Characterisation of two desiccation-linked dehydrins from *Xerophyta humilis*

Cynthia Fan

Supervisor: Professor Jill M. Farrant

Co-supervisor: Dr Suhail Rafudeen

Submitted in fulfillment of the requirements for the degree of Master of Science in the Department of Molecular and Cell Biology, University of Cape Town, South Africa.

September 2016

Declaration

The experimental work described in this MSc thesis was undertaken in the Molecular and Cell Biology Department, University of Cape Town, South Africa under the direct supervision of Professor Jill M. Farrant and Dr Suhail Rafudeen from January 2015 to August 2016.

I hereby declare that this thesis entitled:

"Characterisation of two desiccation-linked dehydrins from *Xerophyta humilis*"

is my own work and has not previously, in its entirety or in part, been submitted at any university for another degree, and that all the sources I have used and cited have been stipulated and acknowledged through complete referencing.

I know the meaning of plagiarism and declare that all of the work in the document, save for that which is properly acknowledged, is my own.

Cynthia Fan, September 2016

Signed by candidate Signature Removed

Acknowledgments

Firstly I would like to thank, most sincerely, my supervisor Professor Jill Farrant for being infinitely encouraging, inspiring and trusting. It is this trust that has been the reassuring force driving me through this MSc. Secondly, I would like to thank my co-supervisor, Dr Suhail Rafudeen – your wisdom is unmatched.

To Dr Nashied Peton, Dr Amelia Hilgart and Dr Jeanne Korsmann – without your timely departures and arrivals throughout this MSc, I would not have arrived at the end feeling somewhat sane and relatively cheerful. I am truly grateful for all the time and effort you have put into helping me, thank you.

Thank you to the departmental wizard, Madhu Chauhan. Without your patience and willingness to teach, this study would not have been possible.

Thank you to members of the plant stress lab for all the laughs and knowledge – Astrid Radermacher, Robyn Waters (honorary), Humaira Lambarey, Hawwa Gabier, Udisha Chathuri, Tim Dennis, Amogelang Tshego Gill, Mehafo Falowo, Halford Dace (honourary), Jo Bentley, Chrissie Madden and the superglue that holds us all together, Keren Cooper.

I would also like to thank Dr Henk Hillhorst, Dr Maria Cecilia Da Costa and Mariana Artur from Wageningen University for their wonderful generosity in providing me with the *Xerophyta viscosa* homologues and upstream promoter region sequences.

A very important thank you to the National Research Foundation for granting me the Innovation Scholarship and to L'Oreal Paris for the Open Research Bursary.

Lastly and most importantly, I would like to thank the LEA proteins to my abiotic stress – my family and the dearest people in my life – without whom I would be a dead plant.

Abbreviations

BLAST	Basic Local Alignment Search Tool
CD	Circular Dichroism
cDNA	Complementary DNA
CFSSP	Chou & Fassman Secondary Structure Predictor
CS	Citrate Synthase
IDP	Intrinsically Disordered Protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria Broth
LEA	Late Embryogenesis Abundant
LDH	Lactate Dehydrogenase
HSP	Heat Shock Protein
kDA	Kilo Daltons
MEME	Multiple Em for Motif Elicitation
mRNA	Messenger RNA
NCBI	National Center for Biotechnology Information
OD	Optical Density
PCR	Polymerase Chain Reaction
pI	Isoelectric Point
PONDR	Predictor of Naturally Disordered Regions
qPCR	Quantitative Polymerase Chain Reaction
RWC	Relative Water Content
SEM	Standard Error of the Mean
TFE	Tetrafluoroethylene
TF	Transcription Factor

Figures and Tables

List of Figures

List of Tables

Abstract

In response to abiotic stresses, organisms throughout the plant kingdom, as well as microorganisms and micro-animals such as nematodes or tardigrades, have been observed to express Late Embryogenesis Abundant (LEA) proteins as protective mechanisms. However, despite two decades of research, little is understood about their physiological functions and this has led to extensive nomenclature, with a large amount of redundancy.

The primary reason for this lack of insight into LEA protein functions is their highly hydrophilic and intrinsically disordered nature. Intrinsically disordered proteins (IDPs) cannot be studied using conventional methods of structural analyses such as X-ray crystallography and, therefore, alternative techniques are required. A combination of transgenic and *in vitro* studies have also shown that LEA proteins are most likely to behave as molecular chaperones by binding water and ions, preventing macromolecular aggregation and protecting enzymatic activity during dehydration.

This study characterized two dehydrins that were expressed during dehydration in the desiccation tolerant plant, *Xerophyta humilis.* From a transcriptome analyses on *X. humilis,* cDNA for the two dehydrins were obtained. These sequences were first analysed using various *in silico* tools in order to identify putative dehydrin-specific characteristics. Subsequently, these two dehydrins were cloned and expressed for production of recombinant dehydrin protein. These proteins were then analysed in terms of structural and functional characteristics. Structurally, through the use of circular dichroism in an *in vitro* system, both dehydrins demonstrated the shift towards being increasingly alpha-helical when placed in environments of decreasing water content. The role of these two dehydrins in stabilizing enzymes during dehydration was subsequently investigated using citrate synthase (CS) and lactate dehydrogenase (LDH). The preservation of enzyme activity was observed in both CS and LDH. This preservation of enzyme activity was further maintained by the presence of trehalose. Anti-aggregation roles were also investigated, however, neither dehydrin demonstrated significant ability to minimize the aggregation of LDH.

This study hopes to establish a pipeline for characterizing LEA proteins using structural and functional assays in order to provide alternative means of LEA protein classification.

Table of Contents

Chapter 1
Literature review: The role of Late Embryogenesis Abundant (LEA) proteins in conferring desiccation tolerance

1.1 Introduction

All terrestrial organisms rely on water to survive and the limitation of water at some point within their life cycles poses as a threat to their survival (Battaglia et al., 2008) . While most organisms are able to tolerate very moderate amounts of water loss, the ability to survive almost complete desiccation is rare and known as 'anhydrobiosis' (Hoekstra, Golovina, & Buitink, 2001). Anhydrobiosis is more commonly known as 'desiccation tolerance' and has been demonstrated in numerous organisms across a variety of life forms (Bewley, 1979). In the plant kingdom, desiccation tolerance is fairly common in reproductive tissues such as pollen, seeds (termed 'orthodox seeds'), spores and also in the vegetative tissue of non-tracheophytes such as bryophytes (Farrant, Cooper, & Nell, 2012). With the exception of 'resurrection plants', desiccation tolerance is rare in the vegetative tissues of angiosperms and completely non-existent in gymnosperms, perhaps as a result of their size. 'Resurrection plant' is the term used for a small group of angiosperms that are desiccation tolerant in their vegetative tissues. This evolutionary adaptation has been hypothesized as a crucial point in the colonization of land by primitive plants that originated from fresh water (Oliver, Tuba, & Mishler, 2000).

More recently, desiccation tolerance has been observed in prokaryotes and also in invertebrates such as nematodes, tardigrades, brine shrimp embryos and midge larvae (Hatanaka et al., 2013). Whilst these organisms appear to differ drastically from one another, many of the protective mechanisms that they adopt in order to survive extreme water deficit are fairly conserved (Costa et al., 2015). Considering the multiple roles that water plays in supporting and enabling life, this would be expected since the consequences of water loss would be the same (Hoekstra et al., 2001). These include the loss of the intermolecular interactions mediated by water molecules, the mechanical stabilization provided on a cellular level and the medium established for diffusion of substances and enzymatic reactions that are vital to metabolism (Farrant et al., 2012). Therefore, it is unsurprising that in almost all desiccation tolerant organisms, the response to extreme water loss is met with a similar repertoire of protective mechanisms in attempts to maintain cellular integrity – the production of considerable amounts of compatible solutes, primary metabolites, antioxidant enzymes,

heat shock proteins (HSPs) and late embryogenesis abundant proteins (LEA) proteins (Dinakar & Bartels, 2013; Farrant, Brandt, & Lindsey, 2007).

On a planet that is facing increasing periods of intense drought, understanding the mechanisms that allow organisms to be desiccation tolerant is crucial (Dai, 2013). Plants, in particular, face a variety of environmental stresses throughout their life cycle as a result of their sessile nature (Amara et al., 2014). These environmental stresses, which include drought, high salinity and extreme temperatures, affect most areas of the world and their impact on plant survival is highly significant. Crop species are particularly sensitive to these abiotic stresses and it is estimated that approximately half the annual crop yield across the world is lost as a result of these abiotic stresses (Jones & Thornton, 2003). Therefore, establishing a greater understanding of the mechanisms adopted by organisms with better protection strategies against abiotic stresses would aid in the development of stronger crops and maintaining food security.

1.2 Protective mechanisms that enable desiccation tolerance

As water availability is restricted, cytoplasmic water is lost from inside the cell ('moderate dehydration') and the intracellular space becomes increasingly crowded with the volume that remains available decreasing drastically (Hoekstra et al., 2001). This decrease in cellular volume with an increase in low molecular weight molecules is known as a state of 'molecular crowding' (Chebotareva, Kurganov, & Livanova, 2004). As a consequence, molecular interactions between the cellular components become increasingly likely and these may result in protein denaturation, aggregation or membrane fusion. These events are deleterious to the recovery of the cell. In plants that are drought tolerant, compatible solutes such as proline, glutamate, mannitol, sorbitol, sucrose and oligosaccharides are produced as a response mechanism to dehydration. The compatible solutes that are produced do not interfere with cellular structures or functions but simply exist to maintain a hydration layer around the surfaces of the intracellular proteins (Hoekstra et al., 2001).

Once cytoplasmic water drops below 0.3g H_2O per g dry weight^{-1}, the protective mechanisms of compatible solutes alone are unable to protect membranes and proteins against the damages associated with drying to an air-dry state (Hoekstra et al., 2001). Survival beyond drought conditions requires desiccation tolerance (Ingram & Bartels, 1996). The onset of

desiccation tolerance is commonly associated with the accumulation of high concentrations of LEA proteins. While numerous roles have been proposed for LEA proteins, the exact functions that they carry out within a molecularly crowded cell remain unknown. Nonetheless, they are thought to play a crucial role in conferring desiccation tolerance as their accumulation has been observed in all desiccation tolerant organisms (Chakrabortee et al., 2007; Furuki et al., 2012; Hatanaka et al., 2013; Collett et al., 2005; Iturriaga et al., 1992). The particular link between the desiccation tolerant resurrection plants and LEA proteins will be explored in more detail at a later point in this review.

1.3 Discovering the abundance of LEA proteins

The existence of the 'enigmatic" LEA protein was first demonstrated in the cotton plant *Gossypium hirsutum* by Dure et al., (1981). These were a small group of proteins that accumulated at high levels during the mature phase of embryo development in the cotton plant seeds, thus acquiring the name "Late Embryogenesis Abundant". This stage of seed development was known to coincide with the period during which orthodox seeds gained the ability to survive conditions of extreme dehydration (Tunnacliffe & Wise, 2007). Therefore, this was the first indicator of the role that LEA proteins played in conferring desiccation tolerance. It was subsequently discovered that LEA protein expression was found across a large number of species including plants, invertebrates and microorganisms and even in organisms that were desiccation sensitive. Whilst it was initially assumed that LEA protein expression occurred as a response to water deficit stress, it was later observed that they could be induced by a variety of abiotic stresses such as cold, osmotic and salinity. As a result, the small group of proteins that had originally been considered to be seed specific grew into a huge protein family consisting of multiple subgroups.

Classification of these various LEA protein groups proved to be a monstrous challenge as it could not be done solely on the basis of expression profiles since expression varied according to the type of abiotic stress being experienced (Tunnacliffe & Wise, 2007). In addition, variation in expression was observed across different tissue types and not restricted to embryonic tissue. Nonetheless, a few unifying characteristics were identified and these are often considered as the necessary criteria for a protein to be classified as a 'LEA protein'. Firstly, LEA proteins are generally extremely hydrophilic and this quality is likely to be responsible for their second characteristic, the intrinsically disordered nature in a hydrated

environment (Bies-Ethève et al., 2008). LEA proteins are also considered to be fairly heat stable. However, even characteristics as minimal as these have been contradicted by the 'atypical' Group_2 LEAs (also known as Group 5) which are more hydrophobic, lack heat stability and possess fairly stable structures in solution (Dang, Popova, Hundertmark, & Hincha, 2014). To date, LEA proteins have been characterized according to the presence of sequence-specific motifs that have been deemed signatures of a particular group (Tunnacliffe & Wise, 2007). Therefore, this indicates that there exists a disconnect between the classification systems and the functions that can be inferred from each group.

Numerous *in vitro* experiments have been developed in attempts to understand the behaviour and functions of LEA proteins. In terms of structural analyses, LEA proteins are challenging to study because they are intrinsically disordered proteins (IDPs). IDPs lack stable secondary or tertiary structures under standard physiological conditions, despite being able to carry out biological functions (Sun, Rikkerink, Jones, & Uversky, 2013). This defies the traditional understanding that associated protein function with stable three-dimensional structures. Additionally, traditional methods that are used to understand protein structure such as X-ray crystallography cannot be used since the structure of LEA proteins cannot be crystallized in water. The most commonly used analysis for understanding LEA protein structure in-solution has been circular dichroism (CD) (Fallis, 2013). CD spectroscopy detects the presence of secondary structures in proteins by observing the characteristic spectra displayed by α-helices, β-sheets and random coils (Mouillon, Eriksson, & Harryson, 2008). This technique also allows assumptions to be made regarding hypothetical structural changes that a LEA protein would undergo in different conditions. Since the production of compatible solutes such as proline and sucrose are produced in response to drought, Mouillon et al., (2008) monitored the structural changes in the presence of 0% and 80% sucrose solutions. However, minimal changes in secondary structure were observed and no physiological functions could be inferred. In the same study, the addition of glycerol resulted in a strongly α-helical response, suggesting that glycerol acted as a secondary structure inducer, generating a similar response to tetrafluoroethylene (TFE).

To understand changes in LEA protein structure between hydrated and dehydrated conditions, it is possible to compare the structure of a protein when it is in dH_2O and then following desiccation on a CaF_2 window (Thalhammer, Hundertmark, Popova, Seckler, &

Hincha, 2010). Thalhammer et al., (2010) observed that two intrinsically disordered stress proteins from *Arabidopsis thaliana*, COR15A and COR15B, appeared predominantly unstructured in solution but became 65% and 57% α-helical after drying. It is evident from these studies that CD spectroscopy provides significant insight into the possible changes undergone by LEA proteins in the presence of various cellular conditions.

Having established an improved understanding of the changes that occur structurally during the transition from hydrated to dehydrated, the next research question into LEA proteins targeted the associations between structural changes and responses in physiological functions. The general assumption made regarding the functions of LEA proteins revolve around their likelihood of behaving as molecular chaperones in protecting the contents of a dehydrating cell (Tunnacliffe & Wise, 2007). Goyal et al., (2005) first tested these predicted functions using aggregation and functional enzyme assays with citrate synthase (CS) and lactate dehydrogenase (LDH). These two enzymes were selected due to their reported sensitivity to water loss. When these two enzymes were dried down in the presence of AavLEA1, a group 3 LEA from an anhydrobiotic nematode, and Em, a group 1 LEA from wheat, it was observed that both LEAs were able to protect enzyme activity in CS and LDH after undergoing desiccation and freezing. These two LEAs also displayed anti-aggregation roles in the presence of both heat and desiccation, although the effect was strengthened in the presence of the putative chemical chaperone, trehalose. An IDP from *A. thaliana,* LEA7, was also observed to protect LDH activity from desiccation and freezing (Popova, Rausch, Hundertmark, Gibon, & Hincha, 2015). Similarly in MtPM25, the Group 5 LEA from th orthodox seed *Medicago truncatula*, *in vitro* enzyme activity was maintained despite heat, freezing and desiccation stress in CS and LDH (Boucher et al., 2010). Protein aggregation was also minimized by the presence of MtPM25. However, when two desiccation-induced Group 1 LEAs from *X. humilis* were tested for chaperone activity with LDH, only one was able to maintain LDH activity despite the two LEAs having 50% sequence similarity (Ginbot, 2011). Therefore, this suggests that LEA protein function is not necessarily conserved within the groups as classified by the existing nomenclature. In addition, the ability to behave as a protective chaperone was also not limited to LEAs that come from plants or organisms that are desiccation tolerant.

1.4 An overview of the different classes of LEA proteins

Over time, extensive classification systems have been developed in order to categorize the family of LEA proteins. The classification scheme that will be used in this overview of the different classes of LEA proteins is the original system of Group 1 to Group 6, proposed by Bray (1993 & 1994), with the additions contributed by Amara et al., (2014), Battaglia et al., (2008) and Hong-Bo et al., (2005).

The Group 1 LEA proteins, originally labeled by Dure et al., (1993) as D-19, is a group of highly conserved and hydrophilic LEA proteins that contain a characteristic motif of 20 amino acids – GGQTRREQLGEEGYSQMGRK (Close, 1996). These LEAs have been reported to have enhanced water-binding capacity (Bray, 1993). Experimentally, it was shown that the presence of PMA1959, a Group 1 LEA from *A. thaliana*, was able to minimize electrolyte leakage in the leaves of transgenic plants grown under high salt stress (Cheng, Targolli, Huang, & Wu, 2002). In yeast, over-expression of a wheat Group 1 LEA, *EM*, conferred osmotic tolerance (Swire-Clark & Marcotte, 1999). The experimental findings suggest that Group 1 LEAs play a protective role against both desiccation and osmotic stress caused by high salt concentrations.

Group 2 LEAs, or dehydrins (dehydration induced proteins), were originally known as D-11 in cotton embryos (Dure et al., 1993). Dehydrins are one of the most extensively studied LEA protein groups. The three most commonly conserved motifs found within the dehydrin family are the Y-, S- and K-segments (Close, 1996). The presence of the K-segment (EKKGIMDKIKEKLPG) is generally regarded as what defines a dehydrin, although conservation of this sequence may not be absolute (Graether & Boddington, 2014). The S-segment consists of 5-7 consecutive Ser residues and is regarded as a possible phosphorylation site. The third conserved motif in dehydrins is the Y segment, which consists of the sequence motif (V/T)D(E/Q)YGNP. The arrangement and number of Y-, S- and K-segments determine the ‘dehydrin architecture’. There are commonly observed dehydrin architectures and this allows inferences to be made regarding possible functions and subcellular localisation (Graether & Boddington, 2014). Located between Y-, S- and K-segments are motifs known as Φ segments – these are generally rich in Gly, Gln and Thr while Phe, Cys and Trp are rarely observed. Such amino acid biases within the Φ segments contribute to the hydrophilic nature of LEAs (Allagulova, Gimalov, Shakirova, & Vakhitov,

2003; Graether & Boddington, 2014). Several other motifs have been observed in dehydrins, though their conservation across the class has not yet been established (Graether & Boddington, 2014). These include ChP segments (Lys-rich region followed by a Glu or Asp), which have been suggested to function in nuclear targeting and DNA binding, and His-segments (either HKGEHHSGDHH or His-His/His flanking the K-segment), which may play a role in membrane binding (Eriksson, Kutzer, Procek, Gröbner, & Harryson, 2011; Hara, Fujinaga, & Kuboi, 2005). Up-regulation of PpDHNA, a group 2 LEA, was observed in response to salt and osmotic stress (Saavedra et al., 2006). The ectopic expression of a wheat dehydrin (DHN-5*)* in *A. thaliana* also displayed improved tolerance to high salinity and water deficit (Kosová et al., 2008). These results indicate that dehydrins respond to a variety of abiotic stresses.

Group 3 LEA proteins were originally identified as D7 by Dure et al., (1993) and contain the sequence motif TAQAAKEKAGE, which can be found in repeats of up to thirteen times (Bray, 1993). This conserved motif is predicted to form an amphiphilic α-helix. This group has also been predicted to play a role in the sequestration of ions during dehydration. Group 3 LEAs are known to be the most widely distributed across the plant kingdom and also in other organisms that are capable of anhydrobiosis such as the nematode, *Aphelenchus avenae* (Goyal, Pinelli, et al., 2005). Anhydrobiotic organisms were observed to accumulate group 3 LEAs during desiccated states (Gal, Glazer, & Koltai, 2004). A group 3 LEA from *A. avenae*, AavLEA1, has been shown to protect protein aggregation and activity in enzymes after dehydration (Goyal, Walton, & Tunnacliffe, 2005).

Group 4 LEA proteins were classified as D113 by Dure (1993) and identified by their conserved N-terminal motif, which consists of approximately 70-80 residues. This group of LEAs has been predicted to preserve membrane structure by mechanisms of water replacement. The N terminal sequence in this group of LEAs is also known to form α-helical structures (Bray, 1993). Constitutive expression of a Group 4 LEA from *A. thaliana*, AtLEA4-5, showed increased drought tolerance (Battaglia et al., 2008).

Group 5 and Group 6 LEAs, identified by Dure (1993) as D29 and D34, are significantly understudied in comparison to the previous four groups. These two glasses generally consist of the ‘atypical’ LEAs that contain a high portion of hydrophobic amino acids (Boucher et

al., 2010). These proteins are also unable to withstand boiling, unlike the other LEA groups, and have been predicted to be more globular than intrinsically disordered (Amara et al., 2014). MtPM25, a group 5 LEA protein from *M. truncatula*, has been observed to prevent dehydration-associated enzyme denaturation despite being relatively hydrophobic (Boucher et al., 2010). This suggests that, although they do not bear the hydrophilicity and intrinsic disorder characteristic of LEA proteins, they are also able to play a role in protecting molecules during dehydration.

What should be evident from this overview is the lack of insight provided by this classification system into the unique functional or structural characteristics, apart from what is known about the Group 5 LEAs. Classification on the basis of sequence similarity does not allow the inference of functions or structures. This is a result born out of the absence of understanding or unifying experimental evidence for how LEA proteins undergo functional and structural changes in the presence of different abiotic stresses.

1.5 LEA proteins and desiccation tolerance

Gene expression profiles are a desirable way to study changes on a molecular level in the respond to an external stimulus such as extreme water deficit stress (Collett et al., 2004). Transcriptome analyses were conducted on various desiccation tolerant angiosperms species including *Craterostigma plantagineum*, *Xerophyta humilis* and *Boea hygrometrica* (Collett et al., 2004; Rodriguez et al., 2010; Zhu et al., 2015). Transcript profiles in the desiccated vegetative tissues of both *C. plantagineum* and *B. hygrometrica* were dominated by the presence of LEA encoding transcripts (Rodriguez et al., 2010; Zhu et al., 2015). The importance of LEA proteins to surviving desiccation was also emphasized in the transcript levels of desiccated *X. humilis* in which, of the 55 genes that were found to be up-regulated in response to dehydration, 16 were LEAs (Collett et al., 2004). Waters (2015) subsequently monitored the changes in mRNA expression levels of 21 putative LEA genes in *X. humilis* during dehydration and rehydration. It was observed that in all 21 genes, a significant up-regulation of expression in response to early dehydration occurred, particularly between 40-50% and 20-30% relative water content (RWC). Such findings further strengthen the proposed dominant role that LEA proteins play in conferring desiccation tolerance.

Changes in gene expression on a transcript level do not necessarily correspond with changes on a protein level; therefore, results on a transcriptomic level alone are insufficient in providing explanations to how LEA proteins function. In the context of resurrection plants, some mRNAs are known to be stored during dehydration and only translated upon rehydration (Collett et al., 2004). Therefore, proteomes should also be monitored under the same water deficit stress conditions. In the absence of genome sequences for non-model organisms such as resurrection plants, proteome analyses encounter limitations when protein identities cannot be determined (Dinakar & Bartels, 2013). Nonetheless, an advantage that proteomics have over transcriptomics is that it provides information about the portion of the genome that is being actively translated (Ingle et al., 2007). LEA proteins in *C. plantagineum* were observed to accumulate in abundance during dehydration; phosphorylation of two of these LEAs was also shown, suggesting their involvement in subsequent signal transduction pathways. The heat stable proteome of the orthodox seed, *M. truncatula*, was isolated (Chatelain et al., 2012). Of the 135 proteins identified in the heat stable fraction, 38 were identified to be LEA polypeptides. When measured according to intensity, it was calculated that LEA proteins represented 54% of the heat stable protein fraction. This study concluded that, as a result of differing accumulation profiles, it is likely that different subsets of LEAs are responsible for different functions in a dehydrating cell and this depended on the level of water deficit stress experienced.

To date, the sequenced genomes of two resurrection plant species are publically accessible – *B. hygrometrica* and *Oropetium thomaeum* (Van Buren et al., 2015; Xiao et al., 2015). These two species have been reported to contain 67 and 94 LEA motif-containing proteins, respectively. A third resurrection plant genome, that of *Xerophyta viscosa*, has recently been sequenced and is currently under review (Costa et al., unpublished). In *X. viscosa*, 126 putative LEA motif-containing proteins were identified and this is significantly higher than those observed in the two previously sequenced genomes. These LEAs have been regarded as the primary response to extreme water loss. In the genomes of desiccation sensitive species, 51 putative LEA proteins were identified in *A. thaliana* whilst *O. sativa* has been reported to have 33 putative LEAs and *V. vinifera* has 36 (Amara et al., 2012). With such expansive numbers of LEA proteins being identified in various plant species, it can be expected that the nomenclature for the classification systems will merely become more convoluted whilst insight into their precise mechanisms used by LEA proteins remain unexplored. This further

suggests a need to characterize the structure and functions of putative LEA proteins, instead of only identifying the group specific sequence motifs.

1.6 This study

The objective of this MSc study was to establish a pipeline for ultimately characterizing all 21 of the desiccation induced LEA genes identified in *X. humilis* by Waters (2015). In this study, the structural and functional characterisation techniques will be evaluated using two putative dehydrins selected from the 21 LEA genes. Collett et al., (2004) had previously identified these two dehydrins as Xh_RD_30C12 and Xh_RD_19H04. However, in order to highlight their identity as the group of LEA transcripts that were induced during dehydration in *X. humilis,* Waters (2015) developed novel nomenclature for each of the 21 LEAs. This novel classification system adopted the group system originally presented by Battaglia et al., (2008). According to this system, Xh_RD_30C12 and Xh_RD_19H04 were renamed XhLEA2-4 and XhLEA2-5, respectively.

The full-length cDNA sequences for all 21 putative LEA clones were used for analysis with quantitative PCR (qPCR) for both the dehydration and the rehydration profile within leaf tissues (Waters, 2015). The changes in transcript abundance that were observed are shown in Figure 1.1 for XhLEA2-4 and XhLEA2-5. The general trend across the group of dehydrins was early up-regulation during dehydration and down-regulation during rehydration. Whilst XhLEA2-4 is observed to increase in expression levels from 100% to 20-30% RWC, XhLEA2-5 is shown to be most highly expressed at 70-80% RWC and then the expression levels drop gradually until the plant returns to 100% RWC following rehydration. This differential expression suggests the likelihood of a difference in roles played by the two dehydrins during dehydration and rehydration in *X. humilis.* Despite both being dehydrins, XhLEA2-4 and XhLEA2-5 are ideal candidates for functional and structural characterisation as their differential transcript levels indicate that they are likely to be responsible for different protective responses. Both the nucleotide sequences and the amino acid sequences of the two dehydrin sequences were analysed using variety of *in silico* tools. This aimed to distinguish the differences between sequence similarities, along with possible differences in other characteristics that could explain the transcript expression patterns. Hydrophilicity and the amount of disorder present in the proteins were also predicted. These analyses were used

collectively to identify LEA and dehydrin specific motifs in order to support their identity as putative dehydrins.

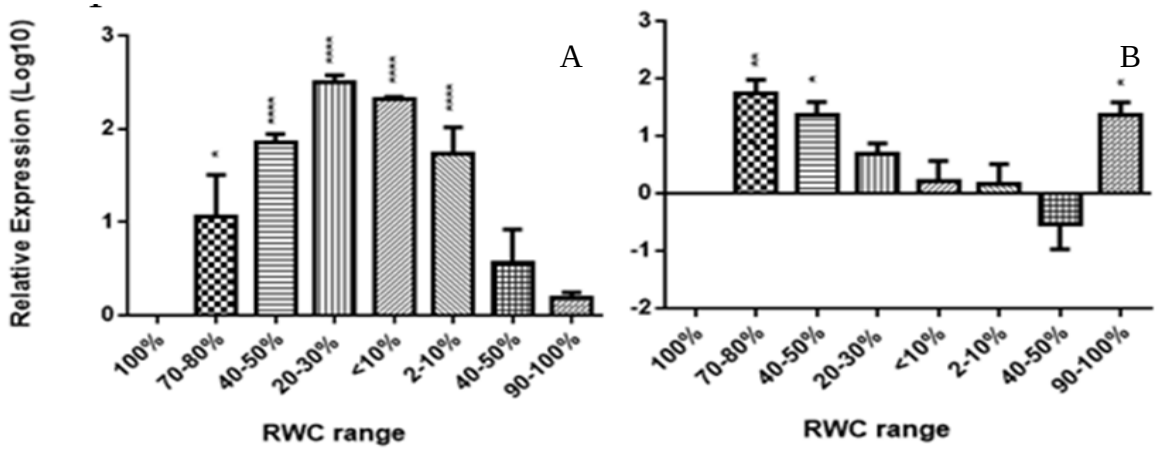


Figure 1.1 Relative gene expression levels of the two putative dehydrins, XhLEA2-4 (A) and XhLEA2-5 (B). Each data point represents the mean and the error bars represent the standard error of the mean (SEM), of relative expression levels of each LEA-like gene (N = 4), compared to 18S ribosomal RNA, calculated according to Pfaffl (2001). The Y-axis represents the log10 change in expression levels and the % RWC values are represented on the X-axis. The level of significance in transcript level change is indicated by the asterisks (*). The graph and SEM values were generated using the GraphPad Prism software (Version 6). This image was adapted from Waters (2015).

The cDNA sequence of each dehydrin was subsequently cloned into bacterial expression systems for recombinant protein production and purification. The in-solution changes in secondary structure of each dehydrin were measured using circular dichroism (CD) spectroscopy in the presence of dH_2O and two concentrations of an inorganic solvent. The aim of this was to mimic a transition into dehydration, with decreasing concentrations of water being present (Dennis, 2015). The predictions made by the *in silico* programs were then compared to the CD spectra for validation. Each recombinant protein was then tested for its possible role as molecular chaperones in protecting enzymatic activity during desiccation. Their possible role in minimizing aggregation induced by heat stress was also investigated. The findings from this work have been discussed in the subsequent chapters.

The aim of this study was to characterize two LEA proteins within the same group (dehydrins). The characterisation was done to evaluate the potential differences in structural and functional traits, despite these two dehydrins being categorized according to similar

sequence motifs. The existing techniques that are most commonly used to infer functional and structural characteristics were used to determine their efficacy at revealing the roles of each LEA protein. With the abundance of *in silico* analyses programs that exist, this study aims to assess the accuracy with which these programs are able to predict *in vitro* results. Since LEA proteins are intrinsically disordered and unstructured, it would be expected that these existing programs are geared towards predicting behaviours and traits of conventional proteins. Therefore, their reliability and accuracy can be questioned using *in vitro* techniques such as CD and functional enzyme assays. In doing so, this study aims to establish a structural and functional characterisation pipeline that can be applied to the remaining 21 desiccation-induced LEA proteins found in *X. humilis* and potentially to other LEA proteins in general.

Chapter 2
In silico analysis of two putative dehydrin sequences

2.1 Introduction

What started off as a few families of proteins classified according to their patterns of accumulation during the mature phase of cotton has rapidly expanded over the past twenty years into the group known as 'LEA proteins'. As a result, a number of classification systems with differing nomenclature have been developed (discussed in Chapter 1). With more than eight different methods of classification, the grouping of LEAs has proven itself to be overlapping or even contradictory at times (Wise, 2003). With bioinformatics analyses, Wise (2003) was able to demonstrate that, on the basis of common sequence domains, LEAs previously classified, as Group 5 should have been included in Group 3. This was achieved using a statistically based bioinformatics tool and demonstrates the usefulness of modern bioinformatics techniques in updating existing knowledge about LEA proteins.

Bioinformatic tools allow large sets of biological data to be analysed and categorized. This is an ability that is highly appealing when it comes to attempting to understand and characterize large numbers of LEA proteins. *In silico* methods of analyzing the sequences of putative LEA proteins have been highly beneficial as it is possible to predict the biochemical properties such as molecular weight, hydrophobicity, state of disorder and pI (Amara et al., 2014; Tunnacliffe & Wise, 2007). Since LEA proteins, and dehydrins in particular, are known to have distinct sequence specific characteristics and amino acid biases, bioinformatics tools enable these traits to be identified in a putative LEA before any *in vitro* or transgenic studies are performed.

One of the dominating challenges in the attempts to understand LEA protein structure and function is their lack of a stable three-dimensional conformations in the hydrated state (Habchi et al., 2014; Sun et al., 2013). Their intrinsically disordered nature has been hypothesized to be key to their ability to perform specific cellular functions and protective roles as molecular chaperones (Kovacs, Agoston, & Tompa, 2008). As a result, bioinformatics tools that are able to predict secondary structure from amino acid sequences are particularly useful. In a study looking at two stress-induced IDPs from *A. thaliana*, COR15A and COR15B, thirteen different secondary structure prediction programs were used

to generate a consensus of the most likely structures that each IDP would adopt (Thalhammer et al., 2010). From the thirteen programs, it was noted that all the investigated secondary prediction programs, with the exception of SOPMA, predict the secondary structure of the proteins in the dry state rather than in a fully hydrated state. This suggests that interpretation of results from such secondary structure prediction programs require knowledge about the hypothetical conditions used by the program. Furthermore, it is important to consider that these prediction programs are often designed with a bias towards 'regular' proteins and their structural changes. Therefore, the results generated by these programs should be run in conjunction with *in vitro* analyses to confirm predicted changes in secondary structure.

2.2 Aim

The objectives of the work described in this chapter are as follows:

a) Analyse the sequences of the two putative dehydrins from *X. humilis* using ExPASy, ProtParam and Clustal Omega to identify dehydrin specific characteristics.
b) Predict structural characteristics using disorder predictors such as PONDR-VL-XT and PONDR-FIT and various secondary structure prediction tools such as CFSSP, GOR V, PELE, SCRATCH and SOPMA.
c) Infer potential functions of the two putative dehydrins using BLAST, NetPhos 2.0 and Plant-mLOC and;
d) Identify putative transcription factor binding sites using the MEME suite with the 1kb upstream promoter region for the *X. viscosa* homologues of XhLEA2-4 and XhLEA2-5.

2.3 Methods and materials

2.3.1 Analysis of the two putative dehydrin sequences for LEA-like characteristics

The full-length nucleotide sequences of two putative dehydrins from *X. humilis* were obtained from the National Centre of Biotechnology Information (NCBI) database and translated into amino acid sequences using the ExPASy translation tool (http://web.expasy.org/translate/) (Appendix A1). The amino acid sequences were then analysed with ProtParam (http://web.expasy.org/protparam/) to obtain physico-chemical properties such as molecular weights, GRAVY scores (degree of hydropathicity) and aliphatic indices, as these would indicate whether or not the putative dehydrins bear LEA-like characteristics. The presence or absence of dehydrin-specific motifs such as the K-segment, the Y-segment motif or the S segment were also identified (Allagulova et al., 2003). ProtParam was then used to identify possible amino acid biases. Lastly, Clustal Omega was used to evaluate the sequence conservation between the two dehydrins since it is known that the degree of conservation of the K-segment sequence varies between dehydrins (Graether & Boddington, 2014).

2.3.2 Secondary structure predictions using various online prediction programs

The degree of disorder present in each dehydrin was predicted using two online platforms for predicting disorder in IDPs: PONDR-VL-XT (http://www.pondr.com) and PONDR-FIT (http://www.disprot.org/metapredictor.php). In a study by Hincha et al., (2011), it was concluded that the SOPMA algorithm predicted the structures of the proteins in the dry state, or *in vacuo,* and this was their preferred secondary structure prediction program *(*https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html). For XhLEA2-4 and XhLEA2-5, prediction programs using both the hydrated and dry states were investigated in order to compare differences in predictive accuracy. Secondary structures were predicted using a number of online tools, namely: CFSSP (http://www.biogem.org/tool/chou-fasman/), GOR IV (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html), SCRATCH (http://scratch.proteomics.ics.uci.edu) and SOPMA. The online prediction program PELE, available on the SDSC Biology Workbench (http://workbench.sdsc.edu), was also used with the following algorithms: DSC, GGR, GOR, H-K and K_S.

2.3.3 Functional predictions using available online tools

Homology searches were conducted using the Basic Local Alignment Search Tool (BLAST) against plant amino acid sequences within the NCBI databases. Since LEA proteins have been hypothesized to have a wide range of functions, comparisons made between XhLEA2-4, XhLEA2-5 and LEA proteins with known functions allow for predictions of the possible functions prior to *in vitro* analyses. Homologues of XhLEA2-4 and XhLEA2-5 found in the *X. viscosa* genome were also identified. Possible phosphorylation sites were predicted using NetPhos2.0 and Plant m-LOC was used to predict subcellular localisation.

2.3.4 Identification of putative *cis*-regulatory elements

Utilizing the *X. viscosa* genome and the homologues for XhLEA2-4 and XhLEA2-5, the 1kb regions upstream of the promoter sequences were obtained and subsequently analysed (Appendix A2). The MEME (Multiple Em for Motif Elicitation) suite, a motif based sequence analysis tool, was used to identify possible motifs within the promoter sequences. The MEME suite analysis was chosen to analyse the promoter regions because of the emphasis on the statistical significances of the motifs discovered. Alternative software programs such as PlantCARE and softberry lack reporting of the validity of their search results, making them less desirable. MEME identifies novel motifs from submitted sequences and selects the most likely patterns using statistical modeling (Bailey & Elkan, 1994). These identified motifs are then compared to existing databases of known regulatory elements or transcription factor binding sites using the TOMTOM Motif Comparison tool. Based on the motifs found in *A. thaliana* transcription factor (TF) binding sites, the motifs are investigated using the UniPROT database (http://www.uniprot.org/uniprot/) and then reported according to statistical significance.

2.4 Results and discussion

2.4.1 Sequence analysis of putative dehydrins for LEA-like characteristics

2.4.1.1 Physico-chemical properties

The nucleotide sequence of each dehydrin was translated using ExPASy and the resultant amino acid sequence is shown in Appendix A1. The amino acid sequence of each dehydrin was also analysed using ProtParam (http://web.expasy.org/translate/) for predicted molecular weight (MW), theoretical pI, GRAVY score and aliphatic index. The results obtained have been summarized in Table 2.1. The predicted molecular weights of both XhLEA2-4 and XhLEA2-5 were approximately 15kDa. Although this is consistent with the idea of LEA proteins being small and hydrophilic, LEA proteins have in fact been reported to span a range of sizes from 9.6kDa to 70kDa (Graether & Boddington, 2014). Nonetheless, one of the most unifying characteristics of LEA proteins is their hydrophilic nature. The most common method to confirm hydrophilicity, *in silico*, is through a GRAVY score that can be generated by the ExPASy server (Tunnacliffe & Wise, 2007). The scores range from 2 (hydrophobic) to -2 (hydrophilic). From the GRAVY scores of -1.43 and -1.25 for XhLEA2-4 and XhLEA2-5, respectively, it is evident that these two putative dehydrins are significantly hydrophilic.

Table 2.1 Physico-chemical properties of XhLEA2-4 and XhLEA2-5 using ProtParam

	MW (kDa)	Theoretical pI	GRAVY score	Alipathic index
XhLEA2-4	14.80	6.64	-1.43	30.29
XhLEA2-5	14.50	9.60	-1.25	34.06

2.4.1.2 Identification of dehydrin specific motifs and amino acid biases

Dehydrin-specific motifs that are present within each of the dehydrins have been highlighted in Figure 2.1. ProtParam was then used to calculate the amino acid content to identify any trends in amino acid composition and the results can be found in Table 2.2. XhLEA2-4 only contains one K-segment and this is followed by a His-His motif. Additionally, while Gly, Gln and Thr residues make up 41% of the protein, there is no defined Φ segment (Figure 2.1). The presence of only one K-segment does not provide any information about dehydrin architecture; therefore, functions or abiotic stress responses cannot be inferred. However, the His-His motif has been linked to a role in membrane binding suggesting that this may be a dehydrin involved in membrane stabilization roles (Eriksson et al., 2011). XhLEA2-5, on the other hand, appears to be more dehydrin-like since it contains two copies of the K-segment

and one copy of the S-segment. The K-segment is likely to be an inducer of amphipathic α-helical formation while the S-segment is known to be a phosphorylation site in maize (RAB17) and tomato (TAS16) dehydrins (Allagulova et al., 2003). This arrangement of the two motifs gives the dehydrin an SK_n architecture (Close, 1996). The SK_n architecture has been observed in dehydrins from Ruby Grapefruit (cpDHN), Poplar trees (Peudhn1) and Arabidopsis (ERD14) and, while localisation cannot be inferred, all three dehydrins are known to respond to cold stress (Caruso, Morabito, Delmotte, Kahlem, & Carpin, 2002; Kovacs, Kalmar, Torok, & Tompa, 2008; Porat et al., 2004). A single His-His motif is also observed, flanking the second K-segment, suggesting a role in possible membrane stabilization. From the amino acid composition, XhLEA2-5 contains a large number of Gly, Gln and Thr residues (31.2%) with a notable Φ segment from amino acid position 101, which spans 18 residues. There is also an absence of Cys and Trp residues. The high percentage of hydrophilic amino acids supports the hydrophilic nature of the dehydrin predicted by the GRAVY score (section 2.4.1.1).

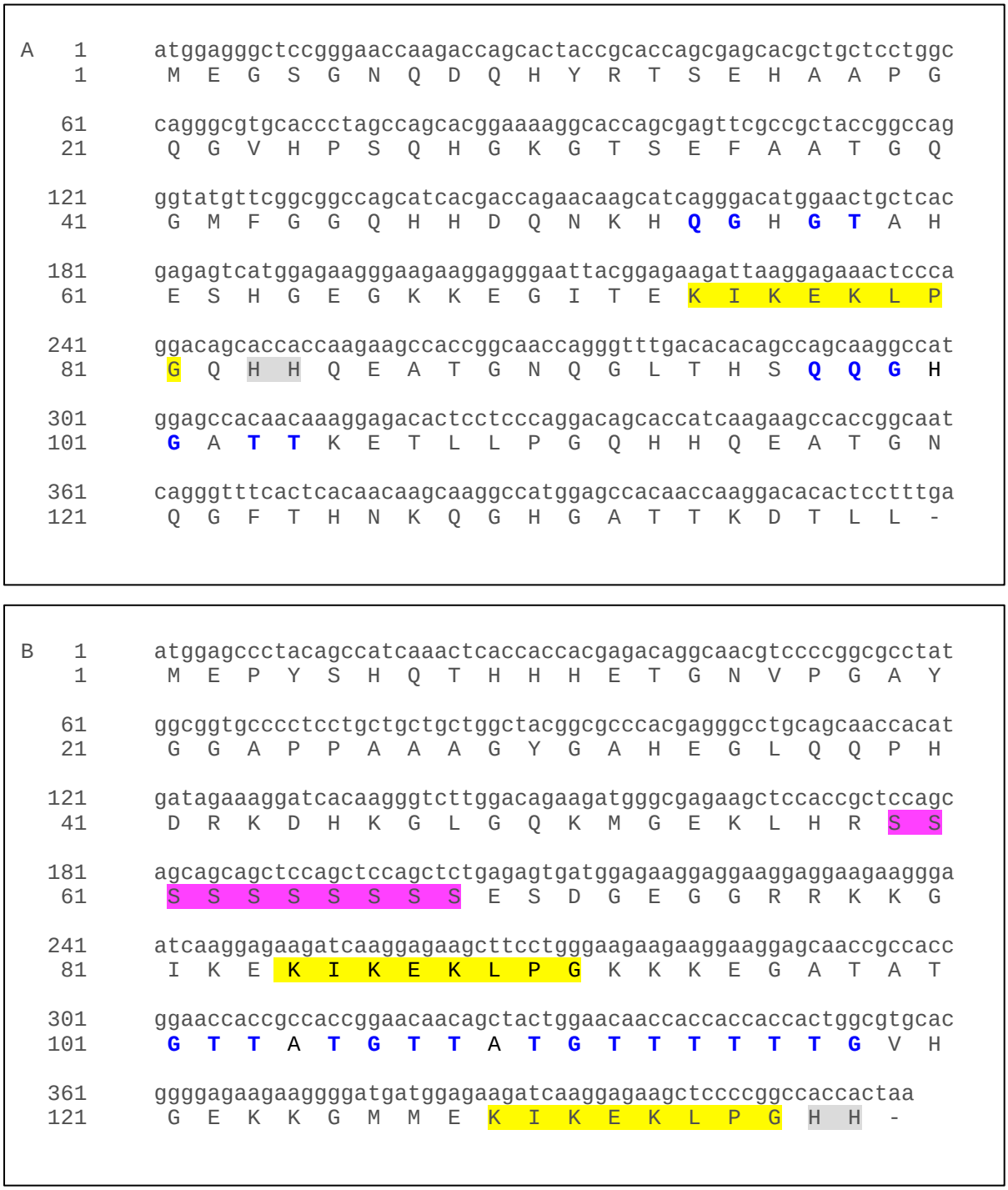


Figure 2.1 Translated sequences of XhLEA2-4 (A) and XhLEA2-5 (B) with conserved motifs highlighted. The K-segment has been highlighted in yellow and the S-segment in purple. His-His tags flanking the K-segments have been highlighted in grey. The Φ segments are in blue.

Table 2.2 Amino acid composition for each dehydrin generated by ProtParam.

Group	Amino Acid	Percentage Composition (%)			
		XhLEA2-4		XhLEA2-5	
Non-polar (Hydrophobic)	Ala (A)	5.8	22.4	7.2	19.4
	Ile (I)	1.4		2.2	
	Leu (L)	4.3		3.6	
	Met (M)	1.4		2.9	
	Phe (F)	**2.2**		**0.0**	
	Pro (P)	3.6		5.1	
	Trp (W)	0.0		0.0	
	Val (V)	0.7		1.4	
Polar (Hydrophilic)	Asn (N)	3.6	40.6	0.7	48.9
	Cys (C)	**0.0**		**0.0**	
	Gln (Q)	**12.2**		**2.9**	
	Gly (G)	**18.0**		**16.7**	
	Ser (S)	4.3		8.7	
	Thr (T)	**10.8**		**11.6**	
	Trp (W)	**0.0**		**0.0**	
Acidic	Asp (D)	2.2	10.9	2.2	10.1
	Glu (E)	7.9		8.7	
Basic	Arg (R)	0.7	23.9	2.9	20.8
	His (H)	12.9		8.0	
	Lys (K)	7.2		13.0	

From the two analyses conducted, it is evident that both of the dehydrin sequences contain dehydrin-specific motifs and amino acid biases. These characteristics allow predictions to be made regarding identity, as well as function or responses to abiotic stress. This will be explored further in subsequent sections.

2.4.1.3 Sequence conservation between the two dehydrins

Clustal Omega was used to compare the percentage similarity between XhLEA2-4 and XhLEA2-5. The two sequences showed significant similarity on the amino acid level (47%, E value = 2e-7) but no similarity on the nucleotide level. The amino acid sequence alignment generated is shown in Figure 2.2. As expected, the K-segment sequence ‘EKIKEKLPG’ is most conserved between the two dehydrins. It is also important to note that in regions where the specific amino acids residues are not conserved, there is still conservation between amino acid groups with similar properties.

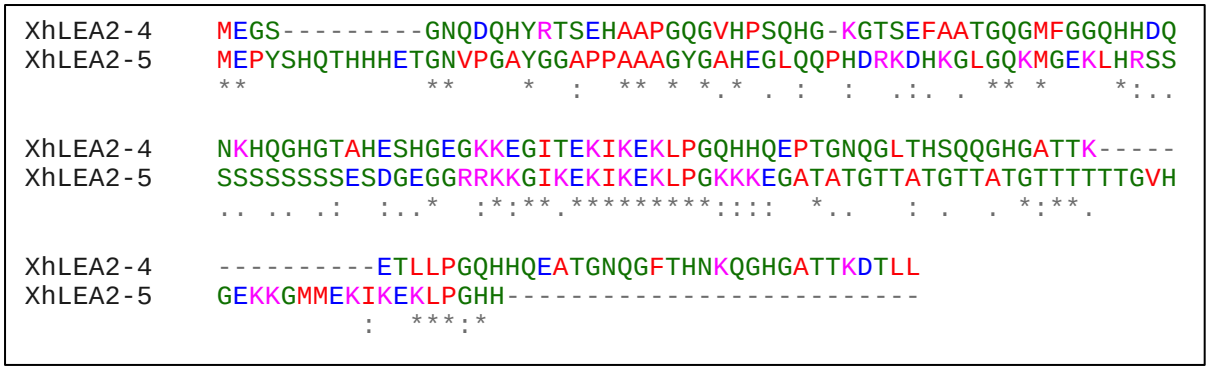


Figure 2.1 CLUSTAL O (1.2.1) Sequence Alignment between XhLEA2-4 and XhLEA2-5. An asterisk (*) indicates a position with a single, fully conserved residue. The colon (:) indicates conservation between groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix. A period (.) indicates conservation between groups of weakly similar properties - scoring = < 0.5 in the Gonnet PAM 250 matrix.

2.4.2 Predictions of structural characteristics using online prediction programs

2.4.2.1 Disorder predictions using PONDR and PONDR-FIT

In previous studies by Ginbot (2011) and Waters (2014), the PONDR-VL-XT platform was used to predict disorder properties for group 1, 2 and 3 LEA proteins. Whilst many versions of PONDR exist, the VL-XT combination of algorithms was selected as the most accurate since it merges three disorder predictors – one for characterizing long disordered regions (>40 residues) and two for X-ray characterized regions at both the C and N terminal (Romero, Obradovic, & Dunker, 1997). Each amino acid is scored according to its level of disorder – above 0.5 for disordered and below 0.5 for ordered. Plots are then generated with the disorder scores for each residue. PONDR-VL-XT was used to generate disorder plots for XhLEA2-4 and XhLEA2-5 (Figure 2.3 A and B, respectively). From the results, it is suggested that the two dehydrins differ in terms of disorder. Whilst XhLEA2-4 showed disorder in only two small regions, each containing approximately 20 residues, XhLEA2-5 appeared to be predominantly disordered, with a continuous region of disorder spanning 90. These predictions support the results observed from in-solution studies on LEA proteins using CD spectroscopy which suggest that, while dehydrins are generally disordered proteins, the K segments have been shown to be responsible for the formation of transient or weak helical structures (Graether & Boddington, 2014). However, since over 50 programs are currently available online for predicting disorder in proteins, it is necessary to compare results from more than one predictor (Xue, Dunbrack, Williams, Dunker, & Uversky, 2010). According to Xue et al., (2010), PONDR-FIT is a meta-predictor that incorporates newly

developed meta-predictors that aimed to improve disorder prediction accuracy with existing PONDR predictors. This has resulted in a meta-predictor that makes use of previously unexplored combinations of prediction algorithms. According to PONDR-FIT, both XhLEA2-4 and XhLEA2-5 were predicted to be highly disordered proteins (Figure 2.4 A and B, respectively) as all residues scored between 0.75 and 1. This finding differs from the results generated by PONDR-VL-XT and will require validation from experimental data in section 3.4.18, using CD spectroscopy.

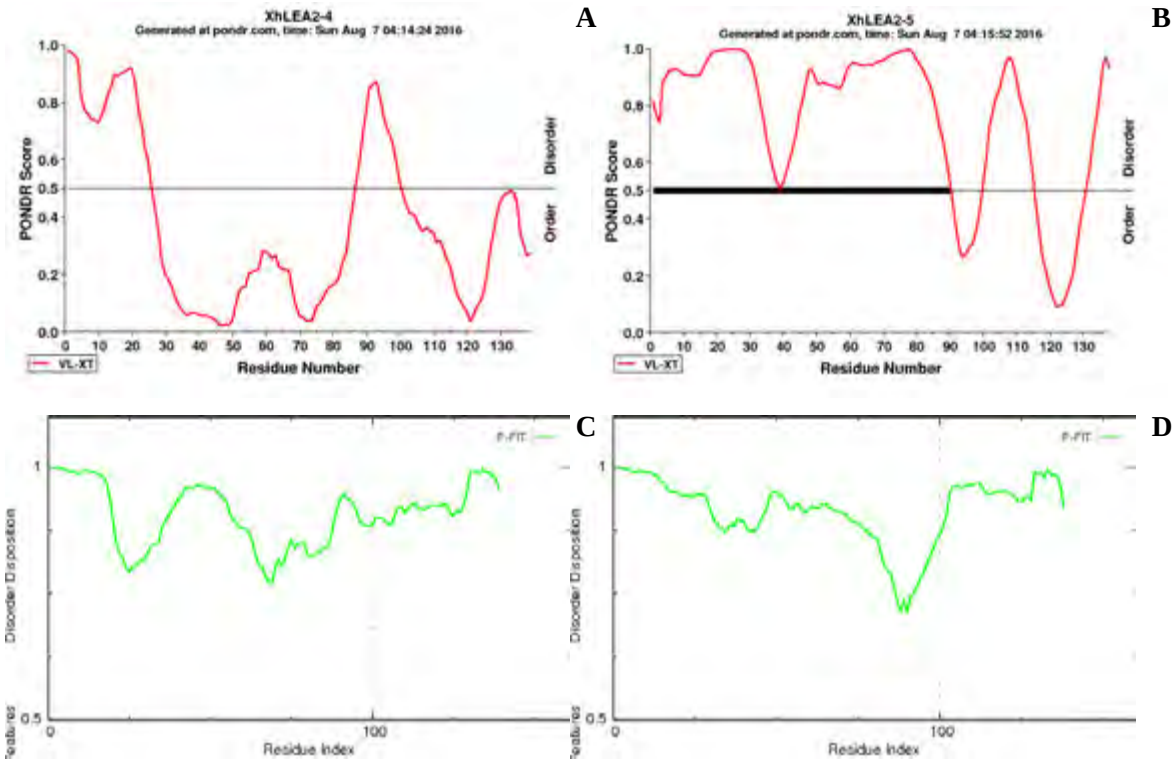


Figure 2.3 Disorder plots generated for XhLEA2-4 (A and C) and XhLEA2-5 (B and D) using PONDR-VL-XT (A and B) and PONDR-FIT (C and D). Amino acids that score above 0.5 are considered to be disordered. Regions of disorder are indicated by a bold black line.

2.4.2.2 Secondary structure predictions using various online programs

2.4.2.2.1 Secondary structure predictions in the hydrated state

SOPMA predicted that XhLEA2-4 was 11.5% α-helical in the hydrated state and XhLEA2-5 was 18.12% α-helical (Figure 2.4). Both dehydrins were predicted to consist predominantly of random coils. In Figure 2.1, it was identified that Xh-LEA2-5 contained two α-helical-forming K-segments and this could explain the slight difference in the predicted α-helix percentage.

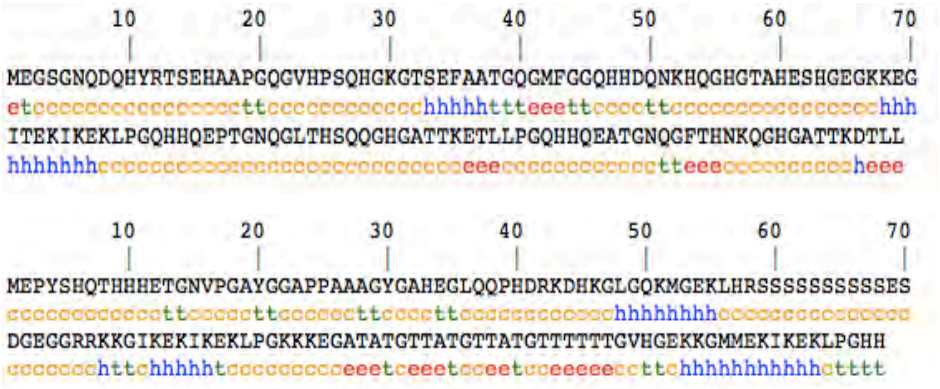


Figure 2.4 Secondary structure predictions for XhLEA2-4 and XhLEA2-5 by SOPMA.

2.4.2.2.2 Secondary structure predictions in *vacuo*

The Chou & Fasman Secondary Structure Prediction (CFSSP) server predicted that XhLEA2-4 and XhLEA2-5 were 32.4% and 45.1% α-helical, respectively (Table 2.3). This method is one of the earliest prediction techniques and implements empirical rules for predicting the initiation and termination of α-helices and β-sheets (Chou & Fasman, 1974). However, it was later noted that this method tended to over predict the percentage of secondary structures present (Chen, Gu, & Huang, 2006).

Table 2.3 Percentage of random coils (C), extended sheets (E), α helices (H) and β turns (T) in XhLEA2-4 and XhLEA2-5 predicted by four online prediction tools.

	XhLEA2-4 (% Secondary structures)				XhLEA2-5 (% Secondary structures)			
	C	E	H	T	C	E	H	T
CFSSP	8.6	37.4	32.4	21.6	17.8	23.5	45.1	13.6
GOR V	80.5	14.5	5	-	80	3.5	16.5	-
SCRATCH	88	2	10	-	80	4	16	-
PELE	68.5	18.5	13	-	56.5	17	26.5	-

The Gor V protein secondary structure prediction server, on the other hand, predicted XhLEA2-4 and XhLEA2-5 to be 5% and 16.5% α-helical, respectively (Table 2.3). The algorithms used by this server combine various theories and statistical methods to provide a predictive accuracy of 73.5% (Sen, Jernigan, Garnier, & Kloczkowski, 2005). Similarly, the SCRATCH protein predictor server predicted that XhLEA2-4 and XhLEA2-5 were 10% and

16% α-helical, respectively. Both dehydrins were predicted to be predominantly random coils by GOR V and SCRATCH. The prediction tool PELE from the SDSC Biology Workbench was also used to predict secondary structure in the two dehydrins. Five of the available algorithms (DSC, GGR, GOR, H_K and K_S) were used to run the predictions and the results were averaged. PELE predicted slightly higher percentages of α-helices in XhLEA2-4 and XhLEA2-5 (13% and 26.5%, respectively).

2.4.3 *In silico* analysis of elements that may infer function

Homology searches were conducted for the amino acid sequence of each dehydrin using the NCBI tool, BLASTp. Whilst the search for XhLEA2-4 did not identify any putative conserved domains, the most homologous proteins appear to all be dehydrins proteins (Table 2.4). Two dehydrins from *Triticum aestivum* were the most closely related proteins, sharing around 50% sequence similarity. XhLEA2-5, on the other hand, showed significant similarity to various dehydrins, including the dehydrin, Rab 18, from *Tarenaya hassleriana* at 64% sequence similarity. The search also identified a conserved domain belonging to the dehydrin superfamily found within XhLEA2-5 (e value = 2.65e-4). These results further confirm the identity of the two LEA proteins as putative dehydrins.

Table 2.4 Homology search against the NCBI database for XhLEA2-4 and XhLEA2-5.

XhLEA2-4	Total Score	Query Cover	E value	% Identity	Accession
Dehydrin [Triticum aestivum]	47.0	45%	3e-4	49%	CAY85463.1
Dehydrin [Triticum aestivum]	44.6	45%	2e-2	47%	BAF30987.1
XhLEA2-5	**Total Score**	**Query Cover**	**E value**	**% Identity**	**Accession**
Rab18 [Tarenaya hassleriana]	75.1	60%	1e-14	64%	XP_010550431.1
DHN1-like [Cucumis sativus]	74.3	60%	4e-14	64%	XP_011653150.1
DHN1 [Cucumis melo]	72.4	60%	3e-13	63%	XP_008452109.1
DHN2 [Triticum urartu]	70.1	58%	3e-13	65%	EMS50665.1
Rab18-like [Jatropha curcas]	70.9	86%	6e-13	55%	NP_001295638.1

The homologues of XhLEA2-4 and XhLEA2-5 present in the *X. viscosa* genome are shown in Table 2.5. XhLEA2-4 showed high sequence similarity to two dehydrin from *X. viscosa,* with a percentage similarity of 86 and 89%. Since it is known that whole genome duplications (WGD) played a critical role in the expansion of LEA families in *X. viscosa*, this could explain the increased number of *X. viscosa* homologues. XhLEA2-5 also showed

significant similarity to two dehydrins from *X. viscosa,* however, the percentage of sequence similarity is noticeably less (39 and 56%). Nonetheless, it is useful to note the presence of homologues to XhLEA2-4 and XhLEA2-5 in other resurrection plant species as this may increase the likelihood of obtaining a more unified understanding of the physiological roles or functions of these elusive LEA proteins (Hundertmark & Hincha, 2008). In addition, having access to the sequences of these known homologues enable further bioinformatics analyses to be conducted, with subsequent inferences to be made. This has been done for the promoter region analyses in section 2.4.4.

Table 2.5 Homology search within the *X. viscosa* genome for XhLEA2-4 and XhLEA2-5.

X. humilis	Homolog in *X. viscosa*	Score	Identities (Query length)	%	E value	Pfam Group
XhLEA2-4	Xvis02_20155_PA	176	124/139 (138)	89	1e-54	Dehydrin
	Xvis02_13077_PA	158	94/109 (138)	86	4e-47	Dehydrin
XhLEA2-5	Xvis02_01738_PA	108	71/126 (139)	56	1e-28	Dehydrin
	Xvis02_23543_PA	42.7	42/104 (139)	39	3e-06	Dehydrin

It is known that phosphorylation of a protein can regulate or modify its function in various ways (Cohen, 2000). Possible sites of phosphorylation on proteins can provide insight into function since in LEA proteins, it is known that phosphorylation of the S-segment can result in a dehydrin moving from the cytosol to the nucleus (Graether & Boddington, 2014). NetPhos 2.0 predicts the phosphorylation potential of serine, threonine and tyrosine residues against a threshold. Although XhLEA2-4 does not contain an S-segment, NetPhos2.0 found seven phosphorylation sites at serine and threonine residues (Figure 2.5). The likelihood of these residues being phosphorylation sites suggests that the dehydrin itself may be involved in interactions with other signal peptides. XhLEA2-5 contains an S-segment and this region of eleven serine residues, along an additional tyrosine residue and two threonine residues, have been predicted to be phosphorylation sites. Similar to the maize dehydrin Rab17, XhLEA2-5 may require phosphorylation of the S-segment in order to allow the dehydrin to translocate into the nucleus for subsequent cell signaling functions (Graether & Boddington, 2014). The presence of possible phosphorylation sites indicates that these two dehydrins may be involved in signal transduction pathways and cell signaling.

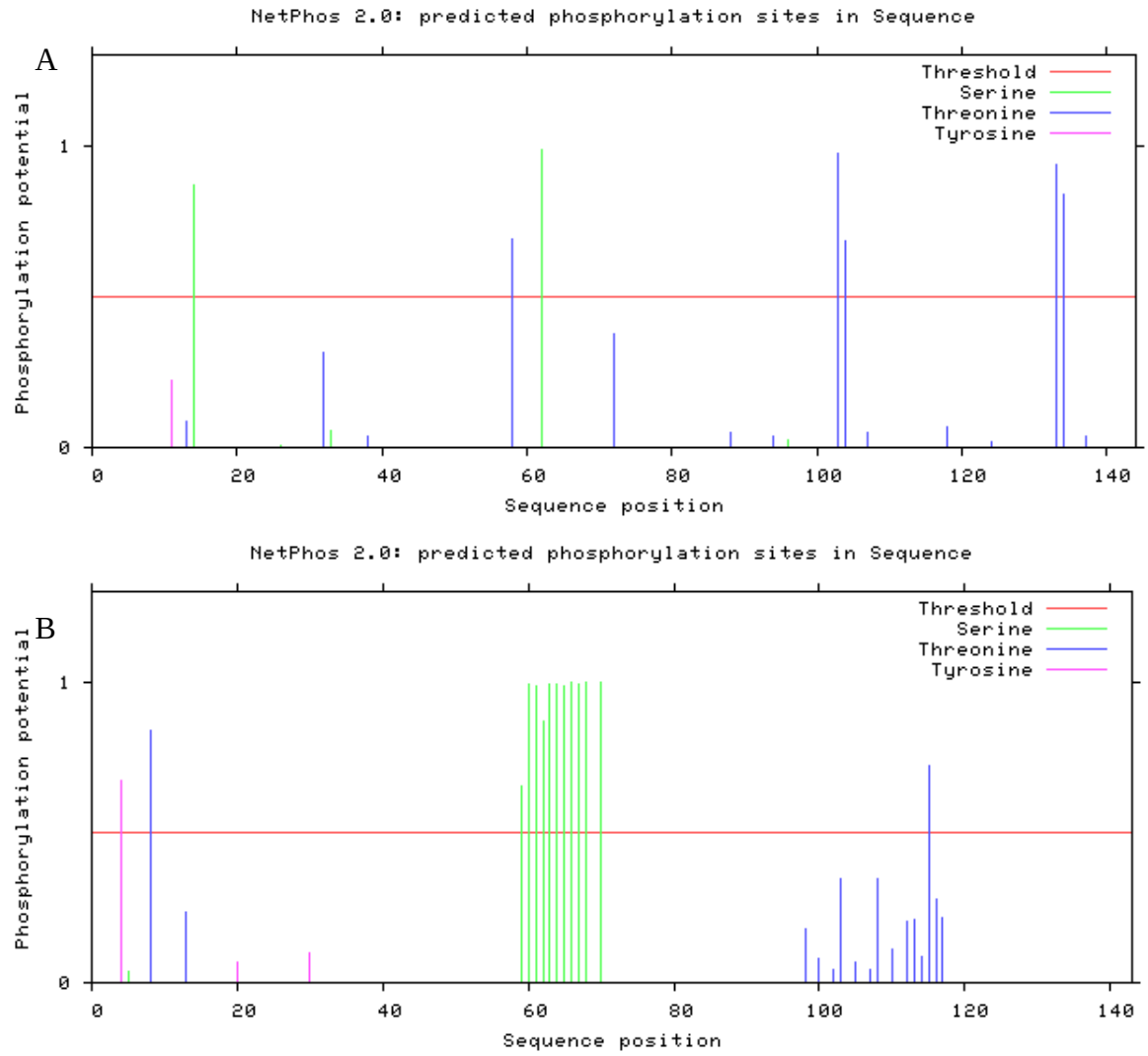


Figure 2.5 Possible phosphorylation sites on XhLEA2-4 (A) and XhLEA2-5 (B) predicted by NetPhos 2.0.

2.4.4 Identification of putative *cis*-regulatory elements using online databases

The 1kb upstream promoter regions of the homologue Xvis02_1738 was used for XhLEA2-4 and the upstream promoter regions of Xvis02_20155 and Xvis02_13077 were used for XhLEA2-5. Two homologues of XhLEA2-4 were used since both sequence similarity percentages were above 85%. Within the upstream promoter region of Xvis02_20155, one statistically significant motif was found (Figure 2.6). When this motif was scanned against the database of *A. thaliana* promoter elements, four transcription factor (TF) binding sites were identified – ARR11, CCA1_2, DREB2C_2 and TOE2. The most noteworthy TF binding site discovered in this search is the DREB2C_2 site. DREB2 (dehydration-responsive element binding) sites in *A. thaliana* are known to play an important role in regulating expression of stress-induced genes in a manner independent of ABA (Lata & Prasad, 2011).

Additionally, they are known to be responsible for inducing responses to abiotic stresses such as high salinity. TOE2 is also a transcription factor known to be involved in regulating gene expression during stress responses. These two putative transcription factor binding sites suggest that this homologue of XhLEA2-4 may play a role in stress response. Similarly, the promoter region for Xvis02_1307 contains a motif that resembles binding sites for HSFB2A_2 and HSFC1 – these being heat stress response transcription factors involved in temperature dependent phosphorylation. The presence of temperature responsive regulatory elements can be explained by the chasmophytic nature of *X. viscosa.* The term 'chasmophyte' refers to the ability to tolerate the extreme temperatures, which *X. viscosa* does since the temperatures in its environment that ranges from below zero to over 50ºC (Farrant et al., 2015). The presence of the regulatory element involved in lateral root formation (LBD16) suggests that this dehydrin may be a LEA protein that is localized to the roots. However, the presence of the DAG2 transcription factor binding site also suggests that the dehydrin may play a role in seed development. Results such as these should be interpreted as being speculative and would require further investigation using assays to identify transcription factor binding to these promoter elements *in vitro.*

The upstream promoter region of Xvis02_1738 was predicted to have regulatory sites for transcription factors similar to the promoter region of XhLEA1307 – two heat stress response transcription factors (HSFC1 and HSFB2A_2) and DAG2, a transcription factor involved in maternal seed germination. A transcription factor binding site for GATA12 was also discovered in this promoter sequence. GATA12 is involved in dictating tissue specific cell differentiation as well as responses to light responsiveness. Light responsive elements are interesting because of their possible role in protecting the plant from photosynthesis-related damage as the water content drops and photosynthetic machinery is broken down.

This investigation into the possible *cis*-regulatory elements present in the two dehydrins from *X. humilis* uses the upstream promoter regions of three homologues in a close relative, *X. viscosa.* Particularly in the case of XhLEA2-5, which has a fairly low sequence similarity to its homologue Xvis02_1738, these results are considered to be suggestive. As these *in silico* analyses are done based on sequence similarity, it is important to interpret the results as preliminary predictions of how the protein may behave in *in vitro* or *in vivo* conditions.

Table 2.6 Upstream promoter sequence analysis for XhLEA2-4 and XhLEA2-5 homologues using the MEME suite. Two homologues from *X. viscosa* were used for XhLEA2-4: Xvis02_20155 and Xvis-2_1307; and Xvis02_1738 was used for XhLEA2-5.

Identified motif element sequence		**p value**
Xvis02_20155		
		4.53e-19
Matched *A. thaliana* regulatory elements		
ARR11	Regulator in His-to-Asp phosphor-relay signal transduction.	4.67e-03
CCA1_2	Involved in circadian and phytochrome regulation.	6.52e-02
DREB2C_2	Involved in ABA-induce transcription. Mediator of high salinity.	6.99e-02
TOE2	Regulates gene expression by stress factors.	7.49e-02
Xvis02_1307		
		3.35e-21
Matched *A. thaliana* regulatory elements		
HSFB2A_2	Involved in temperature dependent phosphorylation.	1.33e-02
DAG2	Involved in maternal control of seed germination.	1.66e-02
HSFC1	Exhibits temperature dependent phosphorylation.	3.25e-02
LBD16	Involved in lateral root formation.	6.16e-02
Xvis02_1738		
		7.70e-25
Matched *A. thaliana* regulatory elements		
HSFC1	DNA binding. Exhibits temperature dependent phosphorylation.	3.38e-02
GATA12	Involved in cell differentiation and light responsiveness.	4.62e-02
DAG2	Involved in maternal control of seed germination.	5.00e-02
HSFB2A_2	Involved in stress response, localized to the nucleus.	5.00e-02
AHL12	DNA binding, protein self association	8.68e-02

Chapter 3
Cloning, expression and purification of two recombinant His-tagged dehydrin proteins

3.1 Introduction

Studies conducted on recombinant LEA proteins are often favoured as they enable *in vitro* analyses that can provide a great amount of detail. Numerous methods of producing recombinant LEA proteins have been documented – these include expression and production in *E. coli* and yeast with a great variety of cloning strategies and use of vectors (Amara et al., 2012; Dang, Popova, Hundertmark, & Hincha, 2014; Goyal, Walton, & Tunnacliffe, 2005; Thalhammer et al., 2010). Previous studies by Ginbot (2011) and Waters (2014) have shown successful expression and purification of two group 1, two group 2 and one group 3 recombinant LEA proteins originally identified in *X. humilis.* The recombinant proteins produced by these in vitro systems enable comparisons to be made to the *in silico* predictions. In particular, the proteins can be monitored using circular dichroism (CD) which allows changes in the percentages of secondary structures present in a protein to be observed under various states of hydration (Thalhammer et al., 2010). Using CD spectroscopy, Thalhammer et al., (2010) showed that while COR15A and COR15B were unstructured in dH_2O, undergoing dehydration promoted formation of α-helical structures. Similarly, Ginbot (2011) showed that in the presence of a secondary structure inducer such as trifluoroethylene (TFE) resulted in both XhLEA1-1S2 and XhLEA1-4S1 adopted an α-helical shape.

3.2 Aim

The experimental objectives described in this chapter are as follows:

a) Clone two LEA dehydrin genes, XhLEA2-4 and XhLEA2-5, into *E. coli* using a modified pET21a(+):His vector for protein expression, followed by the evaluation of soluble, expressed protein.
b) Conduct preliminary secondary structural analyses on the two purified dehydrins using CD spectroscopy.

3.3 Methods and materials

3.3.1 *Escherichia coli* as an expression host for recombinant LEA protein production

Escherichia coli (*E. coli*) is the preferred expression host used in laboratories conducting high-throughput cloning, expression and purification of recombinant proteins for structural protein analyses (Rosano & Ceccarelli, 2014). As an expression host, *E. coli* has been extensively studied in terms of its genetics and physiology, and in the use of genetic tools for, *inter alia*, accelerated growth rates high protein production yields and cost efficacy. However, several drawbacks of expression in this prokaryotic system are also to be noted and have been summarized in Table 3.1. Whilst expression of mammalian proteins in bacterial systems has shown to be problematic due to the lack of specialized post-translational modifications (PTMs) necessary for folding, expression of plant proteins has been significantly more successful. In particular, Goyal et al., (2005) showed that recombinant LEA proteins expressed by the *E. coli* BL21 (DE3) system could be purified and subjected to functional studies. Subsequently studies by both Ginbot (2011) and Waters (2015) showed that soluble recombinant LEA protein expression was possible in *E coli* BL21 DE3 *pLysS* (Novagen, USA). Therefore, the *E. coli* BL21 DE3 *pLysS* expression system was selected for recombinant expression of the LEA dehydrin proteins in this study.

Table 3.1 Advantages and disadvantages of utilizing *E. coli* as a heterologous expression host, adapted from Peton (2013).

System	Advantages	Disadvantages
Escherichia coli	- Rapid proliferation - Easy insertion of foreign DNA - Recombinant cultures are economically and easily produced - Recombinant proteins can be produced in high quantities - Adequate tolerance of additives	- Post-translational modifications (PTMs) required for eukaryotic proteins are not possible - Codon bias - Possibility of endotoxin contamination - Inclusion bodies may make purification of recombinant protein difficult - Amount of functional proteins produced is proportionally smaller - Eukaryotic proteins are often insoluble following expression in *E. coli*

Having selected *E. coli* as the expression host, a cloning and expression strategy was designed to generate soluble recombinant LEA dehydrin proteins.

3.3.2 Selection and modification of plasmid vector

The expression vector selected for cloning was the pET21a(+) plasmid (Novagen, USA). As shown in Figure 3.1, this *E. coli* expression vector consists of the following elements: a T7 promoter/lac operator element which allows high-level protein production and is tightly regulated to reduce metabolic stress and toxic effects; a coding sequence for a T7 tag which allows for easy detection of recombinant protein in subsequent steps and a β-lactamase coding sequence which confers ampicillin resistance. Previous unpublished work in our lab showed that the T7-tag produced difficulties in the process of protein detection via western blots. Therefore, this tag was deleted and a 6x Histidine tag was introduced at the N-terminus via inverse-PCR mutagenesis.

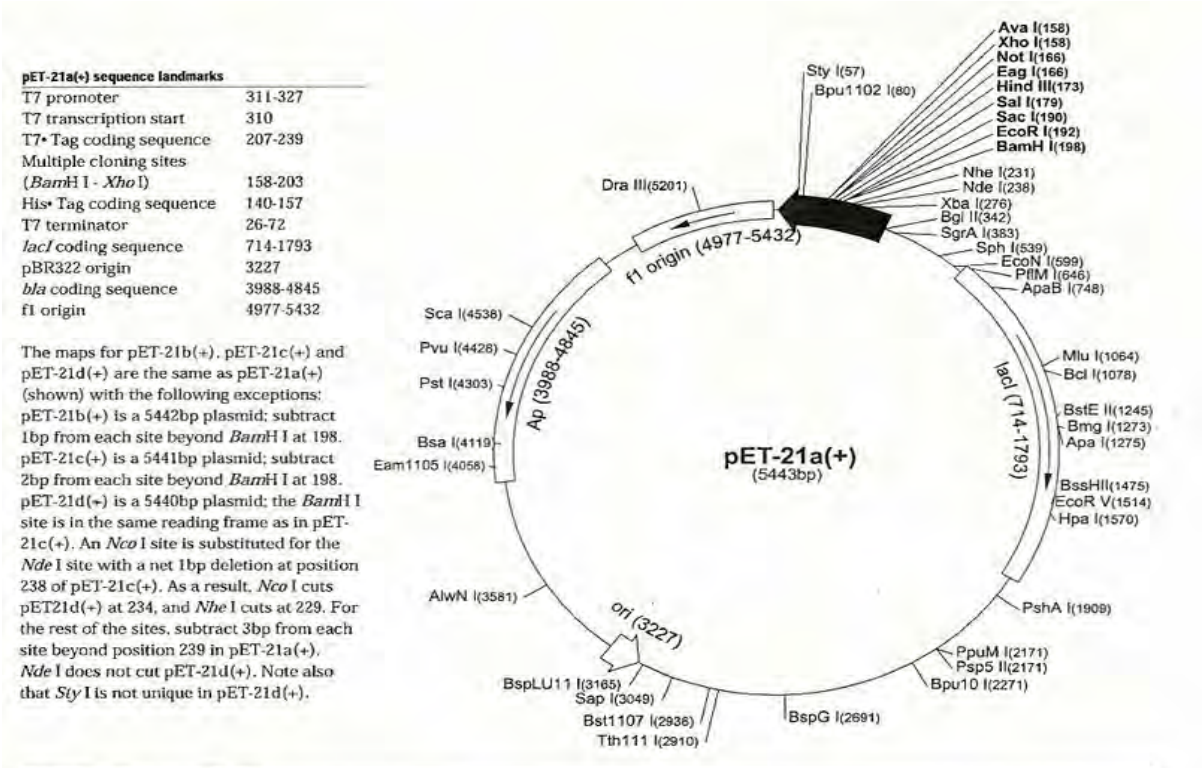


Figure 3.1 Plasmid map of the pET21a(+) vector backbone (Novagen, USA).

Large quantities of this plasmid were generated for cloning as follows: glycerol stocks were streaked out onto LB Agar plates containing the antibiotic ampicillin (final concentration 100 μg/ml) and incubated overnight at 37ºC. Individual colonies were then inoculated into several 5ml LB cultures containing ampicillin and further grown overnight at 37ºC, with constant

shaking. The plasmid DNA was then extracted from these cultures using the Wizard® Plus SV Miniprep DNA Purification System (Promega, USA) with the addition of Alkaline Protease (Thermoscientific, USA) to minimize endonuclease activity. Following elution, isolated plasmid DNA was pooled and quantitated at 260nm using the NanoDrop® ND-100 Spectrophotometer (Thermo Scientific, USA). The modified plasmid pET21a(+) will now be referred to as pET21a(+):His.

3.3.3 Analysis of the vector and two target genes for restriction enzyme site selection.

The multiple cloning site (MCS) of the pET21a(+):His plasmid was analysed for suitable restriction enzyme sites. Two specific restriction sites, *SalI* and *NotI*, were selected for the downstream cloning. The NEB cutter program (http://nc2.neb.com/NEBcutter2/) was used to ensure that the selected restriction sites were not present within the nucleotide sequences of the two dehydrins.

3.3.4 Primer design and synthesis for XhLEA2-4 and XhLEA2-5

The full-length nucleotide sequences for XhLEA2-4 and XhLEA2-5 were previously obtained from the NCBI server (http://www.ncbi.nlm.nih.gov/). In order to facilitate cloning of the PCR products into the MCS of the expression vector pET21a(+):His, restriction enzyme sites for *SalI* and *NotI* were incorporated into the 5' ends of the primer sequences. The OligoAnalyzer® Tool (SciTools®) was used to determine the primer set with melting temperatures (T_M) closest to 60ºC to minimize non-specific binding. These primers were synthesized using standard methods provided by the Synthetic DNA Laboratory (Molecular and Cell Biology Department, University of Cape Town). The primer sets were designed to amplify the full-length gene sequencs of each LEA dehydrin from the storage vector, pBluescript SK, and are listed in Table 3.2.

Table 3.2 Primer sequences for XhLEA2-4 andXhLEA2-5. Restriction enzyme sites for *SalI* and *NotI* are underlined.

Primer	Sequence 5'➔3'	T_m (ºC)
XhLEA2-4 Forward	**CTC CGT CGA CAA** ATG GAG GGC TCC GGG AAC	68.3
XhLEA2-4 Reverse	**GCT TGC GGC CGC** TCA AAG GAG TGT GTC CTT G	68.4
XhLEA2-5 Forward	**CTC CGT CGA CAA** ATG GAG CCC TAC AGC CAT CAA ACT CAC	69.7
XhLEA2-5 Reverse	**GCT TGC GGC CGC** TTA GTG GTG GCC GG	68.9

3.3.5 Plasmid DNA extraction and quantification

The two dehydrin genes were maintained in pBluescript SK vectors in *E. coli* DH5α cells. Freshly plated colonies grown on LB-Amp agar were used to inoculate 5ml of LB-Amp media. Cultures were grown overnight at 37ºC, shaking, plasmid DNA was subsequently extracted and quantified according to the previously described method (section 3.3.2).

3.3.6 Polymerase Chain Reaction (PCR) amplification of full-length dehydrins

PCR amplification was performed using the KAPA High Fidelity PCR Kit (Kapa Biosystems, South Africa) in the following 50µl reaction: 0.5µM Primers, 50ng plasmid DNA, 0.5mM $MgCl^{2+}$, 0.4mM dNTPs and 1U of KAPA HiFi polymerase. The reactions were run according to the parameters found in Table 3.3 below. The PCR products were resolved by gel electrophoresis on a 1% agarose gel at 90V for 30 minutes. Visualisation was performed using the Molecular Imager ChemiDoc XRS+ system (Bio-Rad, Germany).

Table 3.3 Parameters for the amplification of the full-length dehydrin genes. Annealing temperatures (T_A) for XhLEA2-4 and XhLEA2-5 are 62ºC and 63ºC, respectively.

Step	Temperature (ºC)	Duration	Cycles
Initial denaturation	95	3 minutes	1
Denaturation	98	20 seconds	35
Annealing	T_A	15 seconds	
Extension	72	30 seconds	
Final extension	72	3 min	1

3.3.7 Restriction enzyme digestion of PCR product and pET21a(+):His vector DNA

Prior to the restriction enzyme (RE) digest, a PCR clean up was performed using the Wizard® SV Gel and PCR Clean-up System Kit (Promega, USA). Products were eluted, pooled and quantified using methods as previously described. The purified PCR products for each dehydrin along with the vector DNA were subjected to RE double digestion such that the PCR products could be uni-directionally ligated into the plasmid vector. FastDigest *SalI* and *NotI* (Thermo Scientific, USA) restriction enzymes were used according to standard protocols and the manufacturer's instructions. Each digest consisted of 2µg of DNA, 10 U of each enzyme, 1 x RE buffer and distilled water up to a final volume of 60µl. Digests were

incubated at 37ºC for 30 minutes and subsequently resolved on a 0.8% agarose gel for 35 minutes at 90V. Gel fragments containing the restriction product of interest were then excised and the DNA was purified using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, USA) according to the manufacturer's instructions.

3.3.8 DNA ligations

The gel-purified dehydrin DNA fragments were ligated into the pET21a(+):His vector using a 3:1 insert to vector molar ratio with T4 DNA ligase (#EL0011, Thermo Scientific). The ligation reaction contained approximately 1µg of DNA, 2U of ligase, 1x ligase buffer, 5µM ATP and water, in a final volume of 20µl and was incubated at 22ºC for one hour.

3.3.9 Generation and screening of transformed *E. coli* with recombinant DNA.

Competent DH5α *E. coli* cells were prepared using the Rubidium Chloride (RbCl) method (Promega Protocols and Applications guide, 3rd edition, p45-46). A 10µl volume of ligation reaction was added to 100µl of DH5α competent cells, the tube was mixed briefly and then incubated on ice for 30 minutes. The transformation mix was subjected to a heat shock step at 42ºC for 90 seconds, followed by another incubation step on ice for 2 minutes. Subsequently, 900µl of pre-warmed LB was added to each tube and the tubes were incubated at 37ºC, shaking, for four hours. The cells were then plated onto LB-Amp agar plates and incubated overnight at 37ºC. To screen for positive transformants that could be sent for sequence verification, a colony PCR was set up using previously described parameters (Table 2.2) and the KAPA Taq PCR Kit (Kapa Biosystems, Cape Town). Positive colonies were identified by analysis of the PCR products on a 1% agarose gel and subsequently inoculated into 5ml LB-Amp cultures which were grown overnight at 37ºC, shaking. From each 5ml culture, 500µl was used with an equal amount of 50% glycerol to make a glycerol stock and the rest used to extract and quantify recombinant plasmid DNA using the previously described methods (section 3.3.2). From the extracted pET21a(+)His-LEA DNA, 20µl of a 100ng/µl sample was sent for sequencing by Macrogen (Netherlands) using the T7 promoter and T7 terminator primers. Sequence data obtained was analysed to ensure that the recombinant plasmid had been cloned in-frame using ExPASy (http://web.expasy.org/translate/).

3.3.10 Recombinant DNA transformation for protein expression

The recombinant plasmid DNA that had correctly inserted dehydrin genes were transformed into the selected expression *E. coli* strain – BL21 DE3 *pLysS* – using previously described methods (section 3.3.9). Successful transformation was then confirmed using the previously described colony PCR method (section 3.3.9).

3.3.11 Protein expression

Bl21 (DE3) pLysS *E. coli* glycerol stocks containing each dehydrin gene of interest were streaked out onto LB agar plates containing ampicillin and chloramphenicol at final concentrations of 100mg/ml and 50mg/ml, respectively. Plates were incubated at 37ºC overnight. Single colonies from each plate were inoculated into 10ml starter cultures of LB-Amp/Chloramp and grown overnight at 37ºC, shaking. These were subsequently used to inoculate 500ml LB-Amp/Chloramp flasks, which were grown at 37ºC (shaking) for approximately three hours, until an OD_{600} reading of 0.6 was reached. Protein expression was induced with the addition of isopropyl-β-thiogalactoside (IPTG) to a final concentration of 1mM and cultures were incubated at 30ºC for four hours with gentle shaking, as previously described by Waters (2015).

3.3.12 Preparation of crude lysates

Cells were harvested by centrifugation for 10 minutes at 4ºC, 10 000 x g. Total protein from the cell pellet was then extracted, according to protocol supplied by the Protino® Nickel TED (Ni-TED) Histidine Tag Affinity Purification Kit (Macherey-Nagel, Germany) with re-suspension in 1 x Lysis Equilibration Wash (LEW) Buffer, the addition of lysozyme to a final concentration of 1mg/ml and 30 minutes incubation on ice. This was followed by 3 sonication cycles (40% cycle duty, 20 seconds on and 20 seconds off per cycle). To remove cell debris, the lysate was then centrifuged at 10 000 x g for 30 minutes at 4ºC. The supernatant was subsequently heated to 90ºC for 15 minutes and centrifuged at 10 000 x g for 10 minutes to isolate the heat-stable fractions containing the dehydrin proteins.

3.3.13 Purification of His-tagged dehydrin proteins

The His-tagged recombinant dehydrin proteins were extracted and purified from the crude lysate using the Protino® Nickel TED (Ni-TED) Histidine Tag Affinity Purification Kit (Macherey-Nagel, Germany). The purification procedure followed the protocol as described

in the Protino® Ni-TED Kit with the following adjustments: an additional 4ml wash step with 1 x LEW buffer was added and two 3ml elution fractions of 1 x elution buffer were collected instead of three.

3.3.14 Protein concentration and buffer exchange

To remove any compounds present in the elution buffer that could interfere in downstream assays on these dehydrin proteins, the purified samples were subjected to concentration and buffer exchange into dH_2O using the Amicon Ultra Centrifugal Filters (3K MWCO, Merck Millipore). The 6 ml eluate was centrifuged at 8000 x g for one hour to concentrate the protein, followed by two 5ml wash steps with dH_2O with the previously described centrifuging parameters and eluted with 1ml of dH_2O.

3.3.15 Total protein quantification

Quantification of protein present in the various fractions collected was performed using the BioRad Bradford Microassay (BioRad, USA) according to the manufacturer's instructions. Bovine serum albumin (BSA; Sigma Aldrich, USA) was used to generate standards for the analysis.

3.3.16 Protein analysis using SDS-PAGE and western blot

Twenty µg of each sample was run in duplicate on 12% SDS-PAGE at 90V for 2 hours with a Colour Prestained Protein Standard Ladder (New England Biolabs, USA). The first gel was stained using Coomassie Blue Solution (0.25% w/v Coomassie Blue, 50% v/v methanol, 10% v/v acetic acid) for an hour, followed by destaining overnight with destaining solution (45% v/v methanol and 10% v/v acetic acid); both steps were conducted with gentle shaking at room temperature. The second gel was utilized for western blot analysis and transferred onto a nitrocellulose membrane for one hour at 100V, 4ºC. Successful protein transfer was confirmed using Ponceau-S solution (0.1% w/v Ponceau-S and 5% v/v acetic acid). To remove the Ponceau-S stain, the membrane was washed with 1 x TBS buffer containing 0.05% Tween-20 (1xTBST). The membrane was then blocked with 10% fat-free milk powder in 1xTBST buffer for 1 hour at room temperature, with gentle shaking. This was followed by an hour incubation with the monoclonal anti-His HRP-conjugated antibody (1:40 000; Sigma Aldrich, USA) in 10% fat-free milk powder in 1 x TBST. Membranes were then washed for 3 x 5 minute washes with 1 x TBST buffer to remove unbound antibody.

Chemiluminescent detection of the antibody was done using the Advansta WesternBright ECL HRP Substrate and visualised using the Molecular Imager ChemiDoc XRS+ system (Bio-Rad, Germany) according to the manufacturer's instructions.

3.3.17 In-solution secondary structural investigation

CD spectra were obtained using a JASCO J-810 Spectropolarimeter (JASCO Analytical Instruments, Japan). Each purified dehydrin protein in dH_2O was measured at a concentration of 0.2 mg/ml in a 0.2mm cuvette. As described by Ginbot (2011), spectral data were collected from 240 to 185 nm, with 5 accumulations per run. Measurements of millidegrees obtained from the results were subsequently converted into mean residue (θ) and plotted against the wavelength range (nm). To investigate the likelihood of desiccation-induced structural changes, the proteins were dissolved in 50% and 90% acetonitrile (ACN) to simulate a dehydrated environment. BSA (Sigma-Aldrich, USA) was used as a control for the spectra of an α-helical structure, at a concentration of 0.375 mg/ml.

3.4 Results and discussion

3.4.1 PCR amplification of dehydrin genes from pBluescript SK vector

The primers designed in section 3.3.4 were used in PCRs for the amplification of XhLEA2-4 and XhLEA2-5 cDNA from the maintenance vector, pBluescript SK. The amplified fragments were predicted to be 420 and 417bp, respectively. From the gel in figure 3.2, both genes appear to be the correct size. Since 50ng of plasmid was used in the PCR reactions, residual plasmid can be seen at a higher MW on the gel.

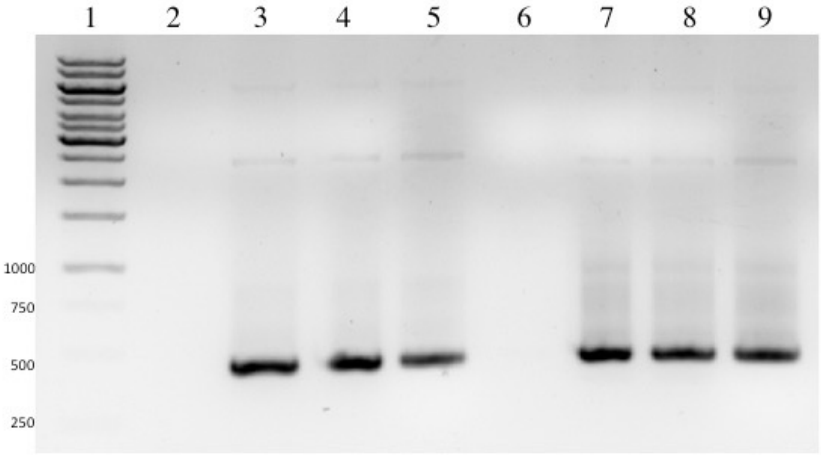


Figure 3.2 Amplification of XhLEA2-4 and XhLEA2-5 from pBluescript SK. Lane 1, O'Gene Ruler 1kb ladder (ThermoFisher Scientific, USA). Lane 2, XhLEA2-5 no template control. Lane 3-5, XhLEA2-5 PCR fragments. Lane 6, XhLEA2-4 no template control. Lane 7-9, XhLEA2-4 PCR fragments.

3.4.2 Restriction enzyme digest using *SalI* and *NotI*

The expression vector and the amplicons for XhLEA2-4 and XhLEA2-5 were digested with *SalI* and *NotI* for directional cloning of the genes for expression using the pET system. Successful RE digestion of the pET21a(+):His vector with *SalI* and *NotI* was confirmed using the gel electrophoresis and can be seen in lanes 2-4 in Figure 3.3. XhLEA2-4 and XhLEA2-5 fragments, previously amplified out of pBluescript SK using PCR (Figure 3.2), were also subjected to a RE double digest and successful digestion is observed in Figure 3.3, lanes 5-10. Residual plasmid can once again be observed faintly at higher MWs; however, the optimal ratios of pET21a(+):His plasmid to LEA insert were used for cloning. The residual backbone was highly diluted and screening techniques (section 3.4.3) excluded any possible re-ligations of insert into pBluescript SK. Digested fragments were excised using gel purification and purified for subsequent ligation reactions.

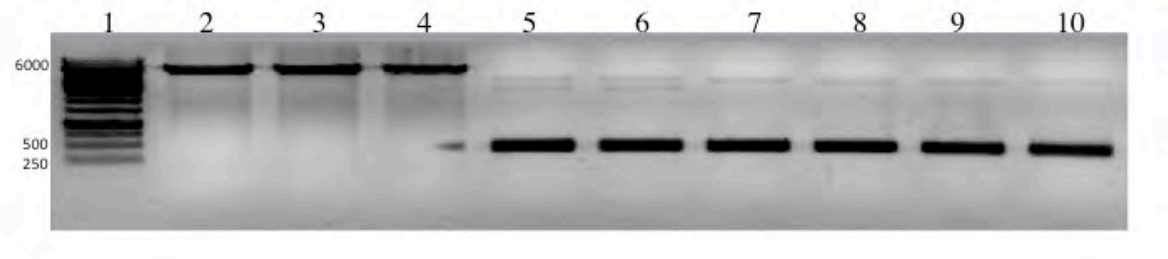


Figure 3.3 RE double digest using *SalI* and *NotI* for both the vector and dehydrin genes. Lane 1: O'GeneRuler 1kb ladder (ThermoFisher Scientific, USA). Lanes 2-4: double digest of pET21a(+):His plasmid. Lanes 5-7: double digest of XhLEA2-5. Lanes 8-10: double digest of XhLEA2-4.

3.4.3 Transformation of recombinant plasmid into *E. coli* for protein expression

Recombinant plasmids containing the genes for XhLEA2-4 and XhLEA2-5 were then transformed into *E. coli* hosts for protein expression. Following the RE double digest, fragments were ligated and transformed into *E. coli* DH5α competent cells. Successful transformation of XhLEA2-4 and XhLEA2-5 was identified using colony PCR and positive clones can be observed in Figure 3.4 A and B, respectively. These results were further confirmed by comparing sequencing data generated by Macrogen (Netherlands, Appendix B1) to the predicted amino acid sequence (Figure 2.1). Recombinant plasmids were extracted from the selected colonies and transformed into *E. coli* BL21 (DE3) pLysS for subsequent protein expression. Successfully transformed colonies were confirmed using colony PCR and can be seen in Figure 3.4, C and D.

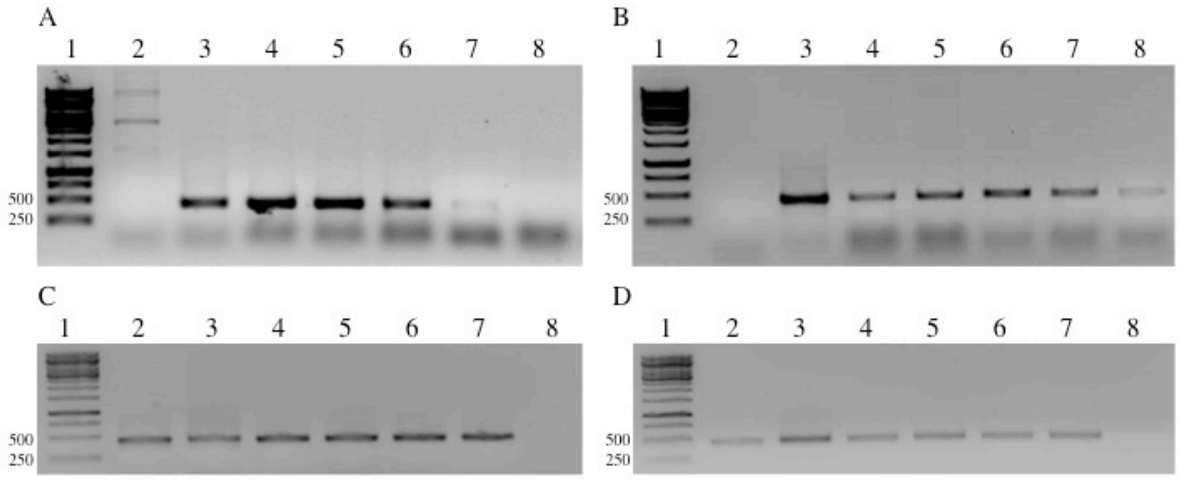


Figure 3.4 Colony PCR screening for successfully transformed *E. coli* DH5α cells (A&B) and *E. coli* BL21 (DE3) pLysS cells (C&D) for XhLEA2-4 (A&C) and XhLEA2-5 (B&D). Lanes A1, B1, C1 and D1: O'GeneRuler 1kb ladder (ThermoFischer Scientific, USA). Lanes A2, B2, C8 and D8: vector only control. Lanes A3, B3, C2 and D2: positive control using digested gene fragments for XhLEA2-4 and XhLEA2-5. Lanes A4-6 and B4-7: positive recombinant colonies for XhLEA2-4 and XhLEA2-5 in DH5α. Lanes C2-7 and D2-7: positive recombinant colonies for XhLEA2-4 and XhLEA2-5 in BL21 (DE3) pLysS.

3.4.4 Expression and purification of recombinant His-tagged XhLEA2-4 and XhLEA2-5 protein

Successfully transformed *E. coli* colonies were used to express recombinant XhLEA2-4 and XhLEA2-5 protein for purification. The predicted sizes of XhLEA2-4 and XhLEA2-5 were 14.8kDa and 14.5kDa (Figure 2.1). However, since the addition of a His-tag onto a protein is known to increase its size by 0.84kDa, the expected protein sizes would be approximately 15.6kDa and 15.3kDa, respectively (Terpe, 2006). On the SDS-gel, a protein band that was approximately 22kDa in size was observed for XhLEA2-4 (Figure 3.5, A). Similarly, for XhLEA2-5, a protein band was observed around 21kDa (Figure 3.5, C). The identities of these two bands were confirmed to be the His-tagged dehydrin proteins using western blot analysis (Figure 3.5, B and D). Early research into LEA proteins observed their tendency to migrate at larger molecular masses during SDS-PAGE analyses (Battaglia et al., 2008). Habchi et al. (2014) observed that intrinsically disordered proteins typically run at molecular weights approximately 1.2-1.5 times greater on an SDS gel. If this ratio is taken into consideration then the apparent molecular weight for both XhLEA2-4 and XhLEA2-5 would

be between 18kDa and 23kDa. This estimation would explain the observed molecular weights of the proteins.

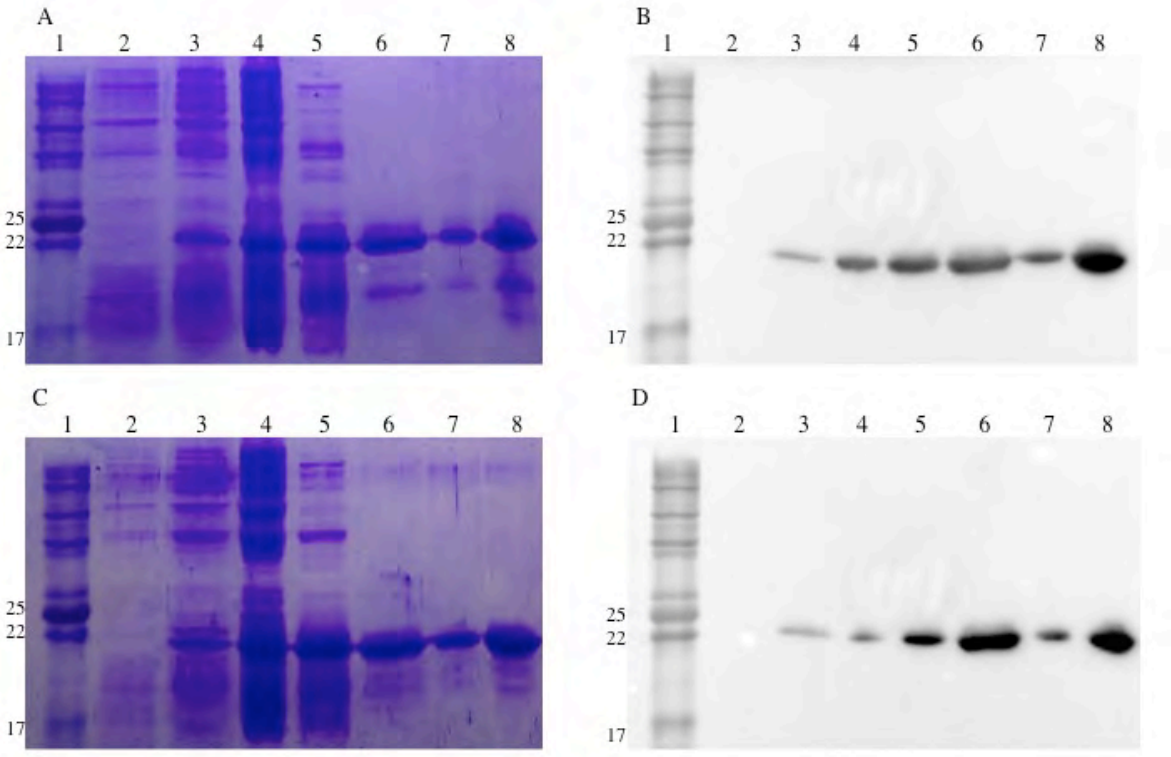


Figure 3.5 Coomassie-stained 12% SDS-PAGE and western blot confirming the presence of XhLEA2-4 (A&B) and XhLEA2-5 (C&D). A, B, C and D – Lane 1: Colour Prestained Protein Standard, Broad Range (NEB, USA). Lane 2: uninduced fraction (T_o). Lane 3: induced fraction (T_4). Lane 4: crude protein fraction. Lane 5: crude fraction after boiling. Lane 6: purified protein fraction from Protino® Ni-TED Histidine Tag Affinity column. Lane 7: purified protein fraction following buffer exchange into dH_2O using the 10K Amicon Filters. Lane 8: purified protein fraction following buffer exchange into dH_2O using the 3K Amicon Filters.

In comparison to the methods described by Waters (2015), one adjustment was made to the protocol that can be observed in Figure 3.5. Instead of using a 10K Amicon Filter to concentrate and buffer exchange the dehydrin proteins into dH_2O, a 3K Amicon Filter was used. The observed result was an increase in recombinant protein yield (lane 7 and 8). The hypothesized reason for this was that the membrane filter on the 10K columns could have allowed recombinant proteins to get trapped inside as a result of their intrinsically disordered structures and molecular weights so similar to the cutoff threshold. It is also important to point out the presence of the feint bands (approximately 19kDa in size) representing smaller

proteins below the purified dehydrin proteins (lane 6 to 8). Since the two sizes of bands do not exist as integral multiples of each other, it is unlikely that they are the result of dimer or multimer formation. These proteins also did not generate signals on the western blot. However, since the antibody targeted the His-tag only, it is not possible to rule out the fact that those bands may represent fragmented recombinant proteins.

3.4.5 Investigation of in-solution structure using circular dichroism (CD)

The recombinant proteins of XhLEA2-4 and XhLEA2-5 were then monitored for changes in secondary structure content using CD. Absorbance results were converted from millidegrees (mdeg) to molar ellipticity or [θ] using the equation provided by Fallis (2013). These values were plotted against the wavelength range of 190nm to 250nm. CD spectra for XhLEA2-4 in dH_2O produced a distinct signal at around 220 and 210nm, followed by a peak at 190nm (Figure 3.6, A). Although the magnitude of the spectra generated by XhLEA2-4 was significantly smaller in comparison to the albumin standard, the similarity in shape suggested that even in dH_2O, the dehydrin adopted a slightly α-helical structure. When the dehydrin was placed in solutions that simulated a dehydration event – 50% acetonitrile (ACN) and 90% ACN – the change in spectral data observed was minimal. Nonetheless, the dehydrin became slightly more α-helical over the two ACN concentrations. A possible explanation for what was observed with XhLEA2-4 is the understanding that, while it is generally expected for LEA proteins to be intrinsically disordered in the hydrated state, dehydrins have been known to adopt weak or transient helical structures under influence of the K-segment (Graether & Boddington, 2014). Since PONDR predicted XhLEA2-4 to be fairly ordered (Figure 2.3), it would be expected that the protein would start off significantly structured in solution.

XhLEA2-5, on the other hand, appeared to be predominantly random coiled in dH_2O (Figure 3.6, B). Random coils are characterized by the presence of a signal minimum located around 200nm and then a considerably weaker signal around 222nm (Hand, Menze, Toner, Boswell, & Moore, 2011). However, when XhLEA2-5 was placed into 50% acetonitrile, the spectrum shifted towards becoming slightly more α-helical as the signals at 220nm and 210nm appeared. In the presence of 90% acetonitrile, the "α-helical"-ness of the dehydrin surpassed the spectra generated by the albumin standard, with a significant peak at 190nm. However, the two dipped signals expected at 210nm and 220nm are not both present. This,

once again, suggests the transient or imperfect nature of the α-helical structures that may be generated by the dehydrin proteins in a variety of aqueous solutions.

The accuracy of *in silico* predictions of secondary structures (Chapter 2, 2.4.2.2.2) should also be noted at this point. Results generated by the various prediction platforms suggested that the two dehydrins existed predominantly as random coils in both the hydrated and dehydrated states. While a more accurate representation of dehydrated structural states could be generated using FTIR spectroscopy and CaF_2 windows, the results produced by the acetonitrile simulated dehydration suggests that a discrepancy exists between *in silico* predictions and *in vivo* observations (Dang et al., 2014). Another possible way to investigate the appearance of secondary structures in LEA proteins with more certainty is through the addition of a molecular crowding agent such as polyethylene glycol (PEG). This has been observed to encourage or speed up protein folding (Christiansen, Wang, Cheung, & Wittung-Stafshede, 2013; Zhou, Rivas, & Minton, 2008).

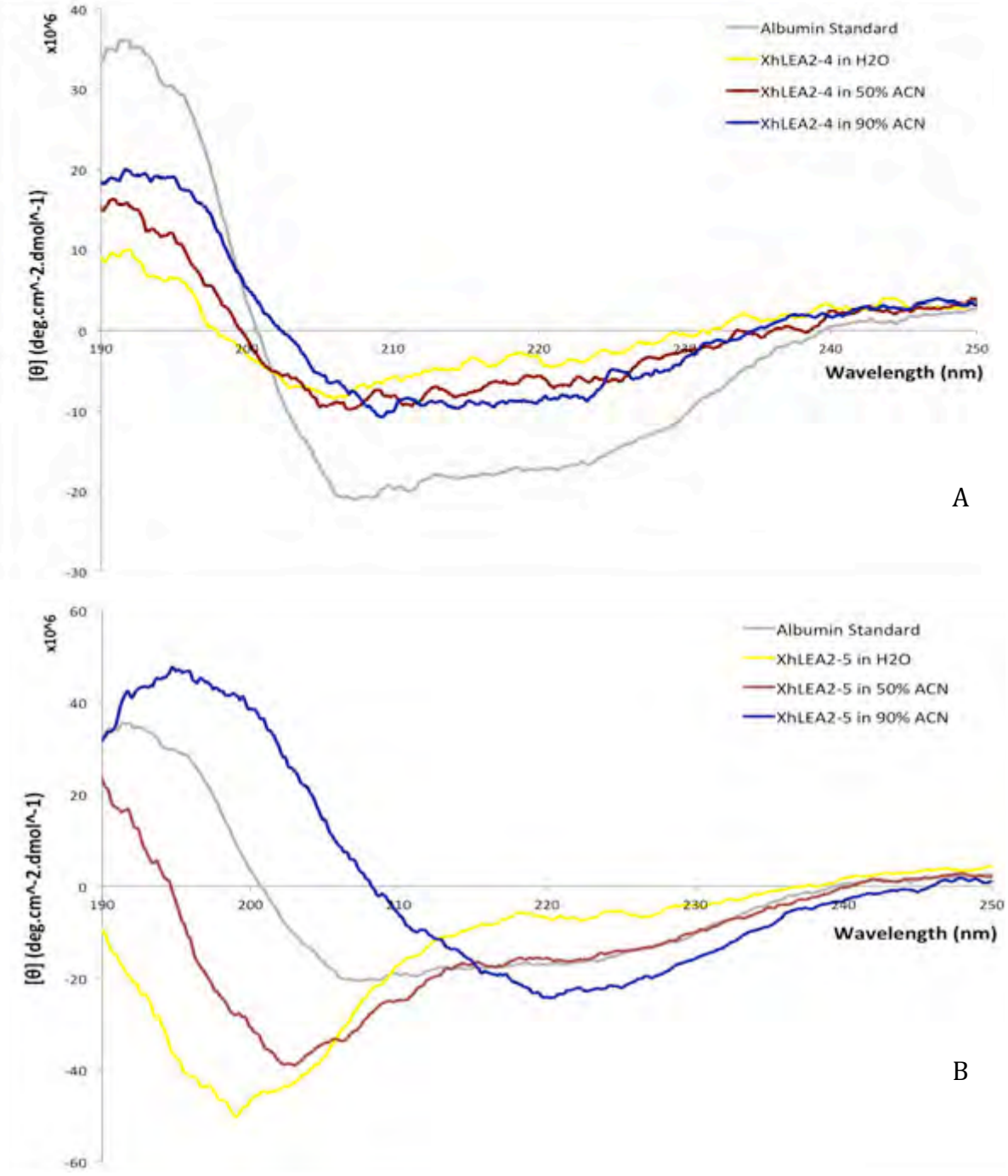


Figure 3.6 CD Spectra for XhLEA2-4 (A) and XhLEA2-5 (B) showing change in secondary structure content. Each protein was monitored for changes in secondary structure in the presence of three solutions – dH_2O, 50% acetonitrile (ACN) and 90% acetonitrile (ACN).

Chapter 4
Functional assessment of the two recombinant dehydrins

4.1 Introduction

It has been frequently postulated, based on Transcriptome studies, that LEA proteins play a role in tolerance of mild to extreme water deficit stress. Further evidence supporting this hypothesized role was provided by transgenic studies in organisms such as yeast, *C. elegans* larvae, cowpea seedlings, tobacco plants etc. (Tunnacliffe & Wise, 2007). Introduction of a barley group 3 LEA (HVA1) enabled transgenic rice to gain an enhanced tolerance of water stress and salt stress (Xu et al., 1996). These transgenic rice plants were able to maintain higher RWCs in their leaves and displayed less electrolyte leakage from their cells, suggesting a membrane-related protective effect. Another group 3 LEA from *Brassica napus* was able to improve salt and drought tolerance within species of Chinese cabbage (Park et al., 2005). Figueras et al., (2004) showed that introducing a potato dehydrin (DHN24) into cucumber plants showed an enhancement in freezing tolerance. The overexpression of a maize dehydrin (Rab17) also conferred improved osmotic stress tolerance in *A. thaliana* (Yin et al., 2006). However, and perhaps unexpectedly, expression of two dehydrins and a group 3 LEA from the resurrection plant *C. plantagineum* did not confer any drought tolerance to transgenic tobacco (Iturriaga et al., 1992). Such variation across these results indicates that the effect of a LEA protein on an organism cannot be predicted. Additionally, there is also insufficient evidence for establishing a link between a specific group of LEAs and the type of abiotic stress tolerance it may confer, suggesting that the precise mechanisms that allow LEA proteins to perform their protective roles are poorly understood.

Various biochemical assays have been developed in attempts to understand the mechanisms of LEA proteins *in vitro.* The most commonly cited experiment investigates the possible molecular chaperone role of LEA proteins by using enzymes such as lactate dehydrogenase (LDH) or citrate synthase (CS) for activity post-dehydration (Goyal, et al., 2005; Place & Hofmann, 2005; Popova et al., 2015). These enzymes are selected for these chaperone activity assays because they are sensitive to both dehydration and freeze-thaw cycles. The ability of LEA proteins to maintain activity in dehydrated CS and LDH enzymes has been documented by various studies (Dang et al., 2014; Ginbot, 2011; Goyal et al., 2005). This ability to preserve activity in desiccation-sensitive enzymes shown to partially stem from

preventing desiccation-induced aggregation (Goyal et al., 2005). In addition, the restorative role of LEA proteins on enzymes undergoing either heating or dehydration has also been shown to be ratio-dependent. Goyal et al., (2005) investigated the effects of a 1:1, 5:1 and 10:1 (LEA protein to enzyme) ratios on the percentage of activity restored in CS and found that 5:1 and 10:1 ratios were most effective in both the group 1 and the group 3 LEA protein. Subsequently, Hincha et al., (2014) demonstrated that two dehydrins (termed 'LEA1' and 'LEA27') were able to preserve a significant percentage of LDH activity at a ratio of 25:1 (dehydrin to LDH). However, another dehydrin investigated in the same study (termed 'LEA26') did not demonstrate significant restoration of LDH activity. This suggests that the potential stabilizing role of LEA proteins on enzyme activity is ratio-dependent but also, once again, varies between the different classes. While the results from these chaperone assays are able to infer a protective role played by LEA proteins during dehydration, the exact interaction between the enzyme and the LEA remains unknown. This protective role could be attained by the LEA proteins binding to enzymes directly to form complexes or they could be behaving as 'molecular shields' by minimizing interactions within a partially denatured polypeptides and preventing aggregation (Tunnacliffe & Wise, 2007).

Taking these caveats into consideration, this chapter explores the possible roles that XhLEA2-4 and XhLEA2-5 may potentially play in functioning as molecular chaperones or stabilizers using the above mentioned classical enzyme assays in the presence of heat and dehydration.

4.2 Aim

The experimental objectives described in this chapter are as follows:

a) Determine if the two dehydrin proteins, with/without the addition of trehalose, play a role in preventing heat-induced protein aggregation using the enzyme lactate dehydrogenase (LDH).
b) Investigate the possible role of the two dehydrin proteins as stabilizers or chaperones to other proteins under dehydration using citrate synthase (CS) and lactate dehydrogenase (LDH). The effect of the addition of trehalose in combination with each dehydrin was also assessed.

4.3 Methods and materials

4.3.1 Production of recombinant proteins

His-tagged recombinant dehydrin proteins were expressed, induced and purified according to the methods described in Chapter 2.

4.3.2 Anti-aggregation role of XhLEA2-4 and XhLEA2-5

To determine whether or not XhLEA2-4 and XhLEA2-5 played a protective role when LDH was heated, two temperatures were selected, 55ºC and 75ºC (Place & Hofmann, 2005). A preliminary study was conducted using 0.1mg of each component (LDH, XhLEA2-4, XhLEA2-5, 0.1M trehalose, BSA) and heating them independently at the two temperatures to ensure that they themselves do not aggregate. BSA was selected as a control because it was shown by Goyal et al., (2005) to have a partial effect on minimizing heat-induced protein aggregation. Additionally, since trehalose has been known to function as a 'chemical chaperone', LDH was also heated in the presence of 0.1M trehalose (Goyal et al., 2005). The possible anti-aggregation role of each dehydrin was then tested by heating it with LDH at a 1:1 ratio (Ginbot, 2011; Goyal et al., 2005). Absorbance readings at 340nm were taken to detect the absence/presence of aggregation. Experiments were performed in triplicate.

4.3.3 Enzyme stabilization role

4.3.3.1 Effect of drying on CS activity, with/without dehydrin proteins

To investigate the role of each dehydrin on the effects of drying on CS, proteins were added to CS at molar ratios of 10:1 (dehydrin to CS). BSA, a protein known to exhibit partial protection to enzymes under dehydration, was used as a control and prepared similarly with CS (Goyal et al., 2005). Trehalose (0.1M) was also added to the dehydration experiment as it has been shown to behave synergistically with LEA proteins in protecting enzyme activity. All proteins were prepared in 40mM Tris-HCl, pH 8.1 and dried for an hour under vacuum at room temperature. After drying, each sample was rehydrated to the starting volume of 20µl. This was repeated for four cycles. From the rehydrated sample at each dry cycle, 2.5U of CS (Sigma Aldrich, USA) was added to 5mM acetyl-CoA, 100mM oxaloacetate, 100mM DTNB, 100mM KCl and 40mM Tris-HCl, pH 8.1 in a final volume of 100µl. The increase absorbance at 412nm was measured every 5 seconds for 3 minutes to monitor enzyme activity. Untreated CS was used as a control for maximum enzyme activity. CS was also dried in the absence of dehydrin proteins to confirm sensitivity towards dehydration. Enzyme

activity was reported as a percentage of control activity in the absence of desiccation. These were repeated in triplicate.

4.3.3.2 Effect of drying on LDH activity, with/without dehydrin proteins

The possible role of each dehydrin on the effects of drying on LDH was subsequently investigated, using methods similar to those adopted for CS. Dehydrin proteins were added to LDH at molar ratios of 10:1 and 20:1 (dehydrin to LDH), BSA was used as a control for partial protection of enzyme activity and the possible synergistic role of trehalose was also investigated with 0.1M trehalose (Goyal et al., 2005). Proteins were prepared in 10mM Tris pH 8.6 and dried for an hour under vacuum at room temperature. The samples were then rehydrated back to the starting volume of 50µl and drying was repeated for a total of four cycles. 1.31U of rehydrated LDH (Sigma Aldrich, USA) from each dry cycle was then assayed using 0.5mM NADH, 240mM pyruvate, 18mM bicarbonate and 10mM Tris-HCl, pH 8.6 in a final volume of 100µl. The decrease in absorbance at 340nm was monitored as an indication of enzyme activity. Untreated LDH was used as a control for maximum activity and LDH was also dried in the absence of dehydrin proteins to ensure that the enzyme was, in fact, sensitive towards dehydration. Enzyme activity was reported as a percentage of control activity in the absence of desiccation. The dehydration experiments were repeated in triplicate.

4.4 Results and discussion

4.4.1 The role of each dehydrin in preventing heat-induced protein aggregation

In the pilot experiment, it was observed that BSA, trehalose, XhLEA2-4 and XhLEA2-5 did not aggregate when heated in isolation to 55ºC and 75ºC for 1 hour (Appendix C1). LDH, on the other hand, displayed aggregation in the absence of any protectant molecules. Aggregation was measured according to light scattering at 340nm (Goyal et al., 2005). This aggregation was observed to increase proportionally with the increase in temperature (Figure 4.1). With the exception of BSA at 75ºC, all four protectants showed a small decrease in the amount of aggregation monitored at both 55ºC and 75ºC. The presence of 0.1M trehalose appeared to produce the most effective minimisation of aggregation in LDH and the 'chemical chaperone' role of trehalose that was described by Goyal et al., (2005). Each dehydrin also demonstrated an ability to minimize LDH aggregation, although the decrease was less than that achieved by trehalose. Since it is known that the efficacy of LEA proteins varies according to the molar ratios that they are present in, it is possible that a different molar ratio of dehydrin to LDH may improve their ability to minimize aggregation.

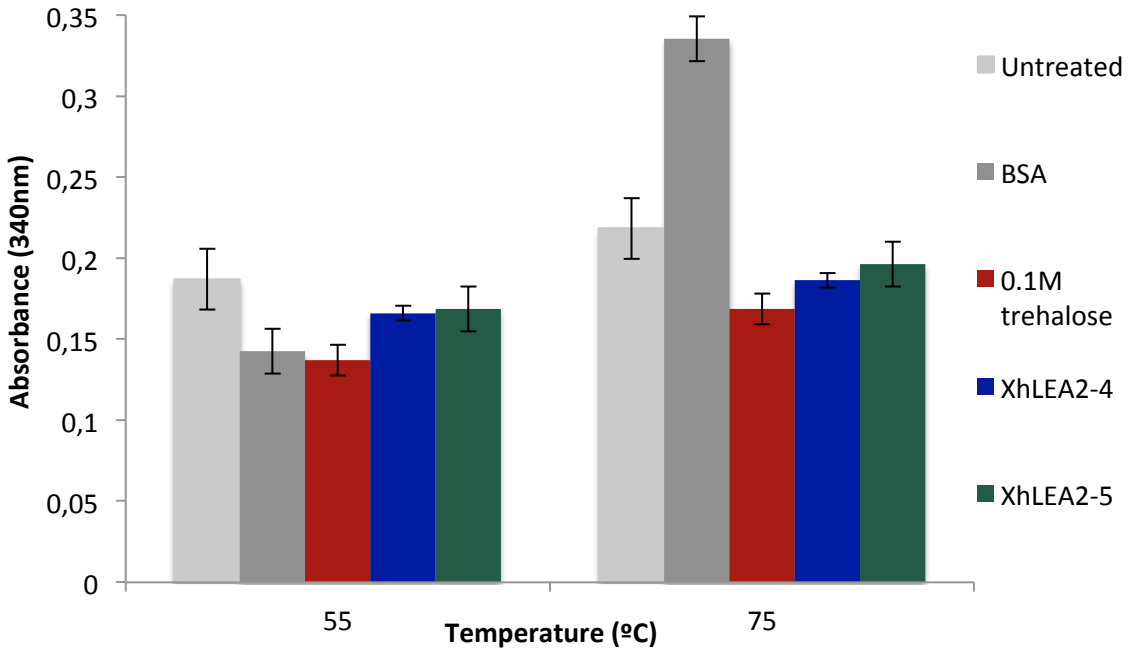


Figure 4.1 LDH aggregation in the presence and absence of BSA, 0.1M trehalose, XhLEA2-4 and XhLEA2-5. BSA, XhLEA2-4 and XhLEA2-5 were all added to LDH at 1:1 molar ratios and aggregation assays at each temperature were done in triplicate. Error bars indicate standard deviation.

4.4.2 Protective role of XhLEA2-4 and XhLEA2-5 in maintaining CS activity

In this assay, CS that had not undergone any dehydration or rehydration was used as the control to denote 100% enzyme activity (Goyal et al., 2005). Despite being reported as desiccation-sensitive, our data showed that CS maintains 75% of its activity after two rounds of dehydration and rehydration (Figure 4.2). After another two rounds of dehydration and rehydration (four cycles of drying), the enzyme activity displayed only 50% of its maximum activity. This suggests that the enzyme is fairly robust and not as sensitive to water loss as previously reported (Goyal et al., 2005). In the data below, the results have been reported as percentages of the control activity.

Figure 4.2 gives data for CS activity when dried in the presence of BSA, each dehydrin (10:1 molar ratios) and with/without 0.1M trehalose. Following two cycles of dehydration, the protective roles of XhLEA2-4, XhLEA2-5, 0.1M trehalose and BSA on CS activity cannot be reported as significant. Despite being reported to work synergistically with LEA proteins to protect enzyme activity, the presence of trehalose alone or with a dehydrin showed a significant decrease in the amount of CS activity restored. BSA was selected as a control for partial protection of enzyme activity. However, its effect independently after two cycles of dehydration appears to be negative whilst the effect in the presence of trehalose alone results in similar enzyme activity to that of the control. This may be attributed to the apparent robust nature of the enzyme. Since CS only loses 25% of its maximum activity, it may only be experiencing minimal amounts of desiccation-induced damage and, consequently, does not facilitate stabilization or protection from chaperone molecules.

Following four cycles of dehydration and rehydration, CS displayed a significant loss in activity, although it is not entirely diminished. Nonetheless, in the presence of XhLEA2-4, the activity of dried CS was restored to approximately 70%. This effect was further increased up to 90% by the presence of both 0.1M trehalose and XhLEA2-4; this observation suggests that the role played by trehalose in enzyme activity restoration may be additive. In contrast to XhLEA2-4, XhLEA2-5 did not restore significant CS activity on its own. However, in the presence of trehalose, XhLEA2-5 was able to restore approximately 95% of CS activity. Additionally, trehalose was also able to restore approximately 70% of CS activity on its own. However, the effect of trehalose was not observed in the presence of BSA, despite BSA displaying a significant protective role on its own.

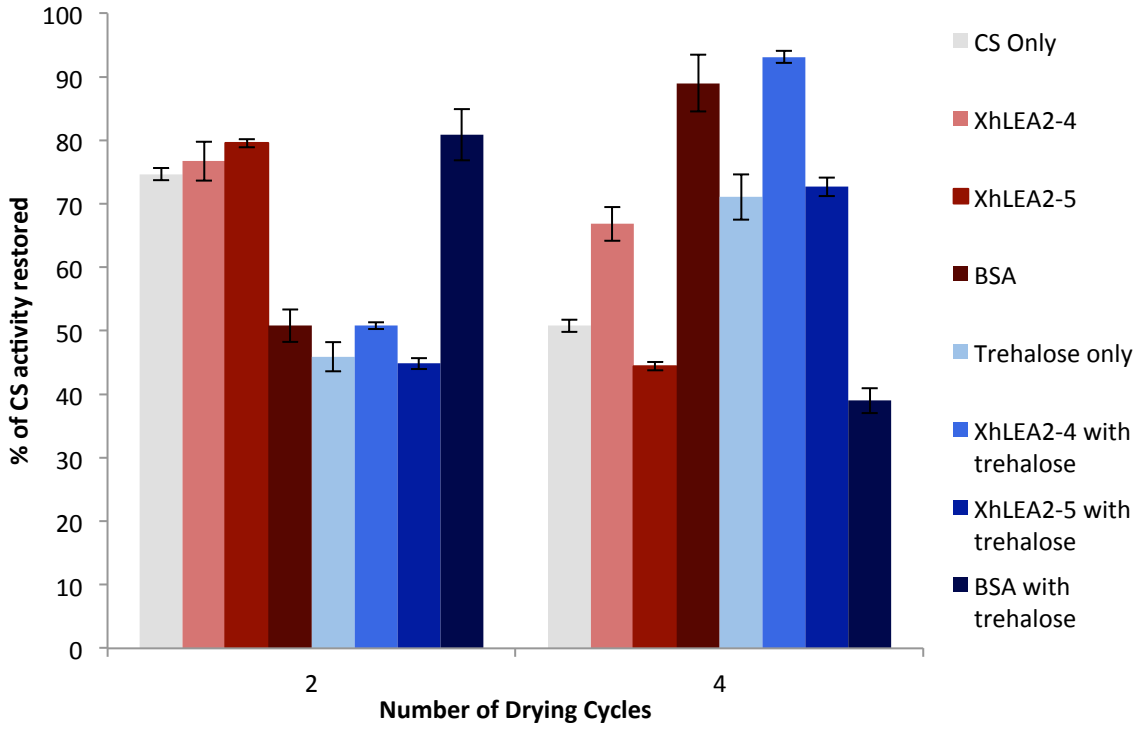


Figure 4.2 Percentage of CS activity restored in the presence of XhLEA2-4, XhLEA2-5, BSA and 0.1M trehalose following 2 and 4 dry cycles. The two dehydrins and BSA were added to CS at 10:1 molar ratios. Enzyme activity was reported as a percentage of control activity in the absence of desiccation. Dehydration experiments were done in triplicate. Error bars show standard deviations.

The considerations that need to be made when interpreting *in vitro* studies on LEA proteins are particularly evident from the results in this assay. Firstly, the experimental variation that exists within this assay may also contribute to both the high standard deviations observed across the assay, as well as the apparent lack of consistency in the behaviours of each assay component. The irregularities observed in the results may also be attributed to the degree of biological relevancy provided by the assay. The conditions of the assay are most likely inadequate for representing a biological environment such as the dehydrating cell. Therefore, the activities observed in the various dehydrins may not necessarily reflect its behaviour or role *in vivo*. Lastly, since the effect of dehydration on CS does not appear to be detrimental to its activity, the enzyme may not be in a condition or state that facilitates assistance from a protectant molecule. Consequently, the assay would not be able to demonstrate the full protective ability of a LEA protein. Therefore, it would be advisable to identify an enzyme that is significantly sensitive to desiccation for investigating the protective abilities of each dehydrin.

4.4.3 Protective role of XhLEA2-4 and XhLEA2-5 in maintaining LDH activity

LDH is another enzyme that has been reported to show sensitivity towards water loss. Therefore, this enzyme was selected to further investigate the possible roles that XhLEA2-4 and XhLEA2-5 may play in protecting or maintaining enzyme activity. The extent to which LDH was sensitive to water loss was first investigated. LDH that had not undergone any dehydration or rehydration was used as the control for maximum enzyme activity. Following two cycles of dehydration and rehydration, LDH appeared to maintain approximately 80% of its control, untreated activity (Figure 4.3). However, after another two dry cycles, it was observed that the activity of LDH decreased considerably to 20%. This indicated a deleterious effect of dehydration on LDH activity and indicated significant sensitivity to water loss. In attempts to minimize the amount of variation within the experimental set up, each sample was assayed for LDH activity following each dry cycle in triplicate. Each dehydration experiment was also repeated in triplicate using different aliquots of LDH enzyme. An additional molar ratio of dehydrin to LDH (20:1) was investigated since the protective role of dehydrins may also dependent on molar ratios (Popova et al., 2015).

The results produced following two cycles of dehydration and rehydration mimic those produced by the CS functional assay. Since there is an absence of significant decrease in LDH activity after two dry cycles, the effects of XhLEA2-4, XhLEA2-5, BSA and 0.1M trehalose cannot be distinguished. Addition of the two dehydrins at 20:1 molar ratios, in the presence of 0.1M trehalose, appears to have a significantly negative effect on LDH activity.

Following four cycles of dehydration and rehydration, the presence of XhLEA2-4, XhLEA2-5, BSA and 0.1M trehalose all appear to restore LDH activity to more than 60% of that exhibited by the activity of the control. The protective roles of XhLEA2-4 and XhLEA2-5 do not appear to be molar ratio dependent since (in the absence of trehalose) the percentage of LDH activity restored does not differ significantly between 10:1 and 20:1. However, in BSA, the molar ratio of 10:1 appears to be more effective at restoring activity than the 20:1 ratio, although this trend does not continue in the presence of trehalose.

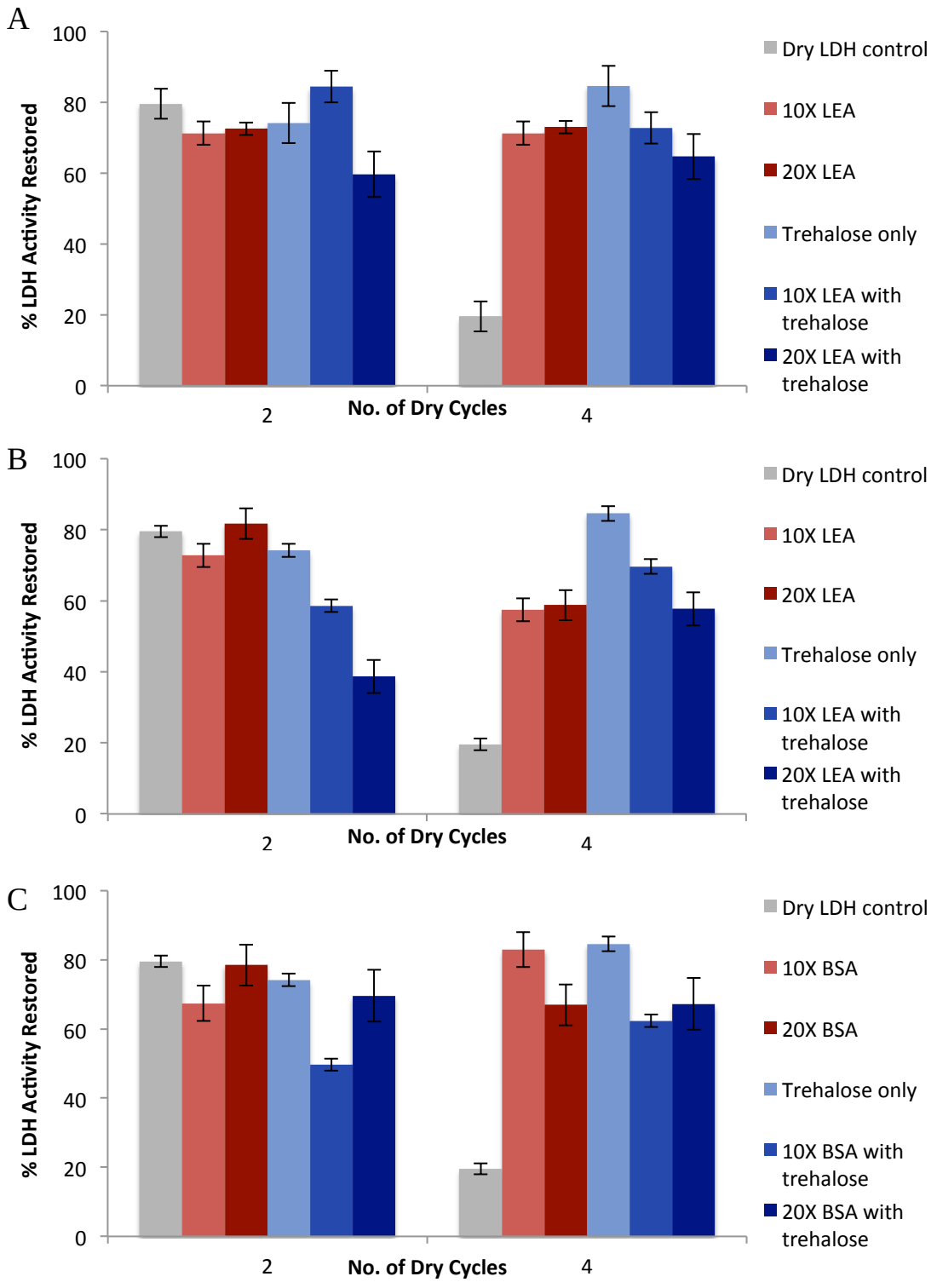


Figure 4.3 Percentage of LDH activity restored in the presence of XhLEA2-4 (A), XhLEA2-5 (B) and BSA (C) with or without 0.1M trehalose, following 2 and 4 dry cycles. The two dehydrins and BSA were added to LDH at molar ratios of 10:1 and 20:1. Enzyme activity was reported as a percentage of control activity in the absence of desiccation. Error bars show standard deviations.

Trehalose is the most effective stabilizing component, restoring 80-90% of LDH activity without any dehydrins or BSA present. This supports the proposed role of trehalose as a 'chemical chaperone' (Tunnacliffe & Wise, 2007). When trehalose is dried down with XhLEA2-4 and XhLEA2-5, the 10:1 molar ratio is more effective than the 20:1 ratio. It is possible that the 20:1 ratio results in a crowded environment that could reduce the efficacy of protection.

From the results generated, it is possible that XhLEA2-4 and XhLEA2-5 play a protective role in stabilizing enzyme activity under dehydrated conditions. However, while these two dehydrins demonstrate a protective role in enzyme activity *in vitro,* this may not translate directly into an *in vivo* environment. This is always important when interpreting *in vitro* experimental data. If the differing trends in mRNA levels during dehydration for each dehydrin are considered (Chapter 1, Figure 1.1), and assuming protein presence is reflective of transcript levels, it is highly likely that they would adopt different roles at different RWCs during dehydration. Therefore, these results should not be interpreted independently. Instead, the results from these *in vitro* experiments should be used in conjunction with bioinformatics predictions and structural characteristics. The combination of these results would enable more informed interpretations to be made about each individual LEA protein. This approach to understanding more about the structural and functional characteristics of LEA proteins will be adopted for XhLEA2-4 an XhLEA2-5 in the concluding chapter.

Chapter 5
Concluding remarks and future work

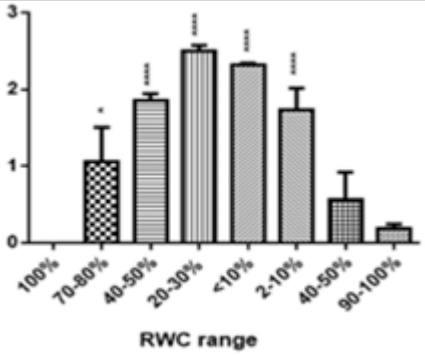
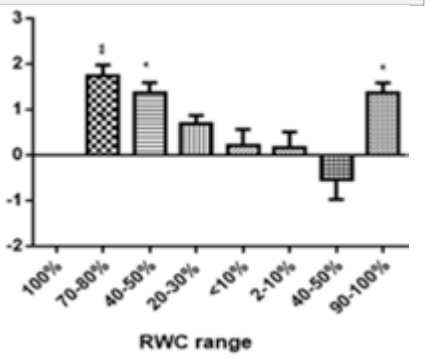
The existing systems that categorize the burgeoning family of LEA proteins contain large amounts of redundancy and little discernibility in terms of characteristics and functions. This study forms part of a bigger initiative that aims to establish a pipeline that could be used to generate more insight into the role that LEA proteins play in facilitating tolerance of dehydration. The intention was to evaluate the most commonly used assays for observing changes in structure and establishing possible roles in enzyme stabilization or protection. Additionally, the results from these assays were compared to *in silico* predictions to assess the degrees of accuracy with which these programs could interpret unconventional proteins such as LEAs. The study concentrated on two desiccation-induced putative dehydrin proteins since these are among the more widely prevalent LEA proteins across plant species.

5.1 Structural and functional characteristics of XhLEA2-4 and XhLEA2-5

The various biochemical, structural and functional characteristics of XhLEA2-4 and XhLEA2-5 identified by this study have been summarized in Figure 5.1. The identification of the K segments confirmed the identities of each putative LEA as dehydrins (Graether & Boddington, 2014). Whilst XhLEA2-4 contained only one K-segment and one HH tag, XhLEA2-5 contained two K segments, an S-segment, a Φ segment and an HH tag. From the various *in silico* prediction programs, XhLEA2-5 was predicted to contain a slightly higher percentage of α-helices in comparison to XhLEA2-4. This may be attributed to the difference in number of K segments. CD spectra show XhLEA2-4 to be fairly structured in dH_2O where as XhLEA2-5 exists in a random coil. This initial observation reflects the PONDR estimations of disorder, in which XhLEA2-4 was predicted to be significantly less disordered than XhLEA2-5. However, as the two dehydrins are placed in solutions of decreasing water content, XhLEA2-5 adopts a strongly α-helical structure whilst the structure of XhLEA2-4 changes only fractionally. Interpretation of this difference in secondary structure could be made with consideration of the mRNA expression levels during dehydration and rehydration (Waters, 2015). Although it is known that mRNA levels do not relate directly to protein levels, the difference in timing of expression indicates that their respective functions are most likely different. XhLEA2-5 appears to be an early response LEA, with maximum expression on dehydration below 80% RWC and expression levels declining below 30% RWC, where as XhLEA2-4 accumulates gradually with maximum

expression occurring below 30% RWC. In addition, XhLEA2-4 remains moderately α-helical throughout the simulated dehydration whilst XhLEA2-5 only adopts a strongly α-helical structure in the presence of 90% ACN.

Table 5.1 An overview of biochemical, structural and functional characteristics of XhLEA2-4 and XhLEA2-5.

	XhLEA2-4	XhLEA2-5
Predicted MW	14.8 kDa	14.5 kDa
Predicted pI	6.64	9.60
GRAVY score	-1.43	-1.25
Dehydrin-specific characteristics	1 x K segment 1 x HH tag	2 x K segment 1 x S segment 1 x Φ segment 1 x HH tag
Disorder prediction	Less disordered	Largely disordered
Changes in mRNA levels (Waters, 2015)		
Changes in secondary structure	dH_2O: slightly α-helical 50% ACN: more α-helical 90% ACN: most α-helical	dH_2O: random coil 50% ACN: transitional structure 90% ACN: strongly α-helical
Role in anti-aggreation	Moderate	Moderate
Role in enzyme protection during dehydration	CS: possible LDH: yes	CS: possible LDH: yes

Given these observations, it is possible that the function of XhLEA2-4 does not involve considerable change in secondary structure for it to be effective in its role in a water scare environment. XhLEA2-5, on the other hand, might be, in part, dependent on the amount and availability of water in the subcellular environment in which it is located. The difference in

the predicted pIs of each dehydrin also suggests that there may be a difference in function. If the assumed internal pH of a plant cell is relatively neutral, XhLEA2-4 is likely to be uncharged in that environment whilst XhLEA2-5 would be in a charged and interacting with membranes or other proteins. This, along with the possible difference in localisation, would greatly vary the behaviours of each dehydrin and also determine whether or not they would behave as molecular chaperones.

5.2 Limitations of this study

The *in vitro* nature of this study can be regarded as the caveat of studying the functional and structural characteristics of LEA proteins. *In vitro* studies are often informative as they enable the number of variables investigated in an experiment to be minimized. The functions performed by LEA proteins in an actively dehydrating cell may be difficult to identify *in vivo* as there are a vast number of additional components within the cell. As a consequence, associating LEA proteins specifically with an observed phenomenon may not be possible. However, the interpretation of results from an *in vitro* experiment requires a significant amount of caution, as the experiment may not directly reflect how a protein would behave inside a cell. Therefore, these assays should not be used in isolation; instead, LEAs should be subjected to a range of different assays and analyses to establish a better understanding of how they function and how the functions change according to environmental factors.

Additionally, since it was not possible to access the appropriate equipment to study LEA protein structure in a fully hydrated and fully dehydrated state, this study used ACN at concentrations of 50% and 80% to simulate a dehydrating environment for structural analysis. Although it was ensured that acetonitrile itself did not interfere with the spectral readings, it is not guaranteed that this artificial desiccant will not affect the proteins.

From an experimental perspective, the nature of the CS and LDH enzymatic assays leave room for technical variation or error. Despite having each sample assayed in triplicate and repeating the entire dehydration experiment three times, the amount of variation observed was still relatively high. Furthermore, as the dehydration procedures did not sufficiently diminish enzyme activity in CS or LDH, the assays themselves did not adequately test the various proteins and compounds for their protective roles.

5.3 Future scope of this study

The key limitation that should be addressed is the lack of *in vivo* affirmation following the *in vitro* experiments. This could be achieved through the production of antibodies that specifically target XhLEA2-4 and XhLEA2-5, as well as the remaining 19 desiccation-induced LEA proteins found in *X. humilis.* Acquisition of these antibodies would be pivotal in facilitating the link between mRNA abundance and protein accumulation during dehydration and rehydration. Since it is known that different groups of LEAs function at different stages of dehydration and rehydration, this would also provide greater insight into the exact roles played by the various groups. This could be extended to observe LEA expression in relation to other stresses. Additionally, it would be possible to determine both tissue specificity and organ specificity of the various LEAs through immunolocalisation studies. In this regard, Ginbot (2011) was able to identify that two group 1 LEAs from *X. humilis* were membrane associated; more specifically, they were localized to the areas between the cell walls and plasma membranes. From this it was assumed that the LEAs were possibly involved in membrane stabilization or the stabilization of other membrane-associated structures. Such knowledge of the subcellular localizations enable improved interpretations of function based on results produced by *in vitro* experiments.

Expression of LEA proteins in desiccation sensitive plant species has shown to confer improved resistance to a range of abiotic stresses. Ectopic expression of two desiccation-induced group 4 LEAs from *B. hygrometrica* in tobacco plants showed increased photosystem II activity and water retention in the leaves of transgenic plants undergoing dehydration (Liu et al., 2009). Furthermore, these transgenic lines displayed greater levels of antioxidant enzyme activity as well as decreased membrane permeability. Transgenic strategies provide insight into whether or not a LEA protein would be suitable as a candidate gene for genetically improving plants that are sensitive to water-deficit stress. Therefore, developing a detailed understanding of how LEA proteins function is crucial for informed selections of genes that could be used to produce of drought resistant crops. With agriculture becoming increasingly vulnerable at the mercy of climate change, the development of drought tolerant crops will undoubtedly be a necessity for ensuring food security.

References

Allagulova, C. R., Gimalov, F. R., Shakirova, F. M., & Vakhitov, V. A. (2003). The Plant Dehydrins: Structure and Putative Functions. *Biochemistry (Moscow)*, *68*(9), 945–951. doi:10.1023/A:1026077825584

Amara, I., Odena, A., Oliveira, E., Moreno, A., Masmoudi, K., Pagès, M., & Goday, A. (2012). Insights into Maize LEA proteins: from proteomics to functional approaches. *Plant & Cell Physiology*, *53*(2), 312–29. doi:10.1093/pcp/pcr183

Amara, I., Zaidi, I., Masmoudi, K., Dolors Ludevid, M., Pagès, M., Goday, A., & Brini, F. (2014). Insights into Late Embryogenesis Abundant (LEA) Proteins in Plants: From Structure to the Functions. *American Journal of Plant Sciences*, *5*(5), 3440–3455. doi:10.4236/ajps.2014.522360

Bailey, T. L., & Elkan, C. (1994). Fitting a Mixture Model by Expectation Maximization to Discover Motifs in Bipolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, 28–36. doi:citeulike-article-id:878292

Battaglia, M., Olvera-Carrillo, Y., Garciarrubio, A., Campos, F., & Covarrubias, A. A. (2008). The Enigmatic LEA Proteins and Other Hydrophilins. *Plant Physiology*, *148*(1), 6–24. doi:10.1104/pp.108.120725

Bewley, J. D. (1979). Physiological Aspects of Desiccation Tolerance. *Annual Review of Plant Physiology*, *30*(1), 195–238. doi:10.1146/annurev.pp.30.060179.001211

Bies-Ethève, N., Gaubier-Comella, P., Debures, A., Lasserre, E., Jobet, E., Raynal, M., … Delseny, M. (2008). Inventory, evolution and expression profiling diversity of the LEA (late embryogenesis abundant) protein gene family in Arabidopsis thaliana. *Plant Molecular Biology*, *67*(1-2), 107–124. doi:10.1007/s11103-008-9304-x

Boucher, V., Buitink, J., Lin, X., Boudet, J., Hoekstra, F. A., Hundertmark, M., … Leprince, O. (2010). MtPM25 is an atypical hydrophobic late embryogenesis-abundant protein that dissociates cold and desiccation-aggregated proteins. *Plant, Cell and Environment*, *33*(3), 418–430. doi:10.1111/j.1365-3040.2009.02093.x

Bray, E. a. (1993). Molecular Responses to Water Deficit. *Plant Physiology*, *103*, 1035–1040. doi:10.1093/jxb/ern087

Caruso, A., Morabito, D., Delmotte, F., Kahlem, G., & Carpin, S. (2002). Dehydrin induction during drought and osmotic stress in Populus. *Plant Physiology and Biochemistry*, *40*(12), 1033–1042. doi:10.1016/S0981-9428(02)01468-7

Chakrabortee, S., Boschetti, C., Walton, L. J., Sarkar, S., Rubinsztein, D. C., & Tunnacliffe, A. (2007). Hydrophilic protein associated with desiccation tolerance exhibits broad protein stabilization function. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 18073–18078. doi:10.1073/pnas.0706964104

Chatelain, E., Hundertmark, M., Leprince, O., Le Gall, S., Satour, P., Deligny-Penninck, S., … Buitink, J. (2012). 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 & Environment*, *35*(8), 1440–55. doi:10.1111/j.1365-3040.2012.02501.x

Chebotareva, N. A., Kurganov, B. I., & Livanova, N. B. (2004). Biochemical Effects of Molecular Crowding, *69*(11).

Chen, H., Gu, F., & Huang, Z. (2006). Improved Chou-Fasman method for protein secondary structure prediction. *BMC Bioinformatics*, *7 Suppl 4*, S14. doi:10.1186/1471-2105-7-S4-S14

Cheng, Z., Targolli, J., Huang, X., & Wu, R. (2002). Wheat LEA genes, PMA80 and PMA1959, enhance dehydration tolerance of transgenic rice (Oryza sativa L.). *Molecular Breeding*, *10*(1-2), 71–82. doi:10.1023/A:1020329401191

Chou, P. Y., & Fasman, G. D. (1974). Prediction of protein conformation. *Biochemistry*, *13*(2), 222–245. doi:10.1002/pro.5560021016

Christiansen, A., Wang, Q., Cheung, M. S., & Wittung-Stafshede, P. (2013). Effects of macromolecular crowding agents on protein folding in vitro and in silico. *Biophysical Reviews*, *5(2), 137–145. doi:10.1007/s12551-013-0108-0*, *5*(2), 137–145. doi:10.1007/s12551-013-0108-0

Close, T. J. (1996). Dehydrins: Emergence of a biochemical role of a family of plant dehydration proteins. *Physiologia Plantarum*, *97*(4), 795–803. doi:10.1034/j.1399-3054.1996.970422.x

Cohen, P. (2000). The regulation of protein function by multisite phosphorylation--a 25 year update. *Trends in Biochemical Sciences*, *25*(12), 596–601. doi:10.1016/S0968-0004(00)01712-6

Collett, H., Shen, A., Gardner, M., Farrant, J. M., Denby, K. J., & Illing, N. (2004). Towards transcript profiling of desiccation tolerance in Xerophyta humilis: Construction of a normalized 11 k X. humilis cDNA set and microarray expression analysis of 424 cDNAs in response to dehydration. *Physiologia Plantarum*, *122*(1), 39–53. doi:10.1111/j.1399-3054.2004.00381.x

Costa, M. C. D., Farrant, J. M., Oliver, M. J., Ligterink, W., Buitink, J., & Hilhorst, H. M. W. (2015). *Key genes involved in desiccation tolerance and dormancy across life forms. Plant Science*. doi:10.1016/j.plantsci.2016.02.001

Dai, A. G. (2013). Increasing drought under global warming in observations and models. *Nature Climate Change*, *3*(1), 52–58. doi:10.1038/nclimate1633

Dang, N. X., Popova, A. V., Hundertmark, M., & Hincha, D. K. (2014). Functional characterization of selected LEA proteins from Arabidopsis thaliana in yeast and in vitro. *Planta*, *240*(2), 325–336. doi:10.1007/s00425-014-2089-z

Dennis, T. (2015). The characterisation of three Group 3 LEA-like proteins from the resurrection plant, Xerophyta humilis, 4040.

Dinakar, C., & Bartels, D. (2013). Desiccation tolerance in resurrection plants: new insights

from transcriptome, proteome and metabolome analysis. *Frontiers in Plant Science*, *4*(November), 482. doi:10.3389/fpls.2013.00482

Dure, L., Crouch, M., Harada, J., Ho, T. H. D., Mundy, J., Quatrano, R., Sung, Z. R. (1989). Common amino acid sequence domains among the LEA proteins of higher plants. *Plant Molecular Biology*, *12*(5), 475–486. doi:10.1007/BF00036962

Eriksson, S. K., Kutzer, M., Procek, J., Gröbner, G., & Harryson, P. (2011). Tunable membrane binding of the intrinsically disordered dehydrin Lti30, a cold-induced plant stress protein. *The Plant Cell*, *23*(6), 2391–404. doi:10.1105/tpc.111.085183

Fallis, A. . (2013). *Circular Dichroism and the Conformational Analysis of Biomolecules*. *Journal of Chemical Information and Modeling* (Vol. 53). doi:10.1017/CBO9781107415324.004

Farrant, J. M., Brandt, W., & Lindsey, G. G. (2007). An Overview of Mechanisms of Desiccation Tolerance in Selected Angiosperm Resurrection Plants.

Farrant, J. M., Cooper, K., Hilgart, A., Abdalla, K. O., Bentley, J., Thomson, J. A., Rafudeen, M. S. (2015). A molecular physiological review of vegetative desiccation tolerance in the resurrection plant Xerophyta viscosa (Baker). *Planta*, *242*(2), 407–426. doi:10.1007/s00425-015-2320-6

Farrant, J. M., Cooper, K., & Nell, H. (2012). Plant Stress Physiology.

Figueras, M., Pujal, J., Saleh, A., Savé, R., Pagès, M., & Goday, A. (2004). Maize Rab17 overexpression in Arabidopsis plants promotes osmotic stress tolerance. *Annals of Applied Biology*, *144*(3), 251–257. doi:10.1111/j.1744-7348.2004.tb00341.x

Furuki, T., Shimizu, T., Chakrabortee, S., Yamakawa, K., Hatanaka, R., Takahashi, T., … Sakurai, M. (2012). Effects of Group 3 LEA protein model peptides on desiccation-induced protein aggregation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1824*(7), 891–897. doi:10.1016/j.bbapap.2012.04.013

Gal, T. Z., Glazer, I., & Koltai, H. (2004). An LEA group 3 family member is involved in survival of C. elegans during exposure to stress. *FEBS Letters*, *577*(1-2), 21–26. doi:10.1016/j.febslet.2004.09.049

Ginbot, Z. G. (2011). Characterisation of two desiccation-linked Group 1 LEA proteins from the resurrection plant, *Xerophyta humilis.* (February 2011).

Goyal, K., Pinelli, C., Maslen, S. L., Rastogi, R. K., Stephens, E., & Tunnacliffe, A. (2005). Dehydration-regulated processing of late embryogenesis abundant protein in a desiccation-tolerant nematode. *FEBS Letters*, *579*(19), 4093–4098. doi:10.1016/j.febslet.2005.06.036

Goyal, K., Walton, L. J., & Tunnacliffe, A. (2005). LEA proteins prevent protein aggregation due to water stress. *The Biochemical Journal*, *388*(Pt 1), 151–7. doi:10.1042/BJ20041931

Graether, S. P., & Boddington, K. F. (2014). Disorder and function: a review of the dehydrin protein family. *Frontiers in Plant Science*, *5*(October), 576.

doi:10.3389/fpls.2014.00576

Habchi, J., Tompa, P., Longhi, S., & Uversky, V. N. (2014). Introducing protein intrinsic disorder. *Chemical Reviews*, *114*(13), 6561–6588. doi:10.1021/cr400514h

Hand, S. C., Menze, M. A., Toner, M., Boswell, L., & Moore, D. (2011). LEA Proteins During Water Stress: Not Just for Plants Anymore. *Annual Review of Physiology*, *73*(1), 115–134. doi:10.1146/annurev-physiol-012110-142203

Hara, M., Fujinaga, M., & Kuboi, T. (2005). Metal binding by citrus dehydrin with histidine-rich domains. *Journal of Experimental Botany*, *56*(420), 2695–2703. doi:10.1093/jxb/eri262

Hatanaka, R., Hagiwara-Komoda, Y., Furuki, T., Kanamori, Y., Fujita, M., Cornette, R., … Kikawada, T. (2013). 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*(11), 1055–1067. doi:10.1016/j.ibmb.2013.08.004

Hoekstra, F. a, Golovina, E. a, & Buitink, J. (2001). Mechanisms of plant desiccation tolerance. *Trends in Plant Science*, *6*(9), 431–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11544133

Hundertmark, M., & Hincha, D. K. (2008). LEA (Late Embryogenesis Abundant) proteins and their encoding genes in Arabidopsis thaliana, *22*, 1–22. doi:10.1186/1471-2164-9-118

Illing, N., Denby, K. J., Collett, H., Shen, A., & Farrant, J. M. (2005). The signature of seeds in resurrection plants: A molecular and physiological comparison of desiccation tolerance in seeds and vegetative tissues. In *Integrative and Comparative Biology* (Vol. 45, pp. 771–787). doi:10.1093/icb/45.5.771

Imai, R., Chang, L., Ohta, A., Bray, E. A., & Takagi, M. (1996). A lea-class gene of tomato confers salt and freezing tolerance when expressed in Saccharomyces cerevisiae. *Gene*, *170*(2), 243–248. doi:10.1016/0378-1119(95)00868-3

Ingle, R.. Schmidt, U. G., Farrant, J. M., & Mundree, S. G. (2007). Proteomic analysis of leaf proteins during dehydration of the resurrection plant. *Environment*, 435–446. doi:10.1111/j.1365-3040.2006.01631.x

Ingram, J., & Bartels, D. (1996). the Molecular Basis of Dehydration Tolerance in Plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, *47*, 377–403. doi:10.1146/annurev.arplant.47.1.377

Iturriaga, G., Schneider, K., Salamini, F., & Bartels, D. (1992). Expression of desiccation-related proteins from the resurrection plant Craterostigma plantagineum in transgenic tobacco. *Plant Molecular Biology*, *20*(3), 555–558. doi:10.1007/BF00040614

Jones, P. G., & Thornton, P. K. (2003). The potential impacts of climate change on maize production in Africa and Latin America in 2055. *Global Environmental Change*, *13*(1), 51–59. doi:10.1016/S0959-3780(02)00090-0

Kosová, K., Holková, L., Prášil, I. T., Prášilová, P., Bradáčová, M., Vítámvás, P., & Čapková, V. (2008). Expression of dehydrin 5 during the development of frost tolerance in barley (Hordeum vulgare). *Journal of Plant Physiology*, *165*(11), 1142–1151. doi:10.1016/j.jplph.2007.10.009

Kovacs, D., Agoston, B., & Tompa, P. (2008). Disordered plant LEA proteins as molecular chaperones, *3*(9), 710–713. doi:10.1104/pp.108.118208.710

Kovacs, D., Kalmar, E., Torok, Z., & Tompa, P. (2008). Chaperone activity of ERD10 and ERD14, two disordered stress-related plant proteins. *Plant Physiology*, *147*(1), 381–90. doi:10.1104/pp.108.118208

Lata, C., & Prasad, M. (2011). Role of DREBs in regulation of abiotic stress responses in plants. *Journal of Experimental Botany*, *62*(14), 4731–4748. doi:10.1093/jxb/err210

Liu, X., Wang, Z., Wang, L., Wu, R., Phillips, J., & Deng, X. (2009). LEA 4 group genes from the resurrection plant Boea hygrometrica confer dehydration tolerance in transgenic tobacco. *Plant Science*, *176*, 90–98. doi:10.1016/j.plantsci.2008.09.012

Mouillon, J.-M., Eriksson, S. K., & Harryson, P. (2008). Mimicking the plant cell interior under water stress by macromolecular crowding: disordered dehydrin proteins are highly resistant to structural collapse. *Plant Physiology*, *148*(4), 1925–37. doi:10.1104/pp.108.124099

Oliver, M. J., Tuba, Z., & Mishler, B. D. (2000). The evolution of vegetative desiccation tolerance in land plants. *Plant Ecology*, *151*(1), 85–100. doi:10.1023/A:1026550808557

Park, B. J., Liu, Z., Kanno, A., & Kameya, T. (2005). Genetic improvement of Chinese cabbage for salt and drought tolerance by constitutive expression of a B. napus LEA gene. *Plant Science*, *169*(3), 553–558. doi:10.1016/j.plantsci.2005.05.008

Place, S. P., & Hofmann, G. E. (2005). Comparison of Hsc70 orthologs from polar and temperate notothenioid fishes: differences in prevention of aggregation and refolding of denatured proteins. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology*, *288*(5), R1195–202. doi:10.1152/ajpregu.00660.2004

Peton, N. (2013). Development of an activity based protien profiling array based assay for quantitative kinase inhibitor selectivity. (August 2013).

Popova, A. V., Rausch, S., Hundertmark, M., Gibon, Y., & Hincha, D. K. (2015). 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. *Biochimica et Biophysica Acta*, *1854*(10), 1517–1525. doi:10.1016/j.bbapap.2015.05.002

Porat, R., Pasentsis, K., Rozentzvieg, D., Gerasopoulos, D., Falara, V., Samach, A., … Kanellis, A. K. (2004). Isolation of a dehydrin cDNA from orange and grapefruit citrus fruit that is specifically induced by the combination of heat followed by chilling temperatures. *Physiologia Plantarum*, *120*(2), 256–264. doi:10.1111/j.0031-9317.2004.0242.x

Rodriguez, M. C. S., Hussain, S. S., Alquezar, D., Rasmussen, M., Gilbert, T., … Mundy, J.

(2010). Transcriptomes of the desiccation-tolerant resurrection plant Craterostigma plantagineum. *Plant Journal*, *63*(2), 212–228. doi:10.1111/j.1365-313X.2010.04243.x

Romero, P., Obradovic, Z., & Dunker, a. K. (1997). Sequence data analysis for long disordered regions prediction in the calcineurin family. *Genome Informatics*, *8*, 110–124. doi:10.11234/gi1990.8.110

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology*. doi:10.3389/fmicb.2014.00172

Saavedra, L., Svensson, J., Carballo, V., Izmendi, D., Welin, B., & Vidal, S. (2006). A dehydrin gene in Physcomitrella patens is required for salt and osmotic stress tolerance. *Plant Journal*, *45*(2), 237–249. doi:10.1111/j.1365-313X.2005.02603.x

Sen, T. Z., Jernigan, R. L., Garnier, J., & Kloczkowski, A. (2005). GOR V server for protein secondary structure prediction. *Bioinformatics*, *21*(11), 2787–2788. doi:10.1093/bioinformatics/bti408

Sun, X., Rikkerink, E. H. a, Jones, W. T., & Uversky, V. N. (2013). Multifarious roles of intrinsic disorder in proteins illustrate its broad impact on plant biology. *The Plant Cell*, *25*(1), 38–55. doi:10.1105/tpc.112.106062

Swire-Clark, G. A., & Marcotte, W. R. (1999). The wheat LEA protein Em functions as an osmoprotective molecule in Saccharomyces cerevisiae. *Plant Molecular Biology*, *39*(1), 117–128. doi:10.1023/A:1006106906345

Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*, *72*(2), 211–222. doi:10.1007/s00253-006-0465-8

Thalhammer, A., Hundertmark, M., Popova, A. V., Seckler, R., & Hincha, D. K. (2010). Interaction of two intrinsically disordered plant stress proteins (COR15A and COR15B) with lipid membranes in the dry state. *Biochimica et Biophysica Acta - Biomembranes*, *1798*(9), 1812–1820. doi:10.1016/j.bbamem.2010.05.015

Tunnacliffe, A., & Wise, M. J. (2007). The continuing conundrum of the LEA proteins. *Die Naturwissenschaften*, *94*(10), 791–812. doi:10.1007/s00114-007-0254-y

VanBuren, R., Bryant, D., Edger, P. P., Tang, H., Burgess, D., Challabathula, D., … Mockler, T. C. (2015). Single-molecule sequencing of the desiccation-tolerant grass Oropetium thomaeum. *Nature*, *527*(7579), 508–11. doi:10.1038/nature15714

Waters, R. (2015). Molecular characterisation of the “LEAome” in the resurrection plant Xerophyta humilis (Baker), (August 2015).

Wise, M. J. (2003). LEAping to conclusions: a computational reanalysis of late embryogenesis abundant proteins and their possible roles. *BMC Bioinformatics*, *4*(1), 52. doi:10.1186/1471-2105-4-52

Xiao, L., Yang, G., Zhang, L., Yang, X., Zhao, S., Ji, Z., … He, Y. (2015). The resurrection

genome of Boea hygrometrica : A blueprint for survival of dehydration. *Proceedings of the National Academy of Sciences*, *112*(18), 201505811. doi:10.1073/pnas.1505811112

Xu, D., Duan, X., Wang, B., Hong, B., Ho, T., & Wu, R. (1996). Expression of a Late Embryogenesis Abundant Protein Gene, HVA1, from Barley Confers Tolerance to Water Deficit and Salt Stress in Transgenic Rice. *Plant Physiology*, *110*(1), 249–257. doi:110/1/249 [pii]

Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K., & Uversky, V. N. (2010). *PONDR-FIT: A meta-predictor of intrinsically disordered amino acids*. *Biochimica et Biophysica Acta - Proteins and Proteomics* (Vol. 1804). doi:10.1016/j.bbapap.2010.01.011

Yin, Z., Rorat, T., Szabala, B. M., Zi????kowska, A., & Malepszy, S. (2006). Expression of a Solanum sogarandinum SK3-type dehydrin enhances cold tolerance in transgenic cucumber seedlings. *Plant Science*, *170*(6), 1164–1172. doi:10.1016/j.plantsci.2006.02.002

Zhou, H.-X., Rivas, G., & Minton, A. P. (2008). Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annual Review of Biophysics*, *37*, 375–97. doi:10.1146/annurev.biophys.37.032807.125817

Zhu, Y., Wang, B., Phillips, J., Zhang, Z. N., Du, H., Xu, T., … Deng, X. (2015). Global transcriptome analysis reveals acclimation-primed processes involved in the acquisition of desiccation tolerance in Boea hygrometrica. *Plant and Cell Physiology*, *56*(7), 1429–1441. doi:10.1093/pcp/pcv059

Appendices

Appendix A: Nucleotide and amino acid sequence information

A1. Full-length nucleotide sequences for XhLEA2-4 and XhLEA2-5 with translated amino acid sequences

Clone ID:
Xh_RD_30C12, renamed XhLEA2-4
Dehydrin, 420bp, ~14.80kDa

Nucleotide sequence:

atggagggctccgggaaccaagaccagcactaccgcaccagcgagcacgctgctcctggccagggcgtgcaccctagc
cagcacggaaaaggcaccagcgagttcgccgctaccggccagggtatgttcggcggccagcatcacgaccagaacaag
catcagggacatggaactgctcacgagagtcatggagaagggaagaaggagggaattacggagaagattaaggagaaa
ctcccaggacagcaccaccaagaagccaccggcaaccagggtttgacacacagccagcaaggccatggagccacaac
aaaggagacactcctcccaggacagcaccatcaagaagccaccggcaatcagggtttcactcacaacaagcaaggccat
ggagccacaaccaaggacacactcctttga

Amino acid sequence:
MEGSGNQDQHYRTSEHAAPGQGVHPSQHGKGTSEFAATGQGMFGGQHHDQ
NKHQGHGTAHESHGEGKKEGITEKIKEKLPGQHHQEPTGNQGLTHSQQGHAT
TKETLLPGQHHQEATGNQGFTHNKQGHGATTKDTLL

Clone ID:
Xh_RD_19H04, renamed XhLEA2-5
Dehydrin, 417bp, ~14.50kDa

Nucleotide sequence:

atggagccctacagccatcaaactcaccaccacgagacaggcaacgtccccggcgcctatggcggtgcccctcctgctgc
tgctggctacggcgcccacgagggcctgcagcaaccacatgatagaaaggatcacaagggtcttggacagaagatgggc
gagaagctccaccgctccagcagcagcagctccagctccagctctgagagtgatggagaaggaggaaggaggaagaa
gggaatcaaggagaagatcaaggagaagcttcctgggaagaagaaggaaggagcaaccgccaccggaaccaccgcca
ccggaacaacagctactggaacaaccaccaccaccactggcgtgcacggggagaagaaggggatgatggagaagatc
aaggagaagctccccggccaccactaa

Amino acid sequence:
MEPYSHQTHHHETGNVPGAYGGAPPAAAGYGAHEGLQQPHDRKDHKGLGQ
KMGEKLHRSSSSSSSSSESDGEGGRRKKGIKEKIKEKLPGKKKEGATATGTA
TGTTATGTTTTTTGVHGEKKGMMEKIKEKLPGHH

A2: 1kb upstream promoter sequence of *Xerophyta viscosa* homologues provided by Dr Maria Cecilia Da Costa (Wageningen University, Netherlands).

X. humilis gene name: XhLEA2-4

X. viscosa homologue ID: Xvis02_20155
1kb upstream promoter sequence:
ggcgtgtgctcgaaactattgaaggagaggtctggtgctttgagattggtcagtaatgaactcgctgtggcggccattctggtttgt
ggacatgtttaccacgccgagtgcttggagattgtgactcctgaaaccgacaagtacgatccctcttgtccagtttgcgcccttgga
ggaaaacttccatataagttactcgcaagatttgattcaaaggcaaagaacaagatttcaaagattgcaattgatatggatattgatg
gggacagttgccgagtatcagaaaagaggaaaagcaccaaagcttggagtaagctgtagcatgaagagttcttttagcaggc
cattctttagaaggcacttcccaattggatcatcacgacatgtgaggtctgtgtcagagagtgaaagctcgagaaaaaaggggttct
gggcaaaatacaggagagattgatcagggatggaaatgttttgaaagctgaaacagatgctggtttttgtattttgctgtagctgcag
tgtatctccacgcatatgcgtgtttagtatttttattgctcccaggatggcgcttttacagctattatcggaaaataaaaatttaattgcg
agtctacttagctatgcttaataatttagaaagctatcattaagtgtgtttcttattagttcacagctacttaaatatgctacttttttagcgt
gaaggtgctagagaaggtttgttttgctcggtactagggtaaaccatattacttataaatacaaaaagatatatatcacgtcatgcaaa
cctcgaatctatccgttttgacgttgtataactattcgcatgtgtgctatttagcatgataaagacagatcacgagcagccactatcca
agtaaagatatcttcagatttttttttttttctttatttcctgagaacagaagacttggtgctcgccacctatgtacgtgccaatctacgac
ccatccgcgtgtctgtcttccaga

X. viscosa homologue ID: Xvis02_13077
1kb upstream promoter sequence:
gtatcattttctgtacattcttcttttgtattccacgtatatgtatattaaggggtgttagtgtttgacttgtcaggaaaaaaaatccaattga
ttttagatgcttactatttccatcaatagaatggataaccaagctttatgattaattgtttatatccatgcgtttttggagaatggaggaga
atataagaaatctcgtataaccaatttttaaaatcacgatttatttcctacaggctctattctgatttttcaaaaaatggaggaactaatcaa
agttataatgactattaactacgttcatatgaaatcgaatccatctttcaaagttaagaatattattttgagtacataaaagagggtgtaa
aagtaattcgtcgagaaaattaaaataaacgtcaccgtgaaatttataacggaacgacatagctcggatgatggaaataatttaactt
ataaattgaaaaagaaaaaaaaataacaatatccttgtagaaaaataaaaaaaataataaccatgtagagttattctttgcaataaac
agatggccatgctctggcaaaaaagcatccaaggcaaccgtccaaacaatgttaagtcttatccaatgacgtcccacaccagacg
agatattggacacatgcgacgatttgaaaccgagcagcgaaaatctggacgcacccgattcacacctctatttttgtgagataacttt
ataactaactagtcaatacaattaatatccgatacgtgttccttccaccgacccccctatcgtgtctcgtatctagactctgcgccatgc
cgcctccgccactgcaccgtgccctcgccatcacaatgccacctactgcacccatcacctccactcttcttgtgtatataaagcaat
ctttatcttccctttcgccatcaacaaagcttcttagctacgattagttatcttcttcaagcggttctacaagtgtttgtttacttttctttcctt
tttgaagaaga

X. humilis gene name: XhLEA2-5

X. viscosa homologue ID: Xvis02_01738
1kb upstream promoter sequence:
aagtatcatatcagtctccatttcaaaaatctaccaaatcagttcctcaaatattgtttgaatttttttcggataaaaatatccctatactt
gttaacagtttggacacaaaaaatatctatatagaagtatttttttaaaaaaagtttgaacatgaaagttgtagagttcgtcgagatcta
caatttttatgttgaacatttttttttaaaaaatactttaaaatatgtgtttttgagatccgaagtgtaaccagctcgacccgaaatcgctg
aaacagtttggacgcaaaaaatacctatttaaagtatttttttgaaaaaattttgaatattaaagttgtagatctcgtcggcgaaacga
ggctgatctcggccttggcgaaacgaggatattttgttcgaaaaaaatacgaaacagtagcggagaggctgatttgtccttttta
aacagagaggctgtccaaagaattcaaaattggaggctgatacgtcattatcccttaataaaatgcgtcctacaatttgtaaaatat
aaattttcggaaaagaattaactataagtaatatttttttactatatttttctgaacttaatttgatattataaacagatgcgtcggtgga
aaagataaagacagctacgttatggccgtacgtaattatcgtatatataaatatacttagaaacgactcatcttacgtgtatgtggtt
ataatgctatgagcaggacgaagacaaataagtctgagtcctgatttttccagatgagagattttcctcggagaagagaaccttc
atcttacttcactctcgccacctaccacgtacatgtctcaccctctagcgaaggcgctacgtggcgccgccgttcgctaccgacc
cctccgtggccagcttccagacgccgctgatgccgtcgtgtcgcggcttctctgacccgtccccgtcagcttatccgctcgccg
cgcccctatataatatcctccatggcgttac

Appendix B: Results from generation of recombinant plasmid with dehydrin genes

B1. Sequencing data from recombinant vector containing XhLEA2-4 and XhLEA2-5 genes provided by Macrogen, Netherlands.

Clone ID:
XhLEA2-4 in pET21a(+):His

Predicted amino acid sequence:
MEGSGNQDQHYRTSEHAAPGQGVHPSQHGKGTSEFAATGQGMFGGQHHDQN
KHQGHGTAHESHGEGKKEGITEKIKEKLPGQHHQEPTGNQGLTHSQQGHATTK
ETLLPGQHHQEATGNQGFTHNKQGHGATTKDTLL

Nucleotide sequence for recombinant plasmid:
>150910-10_g01_30c12_1_t7.ab1 1124
acatcgatagtatatttcccctctagaataattttgtttaactttaagaaggagatatacatatgcaccaccaccaccaccaccgcg
gatccgaattcgagctccgtcgacaaatggagggctccgggaaccaagaccagcactaccgcaccagcgagcacgctgct
cctggccagggcgtgcaccctagccagcacggaaaaggcaccagcgagttcgccgctaccggccagggtatgttcggcgg
ccagcatcacgaccagaacaagcatcagggacatggaactgctcacgagagtcatggagaagggaagaaggagggaatta
cggagaagattaaggagaaactcccaggacagcaccaccaagaagccaccggcaaccagggtttgacacacagccagca
aggccatggagccacaacaaaggagacactcctcccaggacagcaccatcaagaagccaccggcaatcagggtttcactca
caacaagcaaggccatggagccacaaccaaggacacactccttgagcggccgcactcgagcaccaccaccaccaccact
gagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttgg
ggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggattggcgaatgggacgcgccctgtagcgg
cgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctt
tcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgcttta
cggcacctcgaccccaaaaaacttgattagggtgatggttcacttattgggccatcgccctgatagacgggttttcacccttgac
tttggattccactttctttaatattggactcttgttccaactggaacaacacttaacccgattctggatattcttttgattttaagggatttt
ccggatttaggcaatgggttaaaaaagtgctaatttcttaa

Translated amino acid sequence:
5'3' Frame 2
L X I D S I F P L **Stop** N N F V **Stop** L **Stop** E G D I H **Met** H H H H H H R G S E F E L R R Q
Met E G S G N Q D Q H Y R T S E H A A P G Q G V H P S Q H G K G T S E F A A T G Q G
Met F G G Q H H D Q N K H Q G H G T A H E S H G E G K K E G I T E K I K E K L P G Q H H
Q E A T G N Q G L T H S Q Q G H G A T T K E T L L P G Q H H Q E A T G N Q G F T H N K Q
G H G A T T K D T L L **Stop** A A A L E H H H H H H **Stop** D P A A N K A R K E A E L A A A T
A E Q **Stop** L A **Stop** P L G A S K R V L R G F L L K G G T I S G L A N G T R P V A A H **Stop**
A R R V W W L R A A **Stop** P L H L P A P **Stop** R P L L S L S S L P F S P R S P A F P V K L
Stop I G G S L **Stop** G S D L V L Y G T S T P K N L I R V **Met** V H L L G H R P D R R V F T L
Stop L W I P L S L I L D S C S N W N N T **Stop** P D S G Y S F D F K G F S G F R Q W V K K V
L I S **Stop**

Clone ID:
XhLEA2-5 in pET21a(+):His

Predicted amino acid sequence:
MEPYSHQTHHHETGNVPGAYGGAPPAAAGYGAHEGLQQPHDRKDHKGLGQK
MGEKLHRSSSSSSSSSESDGEGGRRKKGIKEKIKEKLPGKKKEGATATGTATGT
TATGTTTTTTGVHGEKKGMMEKIKEKLPGHH

Nucleotide sequence for recombinant plasmid:
>150910-10_M01_19H04_1_T7.ab1 946
taagattcgttcatttcgccgtcgtagagatttttttgtttaactttaagaaggagatatacatatgcaccaccaccaccaccaccga
cggatccgaattcgagctccgtcgacaaatggagccctacagccatcaaactcaccaccacgagacaggcaacgtccccgg
cgcctatggcggtgcccctcctgctgctgctggctacggcgcccacgagggcctgcagcaaccacatgatagaaaggatca
caagggtcttggacagaagatgggcgagaagctccaccgctccagcagcagctccagctccagctctgagagtgatgg
agaaggaggaaggaggaagaagggaatcaaggagaagatcaaggagaagcttcctgggaagaagaaggaaggagcaa
ccgccaccggaaccaccgccaccggaacaacagctactggaacaaccaccaccaccactggcgtgcacggggagaagaa
ggggatgatggagaagatcaaggagaagctccccggccaccactaagcggccgcactcgagcaccaccaccaccaccac
tgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataacccctg
gggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggattggcgaatgggacgcgccctgtagcg
gcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgc
tttcttcccttcctttctcgccacgttcgccggctttccccgtcagctctaaatcgggggctccctttagggttccgaatttagtgc
ttacggcacctcgaccccaaacttggattaggtgatggtcacgtag

Translated amino acid sequence:
5'3' Frame 2
X X E F H S F S P S R I K F V **Stop** L **Stop** E G D I H **Met** H H H H H H R G S E F E L R R Q
Met E P Y S H Q T H H H E T G N V P G A Y G G A P P A A A G Y G A H E G L Q Q P H D R
K D H K G L G Q K M G E K L H R S S S S S S S S S S S E S D G E G G R R K K G I K E K I K E
K L P G K K K E G A T A T G T T A T G T T A T G T T T T T T G V H G E K K G M M E K I K E
K L P G H H **Stop** A A A L E H H H H H H **Stop** D P A A N K A R K E A E L A A A T A E Q **Stop**
L A **Stop** P L G A S K R V L R G F L L K G G T I S G L A N G T R P V S G A L S A A G V V V T
R S V T A T L A S A L A P A P F A F F P S F L A T F A G F P R Q A L N R G A P F K G S D L
V L Y G H L D P K K L **Stop** L V V D G F T **Stop** V G H S P W I D S F F C

Appendix C: Results from the aggregation assay pilot study

C1. Absorbance readings at 340nm detecting the aggregation potential at 55ºC and 75ºC of each individual assay component. Samples were analysed in triplicate.

	Untreated	55ºC	75ºC
LDH	0.069	0.256	0.288
Trehalose	0.072	0.073	0.073
BSA	0.075	0.074	0.073
XhLEA2-4	0.073	0.074	0.074
XhLEA2-5	0.074	0.075	0.073