funding to continue those endeavors. Reports monthly progress reports to the DesiCorp board to maintain transparency regarding product development schedule, cost, and performance.

Presentations –

Oral Presentations

Society for Cryobiology
2020

"Sonoporation-Mediated Trehalose Loading for Red Blood Cell Stabilization"

Society of Integrative and Comparative Biology.
2019 "Liquid-Liquid Phase Separation Behavior of a Crustacean Late Embryogenesis Abundant Protein"

Kentucky Science Center: Eat, Drink, and Do Science.
2018

Graduate Student Research Symposium at University of Louisville.
2018 "Sweet Blood: Using Sugar to Dry and Store Blood at Room Temperature"

Poster Presentations

AABB Annual Meeting.
2018 "Dry-Preservation of Red Blood Cells"

Military Health Systems Research Symposium.
2018 "Dry-Preservation of Red Blood Cells for Far-Forward Settings"

International Protein Society 32nd Annual Conference.
2018 "Conformational Plasticity of Late Embryogenesis Abundant (LEA) Proteins"

NASA Human Research Program Investigator's Workshop.
2018 "Dry-Preservation of Red Blood Cells for Storage at Ambient Temperatures"

International Protein Society 31st annual conference.
2017

"Sequence Analysis of LEA proteins from *Artemia franciscana*"

Graduate Student Research Symposium at University of Louisville.
2017

Society of Integrative and Comparative Biology.

2017 “Impact of Group 3 LEA on Cellular Structure during Desiccation”

Graduate Student Award ceremony at Eastern Illinois University.
2017 “Anhydrobiosis Under a Microscope”

Society of Integrative and Comparative Biology conference.
2016

“A Computational Analysis of LEA Proteins from *Artemia franciscana*”

Grants

NASA Flight Opportunities Program (Essential Personnel). 2018. $275,000.

NSF PFI Grant (Essential Personnel). 2018. $750,000.

NSF National I-Corps Program (Entrepreneurial Lead). 2018: $50,000.

NIH Excite Grant Installment 3 (Co-PI). 2018: $50,000.

NIH Excite Grant Installment 2 (Co-PI). 2018: $50,000.

NSF Local I-Corps Launchit Program (Entrepreneurial Lead). 2017: $2,500.

NIH Excite Grant Installment 1 (Co-PI). 2017: $50,000.

University of Louisville Arts and Sciences Graduate Student Research Grant. 2017: $1,000.

Eastern Illinois University Research and Creative Activity Grant Spring. 2016: $400

Awards and Certificates

Awards

UofL John M. Houchens Prize	2021
UofL J.B. Speed School of Engineering Exemplary Research Scholarship Award	2021
AABB Top Poster Award.	2018
5-Across Pitch Competition Audience Choice.	2018
UofL 3-Minute Thesis Runner-Up.	2018
NASA HRP Poster Competition 3rd Place.	2018
NSF Local I-Corps Launchit Pitch Competition 1st place.	2017
Eastern Illinois University Excellence in Student Research Award.	2016

Certificates

UofL PLAN: High Workshop Attendance Award. 2018

UofL PLAN: Community Engagement Academy. 2018

UofL PLAN: Graduate Teaching Assistant Academy II. 2018

UofL PLAN: High Workshop Attendance Award. 2017

UofL PLAN: Graduate Student Teaching Academy I. 2017

Academic and Community Service

Academic Service

Academic Search Committee 2019

Assisted in screening applicants for a tenure-track position as a molecular biology professor in the Department of Biology at the University of Louisville.

ThinkIR Advisory Board Member 2018–2021

Offer graduate student perspectives and feedback on the ThinkIR website, platform, and policies at the University of Louisville.

Biology Graduate Student Association: Chair of Outreach 2018–2019

Establishes and maintains relationships with local organizations that promote science education and outreach. Connects graduate students with outreach and citizen science volunteer opportunities at the University of Louisville.

Community Service

Scientific Outreach with Maryhurst 2021-present

Develops and leads safe, engaging, and educational scientific demonstrations and activities to promote interest in science and engineering in residents at the Maryhurst Academy.

2019 ACCelerate Festival Presenter
2019

Presented research related to anhydrobiosis and the sonoporation technology developed at the University of Louisville at the Smithsonian Museum of National History in Washington, D.C.

Kentucky Science and Engineering Fair 2017-2019

Judges in local and regional science fairs for grade 5-12 in the state of Kentucky in the subjects of biology, chemistry, chemical engineering, computer sciences, and physics.

Kentucky Academy of Sciences 2017-2019

Judges for an annual science fair in the subjects of biology, chemistry, and physics.

Raptor Rehabilitation of Kentucky Inc.: Volunteer 2016-2018

Assists in regular maintenance and rudimentary treatment of sick and injured birds of prey. Presents for public gatherings about the social and ecological benefits of raptors to inspire respect for birds of prey and spread awareness for policies and history relevant to them. Contributes to seasonal newsletters published and distributed to donors, volunteers, and event attendees. Provides data analysis and grant-writing assistance to acquire funding for general supplies and medical equipment.

Kentucky Science Center (KSC): Volunteer 2016–2018

Presents interactive biology and chemistry labs for the KSC's youth science summits to inspire future STEM students. Performs interactive demos for Holiday events at the KSC, as well as during National Chemistry Week, National Brain Day, and other annual science events.

Research Products

Patents and Disclosures

Menze, M.A., Kopechek, J.A., **Janis, B.** 2017. New Method for Cryopreservation of Cells Using Ultrasound. Reference: 17031.

Publications

Janis, B.R., Priddy, M.C., Otto, M.R., Kopechek, J.A., Menze, M.A. 2020. Sonoporation enables high-throughput loading of trehalose into red blood cells. Online ahead of print. DOI:10.1016/j.cryobiol.2020.12.005

Belott, C.[1], **Janis, B.R.[1]**, Menze, MA. 2020. Liquid-liquid phase separation promotes animal desiccation tolerance. PNAS 117 (44): p 27676-27684. DOI: https://doi.org/10.1073/pnas.2014463117

Centner, C.S., Murphy, E.M., Priddy, M.C., Moore, J.T., **Janis, B.R.**, Menze, M.A., DeFilippis, A., and Kopechek, J.A. 2020. Ultrasound-induced molecular delivery to erythrocytes using a microfluidic system. Biomicrofluidics 14, 024114 (2020);

https://doi.org/10.1063/1.5144617

Rashed, M.Z., Belott, C., **Janis, B.R.**, Menze, M.A., Williams, S. 2019. New insights into anhydrobiosis using cellular dielectrophoresis-based characterization. Biomicrofluidics 13(6): 064113. DOI: 10.1063/1.5126810

Leblanc, B.M., Le, M.T., **Janis, B.R.**, Menze, M.A., Hand, S.C. 2019. Structural properties and cellular expression of *Afr*LEA6, a group 6 late embryogenesis abundant protein from embryos of Artemia franciscana. Cell Stress and Chaperones 24(5). DOI: 10.1007/s12192-019-01025-8

Janis, B.[1], Belott, C.[1], and Menze, M.A. 2018. Role of Intrinsic Disorder in Animal Desiccation Tolerance. Proteomics. DOI: 10.1002/pmic.201800067

Janis, B., Uversky, V., and Michael, M.A. 2017. Potential Functions of LEA Proteins from the Brine Shrimp *Artemia Franciscana* - Anhydrobiosis Meets Bioinformatics. Journals of Biological Structure and Dynamics. Journal of Biomolecular Structure and Dynamics. DOI: 10.1080/07391102.2017.1387177.