FUNCTIONAL CHARACTERISATION OF LEA PROTEINS FROM BDELLOID ROTIFERS

RASHMI TRIPATHI

MPHIL, UNIVERSITY OF CAMBRIDGE

This dissertation is submitted for the degree of

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMICAL ENGINEERING AND BIOTECHNOLOGY

NEWNHAM COLLEGE

UNIVERSITY OF CAMBRIDGE

2012

Preface

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. This dissertation contains 40,116 words including figure legends, tables and bibliography and 46 figures.

Parts of this work were published in:

- Chakrabortee, S., Tripathi, R., Watson, M., Kaminski Schierle, G.S., Kurniawan, D.P., Kaminski, C.F., Wise, M.J., and Tunnacliffe, A. (2012). Intrinsically disordered proteins as molecular shields. Molecular Biosystems *8*, 210-219.
- Tripathi, R., Boschetti, C., McGee, B., and Tunnacliffe, A. (2012). Trafficking of bdelloid rotifer LEA proteins. Journal of Experimental Biology. (recently accepted)

Rashmi Tripathi

Summary

'Anhydrobiosis' or 'life without water' is a fascinating phenomenon that was first described by the eminent Dutch microscopist Antony van Leewenhoek in bdelloid rotifers in 1702. Despite being studied for over three hundred years, our understanding of its molecular basis remains largely elusive. Recently, two group 3 LEA (late embryogenesis abundant) proteins, ArLEA1A and ArLEA1B, have been identified in the desiccation-induced gene set in the bdelloid species *Adineta ricciae* that are hypothesised to protect these animals by preventing protein aggregation and stabilising membranes respectively. In this dissertation, the functional characteristics of bdelloid LEA proteins have been further explored using computational and experimental tools. Analysis of their phylogeny and domain composition reveals that ArLEA1A and ArLEA1B are evolutionarily distinct from other related group 3 LEA proteins. Moreover, unlike some LEA proteins that are unstructured, bdelloid LEA proteins are predicted to form α-helices, with ArLEA1B having an additional propensity to polymerise into tropomyosin like filaments. These proteins are also predicted to localise in the ER (endoplasmic reticulum) and interact with cell signalling molecules. Intracellular localisation analysis of ArLEA1A and ArLEA1B using confocal microscopy confirms that these proteins translocate into the ER as predicted and their distribution within the entire secretory system and the cell exterior is regulated by their N- and C-terminal signals. Preliminary results suggest that over-expressing ArLEA1A fails to provide protection against protein aggregation within the mammalian ER. Lastly, both ArLEA1A and ArLEA1B are found to provide partial protection against desiccation induced damage of fluorescence properties of the red fluorescent protein mCherry. Overall, the results in this dissertation provide important mechanistic insights into the mode of action of bdelloid LEA proteins during anhydrobiosis.

Table of Contents

List of Abbreviations

°C	degree celsius
ABA	abcissic acid
ADDA	Automatic Domain Decomposition Algorithm
ANOVA	Analysis of variance
As	adenosines
ASR	abcissic acid stress ripening
BLAST	basic local alignment search tool
bp	base pair
BSA	bovine serum albumin
CD	circular dichroism
COS-7	CV-1 (simian) in Origin, and carrying the SV40 genetic material
DAPI	4',6'-diamidino-2-phenylindole
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
EDTA	ethylenediaminotetraacetic acid
EGFP	enhanced green fluorescent protein
EGTA	ethylene glycol tetraacetic acid
ER	endoplasmic reticulum
et al.	et alii
FBS	fetal bovine serum
FLAG	phenylalanine (F) leucine (L) alanine (A) glycine (G)
FPLC	fast protein liquid chromatography
FRET	fluorescence resonance energy transfer
FTIR	Fourier-transform infrared spectroscopy
GFP	green fluorescent protein
HBSS	Hank's balanced salt solution
HD	Huntington's disease
HGT	horizontal gene transfer
HMM	hidden Markov model
HRP	horseradish peroxidase
hsp	heat shock protein
IDPs	intrisically disordered proteins
IEX	ion exchange chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	kilobases
KS	Kolmogorov-Smirnov
LB	Luria broth
LEA	late embryogenesis abundant
MALDI	Matrix-assisted laser desorption/ionization
MDTF	mean dihedral transition frequency
ML	maximum likelihood
MP	maximum parsimony
mRFP	monomeric red fluorescent protein
MSA	multiple sequence alignment
mTORR	millitorr

MTS	3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium
MWCO	molecular weight cut-off
N.A.	numerical aperture
NEB	New England Biolabs
Ni-NTA	nickel-nitrilotriacetic acid
NJ	Nearest Joining
NMR	nuclear magnetic resonance
OD_{600}	optical density measured at 600nm wavelength
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PBST	phosphate buffered saline containing 0.2% Tween-20
PCR	polymerase chain reaction
PDB	protein data bank
PES	polyethersulfone
PMS	Phenazine methosulfate
PMT	phomultiplier tube
polyQ	polyglutamine
POPP	protein or oligonucleotide probability profile
PPA	profile-profile threading alignment
PVDF	Polyvinylidene fluoride
Q74	74 glutamine repeats
RNA	ribonucleic acid
RNAi	RNA interference
RPM	rotations per minute
SBMA	spinobulbar muscular atrophy
SC	supercontinuum
SCA1	spinocerebellar ataxia
SD	standard deviation
SDS	sodium dodecyl sulphate
SEM	scanning electron microscope
SH2	Src homology 2
SOC	superoptimal broth with with catabolic repressor
SR	signal receptor
SRP	signal-recognition particles
TAE	Tris Acetate EDTA buffer
TCSPC	time correlated single photon counting
TEMED	N,N,N',N'-Tetramethylethylenediamine
TMAO	Trimethylamine N-oxide
UV	ultra-violet

List of Figures

List of Tables

Acknowledgements

I am highly grateful to Prof. Alan Tunnacliffe for giving me the opportunity to carry out the work described in this thesis. I am also highly grateful to the Cambridge Commonwealth Trust and the European Research Council for funding this study and to my parents for supporting me in the writing period. I would like to thank Drs. Chiara Boschetti, Tom Ellis, Matthew Watson, David Chau, Sohini Chakrabortee and Neil Williamson for sharing their experience in the laboratory. I have been patiently guided by Dr. Chiara Boschetti in using the confocal microscope. Fluorescence and absorbance spectra of mCherry were recorded in collaboration with Dr. Matthew Watson. Drs. Gabriele S. Kaminski Schierle and Claire Michel have been extremely kind in helping me generate the FLIM data. I would also like to thank all the technical and administrative staff for their ever willingness to help. I am indeed grateful to Prof. Nigel Slater for his kind encouragement throughout the course of this research. External advice from Prof. David Macherel, University of Angers, France and GMAP210 constructs from Jesús Cárdenas from CABIMER-CSIC, Seville, Spain is also much appreciated. This work is dedicated to my parents and the Mother.

Introduction

Water is of paramount importance for cell survival and its loss causes death of most organisms. Damage of cells due to water loss has been attributed to protein aggregation and denaturation, generation of reactive oxygen species and breakdown of cellular membranes (Crowe and Clegg, 1978). Intriguingly, the capacity to survive in the absence of water is innate in a number of species including bdelloid rotifers. This phenomenon has been termed 'anhydrobiosis' or 'life without water' (Giard, 1894; Tunnacliffe and Lapinski, 2003). Unlike the majority of anhydrobiotic organisms that apparently rely on non-reducing sugars like trehalose for protection of cells during dehydration (Clegg, 1965, 1967; Gaff, 1989; Quillet and Soulet, 1964; Westh and Ramlov, 1991), bdelloid rotifers lack these molecules (Caprioli et al., 2004; Lapinski and Tunnacliffe, 2003; McGee, 2006). Recently two proteins, ArLEA1A and ArLEA1B, belonging to the LEA protein family, have been discovered in the desiccation-induced gene set in bdelloids (McGee, 2006; Pouchkina-Stantcheva et al., 2007). These proteins are believed to protect cells in the absence of water by preventing desiccation-induced aggregation of other proteins and stabilising phospholipid bilayers, respectively. Moreover, they are predicted to localise in the ER. These findings have resulted in a paradigm shift in our current understanding of the fundamental principles of anhydrobiosis, since sugars that were once considered important in this process are no longer deemed to be absolutely essential.

So far, the functions of ArLEA1A and ArLEA1B have been characterised mainly by *in vitro* experiments and there is little evidence available regarding their functions in the cellular environment. Further, there are a number of fundamental questions related to the phylogeny, domain composition, three dimensional structure and protein-protein interaction motifs of these proteins that remain unanswered. It is also not known whether additional variants of these proteins might exist in the bdelloid genome. Lastly, it has not been elucidated whether these proteins can also prevent structural damage of other proteins caused by loss of water.

In view of these questions, it was decided to perform a comprehensive bioinformatics and experimental analysis of ArLEA1A and ArLEA1B. This research is expected to broaden our understanding of the nature and evolution of these proteins and increase our fundamental knowledge of their mode of action. Moreover, understanding their functions within the physiological context of the ER might enable the design of better therapeutic strategies aimed at countering ER stress-related disorders caused by protein misfolding and aggregation.

Aims and objectives

The primary aim of my dissertation is to investigate the functional characteristics of LEA proteins, ArLEA1A and ArLEA1B, found in the desiccation-tolerant bdelloid rotifer species *Adineta ricciae*. Specific objectives designed to achieve this aim are the following:

- Bioinformatics analysis of ArLEA1A and ArLEA1B phylogenies, domains, structural properties, signalling motifs and protein-protein interaction sites.
- Experimental validation of ArLEA1A and ArLEA1B localisation in the ER in living cells and functional verification of N-terminal and C-terminal signals in these proteins.
- Assessment of the protective effects of ArLEA1A and ArLEA1B in the physiological context of the mammalian ER.
- Assessment of the protective effects of ArLEA1A and ArLEA1B against desiccation-induced structural damage of the mCherry fluorophore.

Taking these objectives into consideration, this dissertation has been organised in the following manner:

Chapter One: Literature review, consisting of background information on bdelloid rotifers, principles of anhydrobiosis and cryptobiosis, general properties and classification of LEA proteins and our current knowledge of ArLEA1A and ArLEA1B functions.

Chapter Two: Materials and Methods describing the bioinformatics and experimental procedures in detail.

Chapter Three: Bioinformatics analysis of ArLEA1A and ArLEA1B proteins: their phylogeny, domains, structural models, localisation sequences and protein-protein interaction sites.

Chapter Four: Confocal microscopy of ArLEA1A and ArLEA1B in living cells and analysis of ER-Golgi trafficking mediated by ArLEA1A/ArLEA1B N-terminal and C-terminal signal sequences.

Chapter Five: Construction of a mammalian cell model with ER-Golgi localisation of aggregation-prone polyQ (polyglutamine) protein and subsequent analysis of the protective effects of ArLEA1A/ArLEA1B on polyQ induced cytotoxicity.

Chapter Six: Study of desiccation induced damage of mCherry absorbance and fluorescence properties and analysis of protective effects of ArLEA1A and ArLEA1B.

Chapter Seven: Conclusions and future work.

Chapter One: Literature Review

1.1 Bdelloid rotifers: evolutionary enigmas

To the naïve eye, bdelloid rotifers under the microscope seem like beautiful, ciliated organisms, swimming around in slow motion, curling up occasionally only to start swimming again. However, these animals are also one of the most extraordinarily robust animals on our planet, surviving a variety of stresses, including desiccation (Tunnacliffe and Lapinski, 2003), high doses of radiation (Gladyshev and Meselson, 2008), starvation and lethal fungal parasitic attacks (Wilson and Sherman, 2010). Since the initial discovery of anhydrobiosis in bdelloids by Antony van Leeuwenhoek in 1702, biologists have endeavoured for centuries to understand the basis of their biostability.

The species currently under study in our laboratory is *Adineta ricciae*, which belongs to the class Bdelloidea in the phylum Rotifera, alongside three additional classes: Seisonidea, Monogononta and Acanthocephala. *A. ricciae* was first isolated from the dry mud of Ryan's billabong in Australia (Segers and Shiel, 2005). Bdelloids reproduce exclusively by ameiotic parthenogenesis (thelytoky) unlike other classes that reproduce either sexually (Seisonidea and Acanthocephala) or exhibit cyclical parthenogenesis with the asexual cycle (amictic females) dominating in their life cycle (Monogononta). In the latter group, sexual reproduction is triggered by specific environmental signals like high population density (Wallace, 2002).

Bdelloid rotifers are perplexing from the evolutionary point of view since they have survived for >35 million years without sex (Ricci and Fontaneto, 2009). It is generally believed that sexual reproduction confers a greater adaptive advantage to species than asexual reproduction as it removes deleterious mutations and facilitates co-evolution with parasites and pathogens (West et al., 1999). However, under exceptional circumstances, organisms might devise means to shed parasites and pathogens and disperse to uninfected habitats (Ladle et al., 1993). It has been recently suggested that bdelloid rotifers have adapted similar means to counter

deadly fungal parasites through anhydrobiosis and wind dispersal as propagules to establish new infection-free populations. These findings might explain the prevalence of asexuality in these animals whereby they are no longer under any evolutionary pressure to counter co-evolving enemies through sex (Wilson and Sherman, 2010).

Moreover, sequence analysis of an approximately 50kb region containing four *hsp82* heat-shock genes in bdelloid species *Philodina roseola* suggests that bdelloids are degenerate tetraploids (Hur et al., 2009; Mark Welch et al., 2008). It is believed that this might have been an outcome of whole genome duplication, followed by segmental deletions.

Another fascinating aspect to the evolutionary mechanisms operating within bdelloid rotifer species has been the acquisition of bacterial, plant and fungal genes through horizontal gene transfer (HGT) (Boschetti et al., 2011; Gladyshev et al., 2008). Horizontal gene transfer involves the movement of genetic material from one organism to another through means other than vertical descent (from parent to offspring). Many of these genes have been incorporated in the sub-telomeric regions of the bdelloid genome and code for enzymes involved in carbohydrate, amino acid and peptide metabolism. It has been proposed that this gene transfer has been facilitated by membrane disruption and DNA fragmentation and repair occurring during anhydrobiosis (Gladyshev et al., 2008). Horizontal gene transfer has contributed immensely towards developing aerobic respiration, nitrogen fixation, photosynthesis and numerous other biosynthetic pathways in prokaryotes (Boucher et al., 2003). Recently it has been shown that bdelloid rotifers can accumulate foreign genes and express them in the desiccated state (Boschetti et al., 2011), by a presumably adaptive response to desiccation in these animals.

1.2 Principles of anhydrobiosis

Before I describe anhydrobiosis in rotifers, I would like to discuss the fundamental principles of this phenomenon in more general terms. Anhydrobiosis or 'life without water' (a term coined by Giard in 1894) is a form of 'cryptobiosis' (Keilin, 1959) which is defined as "the state of an organism when it shows no visible signs of life

and when its metabolic activity becomes hardly measurable or comes reversibly to a standstill". Keilin distinguished this state from dormancy or latent life, which can be marked by some amount of metabolic activity. A variety of stresses including desiccation (anhydrobiosis), low temperature (cryobiosis), lack of oxygen (anoxybiosis) or combinations of these could result in a cryptobiotic state of an animal (Clegg, 2001).

Currently, it is debatable whether metabolism ever comes to a reversible standstill during anhydrobiosis. Keilin and others have recognized the inherent difficulty in proving the absence of metabolism, as being 'unmeasurable' need not imply a lack of it (Clegg, 2001).

Setting the boundaries between the metabolic and the ametabolic state is indeed not trivial. It is speculated that removal of all but 0.1 g g dry weight^{-1} of water should result in cessation of all metabolism in a living cell (Clegg, 1978, 1986, 2001; Potts, 1994). However, in order to prove this, one would have to measure the rate of all the enzymatic reactions and energy transformations in a cell, which is currently beyond the current technological limits of detection. With the advent of sensitive microfluidics platforms (Bousse and Parce, 1994; Douglas et al., 2009), it might one day become possible to monitor subtle metabolic changes during the onset and maintenance of anhydrobiosis in various organisms. For now, it is safe to stick by what Clegg said about the three states of biological organisation- 'alive', 'dead' or 'cryptobiotic' (Clegg, 2001).

Anhydrobiotic organisms demonstrate certain common biochemical and biophysical adaptations. A general feature seems to be the presence of 'compatible solutes' (Yancey et al., 1982) that show a striking similarity in their properties among different organisms. These osmolytes are believed to favor the structural and functional integrity of macromolecules in conditions of stress.

For example, some anhydrobiotic organisms like plants (Gaff, 1989; Quillet and Soulet, 1964), brine shrimps (Clegg, 1965, 1967; Clegg and Evans, 1962) and tardigrades (Westh and Ramlov, 1991) trigger the production of non-reducing disaccharides like trehalose and sucrose during anhydrobiosis (Tunnacliffe and

Lapinski, 2003). In the early stages, when there is an increase in intracellular solute concentration, changes in pH and cellular viscosity, these sugars protect biopolymers by preventing structural changes according to the principle of preferential exclusion (Arakawa and Timasheff, 1982). During later stages of anhydrobiosis when water content is lowered to 0.3g g dry weight^{-1}, it is hypothesised that these sugars become strongly associated with polar groups of proteins and phospholipids by hydrogen bonding and serve as 'water replacement molecules' (Clegg, 1974, 1986; Clegg et al., 1982; Crowe et al., 1992; Crowe and Madin, 1974). In the last phase of anhydrobiosis, these disaccharides contribute towards vitrification or glass formation that results in entrapment of all bio-molecules in a specific space and time (Burke, 1986; Crowe et al., 1998; Sun and Leopold, 1997). These sugars also help prevent deleterious biochemical reactions in the dry state, like the Maillard and Fenton reactions (Loomis et al., 1979), as first proposed by Crowe (Crowe and Clegg, 1973).

Besides sugars, desiccation tolerance has also been associated with the expression of a unique category of proteins called LEA (late embryogenesis abundant) proteins. First discovered late in embryogenesis of cotton seeds (Dure et al., 1981; Galau and Dure, 1981), an analogy has often been drawn between these proteins and chaperones induced during the heat shock response (Tunnacliffe and Wise, 2007). A more thorough description of these proteins is provided in Sections 1.4, 1.5 1.6 and 1.7.

Lastly, one cannot rule out the importance of lipid assisted protein folding during desiccation (Bogdanov and Dowhan, 1999) as pointed out by Clegg (Clegg, 2001). It is postulated that certain 'lipo-chaperones' might be involved in stabilising membranes during desiccation in concert with sugars like trehalose and sucrose (Clegg, 2001).

In summary, cells might employ a variety of biochemical and biophysical strategies to preserve the structure and function of macromolecules during anhydrobiosis. We have only begun to understand a few principles involved in this fascinating phenomenon.

1.3 Anhydrobiosis in bdelloid rotifers

Although Antony van Leeuwenhoek is credited with first discovering anhydrobiosis in rotifers in the 18th century, this phenomenon was properly investigated by an Italian clergyman Lazzaro Spallanzani in his work *Opuscoli fisica animale et vegetabile* (Tunnacliffe and Lapinski, 2003). His description of anhydrobiosis also happens to be the most picturesque: 'When dried, the solids are contracted and distorted, the whole body of the animal is reduced to a hard shapeless atom of matter; pierced by a needle, it flies in pieces like a grain of salt'. Spallanzani further went on to show that rotifers can be stored in the dried state for longer than four years and acquire increased resistance to cold and heat (-12°C and 68°C respectively). In order to explain this phenomenon he postulated that all metabolism must stop in rotifers for them to withstand such extremes.

Previous analysis of the gross morphological changes associated with anhydrobiosis in rotifers has revealed that their body contracts and acquires a 'tun' shape (Ricci et al., 2003). A more detailed study of the dry state morphology of rotifers has been performed only very recently by Marotta et al. using confocal and electron microscopy (Marotta et al., 2010). This investigation provides important glimpses into the morphological changes occurring in rotifer cellular systems during anhydrobiosis for the first time (Fig. 1.1).

As can be observed, in the relaxed hydrated specimen a clearly visible body cavity separates the internal organs from each other and from the body wall (Fig. 1.1 b,c). However, in the hydrated contracted specimen the head and the foot are retracted into the trunk, which acquires a barrel shape; and epidermal folds are present at both extremities (Fig. 1.1e,f). The relative position of the internal organs is similar to that of the relaxed hydrated specimens and there is not much change in the body cavity, although organs appear slightly more condensed. Compared to the contracted hydrated specimens, in anhydrobiotic rotifers the body is flattened dorso-ventrally, and assumes a characteristic dome-shaped morphology (Fig. 1.1h). Several epidermal folds are present at both ends (Fig. 1.1h,i). Moreover, the internal organs appear more compact, with a greatly reduced body cavity. The organs also appear

less resolved than the hydrated specimens, probably due to changes in the biochemical composition of cells.

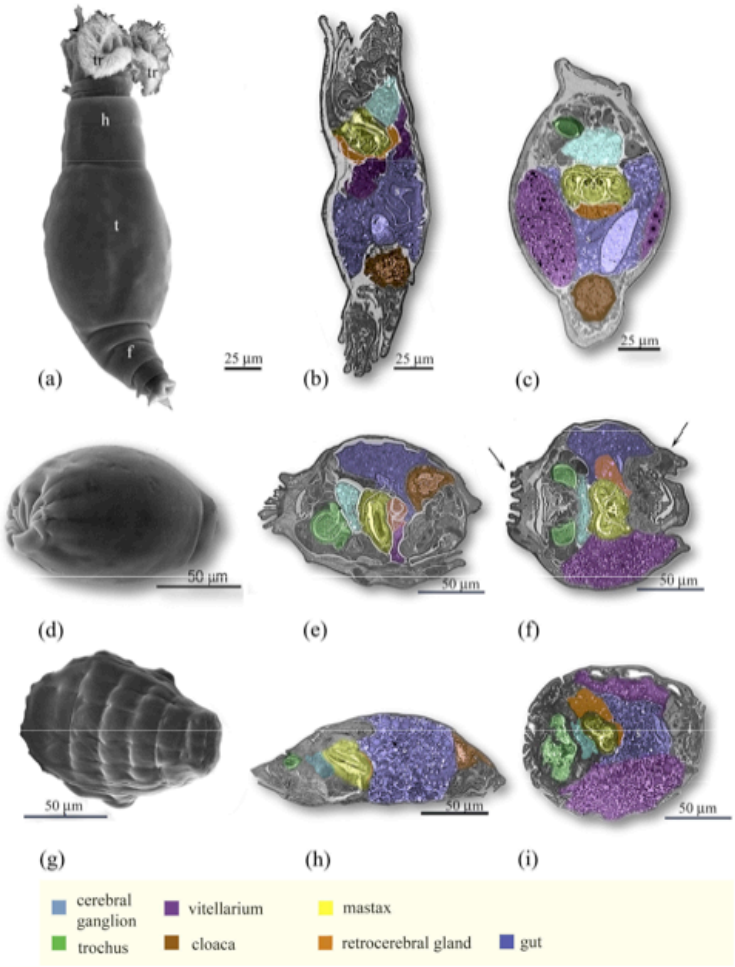


Figure 1.1: Anhydrobiosis in bdelloid rotifers. Scanning electron microscopy and confocal microscopy of hydrated relaxed (a–c), hydrated contracted (d–f) and anhydrobiotic (g–i) specimens of bdelloid rotifer species *Macrotrachela quadricornifera*. (a) Fully relaxed animal at SEM; (b) sagittal section through a relaxed rotifer; (c) parafrontal section through a relaxed specimen; (d) SEM image of a hydrated contracted specimen; (e) sagittal section of a hydrated contracted specimen; (f) parafrontal section through a hydrated contracted specimen; (g) anhydrobiotic specimen at SEM; (h) sagittal section through an anhydrobiotic specimen; (i) frontal section through an anhydrobiont specimen. *Abbreviations:* f, foot; h, head; s, stomach; t, trunk; tr, trochi. Taken from Marotta et al., 2010, with permission.

Currently the molecular basis of biostability in belloid rotifers during anhydrobiosis is not very well understood. Unlike other anhydrobiotic organisms that produce non-reducing disaccharides like trehalose and sucrose (Lapinski and Tunnacliffe, 2003) during desiccation, rotifers are intriguingly deficient in these molecules. In order to explain this deficiency of sugars, Lapinski and Tunnacliffe

hypothesized that these organisms might rely on proteins for cellular protection. The LEA family of proteins, known to be associated with resistance against desiccation in numerous other organisms, were suggested as possible alternatives. The search for desiccation resistance genes in rotifers ultimately led to the identification of two LEA genes, *Ar-lea-1A* and *Ar-lea-1B* in the cDNA isolated from dehydrated rotifers (Pouchkina-Stantcheva et al., 2007). These genes translate into two proteins, ArLEA1A and ArLEA1B, respectively. Before I discuss the specific molecular functions of ArLEA1A and ArLEA1B, I would like to highlight the functional and structural characteristics of the broad family of LEA proteins, their sub-cellular localisation and classification into different groups.

1.4 Functional diversity of LEA proteins

LEA proteins are a large family of proteins that were initially found to be encoded by mRNAs enriched in the later stages of embryo development in cotton seeds by Dure et al. (Dure et al., 1981; Galau and Dure, 1981). Besides being associated with desiccation tolerance in plant seeds, these proteins have also been found in vegetative tissues of drought and cold stressed plants (Joh et al., 1995; Mundy and Chua, 1988) and during abcissic acid (ABA) treatment (Yamaguchi-Shinozaki and Shinozaki, 1993). Moreover, their expression has been found to be correlated with desiccation tolerance in a number of other organisms like the bacterium *Deinococcus radiodurans* (Battista et al., 2001), nematodes *Aphelenchus avenae* (Browne et al., 2002; Browne et al., 2004) and *Caenorhabditis elegans* (Gal et al., 2004), larvae of insects *Polypedilum vanderplanki* (Kikawada et al., 2006) and bdelloid rotifer species *A. ricciae* (Pouchkina-Stantcheva et al., 2007). Additional evidence for the association of LEA proteins with desiccation tolerance is provided by knocking out or knocking down these genes. For example, knocking out LEA protein homologues in *D. radiodurans* impairs desiccation resistance in these bacteria (Battista et al., 2001). Similarly, RNAi silencing of *Ce-lea-1* in *C. elegans* significantly reduces tolerance to drying (Gal et al., 2004).

The functions of these ubiquitous proteins are quite diverse and include lipid association (Pouchkina-Stantcheva et al., 2007; Tolleter et al., 2007), glass stabilisation during anhydrobiosis (Shimizu et al., 2010) and association with actin molecules (Abu-Abied et al., 2006). Some mitochondrial LEA proteins have been shown to protect the inner mitochondrial membrane and mitochondrial enzymes like fumarase and rhodanese (Grelet et al., 2005), thus potentially increasing organelle resistance to water stress.

LEA proteins like AavLEA1 have also been demonstrated to prevent aggregation of desiccation-sensitive enzymes like citrate synthase (Goyal et al., 2005b). Therefore, it is hypothesised that these proteins might act like 'molecular shields' by forming a physical barrier between aggregating proteins during desiccation (Goyal et al., 2005b). Alternatively, it is believed that they could form specific complexes with their client proteins and might act like 'molecular chaperones'. However, recent results suggest that unlike molecular chaperones, LEA proteins cannot associate strongly with their client proteins (Chakrabortee et al., 2012) and are incapable of protecting proteins in harsh conditions of stress like heat. A model has been proposed whereby LEA proteins might partially fold on the surface of misfolded proteins in a manner similar to chaperone proteins containing intrinsically disordered regions (Tompa and Csermely, 2004) allowing a degree of entropy transfer that might facilitate refolding of the client protein during desiccation.

These proteins are also believed to function as 'hydration buffers', whereby they might bind water molecules in their hydration shells. Nuclear magnetic resonance (NMR) has demonstrated that several LEA proteins retain 20-50% more water molecules compared to normal globular proteins like BSA (Bokor et al., 2005). However, in dried plant seeds these proteins might only represent a small fraction of the total cytoplasmic proteins, therefore it is estimated that LEA specific binding would influence the water-loss kinetics of only 2-3% of cell water (Hand et al., 2011). It is unclear whether retaining such a small amount of water might have any biological advantages during the process of desiccation and rehydration. It is possible that by tightly associating with water molecules, LEA proteins might be recruited to

cell structures that normally rely on the presence of a hydration shell to preserve their integrity during desiccation by direct hydrogen bonding.

Other physiological functions of LEA proteins might involve sequestration of charged ions (Svensson et al., 2000; Wise and Tunnacliffe, 2004) as proposed by Dure (Dure, 1993) and scavenging of reactive oxygen species by citrus dehydrins (Hara et al., 2004). These properties might be essential in their role as possible antioxidants.

1.5 Structural transitions in LEA proteins upon drying

Several LEA proteins have been found to be unstructured in solution and have therefore been classified as 'intrinsically disordered proteins' or IDPs (Tompa and Kovacs, 2010; Uversky and Dunker, 2010). However, upon drying many acquire a more folded state. This was first demonstrated for AavLEA1 by Goyal et al. using Fourier-transform infrared spectroscopy (FTIR) (Goyal et al., 2003). AavLEA1 showed gain of structure during dehydration that was fully reversible upon rehydration. It is hypothesised that gain in secondary structure is mostly due to an increase in intra-molecular hydrogen bonds between amino acids in the protein and a corresponding decrease in hydrogen bonding between the protein and water molecules (Li and He, 2009). These changes typically occur below 20% water content as revealed by RMSD (root-mean-square deviation) and MDTF (mean dihedral transition frequency) of all amino acids in the protein. Fig. 1.2 depicts the structural transitions observed in AavLEA1 *in silico*. This structural plasticity is believed to be essential for the function of AavLEA1, allowing it to fold on the surfaces of a wide range of proteins and macromolecular structures during desiccation.

The gain in structure of some randomly coiled LEA protein chains is predicted to be not a completely random event (Gilles et al., 2007). These authors found that introducing two proline residues into the N-terminal helical domain of Group 1 LEA protein Em, compromised the protective ability of this protein during drying. It is hypothesised that the N-terminal domain might serve as a scaffold for mediating

specific intra- and inter-molecular interactions that might occur when the protein folds during drying.

It has also been found that the rate of drying can have a significant effect on the propensity of formation of various structures in proteins. For example, slow drying of poly-L-lysine results in the formation of extensive β-sheet structure, but fast drying precluded the formation of β-sheet (Wolkers et al., 1998).

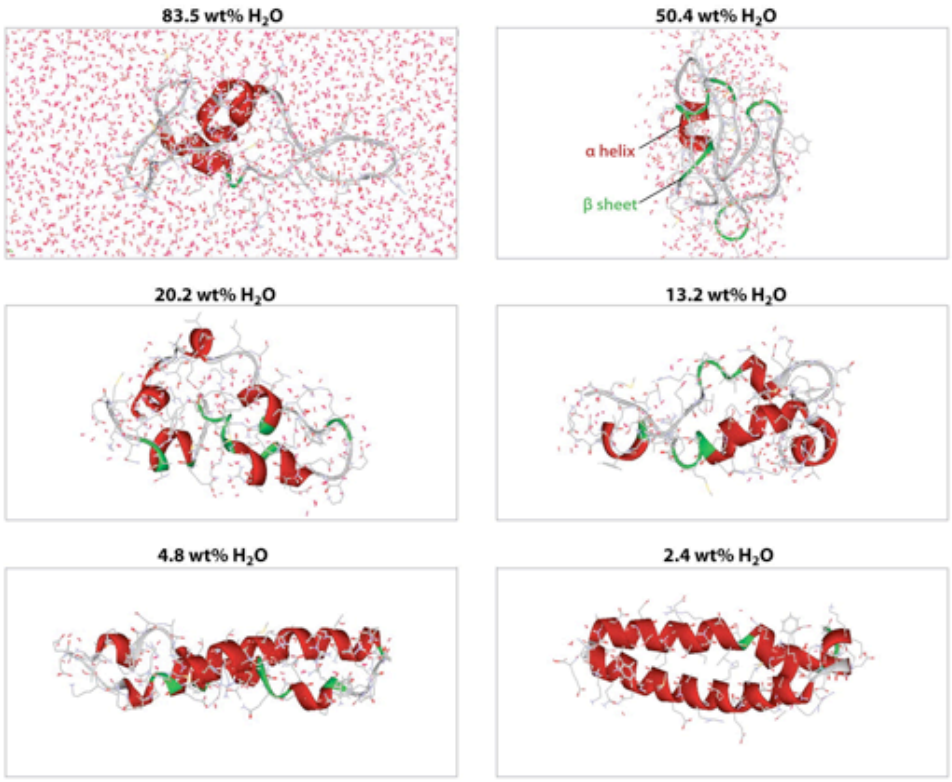


Figure 1.2: Structural modeling of AavLEA1. Representative conformations of the 66-amino-acid fragment of a LEA protein (AavLEA1) from the nematode *Aphelenchus avenae* are shown at different water contents. The smaller water molecules (gray and red) are depicted in the line style, and the larger LEA protein molecules are denoted using the solid ribbon style (α helix, red; β sheet, green; random coil, gray). Taken from Li et al., 2009, with permission.

Structural modelling has revealed that some LEA proteins like LEAM fold into class A amphipathic helices (Fig. 1.3) typical of apolipoproteins that coat low density lipoprotein particles (Tolleter et al., 2007). LEAM also undergoes intrinsic chemical modifications such as deamidation and oxidation that might affect structural features of this protein. It is believed that LEAM interacts with both the interior and exterior

layers of membranes through its alternating hydrophilic and hydrophobic residues respectively.

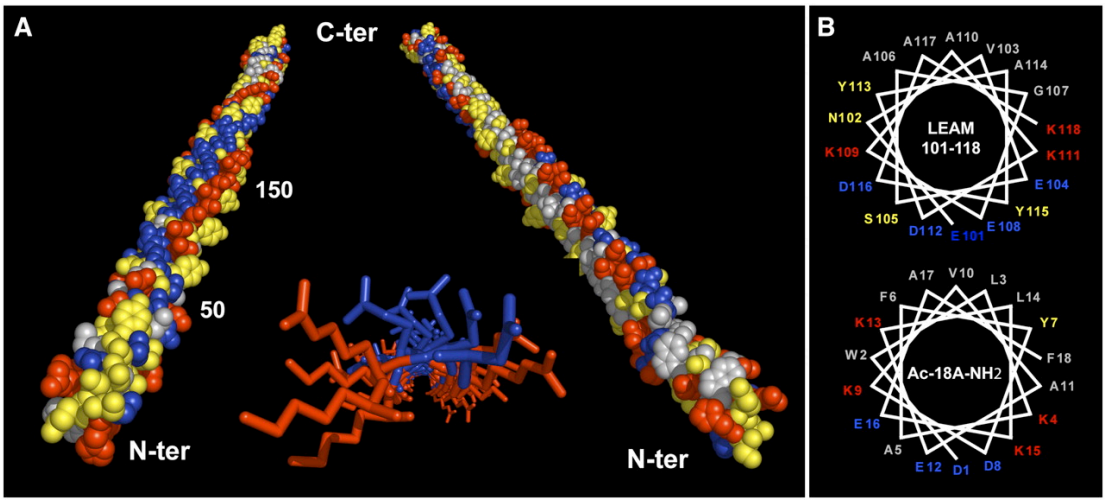


Figure 1.3: Structural modeling of LEAM. LEAM was modeled with the Swiss-PDB Viewer program using a helical template filamin (Protein Data Bank accession number 1GK7). Charged amino acids are indicated in red (basic: H, K, or R) or blue (acid: D or E), and other residues are indicated in yellow (polar or hydrophilic: N, Q, S, T, or Y) or gray (nonpolar or hydrophobic: A, G, I, L, M, V, or W). Two views of the protein are shown on the sides, and the center image shows a front view of the helix from amino acids 50 to 150, displaying only charged residues. Images were generated using PyMol. Taken from Tolleter et al., 2007, free access.

Interestingly, LEA18 is a small seed-specific, basic and highly hydrophilic protein that has also been predicted to be an intrinsically disordered protein (Hundertmark et al., 2011). This protein has been shown to bind to negatively charged membranes. In this protein, the increase in secondary structure is contributed mainly by β-sheets. Contrary to the 'classical' function of LEA proteins as membrane stabilisers, LEA18 has a destabilizing effect on membranes, a property that is compatible with the hypothesis that this protein might be involved in vesicle transport or signalling during development.

Hence, it can be observed that LEA proteins have evolved a diverse repertoire of functions and structures relevant to their role in protecting cells during water loss.

1.6 Sub-cellular localisation of LEA proteins and their role in organelle protection

Localisation of LEA proteins to specific cellular targets might be an essential requirement for protecting various cellular compartments during desiccation. In plants, LEA proteins are found to localise in the cytoplasm, nucleus, mitochondrion, chloroplast, endoplasmic reticulum, vacuole, peroxisomes and plasma membrane (reviewed in Tunnacliffe and Wise, 2007). Some LEA proteins that are targeted to specific organelles like mitochondria contain a hydrophobic pre-sequence at the N-terminal end, which might enable compartmentalisation to specific hydrophobic surfaces of various organelle membranes. For example, the group 3 LEA protein PsLEAm resides in the mitochondrial matrix of seeds and contains largely hydrophilic amino acids with the exception of the N-terminal hydrophobic mitochondrial targeting sequence (Grelet et al., 2005).

1.7 Classifying LEA proteins

Due to the plethora of LEA protein sequences studied over the years, there have been many attempts to classify them. The original classification of LEA proteins was performed by Dure et al. where these proteins (LEA D7, LEA D11 etc.) were sub-divided into three groups according to their sequence similarity and amino acid composition (Dure et al., 1989). Battaglia et al. have extended the work by Dure (Dure et al., 1989) and others (Cuming, 1999) by creating seven different groups of LEA proteins (Battaglia et al., 2008). Groups 1, 2, 3, 4, 6, and 7 correspond to the hydrophilic LEA proteins, whereas hydrophobic LEA proteins have been placed in Group 5. These groups are described in brief below.

Group 1 LEA proteins contain a characteristic 20-mer sequence (Baker et al., 1995; Galau et al., 1992) and have a very large proportion of charged residues, which contributes to their high hydrophilicity, and a high content of glycine residues (approximately 18%). Structural analysis using circular dichroism indicates that these proteins exist mostly as random coils, with a small percentage of the protein

exhibiting a left-handed extended helical or poly-(L-Pro)-type (PII) conformation in the presence of detergents (Soulages et al., 2002). NMR also indicates that these proteins have an unstable structure and are quite flexible (Eom et al., 1996).

Group 2 LEA proteins, also known as dehydrins were originally identified as the D-11 family in developing cotton embryos. These proteins typically contain a high proportion of charged and polar amino acids and a low fraction of hydrophobic residues (Garay-Arroyo et al., 2000). A unique feature of group 2 LEA proteins is a conserved, lysine-rich 15-residue motif, EKKGIMDKIKEKLPG, named the K-segment (Close et al., 1993; Close et al., 1989) that can be found in one to 11 copies within a polypeptide sequence. An additional motif also found in this group is the Y-segment, whose conserved consensus sequence is [V/T]D[E/Q]YGNP, is usually found in one to 35 tandem copies in the N terminus of the protein (Campbell and Close, 1997; Close, 1996; Close et al., 1993). Many group 2 proteins also contain a tract of serine residues, called the S-segment, which in some proteins can be phosphorylated (Plana et al., 1991; Vilardell et al., 1990). Less conserved motifs (Φ-segments), which are usually rich in polar amino acids are present in some proteins of this group (Campbell and Close, 1997). Proteins that only contain the K-segment are in the K-subgroup, and those that include the S-segment followed by K-segment are in the SK-subgroup. In addition, there are the YSK-, YK-, and KS-subgroups (Campbell and Close, 1997). Structural analysis of four group 2 LEA proteins, Dsp16 (YSK2) from resurrection plant (Lisse et al., 1996), 35-kD protein (Y2K) from cowpea (Ismail et al., 1999), rGmDHN1 (Y2K) from soybean (Soulages et al., 2003) and ERD10 (SK3) from mouse-ear cress (Bokor et al., 2005), indicated that these proteins are in an unstructured conformation in aqueous solution. However, neither temperature, metal-ions, nor stabilizing salts could promote ordered structures in either the peptides or the full-length proteins (Mouillon et al., 2006). The K-segment motifs of group 2 LEA proteins are predicted to form amphipathic α-helical structures and are thought to protect membranes (Close, 1996). Although the majority of group 2 LEA proteins accumulate in the cytoplasm, some of them are also localised in the nucleus. In some cases nuclear localised group 2 proteins are shown to be dependent on phosphorylation of amino-acid residues (Plana et al., 1991). However, for other proteins, nuclear localisation seems to be independent of the phosphorylation state

of proteins (Riera et al., 2004). Some of these proteins are also found in the plasma membrane (Danyluk et al., 1998) and mitochondria (Borovskii et al., 2000).

Group 3 LEA proteins consist of an 11-mer amino acid repeat motif that has the potential to form an amphipathic α-helix (Dure, 1993). The consensus sequence for the 11-mer proposed by Dure (Dure, 1993) is as follows: hydrophobic residues (F) in positions 1, 2, 5, and 9; negative or amide residues (E,D,Q) in positions 3, 7, and 11; positive residues (K) in positions 6 and 8; and a random assortment (X) in positions 4 and 10 (FF[E/ Q]XFK[E/Q]KFX[E/D/Q]). The variability in the 11-mer motif led Dure to subclassify group 3 LEA proteins into two subgroups: 3A, represented by the cotton D-7 LEA protein; and 3B, represented by the cotton D-29 LEA protein (Dure, 2001). The similarity of some group 3 proteins with different plasma apolipoproteins suggests that some these might interact with membranes during dehydration (Tolleter et al., 2007). CD analysis and IR spectroscopy of various group 3 LEA proteins indicates that they are mostly devoid of secondary structure, being largely in a random coil conformation in solution. However, in the presence of sucrose, glycerol, ethylene glycol, or methanol, or after fast drying, they adopt an α-helical conformation (Dure, 2001; Goyal et al., 2003; Tolleter et al., 2007; Wolkers et al., 2001). Moreover it has been found that these structural transitions are fully reversible (Goyal et al., 2003; Tolleter et al., 2007). More recently group 3 LEA proteins have been shown to contribute towards glass formation (Shimizu et al., 2010).

Group 4 proteins consist of a conserved region in their N-terminal portion, which is about 70 to 80 residues long and is predicted to form amphipathic α-helices, while the less conserved C-terminal portion is variable in size (Battaglia et al., 2008). A motif that has characterised the proteins in this group is motif 1, located at the N-terminal region with the following consensus sequence: AQEKAEKMTA [R/H]DPXKEMAHERK[E/K][A/E][K/R]. However, four additional motifs can be distinguished in many group 4 LEA proteins. The presence or absence of motif 4 or 5 defines two subgroups within the family. The first subgroup (group 4A) consists of small proteins (80–124 residues long) with motifs 2 and/or 3 flanking motif 1. The other subgroup (group 4B) has longer representatives (108–180 residues) that, in addition to the three motifs in the N-terminal portion, may contain motifs 4 and/or 5

at the C-terminal region. D-113 protein from cotton, the first discovered of this group, belongs to group 4B (Battaglia et al., 2008). *In silico* analysis of group 4 LEA proteins predicts that the first 70 to 80 residues could adopt an α-helical structure, whereas the rest of the protein assumes a random coil conformation (Battaglia et al., 2008).

Group 5 consists of hydrophobic proteins and is therefore quite heterogenous (Battaglia et al., 2008). These proteins are not soluble after boiling, suggesting that they adopt a globular conformation (Baker et al., 1988; Cuming, 1999; Galau et al., 1993). Not much is known about these proteins except for the fact that their transcripts accumulate during the late stage of seed development and in response to stress conditions, such as drought, free radical damage, salinity and wounding (Kim et al., 2005; Kiyosue et al., 1992; Maitra and Cushman, 1994; Park et al., 2003; Stacy et al., 1999; Zegzouti et al., 1997).

Group 6 proteins are characterised by their small size (approximately 7–14 kD) (Battaglia et al., 2008). Four motifs distinguish this group, two of which (motifs 1 and 2) are highly conserved. The sequence LEDYK is present in motif 1 and the proline and threonine residues located in positions 6 and 7, are present in motif 2. In general, these proteins are highly hydrophilic, lack cysteine and tryptophan residues, and do not coagulate upon exposure to high temperature. *In silico* analyses predicts that group 6 LEA proteins are intrinsically unstructured (Garay-Arroyo et al., 2000).

Group 7 consists of the ASR (Abcissic acid stress ripening 1) proteins. These are considered to be members of the hydrophilin family and are small, heat stable, and intrinsically unstructured (Frankel et al., 2006; Goldgur et al., 2007; Silhavy et al., 1995). Transcripts for these genes accumulate during senescence, fruit ripening, and/or seed and pollen maturation. They also respond to environmental stress conditions, such as water deficit, salt, cold, and limited light, and ABA treatment. All known ASR proteins contain five motifs, among which three are highly conserved (motifs 1, 2, and 3). One of these motifs (motif 3) is located within the C-terminal region and contains a putative nuclear localisation signal (Silhavy et al., 1995; Wang et al., 2003; Wang et al., 2007a). Motifs 1,2 and 5 contain stretches of His residues and are predicted to possess zinc dependent DNA-binding activity (Goldgur et al., 2007; Kalifa et al., 2004).

Another classification scheme has been devised by Bies-Etheve et al. where 10 different LEA groups have been constructed on the basis of distinct motifs (Bies-Etheve et al., 2008). However, a drawback of this scheme is the representation of the same motifs in different groups. For example, among the LEA proteins in *Arabidopsis thaliana,* motif 1 (AYDKA-AKD), present in Group 3, is also represented in Group 5 members.

Table 1.1 (taken from Battaglia et al. 2008, with permission) shows a comparison between nomenclature of different LEA protein groups from different classification schemes.

Table 1.1 Correspondence between different nomenclatures given to LEA protein groups

*, Used by Bies-Ethève et al. (2008) to differentiate proteins initially identified as related to group 5 but revealed as belonging to group 3. –, Used to denote groups that were not identified by these authors.

This Work	Dure	Bies-Ethève	PFAM	PFAM No.	Name
1	D-19	1	LEA_5	PF00477	Em1, Em6
2	D-11	2	Dehydrin	PF00257	Dehydrin, RAB
3A	D-7	3	LEA_4	PF02987	ECP63, PAP240, PM27
3B	D-29	3*	LEA_4	PF02987	D-29
4A	–	4	LEA_1	PF03760	LE25_LYCES
4B	D-113	4	LEA_1	PF03760	PAP260, PAP051
5A	D-34	5	SMP	PF04927	PAP140
5B	D-73	6	LEA_3	PF03242	AtD121, Sag21, lea5
5C	D-95	7	LEA_2	PF03168	LEA14
6	–	8	LEA_6	PF10714	LEA18
7	–	–	ABA_WDS	PF02496	ASR

LEA protein groups have also been represented by Pfam domain families (Wise, 2002). Pfam domains have been mainly created by multiple sequence alignment and structural information of proteins wherever possible (Finn et al., 2010). There are obvious limitations to types of classifications that solely rely on multiple-sequence alignment methodologies as LEA proteins have been mainly found to contain low complexity sequences which might be masked out by sequence comparison tools like BLAST (Wise and Tunnacliffe, 2004).

To overcome problems posed by multiple-sequence alignment algorithms, a new computational tool called POPP (Protein or Oligonucleotide Probability Profile) has been developed (Wise and Tunnacliffe, 2004). This tool allows proteins to be compared based on similarities in their peptide compositions. Interestingly some of

the POPP predictions for LEA proteins provide the basis for testable hypotheses that may be verified experimentally. However, Battaglia et al. argue that although POPP analysis might emphasize the importance of compositional biases in LEA proteins, it might neglect information related to sequence motifs that might also be important in determining LEA protein structure and function (Battaglia et al., 2008).

Given the complex nature of LEA protein sequences, it is obvious that although existing schemes of classification might provide some information related to sequence motifs and low-complexity regions, they are not adequate in revealing any information about the actual molecular function of these proteins and their expression trends. This could be partly attributed to the lack of substantial experimental data for expression and function of LEA proteins. Future bioinformatics analysis might take into account a more holistic approach towards classifying LEA proteins by incorporating expression and function information.

1.8 Unique features of bdelloid LEA proteins ArLEA1A and ArLEA1B

The discovery of LEA proteins ArLEA1A and ArLEA1B in the bdelloid *A. ricciae* has provided vital insights into the molecular basis of anhydrobiosis in these animals. These proteins are encoded by two genes, named *Ar-lea-1A* and *Ar-lea-1B*, respectively, and were first identified in the cDNA isolated from desiccating rotifers (Pouchkina-Stantcheva et al., 2007). These genes are believed to be former alleles as their aligned sequences show 13.5% synonymous site divergence (Ks) over the whole gene, a value which is too high to be observed between alleles in sexual animals, but in the range observed for former allele pairs (Birky, 2004; Butlin, 2002; Pouchkina-Stantcheva et al., 2007; Mark Welch et al., 2004; Mark Welch and Meselson, 2000). Loss of genetic recombination and acquisition of mutations by genetic drift are believed to have resulted in divergence of these alleles when rotifers became asexual.

Both proteins share homology with conserved 11-amino acid motifs characteristic of group 3 LEA proteins. ArLEA1A is 420 residues long with an estimated molecular mass of 44.5 kDa, whereas ArLEA1B consists of 376 residues

with an estimated molecular mass of 39.8 kDa. Comparison of ArLEA1A and ArLEA1B protein sequences has revealed that these proteins differ at only 13 amino acid positions. An extra stretch of 44 amino acids, consisting of repeat sequences, is found exclusively in ArLEA1A. These subtle differences in their sequences result in remarkably different biochemical properties of these proteins as described below.

Analysis of secondary structure of ArLEA1A and ArLEA1B in hydrated and dry states by far-ultraviolet circular dichroism spectroscopy (Pouchkina-Stantcheva et al., 2007) has shown that ArLEA1A is predominantly unstructured in solution and acquires an α-helical structure upon drying (although it is observed to possess a higher degree of structure compared to AavLEA1 in the hydrated state) while ArLEA1B is structured both in the hydrated and dry states (Fig. 1.4).

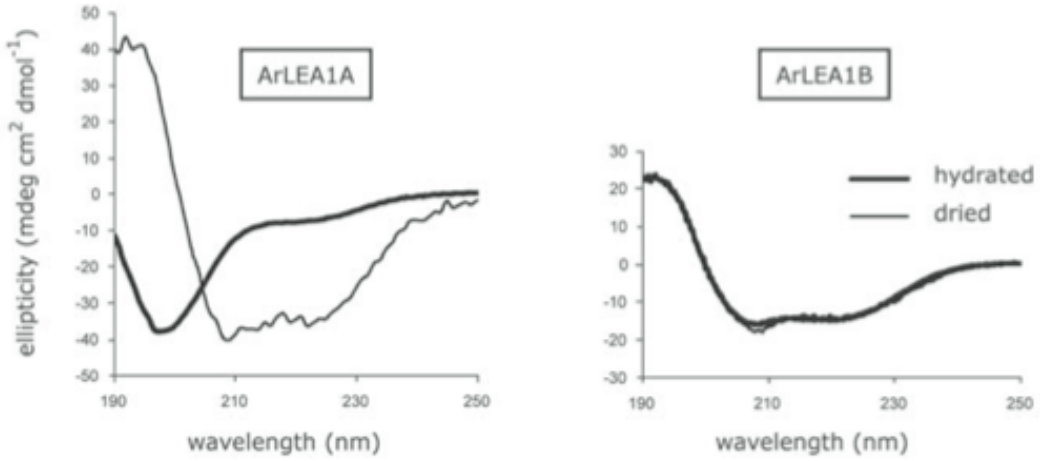


Figure 1.4: Far-ultraviolet circular dichroism spectroscopy of ArLEA1A and ArLEA1B. Taken from Pouchkina-Stantcheva et al., 2007, with permission.

Moreover, like other conventional group 3 LEA proteins, ArLEA1A is able to prevent desiccation sensitive enzymes from aggregating (Fig. 1.5a). In comparison with ArLEA1A, ArLEA1B not only promotes aggregation of desiccation sensitive enzymes, but is itself prone to aggregation both in solution and the dry state (Fig 1.5a). Additionally, unlike ArLEA1A, it demonstrates a strong tendency to interact with dry liposomes (Fig. 1.5b) (Pouchkina-Stantcheva et al., 2007). This feature is common to many other LEAs like LEAM (Tolleter et al., 2007) and dehydrins (LEA D-11) (Koag et al., 2003) that are hypothesised to stabilise cellular membranes when water becomes scarce. For example, LEAM has been found to be natively unfolded in solution,

acquiring an α-helical structure upon desiccation and is predicted to simultaneously integrate into both the outer and inner surfaces of membranes by ionic and hydrophobic interactions respectively (Tolleter et al., 2007). ArLEA1B is different from LEAM in being quite structured both in the hydrated and dry states and in having a tendency to increase the proportion of hydrogen bonding in P=O groups of dry liposomes. Thus, it is believed that ArLEA1B might interact with only the outer monolayers of lipids during drying. More recently, potential novel LEA sequences have been identified but their relationship to the existing known LEAs remains to be clarified (A. Tunnacliffe, personal communication). Thus, the origin of these functionally divergent genes in bdelloids provides a perfect example of how genetic variation precedes acquisition of novel functions during Darwinian evolution.

1.9 Why is further characterisation of ArLEA1A and ArLEA1B proteins necessary?

The characterisation of proteins often requires complete appreciation of the structural and functional complexity of their sequences. So far, much information related to the evolution, three-dimensional conformation and functional motifs of bdelloid LEA proteins has been lacking. The bioinformatics analysis in Chapter Three investigates the features of bdelloid LEA proteins in greater detail.

Previously, the functions of ArLEA1A and ArLEA1B have been deduced by *in vitro* experiments alone and little is known regarding their mode of action within cells. In Chapter Four, intra-cellular localisation analysis of these proteins has been performed to confirm their distribution in the ER in living cells. Moreover, the functions of the N and C-terminal sequences of these proteins in regulating the intra-cellular distribution and secretion of these proteins to the cell exterior have also been studied. The protective role of these proteins in the mammalian ER has been further analysed in Chapter Five.

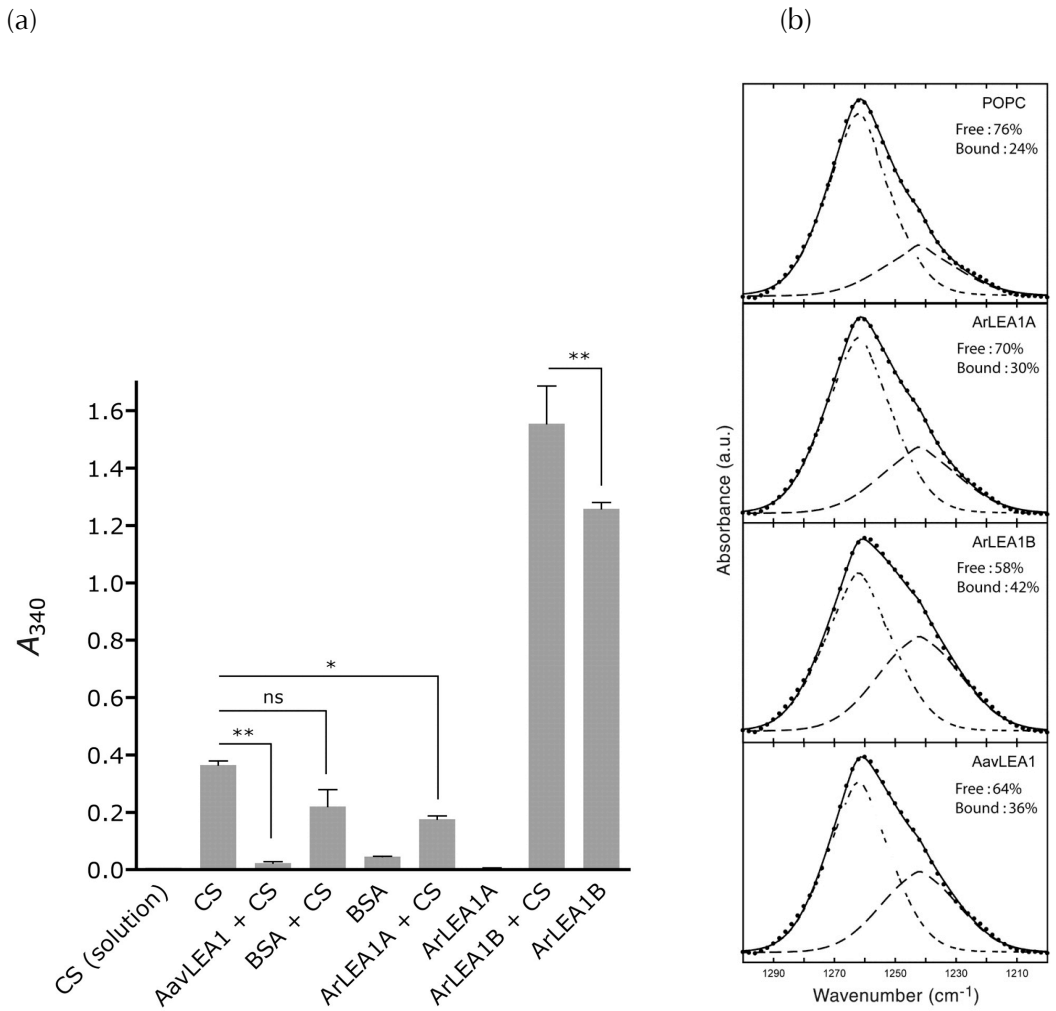


Figure 1.5: Functions of ArLEA1A and ArLEA1B determined *in vitro*. (a) Bdelloid LEA protein anti-aggregation assay. Citrate synthase (CS), with or without LEA proteins or BSA, and the latter proteins alone where indicated, were subjected to two cycles of vacuum drying and rehydration. Light scattering by protein particulates was measured by apparent absorption at 340 nm in the spectrophotometer. Error bars show standard deviation (n = 3); ns, not significantly different ($p > 0.05$); significant values *$p < 0.05$ or **$p < 0.001$. (b) Bdelloid LEA protein membrane association. Infrared spectra of the asymmetric phosphate stretching region of POPC liposomes dried alone or in the presence of ArLEA1A, ArLEA1B, or AavLEA1. Spectra were recorded at 78°C (liquid-crystalline phase). The solid curve comprises both the measured (dots) and fitted absorbance curves. Normalised peaks were fitted into two bands with maxima at 1262 and 1242 cm^{-1} corresponding respectively to P=Oas$_{free}$ (short dashes) and P=Oas$_{bound}$ (long dashes). Taken from Pouchkina-Stantcheva et al., 2007, with permission.

Lastly, the protective capacity of ArLEA1A and ArLEA1B in preventing structural damage of surrounding protein molecules has been assessed using mCherry

as a test protein. The fluorophore or the light emitting region of mCherry has been found to be dependent on the presence of neighboring water molecules (Shu et al., 2006; Topol et al., 2011). Indeed, subjecting this protein to repeated cycles of drying results in a dramatic decrease in its fluorescence and absorbance properties possibly due to loss of its fluorophore structure. When dried in the presence of ArLEA1A or ArLEA1B, mCherry fluorescence properties are only partially protected. These results are discussed in detail in Chapter Six.

Chapter Two: Materials and Methods

2.1 Bioinformatics

The phylogenetic analysis of ArLEA1A and ArLEA1B protein sequences (NCBI Accession numbers ABU62811.1 and ABU62810.1) was carried out using Phylogeny.fr (www.**phylogeny**.fr/). The top 50 hits of these sequences were obtained using the BLAST EXPLORER program (Dereeper et al., 2010) and processed further in the 'One-click' work-flow mode in Phylogeny.fr after removing redundancy from the data-sets. A multiple sequence alignment (MSA) was constructed using the MUSCLE program (Edgar, 2004a, b). Subsequently, the aligned sequences were used to plot the phylograms depicting the phylogenetic distances between these sequences using PhyML (Guindon et al., 2010) and TreeDyn (Chevenet et al., 2006).

For domain searches, protein sequences of ArLEA1A and ArLEA1B, along with lea-1B, lea-1B' and lea-1C from *Adineta vaga*, the LEA-1 protein from *Caenorhabditis elegans*, PvLEA1 protein from *Polypedilum vanderplanki*, LEA protein from *Arabidopsis thaliana*, LEA-76 protein from *Clostridium* and AavLEA1 from *Aphelenchus avenae* (Accession numbers ADD91471.1, ADD91479.1, ADD91460.1, CCA65610.1, BAE92616.1, BAA11017.1, ZP_08128614.1 and AAL18843.1 respectively) were deposited in the Pfam server (http://pfam.sanger.ac.uk/). The search criterion was adjusted to include PfamB domains and the gathering threshold cut-off values. The Pfam domains, along with their HMM alignment positions, bit scores and E-values were compared between different sequences.

For structural modelling of ArLEA1A and ArLEA1B proteins, sequences were deposited in http://zhanglab.ccmb.med.umich.edu/I-TASSER/. Five-top models predicted were analysed further.

For identification of cellular localisation signals, ArLEA1A, ArLEA1B and AavLEA1 protein sequences were proceesed in SignalP 4.0 (http://www.cbs.dtu.dk/services/SignalP/), and TargetP

(http://www.cbs.dtu.dk/sevices/TargetP/) servers. Signalling motifs and protein-protein interaction domains were predicted using Scansite at Massachusetts Institute of Technology (http://scansite.mit.edu/motifscan_seq.phtml).

2.2 Recombinant DNA methods

2.2.1 DNA purity and concentration

DNA concentrations were measured in Nanovue. The 260:280 nm ratios were estimated to assess purity.

2.2.2 PCR amplifications

Ar-lea-1A and *Ar-lea-1B* were PCR amplified using NotI digested cDNA clones obtained from desiccated rotifers using Phusion high fidelity DNA Polymerase (Thermo Scientific). pET15b-AavLEA1 linearised with NdeI was used as a template for amplifying *Aavlea1*. KpnI and NheI sites were engineered at the 5'end and BamHI and FLAG tag were engineered at the 3'end of primers used for these amplifications. In brief, PCR reactions were carried out in the G-storm GS 482 thermal cycler in 50 µl reaction volumes. Each reaction consisted of 35 µl water, 10 µl 5X Phusion HF buffer, 1 µl of 10 mM dNTPs, 1 µl of 10 µM of forward and reverse primers, 1 µl of linearised template (100 pg/µl), 1.5 µl DMSO and 0.5 µl (2 units/µl) of Phusion DNA polymerase. Cycling conditions were set as follows: 98°C for 30 s (template denaturation), followed by 5 cycles of 98°C for 10 s, 47°C for 30 s, 72°C for 1 min 30 s, 25 cycles of 98°C for 10 s, 65°C for 30 s, 72°C for 1 min 30 s, and a final extension of 72°C for 10 min. 'As' were added at the 3'-ends of PCR products by 30 minutes treatment with 1 µl Advantage Taq mix (50X, Clontech) at 72°C. After completion, reactions were stored at 4°C.

mCherry amplification was carried out using linearised pmCherry (Clontech) as template and forward and reverse primers containing BamHI and EcoRI restriction sites respectively. Reaction volumes consisted of 23.4 µl water, 3.0 µl 10X buffer, 0.6 µl 10 mM dNTPs, 0.7 µl of 10 µM forward and reverse primers, 1.0 µl of linearised

template (100 pg/µl) and 0.6 µl of Advantage Taq mix (50X, Clontech). Cycling conditions were set as follows: 95°C for 30 s, followed by 30 cycles of 95°C for 20 s, 60°C for 1 min 30 s, 72°C for 1 min and final extension at 72°C for 15 min. PCR products were stored at 4°C.

The first 22 amino-acids of the N-terminal signal sequence and the C-terminal ATEL, KDEL or the ochre stop codon sequences were fused with mCherry and EGFP by PCR amplification. Linearised pcDNA3.0-mCherry and pET-28a-EGFPHDQ74 were used as templates. Reactions were performed in 20 µl final volume and consisted of 14.2 µl water, 4.0 µl 5X buffer, 0.4 µl 10 mM dNTPs, 0.4 µl of linearised template (100 pg/µl), 0.4 µl of 10 µM forward and reverse primers and 0.2 µl Phusion polymerase (0.4 units). For mCherry constructs, cycling conditions were set as: 98°C for 30 s (template denaturation), followed by 5 cycles of 98°C for 10 s, 55°C for 30 s, 72°C for 1 min, 25 cycles of 98°C for 10 s, 65°C for 30 s, 72°C for 1 min, and a final extension of 72°C for 20 min. 'As' were added at the 3'-ends of PCR products by 30 minutes treatment with 1 µl Advantage Taq mix (50X, Clontech) at 72°C. For constructs with mutations at 5' prime end, the PCR was repeated using a shorter 5'-end primer at 98°C for 30 s, followed by 30 cycles of 98°C for 10 s, 53°C for 1 min, 72°C for 1 min 30 s. Final extension was performed at 72°C for 10 min. 'As' were added at the 3'-ends of PCR products by 30 min treatment with 1µl Advantage Taq mix (50X, Clontech) at 72°C. After completion, reactions were stored at 4°C. For EGFP constructs PCR conditions were: 98°C for 30 s (template denaturation), followed by 5 cycles of 98°C for 10 s, 56°C for 30 s, 72°C for 30 s, 25 cycles of 98°C for 10 s, 70°C for 15 s, 72°C for 15 s, and a final extension of 72°C for 10 min. 'As' were added at the 3'-ends of PCR products by 30 minutes treatment with 1 µl Advantage Taq mix (50X, Clontech) at 72°C. After completion, reactions were stored at 4°C.

The N-terminal signal sequence and the C-terminal ATEL sequence of bdelloid LEAs were fused with EGFPHDQ74 by PCR amplification. Linearised pET-28a-EGFPHDQ74 was used as template. Reactions were performed in 20 µl final volume and consisted of 14.2 µl water, 4.0 µl 5X buffer, 0.4 µl 10 mM dNTPs, 0.4 µl of linearised template (100 pg/µl), 0.4 µl of 10 µM forward and reverse primers and 0.2

µl Phusion polymerase (0.4 units). Cycling conditions were set as follows: 98°C for 30 s (template denaturation), followed by 5 cycles of 98°C for 10 s, 50°C for 20 s, 72°C for 30 s, 25 cycles of 98°C for 10 s, 70°C for 10 s, 72°C for 15 s, and a final extension of 72°C for 10 min. 'As' were added at the 3′-ends of PCR products by a 30 minutes treatment with 1 µl Advantage Taq mix (50X, Clontech) at 72°C. After completion, reactions were stored at 4°C.

Sequences of primers used:

Primer Name	**Primer Sequence**
ArLEA1AForward	5'-TCA GGT ACC GCT AGC GCC ACC ATG AAC AAA ATT CTA AT-3'
ArLEA1AReverse	5'-TCG GAT CCG ACT TGT CAT CGT CGT CCT TGT AAT CTA ATT CAG TAG CTT GTC GTT TCT TTT TC-3'
ArLEA1BForward	5'-TCA GGT ACC GCT AGC GCC ACC ATG AAC AAA ATT ATA TCA AT-3'
ArLEA1BReverse	5'-GTC GGA TCC GAC TTG TCA TCG TCG TCC TTG TAA TCT AAT TCA GTA GCT TG-3'
NSmCherryForward	5'-TCG GAT CCG CCA CCA TGA ACA AAA TTC TAT CAA TTC TTT GTC TGA TAC TTT TCG TTT CTG CTT CAT TAG CTA AGC AGA AAA TGG TGA GCA AGG GCG A-3'
NSmCherryATELReverse	5'-TCG AAT TCT TAT AAT TCA GTA GCC TTG TAC AGC TCG TCC ATG C-3'
NSmCherryKDELReverse	5'-TCG AAT TCT TAG AGT TCA TCT TTC TTG TAC AGC TCG TCC ATG C-3'
NSmCherrystopReverse	5'-TCG AAT TCT TAC TTG TAC AGC TCG TCC ATG CCG CC-3'
NSEGFPForward	5'-TCG GAT CCG CCA CCA TGA ACA AAA TTC TAT CAA TTC TTT GTC TGA TAC TTT TCG TTT CTG CTT CAT TAG CTA AGC AGA AAA TGG TGA GCA AGG GCG A-3'
NSEGFPATELReverse	5'-TCG AAT TCT TAT AAT TCA GTA GCG AGT CCG GAC TTG TAC AG-3'
NSEGFPKDELReverse	5'-TCG AAT TCT TAG AGT TCA TCT TTG AGT CCG GAC TTG TAC AG-3'

mCherryForward	5'-TCG GAT CCG CCA CCA TGG TGA GCA AGG GCG AGG A-3'
mCherryReverse	5'-TCG AAT TCT TAC TTG TAC AGC TCG TCC ATG CCG CC-3'
AavLEA1Forward	5'-TCC GGT ACC GCT AGC GCC ACC ATG TCC TCT CAG CAG AAC CA-3'
AavLEA1Reverse	5'-GTC GGA TCC GAC TTG TCA TCG TCG TCC TTG TAA TCG TCG CGG CCC TTG AC-3'
pET28-EGFPHDQ74Reverse	5'-CCG CAA GCG ACG GAG CTT TAT AAT TCA GTA GCC GAA TTC CCG GTG GTA TAC TGT T-3'

2.2.3 Gel electrophoresis

Electrophoresis was carried out to analyse the PCR products in 1X TAE buffer. Ultra pure agarose (GibcoBRL) was used to prepare 0.8% agarose gels in 1X Tris Acetate EDTA buffer (TAE): 40mM Tris HCl, 20mM acetic acid and 2mM EDTA, pH 8.1, 3 μl of 10mg/ml ethidium bromide was added to 50 ml melted agarose, cooled to around 60°C. Agarose was poured into casting trays and combs were inserted to make wells. All bubbles were removed before leaving the agarose to set. Samples were mixed at a ratio of 5:1 with loading buffer. 0.5 mg of 1kb ladder (Fermentas) was loaded per lane. Gels were run at 80V-100V until the front dye traveled two thirds of the distance through the gel.

2.2.4 TOPO cloning

Ar-lea-1A, *Ar-lea-1B*, *Aavlea1* and mCherry PCR products were cloned into pCR-2.1-TOPO (Invitrogen) by TOPO-TA cloning. mCherry/EGFP/EGFPHDQ74 fusions were cloned into pcDNA3.3 TOPO vector (Invitrogen). In brief reactions consisted of 0.5 μl salt solution (1.2M NaCl, 0.06$MgCl_2$), 0.5 μl vector (10ng/μl), 1.0 μl water and 1.0 μl PCR product. After incubating the reactions at room temperature for 5-10 min, 2.0 μl of reaction was transformed into 25 μl DH10B competent *E. coli* cells (New England Biolabs).

2.2.5 Restriction enzyme digestion

500 ng of pcDNA3.0 and pCR-2.1-TOPO vector containing *Ar-lea-1A/Ar-lea-1B/Aavlea1* inserts were digested with KpnI and BamHI restriction enzymes sequentially in a 20 µl final volume. The reactions were set as follows: 1 µl of 500 ng DNA, 2.0 µl of 10X NEB buffer 2, BSA 0.3 µl (100µg/ml), KpnI 1.0 µl (10 units/µl), 12.7 µl water. After incubation for one hour at 37°C, 1 µl of BamHI (20 units/µl) and 2.0 µl of 1M NaCl were added to the reaction which was further incubated for 1 h at 37°C.

2.2.6 LMP gel ligations

After restriction enzyme digestion pcDNA3.0 and pCR-2.1-TOPO were loaded in 1% low-melting point agarose gel made in 1X TAE. After running the digested products for sufficient duration, the bands were cut out using long-wavelength UV illuminator. Slices were transferred to 1.5 ml eppendorf tubes and melted at 70°C. The DNA fragments were combined in 10 µl final volume (1-2 µl vector, plus 8-9 µl insert DNA). After adding 7.5 µl water, samples were mixed well and cooled to 37°C in a water bath. 2 µl of 10X NEB ligase buffer and 0.5 µl (400 units/µl) of T4 DNA ligase (NEB) were added and the reactions were incubated overnight at 16°C. For transformations, the ligations were melted at 70°C and immediately added to 45 µl competent cells, and incubated on ice for 30 min. mCherry and *Aavlea1*-FLAG-mCherry were sub-cloned into pET-28a(+) from pCR-2.1-TOPO vector and pcDNA3.0 vector respectively using the same protocol as above.

2.2.7 Transformation of *E. coli*

2-3 µl of ligated pCR-2.1/pcDNA3.3 TOPO vectors and 6 µl of melted LMP ligation mixtures were added to 25 µl or 50 µl of DH10B competent *E. coli* respectively. After incubation of *E. coli* on ice for 30 min, cells were heat-shocked at 42°C for 30 s and returned to ice. *E. coli* transformed with ligated plasmid were recovered at 37°C for 1 hour in 1 ml SOC. Subsequently cells were pelleted by spinning cultures at 1500 RPM for 45 s. 200 µl of resuspended *E. coli* were plated on LB plates containing 100

μg/ml ampicillin and incubated overnight at 37°C in an incubator. Colonies observed the next day were screened by colony PCR.

2.2.8 Colony PCR

Single colonies were picked and resuspended in 10 μl distilled water. PCR reactions consisted of: 4.5 μl re-suspended *E. coli*, 5 μl 2X Thermofisher PCR master mix, 0.25 μl of forward and reverse primers (10 μM). Cycling conditions were set as follows: 95°C for 10 min (template denaturation), followed by 30 cycles of 95°C for 30 s, annealing temperature (variable) for 30 s, 72°C for variable time depending upon length of product (1000 bp/ 60 s) and a final extension of 72°C for 10 min. Products were stored at 4°C subsequently and analysed by agarose gel electrophoresis. Positive colonies were inoculated in 5 ml LB containing 100 μg/ml ampicillin at 37°C overnight in a shaker.

2.2.9 DNA extraction and sequencing

Plasmid preps were made using DNA-miniprep kit (QIAGEN). Sequencing of inserts was carried out by the DNA sequencing facility at the Department of Biochemistry.

2.3 Cell culture and transfections

Mammalian COS-7 cells were grown in Dulbecco's Modified Eagle Medium (DMEM, Sigma) supplemented with 10% FBS (Sigma), 1% L-glutamine-penicillin-streptomycin solution (Sigma) in a humidified 5% CO_2 atmosphere at 37°C. Transient transfection was performed using GeneJammer Transfection reagent (Agilent Technologies) according to manufacturer's instructions. For colocalisation analysis, COS-7 cells were allowed to reach 70-80% confluence in 35mm MatTek glass bottom plates. 0.75/1 μg of plasmid DNA was used for single/double transfections with 4.5/6.0 μl of GeneJammer transfection reagent, respectively. Cells were visualised by confocal microscopy after 24 h. For supernatant analysis, COS-7 cells were grown on 10 cm diameter dishes (Nunc), until they reached 70-80% confluence, then transfected with 3 μg of DNA. 12 h post-transfection, medium was changed to DMEM (without FBS).

Supernatants were harvested after overnight incubation. To confirm expression of ArLEA1A-FLAG-mCherry and ArLEA1B-FLAG-mCherry by western blot, 1 μg each of pcDNA3.0-ArLEA1A-FLAG-mCherry, pcDNA3.0-ArLEA1B-FLAG-mCherry, pcDNA3.0-AavLEA1-FLAG-mCherry and pcDNA3.0-mCherry plasmids were transfected in COS-7 cells growing on 6-well plates (Nunc), using 6.0 μl GeneJammer. Cells were harvested 48 h post-transfection. For NS-EGFPHDQ74-ATEL expression analysis cells were transfected with 1 μg of pcDNA3.3-NS-EGFPHDQ74-ATEL alone or with 5 μg of pcDNA3.0-ArLEA1A-FLAG-mCherry/pcDNA3.0-ArLEA1B-FLAG-mCherry using 6.0 μl /36.0 μl of GeneJammer for single/double transfections. For MTS assay, cells were grown in a 96-well plate to around 70-80% confluence, and transfected with 50 ng of pcDNA3.3-NS-EGFPHDQ74-ATEL alone or with 250ng of pcDNA3.0-ArLEA1A-FLAG-mCherry/pcDNA3.0-ArLEA1B-FLAG-mCherry per well. 0.15μl/0.90 μl of GeneJammer was used for single/double transfections. MG132 and TMAO were added to cells expressing NS-EGFPHDQ74-ATEL at a final concentration of 1 μM and 100mM right after transfection.

2.4 ER-Tracker Green and MitoTracker Green staining

ER-Tracker Green (BODIPY® FL glibenclamide) (1mM stock, Invitrogen) stock solution (prepared in DMSO) was diluted to 1 μM in pre-warmed Hank's Balanced Salt Solution (HBSS). After removing media, cells were rinsed with HBSS and pre-warmed staining solution was added, then incubated at 37°C for 30 min. The staining solution was subsequently removed and pre-warmed PBS with 2% FBS was added to view cells under the confocal microscope. For MitoTracker Green FM (1mM stock, Invitrogen) staining, stock solution (prepared in DMSO) was diluted to 100 nM in pre-warmed DMEM medium. The growth medium was replaced with 2 ml staining media and cells were incubated at 37°C for 15 min. The staining medium was subsequently replaced with pre-warmed media and cells were visualised under the confocal microscope.

2.5 Fixation and DAPI staining

Cells co-expressing NS-EGFPHDQ74-ATEL and NS-mCherry-ATEL were grown on coverslips. After 24h/48h/72h post-transfection, cells were washed with 1X PBS once and fixed with 1 ml 4% paraformaldehyde for 15 min at room temperature. Subsequently, paraformaldehyde was removed and cells were washed twice with 1X PBS. The coverslips were subsequently mounted on 15 µl DAPI (Vectashield) on glass slides and sealed with nail polish. After air-drying for 30 min, slides were stored at 4°C in the dark.

2.6 Confocal microscopy

Imaging was performed using a Zeiss LSM 510 META laser scanning confocal microscope (X63, 1.4 N.A., Plan apochromat objective). Excitation wavelengths were 405 nm for DAPI, 488 nm for ER-Tracker Green/EGFP/GFP and 543 nm for mCherry; emission wavelengths were 420-480 nm for DAPI, 505-530 nm for ER-Tracker Green/EGFP/GFP and 560-615 nm for mCherry. For the co-localisation analysis, the voxel dimensions were set according to the Nyquist criteria (Scriven et al., 2008). The x, y and z dimensions of pixel sizes were calculated for 488 nm excitation wavelength as x= 43 nm; y= 43 nm; z= 0.13 µm; digital zoom: 1.6; resolution: 2048; pinhole size: 1 Airy unit. Calculations were made using www.svi.nl/NyquistCalculator. Images were processed using manufacturer's software. 2D pixel analysis of red and green channel intensities was performed using JACoP (Just Another Colocalisation Plug-in) (Bolte and Cordelieres, 2006) on ImageJ 1.45s.

2.7 FLIM

Cells transfected with pcDNA3.3-NS-EGFPHDQ74-ATEL were visualised on an inverted confocal microscope attached to a time correlated single photon counting (TCSPC) device that permits excitation and emission spectra as well as fluorescence lifetime to be recorded in every image pixel (Kaminski Schierle et al., 2011). Images

were taken at 24h and 48h post-transfection. At least 8 sample images were collected for each time point. Cells expressing NS-EGFP-ATEL in the ER and EGFPHDQ74 in the cytoplasm were used as controls. The multi-parametric imaging system used is described in detail in Frank et al. (Frank et al., 2007; Kaminski Schierle et al., 2011). A pulsed supercontinuum source (SC 450, Fianium Ltd., Southampton, UK) was used for excitation, emitting a train of sub 10 ps pulses at a 40 MHz repetition rate. The fluorescence light was band pass filtered and passed onto a fast photomultiplier tube (PMT-100, Becker & Hickl GmbH, Berlin, Germany). Lifetimes were recorded using time correlated single photon counting, TCSPC, circuitry (SPC-830, Becker & Hickl GmbH, Berlin, Germany). An excitation wavelength of 488 nm was used, with an emission longpass filter of 515 nm. Photon count rates were kept below 1% of the laser repetition rate to prevent pulse pile-up. Images were acquired during 100 s, and photobleaching was verified to be negligible during these acquisition times. All TCSPC images were processed using SPCImage (Becker & Hickl GmbH, Berlin, Germany) and fitted with a monoexponential decay function. Pixel binning was increased until approximately 3500 to 5000 photons were obtained per pixel. Image processing and data analysis were carried out using various codes developed in-house using Matlab (The Mathworks Ltd., Cambridge, UK; Chakrabortee et al., 2010; Elder et al., 2009; Esposito et al., 2010; Schlachter et al., 2009). The mean fluorescence lifetime +- 1 SD was calculated for each image.

2.8 Protein methods

2.8.1 Protein concentration determination

Total protein concentration was estimated using DC++ protein assay kit (BioRad) based on the Lowry method according to the manufacturer's instructions. Bovine serum albumin (BSA) standards in the range of 0.1-1 mg/ml were used to plot a calibration curve. Absorbance was measured at 750 nm.

2.8.2 Protein dialysis

Volumes lesser than 300 µl were dialysed in Slide-A-Lyzer mini dialysis cups, 3500MW cutoff (MWCO)(Pierce). Dialysis cups were washed thoroughly in distilled

water 30 minutes before dialysis. Dialysis was carried out at 4°C with stirring and buffer changes every 2h. Larger volumes were dialysed in Spectra/Por dialysis tubing, 3500 MWCO (Spectrum Laboratories). Tubing was washed with distilled water for 30 min before dialysis. Dialysis was carried out at 4°C with constant stirring and buffer changes.

2.8.3 Protein concentration

Protein samples were concentrated using Vivaspin 2 (Sartorius Stedim) with 3000 MWCO polyethersulfone (PES) membranes according to the manufacturer's instructions.

2.8.4 Polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was carried out in Mini Protean 3 gel tank (BioRad) according to the manufacturer's instructions. Either 10% or 15% gels were used according to the following recipe:

10% resolving gel: 3.3 ml 30% 29:1 Acrylamide/bisacrylamide solution, 2.5 ml 1.5M Tris-HCl pH 8.8, 0.1 ml 10% SDS, 50 µl 10% APS and 10 µl TEMED, 4.05 ml water.

15% resolving gel: 5.0 ml 30% 29:1 Acrylamide/bisacrylamide solution, 2.5 ml 1.5M Tris-HCl pH 8.8, 0.1 ml 10% SDS, 50 µl 10% APS and 10 µl TEMED, 2.35 ml water.

4% stacking gel: 1.3 ml 30% 29:1 Acrylamide/bisacrylamide solution, 2.5 ml 0.5M Tris-HCl pH 6.8, 0.1 ml 10% SDS, 50 µl 10% APS and 10 µl TEMED, 6.1 ml water.

2X SDS-PAGE sample buffer: 30% glycerol, 60 mM Tris-HCl pH6.8, 0.5% bromophenol blue, 2% SDS, 5% β-mercaptoethanol. Samples were mixed with loading buffer in a 1:1 ratio and boiled for 5 min. 5 µl of molecular weight ladder (Fermentas) was run along with the samples at 30 mA for sufficient amount of time in 1X SDS running buffer. Gels were visualised by staining in a solution of 0.2% coomassie brilliant blue in 10% acetic acid and 45% methanol for 30 min-12h. Destaining was performed in a solution containing 10% acetic acid and 30% methanol for 12h.

2.8.5 Recombinant protein expression in *E. coli*

pET-28a(+) expression vectors containing ArLEA1A/ArLEA1B (Pouchkina-Stantcheva et al., 2007), AavLEA1-FLAG-mCherry and pET-15b containing AavLEA1 (Goyal et al., 2003), were transformed into *E. coli* BL21 (DE3) cells. 20 µl of cell solution was spread on LB plates containing 50 mg/ml kanamycin. After selecting clones expressing the recombinant protein of interest, single colonies were inoculated in 5 ml of LB with 50 µg/ml kanamycin and incubated overnight. 400 µl of overnight culture was inoculated into 400 ml LB containing 50 µg/ml kanamycin and incubated at 30°C until the OD_{600} reached 0.6. IPTG was added to a final concentration of 1 mM and the cells were cultured overnight. Samples of induced and pre-induced cultures were analysed by SDS-PAGE to check for expression levels.

For mCherry production, *E. coli* were grown at 30°C, and subsequently induced at three different temperatures: 16°C, 30°C and 37°C. Sample lysates were tested for optimum expression of mCherry protein at 1,2,3 and 5 h post-induction. For final purification, cultures grown at 30°C were used.

2.8.6 Protein extraction

Protein extraction was carried out at 4°C. Induced cells were harvested by centrifugation at 4000 RPM for 30 minutes in the Sorvall GSA30 rotor. Cell pellets were re-suspended in lysis buffer (100 mM NaH_2PO_4, 10 mM Tris·Cl, 6 M GuHCl, pH 8.0) at 2-3 ml per gram weight of the pellet. Cells were lysed by sonication (10 s off, 15 s on followed by 90 s off). The lysate was stored on ice for 15 min. Supernatant was collected and retained after centrifugation at 18000 RPM for 20 min in the Sorvall SS34 rotor.

2.8.7 Nickel-nitrilotriacetic acid (Ni-NTA) column purification

Histidine-tagged recombinant proteins were purified on a 5 ml Ni-NTA column at 4°C. The column was washed with 25 ml distilled water, and equilibrated with 25 ml lysis buffer. The lysate was applied on the column and the flow through was collected. Subsequently, the column was washed with 3 x 5 ml wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 0.05% Tween 20, pH 8.0). The protein

was finally eluted 3-5 times with 3 ml elution buffer (50 mM NaH_2PO4 300 mM NaCl 250 mM imidazole 0.05% Tween 20, pH 8.0). The flow-through, washes and eluted proteins were analysed by SDS-PAGE.

2.8.8 Thrombin cleavage

mCherry was cleaved using the Thrombin CleanCleave Kit (Sigmal-Aldrich) according to the manufacturer's instructions. The remaining purified proteins were dialysed into phosphate buffered saline and the His-tag was cleaved by overnight incubation at 4°C with 50 µl thrombin (1 unit/µl, GE Healthcare) according to the manufacturer's instructions. The tag and any uncleaved protein were removed by Ni-NTA sepharose and thrombin was removed using benzamidine sepharose (GE Healthcare).

2.8.9 Gel filtration

The HiPrep 16/60, Superdex 75 gel filtration column was attached to ÄKTAxpress FPLC machine. The column was equilibrated with gel filtration buffer containing 0.05 M sodium phosphate and 0.15M NaCl, pH 6.2. The void volume was calculated by running blue-dextran solution (10 mg dissolved in 10 ml PBS), filtered through 0.22 µm Millipore filter, through the gel filtration column. The column was calibrated by running proteins of known molecular weight through the column (Ribonuclease A- 13.7 kDa, Lysozyme- 14.3 kDa, Ovalbumin- 44kDa and BSA- 66.3 KDa). After re-equilibrating the column, thrombin cleaved mCherry protein was applied to the column after passing it through the 0.22 µm Millipore filter. The fraction containing the elution peak was analysed by SDS PAGE gel.

2.8.10 Ion-exchange chromatography

mCherry protein was dialysed overnight in 20mM Tris HCl pH 8.0 with two changes of buffer. The MonoQ ion-exchange column was attached to the ÄKTAxpress FPLC machine (GE Healthcare) and was equilibrated with 20mM Tris HCl pH 8.0. The protein sample was then applied to the column. Proteins were subsequently eluted

using a linear gradient of NaCl from 0-0.5M NaCl in 20mM Tris HCl pH 8.0. Fractions were analysed by SDS PAGE.

2.9 mCherry drying experiments

Prior to drying experiments all thrombin cleaved proteins were dialysed extensively against water and the concentrations determined by measuring the absorbance at 280 nm using the following molar extinction coefficients: mCherry, 32430 $M^{-1}cm^{-1}$; ArLEA1A, 14080 $M^{-1}cm^{-1}$; ArLEA1B, 12800 $M^{-1}cm^{-1}$; AavLEA1, 8250 $M^{-1}cm^{-1}$; AavLEA1-FLAG-mCherry, 41960 $M^{-1}cm^{-1}$; BSA (Sigma-Aldrich, product number A7906), 43824 $M^{-1}cm^{-1}$. 100 μl samples in water were prepared containing 30 μM mCherry alone or in the presence of other proteins at molar ratios of 1:5. AavLEA1-FLAG-mCherry fusion protein was also used at 30 μM concentration.

For generating mCherry absorbance and emission spectra, mCherry was dried in an Eppendorf 5301 vacuum concentrator, transferred to a Dura-Stop™ microprocessor-controlled vacuum tray drier (FTS Systems, Stone Ridge, NY) for 1h at 500 mTorr with a tray temperature of 25°C, and re-suspended in 100 μl of water; four cycles of drying and rehydration were carried out. mCherry absorbance was measured at 23°C in a Lambda 35 UV/visible spectrophotometer (PerkinElmer, Cambridge, UK), using a 1 cm path-length UV-transparent cuvette. Wavelengths between 350 and 650 nm were measured with a scanning rate of 240 nm/min and a data interval of 1 nm. Fluorescence emission spectra were recorded using a Cary Eclipse fluorimeter (Agilent Technologies). Excitation was set at 543 nm and emission spectra were recorded between 570 and 700 nm with a scan rate of 600 nm/min and a data interval of 1 nm. Emission and excitation slit width was set at +/-10 nm and a photomultiplier tube voltage of 690 V was used.

In the second round of experiments, mCherry was dried with or without LEA proteins/BSA in an Eppendorf 5301 vacuum concentrator and then transferred to Dura-Stop™ microprocessor-controlled vacuum tray drier (FTS Systems, Stone Ridge, NY) for 1h at 500 mTorr with a tray temperature of 25°C, and re-suspended in 100 μl of water; four cycles of drying and rehydration were carried out. Fluorescence and

absorbance measurements were carried out in triplicate using EnVision plate reader (PerkinElmer) in a 96 well OptiPlate. Absorbance was measured at the absorbance maxima of 587nm. Fluorescence emission was measured using excitation wavelength of 544 nm at 615nm.

2.10 Western blot analysis

For confirming the expression of ArLEA1A-FLAG-mCherry and ArLEA1B-FLAG-mCherry 48 h post-transfection, cells were trypsinized and re-suspended in PBS and pelleted at 5000 RPM for 5 min at 4°C. The cells were harvested again in 500 µl PBS, transferred to fresh eppendorf tubes and centrifuged at 5000 RPM for 5 min, 4°C. After discarding the supernatant, pellets were lysed with 100 µl of 1X lysis buffer. 2X lysis buffer consisted of 20mM Tris HCl, 137mM NaCl, 1mM EGTA, 1% TritonX, 10% glycerol and water. 1 tablet containing protease inhibitors with EDTA (Roche) was dissolved in 15 ml water. 142 µl of protease inhibitor, 358 µl water and 500 µl of 2X lysis buffer were finally mixed to make 1 ml of 1X lysis buffer. Protein concentration was estimated using BioRad (Lowry) protein estimation kit. 10 µg of lysate was loaded onto a 10% SDS PAGE gel, along with 5 µl protein ladder. After running the gel, the gel was soaked in 1X transfer buffer containing 48 mM Tris HCl pH 6.8, 39 mM glycine, 0.035% SDS, 20% methanol, for 15 min. The PVDF membrane (Amersham Hybond, GE Healthcare) was treated with methanol (20 s), flipped and soaked in water (5 min) and transfer buffer (10 min). The membrane was then loaded on top of blotting paper pre-soaked in transfer buffer. The protein gel was carefully placed over the membrane, which was covered with another sheet of blotting paper. The transfer was carried out at 15V for 30 min. The membrane was then blocked with 5% milk in PBST for one hour and incubated with primary antibody (1:1000 dilution in 1% PBST) overnight. It was then washed the next day with PBST for 15, 10, and 5 min. Secondary antibody was added in 1:4000 dilution (1% milk in PBST) overnight. It was then washed with PBST and developed with Lumigen Chemiluminiscence reagent A and B. The X-Ray film was exposed to the membrane for 3 min and developed.

Analysis of NS-EGFPHDQ74-ATEL expression with or without ArLEA1A-FLAG-mCherry/ArLEA1B-FLAG-mCherry/MG132/TMAO was performed as follows. 5 μg of lysates were loaded in two identical 10% SDS PAGE gel simultaneously, along with 5 μl protein ladder. After running SDS-PAGE, the proteins were transferred to PVDF membranes from both the gels as described above. After transfer, the membranes were washed twice with 5 ml PBST and blocked in 5ml PBST, containing 5 % milk for 1 h. Subsequently, the membranes were incubated with anti-GFP (1:2500) or anti-ArLEA1A (1:1000) in the blocking solution for 1 h. After washing the membrane in PBST (3×5 min), the membrane was incubated in the HRP-linked secondary antibody (anti-mouse for GFP and anti-rabbit for ArLEA1A) in 1:4000 dilution for 40 min. After washing the membrane in PBST (3×5 min), it was placed on OHP film and 700 μl of ECL Prime (350μl of Solution A and 350 μl of Solution B) was added uniformly on each membrane. After leaving the membranes in the dark for 5 min all liquid was removed and the membranes were exposed in the G-Box western blot developer (Syngene) for 10 min.

2.11 MTS assay

20 μl of MTS reagent (Promega) was added to 100 μl of cell culture medium. The plate was covered in aluminium foil, to prevent exposure to light and incubated at 37°C in a humidified atmosphere containing 5% CO_2 and incubated for 1 h. The absorbance was subsequently measured at 490 nm in an EnVision plate reader (PerkinElmer). Experiments were carried out in triplicate.

2.12 mCherry fluorescence measurements in cell supernatants

Supernatants obtained after transfection experiments were centrifuged at 4000rpm, 4˚C to pellet cell debris. Vivaspin 2 concentrators (Sartorius Stedim) were used to concentrate supernatants following the manufacturer's instructions. Supernatants (4 ml) were concentrated to approximately 200 μl. Protein concentration of supernatants was determined using DC Protein Assay kit (Biorad). Fluorescence

measurements were carried out using normalised amount of protein in all samples, with an EnVision plate reader (PerkinElmer) in a 96-well opaque OptiPlate, using excitation wavelength of 544 nm. Emitted light was collected at 615 nm. Number of flashes was set to 10.

2.13 Statistical analysis

All statistical analysis was performed using the InStat package version 3.1a (Graphpad Software, Inc.). Since, the fluorescence values in supernatants did not pass the Kolmogorov-Smirnov (KS) normality test, square-root transformation of the data was performed. Statistical significance between different groups was determined by one-way ANOVA and Tukey's post-hoc test. Data were plotted using Microsoft Excel. FLIM data was analysed using the one-sided t-test with Welch correction. MTS assay data at 48 h time point was analysed by one-way ANOVA and Tukey's post-hoc test. mCherry drying data was normalised with respect to pre-dried samples and analysed using one-way ANOVA and Tukey's post-hoc test when the data was normal. Otherwise non-parametric ANOVA and Dunn's post-hoc test was applied.

Chapter Three: Bioinformatics of Bdelloid LEA Proteins

3.1 Introduction

Although we have gained significant understanding of the potential functions of ArLEA1A and ArLEA1B as anti-aggregating agents and membrane stabilisers by *in vitro* experiments (Pouchkina-Stantcheva et al., 2007), a knowledge gap regarding their phylogeny, domains, structural features, signalling and interacting motifs persists. In order to fill this gap, it was decided to use bioinformatics to discover novel features and functions of these proteins.

Bioinformatics was a term coined by Paulien Hogeweg and Ben Hesper in 1970 (Hesper and Hogeweg, 1970) and was originally defined as "the study of informatic processes in biotic systems". This definition was inspired by the observation that life processes are information driven, with the information in DNA being translated into intra-cellular and inter-cellular functions of organisms. More recently with the deluge in biological data, the use of this term has expanded to include the development and application of computational methods for large-scale analysis of genomes, transcriptomes, proteomes, interactomes, metabolic and regulatory networks etc. Such methods can also be applied to decipher protein functions as discussed below.

In modern day research, phylogenetic trees built by sequence comparison tools can be used to deduce the evolutionary relationships and functions of novel genes and proteins in newly sequenced genomes (Nierman et al., 2001; Trucksis et al., 1998). The phylogenetic analysis of a protein (Fig. 3.1) usually starts with the detection of other homologous members of its family (Gabaldon, 2005). This is performed by comparing the sequence of the protein of interest with other sequences stored in a database and subsequently, selecting those that are significantly similar. The assumption is that proteins with similar sequences are derived from a common

ancestral protein (Fitch, 1970). Many algorithms can be used to perform this task like Smith-Waterman (Smith and Waterman, 1981) and its faster counterpart, BLAST (Altschul et al., 1997). Once the sequences of a protein family are retrieved, they can be aligned by using multiple sequence alignment (MSA) techniques. MSA aims to display residues for each protein on a single line, such that equivalent residues appear on the same column. Finding an optimal multiple alignment in a set of sequences is a complex mathematical problem and can be performed by using either global, progressive or local alignment techniques (Do and Katoh, 2008).

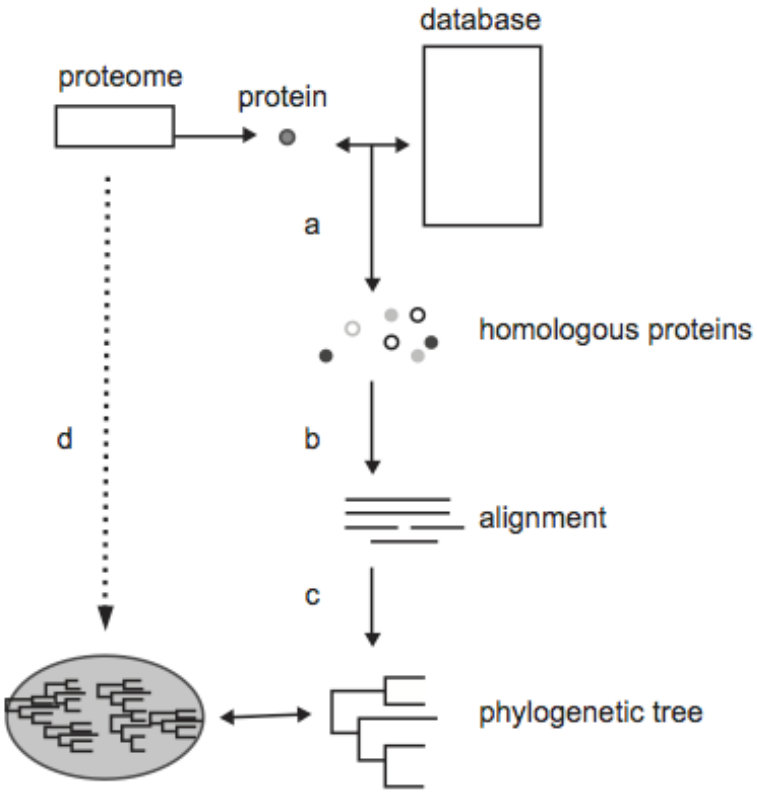


Figure 3.1: Phylogenetic tree construction. (a) The protein sequence of interest is compared to a sequence database to retrieve significantly similar proteins; (b) homologous proteins are aligned to place homologous residues on top of each other under the assumption of an evolutionary model, (c) phylogenetic tree representing the evolutionary relationships among the protein sequences is reconstructed; if this process (a to c) is repeated over all proteins encoded by a genome, the total set of phylogenetic trees or phylome is reconstructed (d). Taken from Gabaldon, 2005, with permission.

Subsequently, a specific evolutionary model can be applied to explain the amino acid substitutions observed in the MSAs and the evolutionary distances between all pairs of proteins can be computed (Gabaldon, 2005). The evolutionary distance, which reflects the expected mean number of changes per site that have occurred since two sequences diverged from their common ancestor, can be derived by using distance-methods for phylogenetic inference such as the Neighbor Joining method (NJ) (Saitou and Nei, 1987) or by using the Maximum Likelihood (ML) (Schadt et al., 1998) or Maximum Parsimony (MP) (Felsenstein, 1996) methods. These trees can then be used to establish orthology and paralogy relationships among proteins (Fitch, 1970) and to detect horizontal gene transfers (HGT) (Bapteste et al., 2004) between different species. The evolutionary relationships and functions of bdelloid LEA proteins have been deduced by phylogenetic tree construction in Section 3.2.1.

Another approach to analyse protein functions involves grouping them into specific families, in a manner similar to classification of elements in the periodic table, using a database called Pfam (Sammut et al., 2008). This database was first founded in 1995 by Erik Sonnhammer, Sean Eddy and Richard Durbin (Sonnhammer et al., 1997) and originated from the idea that proteins can be clustered into families whose members have diverged from a common ancestor and hence have similar sequences and folds and possibly similar functions (Chothia, 1992). Since the original estimates of around 1000 families covering the entire protein universe, the latest release of Pfam has around 13,672 manually curated families (Finn et al., 2010). To perform a search in Pfam, a query sequence is converted into a 'hidden Markov model' (HMM) profile, which is a position-specific probabilistic model of a multiple sequence alignment derived with the aid of the BLOSUM62 standard scoring matrix plus the empirically set insertion/deletion transition probabilities of aligned sequences. The current version of Pfam employs the HMMER (version 3) program that designates proteins into families by computing local HMM alignments. Moreover, each family consists of sequences with manually defined threshold values (also referred to as the 'gathering threshold cut-off values') such that each sequence in the family exceeds these values, thereby reducing the number of false positives. Pfam entries have been divided into two parts: Pfam-A, that consists of high quality,

manually curated matches; and Pfam-B, that consists of automatically generated alignments from various databases like ADDA (Automatic Domain Decomposition Algorithm) (Finn et al., 2010), and which are considered to be less reliable. However, Pfam-B families can also hint towards possible functions of newly identified domains in novel sequences. The Pfam domains in ArLEA1A and ArLEA1B have been characterised in Section 3.2.2.

Protein functions are often determined by their structures as well. Previous analysis of secondary structure of ArLEA1A and ArLEA1B in hydrated and dry states by far-ultraviolet circular dichroism spectroscopy (Pouchkina-Stantcheva et al., 2007) has shown that ArLEA1A is predominantly unstructured in solution (with 29% α-helical content) and acquires a greater proportion of α-helices upon drying (84% α-helical content), while ArLEA1B is structured both in the hydrated (82% α-helical content) and dry states (87% α-helical content). However recently, Boschetti et al. have reported that bdelloid LEA proteins are unusual compared with group 3 LEA proteins found in other species since they possess a greater degree of order and folding as evident by their high GRAVY and FoldIndex unfoldability scores (Boschetti et al., 2011). Hence, it was decided to unravel the possible three-dimensional structures of ArLEA1A and ArLEA1B using an online modeling tool called I-TASSER (Roy et al., 2010; Zhang, 2008). I-TASSER employs a hierarchial protein structure modeling approach based upon secondary structure enhanced profile-profile threading alignment (PPA). Target sequences are threaded through a representative PDB structure library (with a pair-wise sequence identity cut-off of 70%) to search for possible folds. Models are further refined by Monte-Carlo simulations. The top five models are finally reported after multiple cycles of clustering, statistical modeling and energy optimizations. These models are described in Section 3.2.3.

Determining sub-cellular localisation is also important for elucidating protein functions. Pouchkina-Stantcheva et al. have reported that both ArLEA1A and ArLEA1B proteins possess a 19-residue hydrophobic sequence at the N terminus for ER translocation, and a putative variant ER retention signal, ATEL, at the C terminus (Pouchkina-Stantcheva et al., 2007). To confirm the identity of the N-terminal signal, two sub-cellular localisation prediction tools: SignalP and TargetP (Emanuelsson et

al., 2007; Petersen et al., 2011) were applied. SignalP 4.0 is a neural network based program that has been trained on transmembrane sequences as negative data to improve its performance in detecting true signal peptides. It is an improvement over existing prediction methods for signal peptide detection like SignalP 3.0, PrediSi, SPEPlip, Signal-CF, Signal-3L and Signal-BLAST which have only limited capacity to distinguish between signal peptides and N-terminal transmembrane helices (Petersen et al., 2011). TargetP 1.1 (Emanuelsson et al., 2007) is another tool for subcellular localisation predictions of newly synthesised proteins. Using the N-terminal sequence information, it discriminates between proteins destined for the mitochondrion, the chloroplast (for plant sequences) or the secretory pathway. These two tools have been applied to confirm the presence of localisation signals in ArLEA1A and ArLEA1B in Section 3.2.4

Moreover, as most biological phenomena are dependent on multiple interactions between different proteins that might be affected by specific environmental or metabolic signals, it was speculated that ArLEA1A and ArLEA1B might interact with other proteins during anhydrobiosis. In order to test this hypothesis, bdelloid LEA sequences were searched for the presence of interacting motifs using Scansite in Section 3.2.5.

3.2 Results

3.2.1 Sequence comparison and phylogenetic analysis of bdelloid LEA proteins

In order to infer the phylogenetic relationships of ArLEA1A and ArLEA1B, a 'local alignment' based BLAST search (Dereeper et al., 2010) was carried out using the Phylogeny.fr analysis suite (Dereeper et al., 2008). There were 597 and 467 hits obtained for ArLEA1A and ArLEA1B sequences, respectively, corresponding to sequences from diverse taxa including bacteria, yeast, fungi, plants, worms and insects. A majority of these sequences corresponded to LEA proteins, although many hypothetical proteins were pulled out in the search as well. The top 50 hits from various species ranging from eukaryotes like *A. vaga*, *C. elegans*, *Polypedilum*

vanderplanki, *A. thaliana* and prokaryotes like *Nostoc* and *Clostridium* were processed in the 'One-click' work-flow mode in Phylogeny.fr. Redundant sequences were filtered out and an MSA was constructed using the MUSCLE program (Edgar, 2004a, b). Subsequently, the aligned sequences were used to plot the phylograms depicting the phylogenetic distances between these sequences using PhyML and TreeDyn (Fig. 3.2, 3.3) (Chevenet et al., 2006; Guindon et al., 2010).

As can be observed in the phylograms (Fig. 3.2, 3.3), ArLEA1A and ArLEA1B form a distinct evolutionary clade that is separate from other clades consisting of LEA protein isoforms from organisms like *P. vanderplanki*, *C. elegans*, *C. brenneri, A. thaliana* etc. The approximate likelihood ratio test scores for clades consisting of ArLEA1A and ArLEA1B matches are 0.98 and 0.99 respectively. These results suggest that bdelloid LEA proteins have evolved independently of other related group 3 LEA proteins and therefore might possess unique functions.

The observed phylogeny of ArLEA1A and ArLEA1B can also be used to deduce novel functions of these proteins. For example, ArLEA1A and ArLEA1B are also found to be homologous to dehydration-induced group 3 LEA proteins found in *P. vanderplanki* (PvLEA1, gi_90959527 and PvLEA3, gi_90959531) (Kikawada et al., 2006). A recent study has demonstrated that the 11-mer repeat units of these proteins can contribute towards glass formation with a high glass transition temperature (>100 degrees C) and a low enthalpy relaxation rate (Shimizu et al., 2010). It is tempting to speculate that bdelloid LEA proteins might vitrify in the absence of water in a manner similar to these proteins.

Interestingly, ArLEA1A and ArLEA1B also exhibit homology to ankyrin proteins found in certain ant species like *Acromyrmex* and *Camponotus floridanus* (gi_332022332 and gi_307184783). Ankyrin proteins are responsible for anchoring various receptors and other scaffold proteins to the plasma membrane (Bennett and Baines, 2001) and are hypothesised to play a role in preserving cell and tissue integrity in various metazoa species. One can suppose that ArLEA1A and ArLEA1B might provide structural stability to membranes in a manner similar to ankyrin proteins.

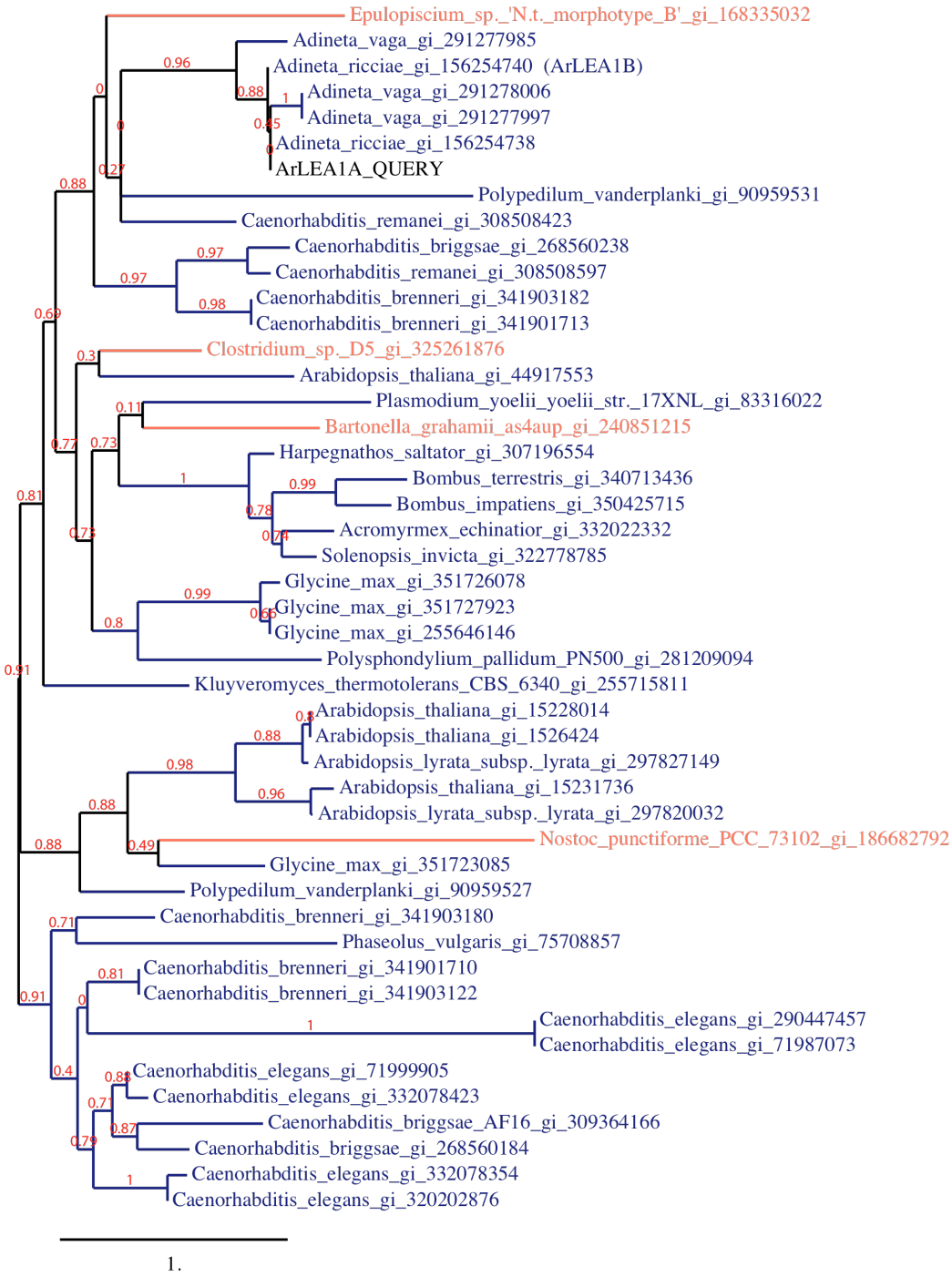


Figure 3.2: Phylogram consisting of ArLEA1A homologs, plotted with PhyML (Maximum likelihood approach) and TreeDyn in Phylogeny.fr. The scale-bar represents the amino acid changes per site, while the likelihood scores for each branch are shown in red. As can be observed, bdelloid LEA proteins form a distinct evolutionary clade. Purple: Eukaryotes, orange: prokaryotes.

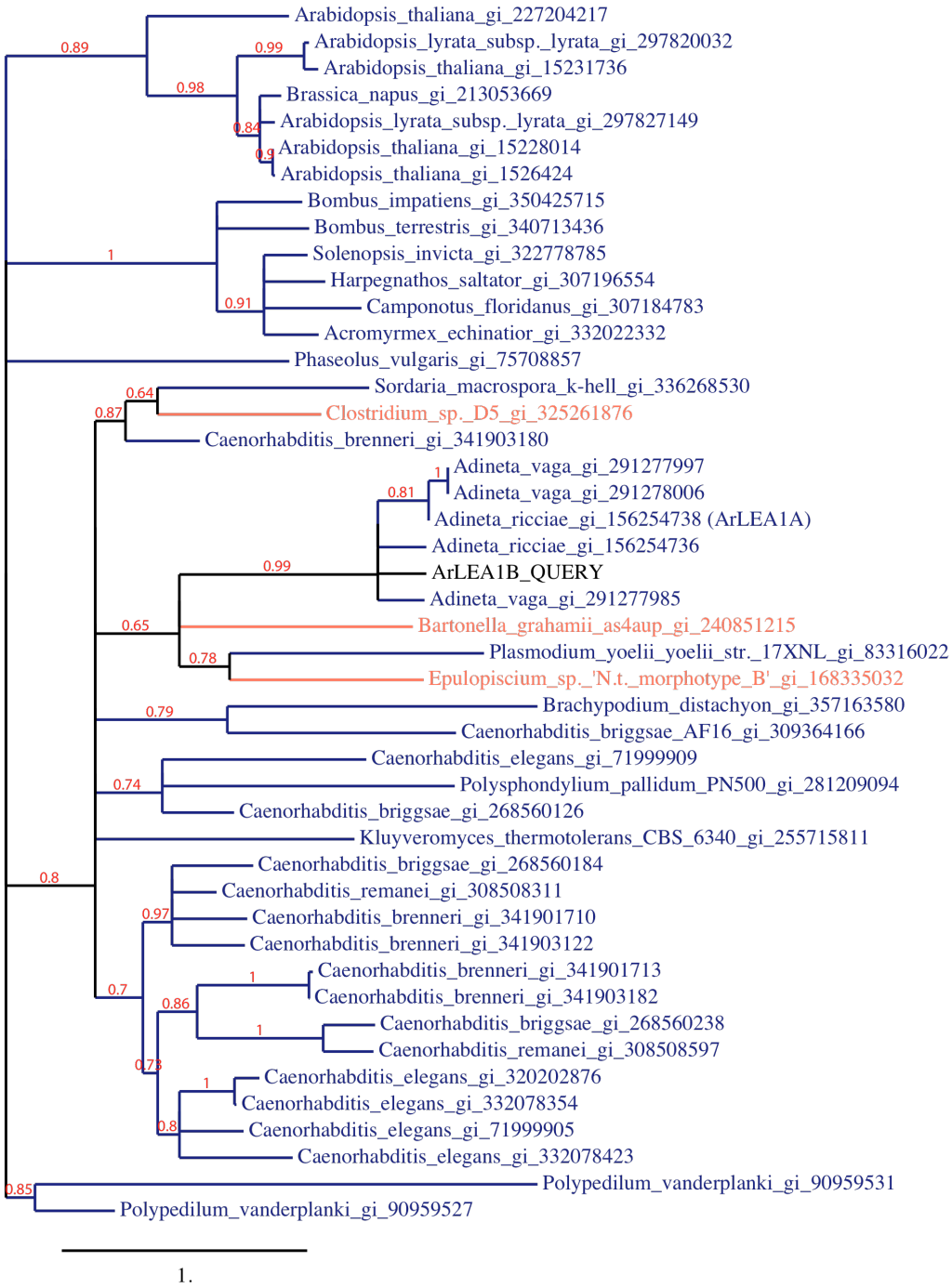


Figure 3.3: Phylogram consisting of ArLEA1B homologs, plotted with PhyML (Maximum likelihood approach) and TreeDyn in Phylogeny.fr. The scale-bar represents the amino acid changes per site, while the likelihood scores for each branch are shown in red. As can be observed, bdelloid LEA proteins form a distinct evolutionary clade. Purple: Eukaryotes, orange: prokaryotes.

Moreover, ArLEA1B has also been found to be homologous to a stress related protein KLTH0E16258p found in yeast *Lachancea thermotolerans*, with a LEA-1-like domain and a cell envelope integrity inner membrane protein domain. These similarities provide further evidence related to the potential function of ArLEA1B in stabilising membranes during anhydrobiosis (Pouchkina-Stantcheva et al., 2007). However, the fact that ArLEA1A also exhibits homology to ankyrin implies that it might interact with membranes as well. However, this hypothesis is only partially supported by the FTIR data obtained by Pouchkina-Stantcheva et al. which demonstrates that only 30% of ArLEA1A is found in the bound fraction of P=O groups of dried membranes compared to 42% of ArLEA1B (Pouchkina-Stantcheva et al., 2007). These results suggest that ArLEA1A might bind weakly to membranes, in comparison with ArLEA1B that might demonstrate a stronger interaction with membranes during anhydrobiosis.

Bdelloid LEA proteins are also found to be homologous to the LEA-76 protein found in the genus *Clostridium* which consists of Gram-positive bacteria that are capable of producing desiccation-resistant endospores. The LEA-76 family belongs to group 3, first identified in desiccation-induced gene sets in various plants (Curry and Walker-Simmons, 1993; Dure, 1993; Hsing et al., 1995). Besides plants, at least two proteins of this family (DR0105 and DR1172) are also present in extremophiles like *Deinococcus radiodurans* (Makarova et al., 2001). Hence, it is likely that rotifer LEA proteins might share functions with these proteins.

Interestingly, ArLEA1A and ArLEA1B demonstrate close phylogenetic linkage with some prokaryotic proteins. For example, ArLEA1A is found to be related to a 5'-nucleotidase/2' 3'-cyclic phosphodiesterase, (gi_168335032) found in *Epulopiscium sp.*, a resident of the intestinal tract of tropical fish. ArLEA1B, on the other hand clusters with Tol-Pal system protein YbgF (gi_240851215), responsible for maintaining membrane integrity of parasitic *Bartonella sp*. Similarity in sequences of proteins of distantly related organisms, can sometimes indicate the possibility of horizontal gene transfer (HGT) between these species.

However, one must be very careful in interpreting results from this phylogenetic analysis since phylogenetic tree accuracy is dependent on the quality of

multiple sequence alignments which might sometimes be compromised by the presence of repeats, deletions and insertions in protein sequences. Such repeats, deletions and insertions are often observed in LEA protein sequences as well (for review see Battaglia et al., 2008). Further, Bapteste et al. have pointed out the inherent difficulties in providing evidence for or against HGT, since for a given set of proteins, it is possible to construct rooted trees in many different ways (Bapteste et al., 2004).

3.2.2 Pfam domains in ArLEA1A and ArLEA1B

In order to identify the domains in ArLEA1A and ArLEA1B, their protein sequences along with their few selected homologues from *Ad. vaga*, *C. elegans*, *A. thaliana*, *P. vanderplanki* and *Clostridium* sp. were deposited in the Pfam server. The AavLEA1 sequence from the worm *A. avenae* was also used for comparative purposes. The search criteria were adjusted to include the gathering threshold cut-off values and Pfam-B families.

Table 3.1 shows the Pfam classifications of rotifer LEA proteins, along with their eukaryotic and prokaryotic homologues. As can be observed, ArLEA1A and ArLEA1B (designated lea-1A and lea-1B respectively) do not show any significant Pfam-A matches. Instead, they contain three Pfam-B domains: PB00621, PB006097 and PB008237. *Ad. vaga* homolog lea-1B also possesses these domains, while lea-1B' lacks PB008237 and lea-1C lacks both PB006097 and PB008237. Further, lea-1C contains PB00621, which is common to other rotifer LEAs as well. However, it also possesses PB002305 that is absent in the rotifer set, but present in AavLEA1.

Pfam-B domain PB00621 is present in all the bacterial, worm, plant and insect homologues of rotifer LEA proteins and is perhaps critical for the function of these proteins. Further investigation into the properties of this domain using the ADDA database reveals that it is widely distributed in signalling kinases in a number of bacteria and fungi like *Bacillus*, *Nostoc* and *Trichomonas* species. The HMM profile of the second Pfam-B domain PB006097 found in rotifer LEA proteins, is observed solely among prokaryotic proteins and has no representative sequences in

eukaryotes. These results suggest common evolutionary origins of eukaryotic and prokaryotic LEA proteins. The functions of other Pfam-B domains like PB008237 and PB002305 remain unknown.

In contrast with ArLEA1A and ArLEA1B, LEA protein homologues from other eukaryotic species like *C. elegans*, *P. vanderplanki* and *A. thaliana* contain a unique Pfam-A domain called LEA_4 (PF02987.11). This domain is prevalent in 777 protein sequences curated in Pfam, spanning 80 different species ranging from archaea, bacteria, plants, worms and insects and is found to be related to a bacterial stress protein CsbD (Pragai and Harwood, 2002). The fact that this domain has been preserved during the course of evolution of LEA proteins in various species, seems to suggest that it might be important in determining the overall architecture and function of many eukaryotic group 3 LEA proteins. However, bdelloid LEA proteins lack this domain, raising the possibility that these proteins might have evolved independently in these organisms. Moreover, this domain is also absent in LEA-76 protein found in *Clostridium* species, suggesting that bdelloid LEA proteins are more closely related to prokaryotic LEA proteins than eukaryotic LEA proteins.

Table 3.1: Pfam (version 26.0) families for ArLEA1A and ArLEA1B and their homologs

Species	*Protein Name*	*NCBI Accession*	*HMM accession*	*HMM Name*	*HMM Start*	*HMM End*	*Bit-Score*	*E-value*
Adineta ricciae	lea-1A	gi|156254742|gb|ABU62811.1|	PB006621	Pfam-B_6621	94	184	22.7	7.20E-05
		gi|156254742|gb|ABU62811.1|	PB006097	Pfam-B_6097	74	274	33	4.30E-08
		gi|156254742|gb|ABU62811.1|	PB006621	Pfam-B_6621	125	184	22.7	7.40E-05
		gi|156254742|gb|ABU62811.1|	PB008237	Pfam-B_8237	281	455	23.8	1.80E-05
Adineta ricciae	lea-1B	gi|156254740|gb|ABU62810.1|	PB006097	Pfam-B_6097	13	263	30.7	2.10E-07
		gi|156254740|gb|ABU62810.1|	PB006621	Pfam-B_6621	94	184	21.2	0.0002
		gi|156254740|gb|ABU62810.1|	PB006621	Pfam-B_6621	125	216	21.8	0.0001
		gi|156254740|gb|ABU62810.1|	PB008237	Pfam-B_8237	281	455	23.6	1.90E-05
Adineta vaga	lea-1B	gi|291277997|gb|ADD91471.1|	PB008237	Pfam-B_8237	239	459	18.1	0.0009
		gi|291277997|gb|ADD91471.1|	PB006621	Pfam-B_6621	94	184	20.7	0.0003
		gi|291277997|gb|ADD91471.1|	PB006097	Pfam-B_6097	62	265	27	2.70E-06
		gi|291277997|gb|ADD91471.1|	PB006621	Pfam-B_6621	104	224	21.4	0.0002
Adineta vaga	lea-1B'	gi|291278006|gb|ADD91479.1|	PB006621	Pfam-B_6621	94	176	19.4	0.0008
		gi|291278006|gb|ADD91479.1|	PB006097	Pfam-B_6097	60	265	27.9	1.50E-06
		gi|291278006|gb|ADD91479.1|	PB006621	Pfam-B_6621	104	224	24	3.00E-05
Adineta vaga	lea-1C	gi|291277985|gb|ADD91460.1|	PB002305	Pfam-B_2305	28	115	29.4	6.90E-07
		gi|291277985|gb|ADD91460.1|	PB006621	Pfam-B_6621	114	182	19.4	0.0007
Caenorhabditis elegans	LEA-1 isoform f	gi|332078354|emb|CCA65610.1|	PB006621	Pfam-B_6621	120	214	28.2	1.60E-06
		gi|332078354|emb|CCA65610.1|	PB006097	Pfam-B_6097	68	239	21.3	0.0002
		gi|332078354|emb|CCA65610.1|	PB006621	Pfam-B_6621	59	200	33.8	3.10E-08
		gi|332078354|emb|CCA65610.1|	PB006420	Pfam-B_6420	39	146	20.4	0.0006
		gi|332078354|emb|CCA65610.1|	PB006621	Pfam-B_6621	63	186	45.6	7.70E-12
		gi|332078354|emb|CCA65610.1|	PF02987.1111	LEA_4	1	38	13.6	0.041
		gi|332078354|emb|CCA65610.1|	PB006621	Pfam-B_6621	107	208	41.3	1.50E-10
		gi|332078354|emb|CCA65610.1|	PF02987.1111	LEA_4	1	39	14.9	0.016
		gi|332078354|emb|CCA65610.1|	PF02987.1111	LEA_4	4	39	21.4	0.0002
		gi|332078354|emb|CCA65610.1|	PF02987.1111	LEA_4	7	39	19.7	0.0005
		gi|332078354|emb|CCA65610.1|	PB006621	Pfam-B_6621	108	219	48.8	7.90E-13
		gi|332078354|emb|CCA65610.1|	PF02987.1111	LEA_4	17	42	14.5	0.021
		gi|332078354|emb|CCA65610.1|	PB006621	Pfam-B_6621	63	223	62	7.30E-17
		gi|332078354|emb|CCA65610.1|	PF02987.11	LEA_4	1	40	14.5	0.022
		gi|332078354|emb|CCA65610.1|	PF02987.11	LEA_4	2	37	15	0.014
		gi|332078354|emb|CCA65610.1|	PB006097	Pfam-B_6097	34	266	23.9	2.50E-05
		gi|332078354|emb|CCA65610.1|	PB006621	Pfam-B_6621	92	225	65	9.50E-18
		gi|332078354|emb|CCA65610.1|	PF02987.11	LEA_4	1	39	18.5	0.0012
		gi|332078354|emb|CCA65610.1|	PF02987.11	LEA_4	2	41	25.6	7.50E-06
		gi|332078354|emb|CCA65610.1|	PB006621	Pfam-B_6621	120	222	39	7.90E-10
		gi|332078354|emb|CCA65610.1|	PB006621	Pfam-B_6621	53	224	57.6	1.70E-15
		gi|332078354|emb|CCA65610.1|	PB006097	Pfam-B_6097	70	265	19	0.0008
		gi|332078354|emb|CCA65610.1|	PB006621	Pfam-B_6621	58	218	45.4	8.50E-12
		gi|332078354|emb|CCA65610.1|	PF02987.11	LEA_4	2	42	12.4	0.097
		gi|332078354|emb|CCA65610.1|	PB006621	Pfam-B_6621	70	233	33.6	3.40E-08
		gi|332078354|emb|CCA65610.1|	PF02987.11	LEA_4	2	43	18.2	0.0015
		gi|332078354|emb|CCA65610.1|	PB002940	Pfam-B_2940	263	443 1	16.7	0.001

(Continued on next page)

Table 3.1: Pfam (version 26.0) families for ArLEA1A and ArLEA1B and their homologs (continued)

Species	*Protein Name*	*NCBI Accession*	*HMM*	*HMM Name*	*HMM*	*HMM*	*Bit-*	*E-value*
Polypedilum vanderplanki	PvLEA1	gi|90959527|dbj|BAE92616.1|	PF02987.11	LEA_4	13	38	16.3	0.0059
		gi|90959527|dbj|BAE92616.1|	PB000272	Pfam-B_272	696	1214	2 20	0.0001
		gi|90959527|dbj|BAE92616.1|	PF02987.11	LEA_4	1	44	47.9	7.90E-13
		gi|90959527|dbj|BAE92616.1|	PB006621	Pfam-B_6621	104	222	43.7	3.00E-11
		gi|90959527|dbj|BAE92616.1|	PF02987.11	LEA_4	3	39	22.9	5.10E-05
		gi|90959527|dbj|BAE92616.1|	PF02987.11	LEA_4	4	44	38.6	6.40E-10
		gi|90959527|dbj|BAE92616.1|	PF02987.11	LEA_4	2	44	35.7	5.00E-09
		gi|90959527|dbj|BAE92616.1|	PB006621	Pfam-B_6621	71	221	42.1	9.10E-11
		gi|90959527|dbj|BAE92616.1|	PF02987.11	LEA_4	4	44	34.2	1.50E-08
		gi|90959527|dbj|BAE92616.1|	PF02987.11	LEA_4	1	44	48	7.20E-13
		gi|90959527|dbj|BAE92616.1|	PF02987.11	LEA_4	1	43	22.9	5.00E-05
		gi|90959527|dbj|BAE92616.1|	PB006621	Pfam-B_6621	58	232	51.4	1.30E-13
		gi|90959527|dbj|BAE92616.1|	PF02987.11	LEA_4	1	43	23.4	3.50E-05
		gi|90959527|dbj|BAE92616.1|	PF02987.11	LEA_4	2	44	34.6	1.10E-08
		gi|90959527|dbj|BAE92616.1|	PF02987.11	LEA_4	1	39	28.2	1.10E-06
		gi|90959527|dbj|BAE92616.1|	PF02987.11	LEA_4	15	42	13.5	0.043
		gi|90959527|dbj|BAE92616.1|	PF02987.11	LEA_4	1	44	17.6	0.0023
Arabidopsis thaliana	LEA (Group 3)	gi|1526424|dbj|BAA11017.1|	PB006621	Pfam-B_6621	58	188	22.9	6.60E-05
		gi|1526424|dbj|BAA11017.1|	PF02987.11	LEA_4	4	29	25.1	1.00E-05
		gi|1526424|dbj|BAA11017.1|	PF02987.11	LEA_4	1	44	48.9	3.80E-13
		gi|1526424|dbj|BAA11017.1|	PB006621	Pfam-B_6621	104	206	35.2	1.10E-08
		gi|1526424|dbj|BAA11017.1|	PF02987.11	LEA_4	1	44	56.2	2.00E-15
		gi|1526424|dbj|BAA11017.1|	PF02987.11	LEA_4	2	44	54.5	6.80E-15
		gi|1526424|dbj|BAA11017.1|	PF02987.11	LEA_4	2	44	63.2	1.30E-17
		gi|1526424|dbj|BAA11017.1|	PB006621	Pfam-B_6621	102	197	32.1	1.00E-07
		gi|1526424|dbj|BAA11017.1|	PF02987.11	LEA_4	3	43	28.7	7.60E-07
Clostridium sp. D5	LEA-76	gi|325261876|ref|ZP_08128614.1	PB001499	Pfam-B_1499	71	275	21.3	0.0002
		gi|325261876|ref|ZP_08128614.1	PB000693	Pfam-B_693	325	516	19.6	0.0002
		gi|325261876|ref|ZP_08128614.1	PF13248.1	zf-ribbon_3	3	25	24.9	7.10E-06
		gi|325261876|ref|ZP_08128614.1	PB006621	Pfam-B_6621	84	220	34.5	1.90E-08
		gi|325261876|ref|ZP_08128614.1	PB006097	Pfam-B_6097	73	248	39.4	4.80E-10
		gi|325261876|ref|ZP_08128614.1	PB006621	Pfam-B_6621	114	189	25.6	9.70E-06
Aphelencus avenae	AavLEA1	gi|16417737|ref|AAL18843.1	PF02987.11	LEA_4	2	42	22.6	6.10E-05
		gi|16417737|ref|AAL18843.1	PF02987.11	LEA_4	14	43	19.6	0.0005
		gi|16417737|ref|AAL18843.1	PF02987.11	LEA_4	1	43	29.4	4.60E-07
		gi|16417737|ref|AAL18843.1	PB006621	Pfam-B_6621	148	190	20.1	0.0005
		gi|16417737|ref|AAL18843.1	PB006621	Pfam-B_6621	117	198	31.5	1.50E-07
		gi|16417737|ref|AAL18843.1	PB002305	Pfam-B_2305	35	109	21.6	0.0002

3.2.3 Predicted three-dimensional structure of ArLEA1A and ArLEA1B

ArLEA1A and ArLEA1B protein structures were modeled using I-TASSER (Roy et al., 2010; Zhang, 2008). The reliability of predicted models can be assessed by the C-score. The C-score is a function of relative clustering density of models and is found to correlate with the reliability of the predicted models. Values greater than -1.5 are considered to be significant. I-TASSER predicts α-helical models for ArLEA1A (Fig. 3.4). The C-scores for models 1,2 and 3 are -1.08, -1.08 and -1.11, while models 4 and 5 show a relatively lower C-score of -2.19 and -2.42 respectively making the latter less reliable.

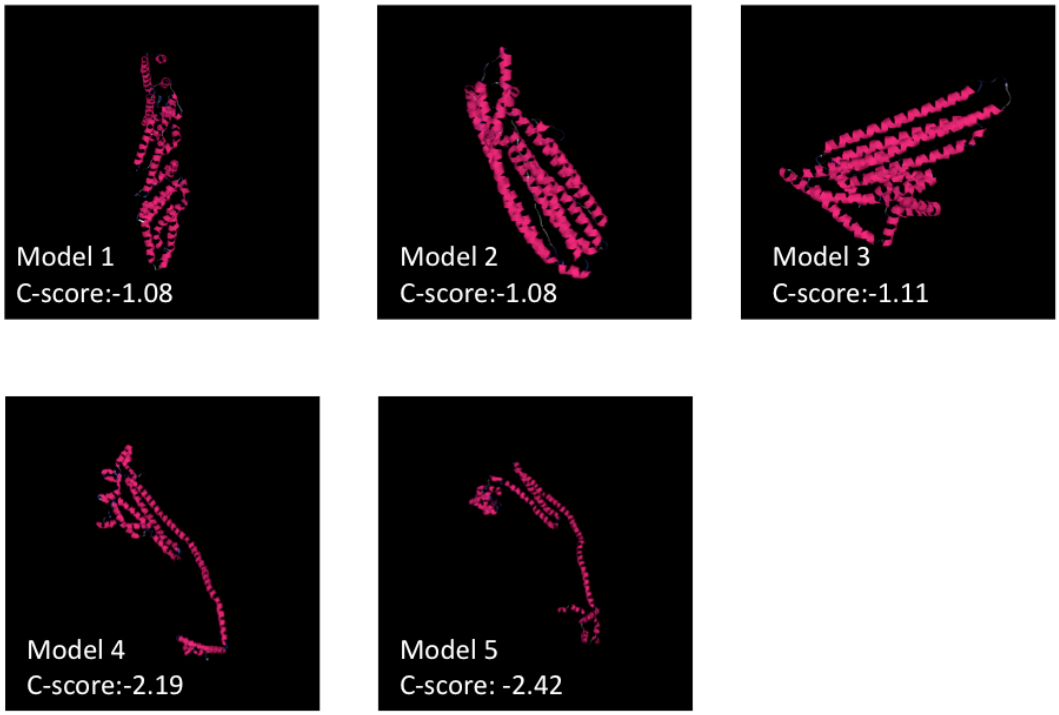


Figure 3.4: I-TASSER models of ArLEA1A.

For ArLEA1B (Fig. 3.5), besides observing α-helical structures (models 1,3,4 and 5), one also observes the formation of a possible filamentous structure resembling tropomyosin filaments (model 2, magnified in Fig. 3.6). Tropomyosin is a 40 nm-long coiled-coil protein that polymerises to form a continuous filament and

associates with actin in muscle and numerous non-muscle cells (Whitby et al., 1992; Whitby and Phillips, 2000). It is known that that ArLEA1B is prone to aggregation in solution (Pouchkina-Stantcheva et al., 2007). Hence, one might speculate that it polymerises in a manner analogous to tropomyosin filaments during desiccation. This polymerisation might be essential for its function in protecting membranes and possibly other proteins. The C-score for model 2 (-1.16) is within the range of significance. However, of the remaining predicted models of ArLEA1B, only models 1 and 3 seem to be reliable by virtue of their C-scores (-1.16) while models 4 and 5 have C-scores less than -1.5, making the latter less significant. In order to address why ArLEA1A and ArLEA1B might be structurally different, a 44 amino acid deletion corresponding to ArLEA1B was made in ArLEA1A and the sequence was deposited in I-TASSER. Interestingly, a tropomyosin like structure was retrieved for ArLEA1A (Fig. 3.7, model 2 magnified in Fig. 3.8) suggesting that this deletion might be responsible for the observed differences in these two proteins that are hypothesised to have originated from a common allele. I-TASSER predictions for AavLEA1 failed to yield statistically significant models, as their C-scores fell below -1.5 (Fig. 3.9). This could be attributed to the unstructured nature of AavLEA1 that might fail to form any discrete folds.

Figure 3.5: I-TASSER models of ArLEA1B.

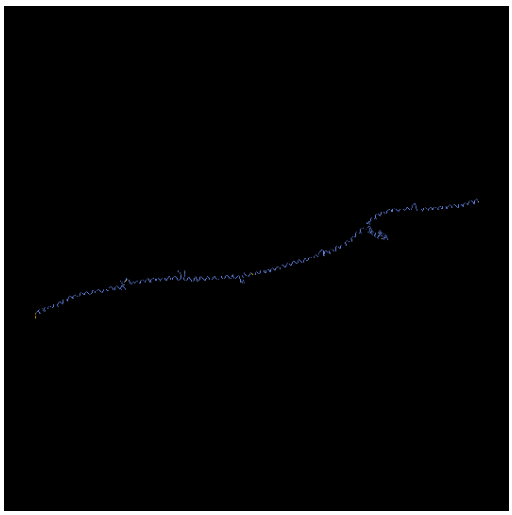


Figure 3.6: Magnified I-TASSER model 2 of ArLEA1B.

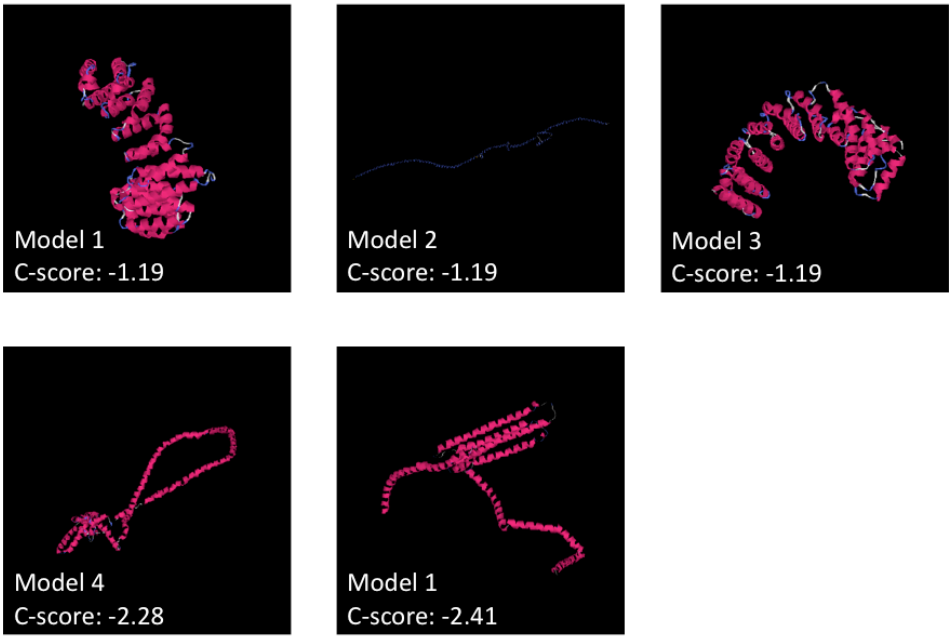


Figure 3.7: I-TASSER models of ArLEA1A with 44 amino acid deletion.

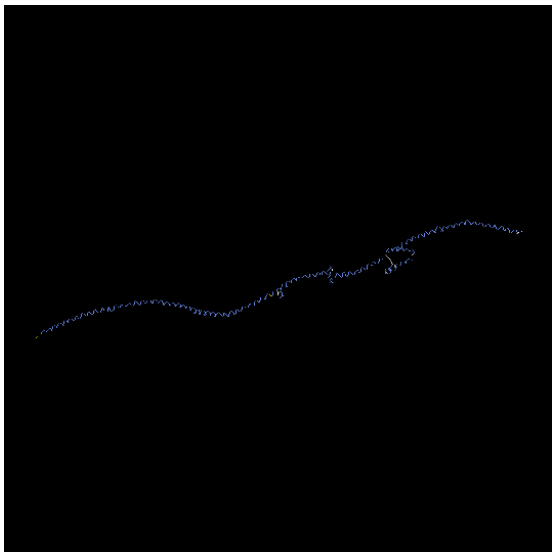


Figure 3.8: Magnified I-TASSER model 2 of ArLEA1A with 44 amino acid deletion.

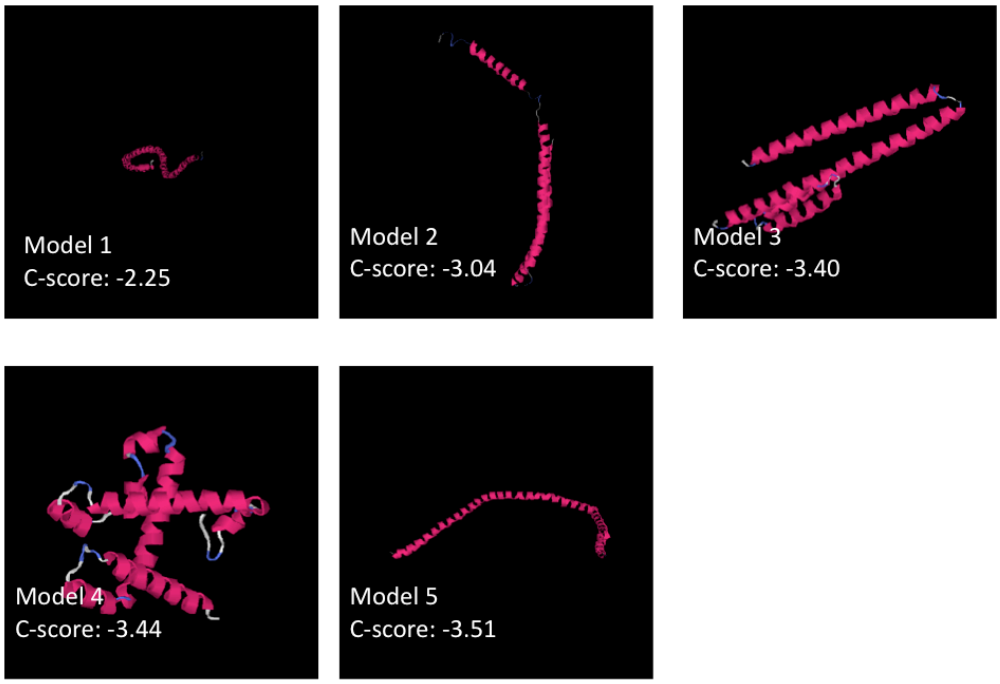
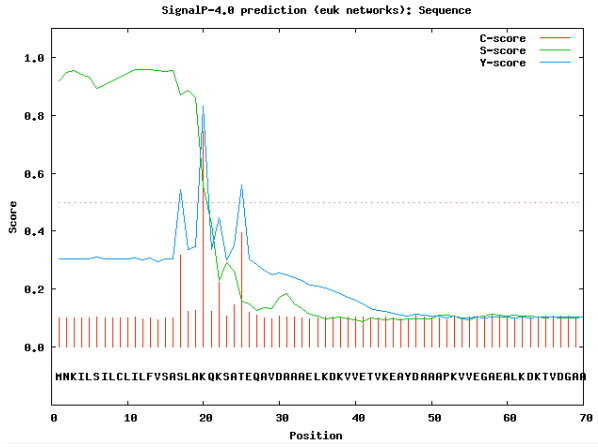


Figure 3.9: I-TASSER models for AavLEA1.

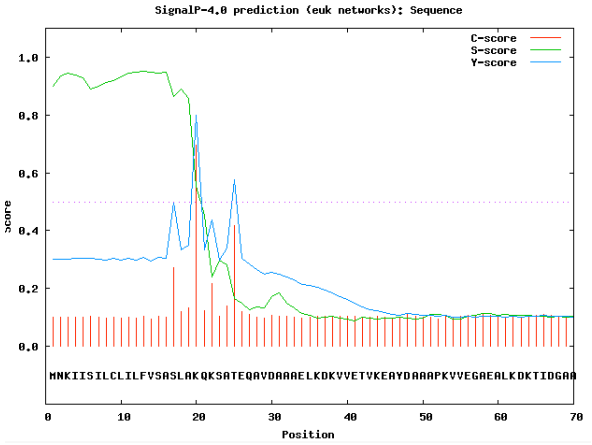
3.2.4. Predicting cellular localisation signals in ArLEA1A and ArLEA1B

In order to accurately confirm the identity of the ER localisation signal in ArLEA1A and ArLEA1B proteins, the SignalP 4.0 program was applied (Petersen et al., 2011). The most significant signal peptidase cleavage site is predicted between the 19th and 20th amino acid positions in both ArLEA1A and ArLEA1B (Fig. 3.10a and 3.10b) with a maximum S-score (signal sequence prediction score) of 0.95 and a C-score (cleavage site prediction score) of 0.695 due to sequence identity in their N-terminal region. In contrast, AavLEA1, derived from anhydrobiotic nematode *A. avenae*, lacks any N-terminal signal sequences or cleavage site, as evidenced by low S- and C-scores (Fig. 3.10c). Using TargetP both ArLEA1A and ArLEA1B show high scores (0.95) for localisation in the secretory pathway and correspondingly low scores for mitochondrial targeting (0.029) (Table 3.2). AavLEA1 on the other hand gives a very low score for localisation in the secretory pathway and mitochondria (0.043 and 0.156, respectively). Hence both SignalP and TargetP indicate a high probability of translocation of ArLEA1A and ArLEA1B into the ER via N-terminal signal sequences.

a.



b.



c.

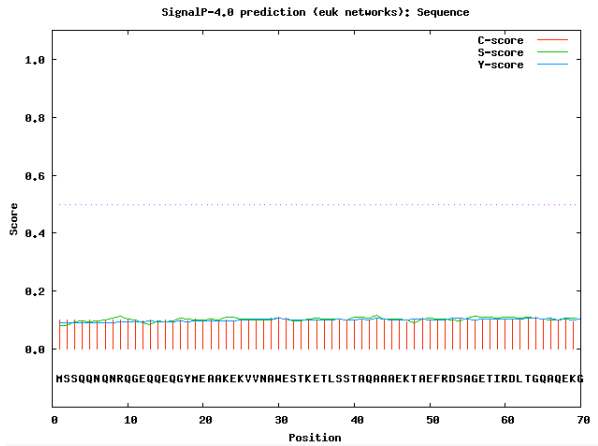


Figure 3.10: SignalP 4.0 predictions for (a) ArLEA1A, (b) ArLEA1B and (c) AavLEA1. C-score (red) gives the "cleavage site" score, S-score (green) gives the score for predicting the signal peptide sequence and Y-score (blue) is a derivative of the C-score and S-score combined. Both ArLEA1A and ArLEA1B are predicted to contain the N-terminal signal sequence and undergo cleavage between 19th and 20th amino acid positions. On the contrary, AavLEA1 lacks any signal sequence or cleavage site. The D-scores (weighted average of S-mean and Y-max score) for ArLEA1A, ArLEA1B and AavLEA1 are 0.883, 0.865 and 0.105 respectively. The dotted line depicts the D-cutoff value of 0.45.

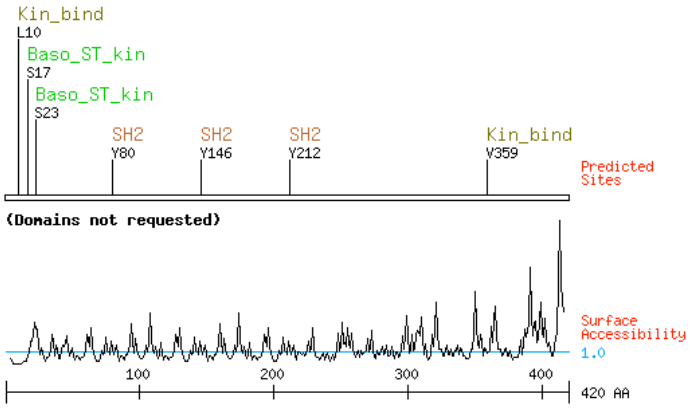
Table 3.2: TargetP predictions for sub-cellular localisation			
	mTP	SP	Other
ArLEA1A	0.029	0.951	0.049
ArLEA1B	0.029	0.951	0.05
AavLEA1	0.156	0.043	0.879

mTP: Mitochondria, SP: Secretory Pathway

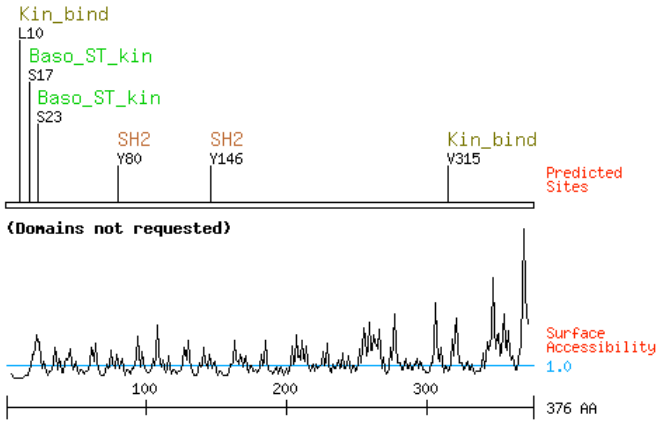
3.2.5 Protein-protein interaction sites in ArLEA1A and ArLEA1B

In order to identify protein-protein interaction motifs in ArLEA1A and ArLEA1B, their sequences were deposited in Scansite (http://scansite.mit.edu), which is an online computational tool built on experimental binding and/or substrate information from oriented peptide library screening and phage display experiments (Obenauer et al., 2003). It consists of a weight matrix-based scoring algorithm that predicts protein–protein interactions and sites of phosphorylation. Although this program has been built using a database of vertebrate sequences, the high level of evolutionary conservation observed between vertebrate and invertebrate signalling mechanisms (Lomberk et al., 2010; Manning et al., 2003; Stout et al., 2010) makes this tool a viable option for detecting possible interacting sites in bdelloid protein sequences. This program can be used at high (top 0.2%), medium (top 1%) and low (top 5%) stringency levels to identify binding sites. For identifying interacting motifs in ArLEA1A and ArLEA1B, it was decided to rely on high and medium stringency level predictions. As can be observed in Fig. 3.11, both ArLEA1A and ArLEA1B sequences are enriched for various kinase binding serines and SH2 (Src homology 2) binding tyrosine residues. However, AavLEA1 (Fig. 3.11c) fails to yield any Scansite predictions at either high or medium levels of stringency. These results suggest that bdelloid LEA proteins might interface with the signalling machinery inside cells during anhydrobiosis.

a.



b.



c.

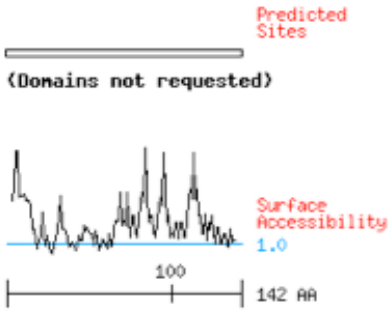


Figure 3.11: Scansite predictions. Protein-protein interaction motifs corresponding to kinase binding, basophilic serine-threonine kinase binding and SH2 binding domains are observed in (a) ArLEA1A and (b) ArLEA1B (at high and medium stringency levels). (c) AavLEA1 was used as a control and does not show any predicted interacting motifs.

3.3 Discussion

The fact that bdelloid rotifers do not rely on the production of non-reducing sugars like trehalose for survival in the desiccated state, has led to the development of an alternative hypothesis for the involvement of other molecules, mainly proteins, in protecting these animals during anhydrobiosis (Tunnacliffe and Lapinski, 2003). It is therefore hardly surprising that bdelloid rotifers express ArLEA1A and ArLEA1B (Pouchkina-Stantcheva et al., 2007), group 3 LEA proteins of the type thought to confer desiccation tolerance in a variety of other anhydrobiotic organisms. Although, considerable effort has been invested in understanding the functions of these proteins (Pouchkina-Stantcheva et al., 2007), there is a huge gap in the literature regarding their evolutionary, structural and functional relationships. To begin to fill this gap, a comprehensive bioinformatics analysis was performed.

The phylogenetic analysis of ArLEA1A and ArLEA1B protein sequences reveals that although these proteins are found to be homologous to a number of LEA proteins from different organisms like *C. elegans*, *A. thaliana* and *Clostridium* species, these proteins along with their closest homologues in *Ad. vaga* have evolved independently as a unique evolutionary clade. These results also support recent findings that have reported the unique evolution of bdelloid rotifers as an independent asexual clade through morphological and genetic analysis (Fontaneto et al., 2007).

Furthermore, the phylogeny of ArLEA1A and ArLEA1B reveals novel insights regarding their possible functions as glass forming agents and membrane anchors during anhydrobiosis. Verifications of these functions by experiments might provide important insights into their roles during anhydrobiosis.

It has been recently shown that bdelloid rotifers have acquired many foreign genes of bacterial, plant and fungal origin through this mechanism (Gladyshev et al., 2008). It is hypothesised that this gene transfer is facilitated by membrane disruption and DNA fragmentation and repair occurring during anhydrobiosis and is presumably an adaptive response to desiccation in these animals. However, from the

existing phylogenetic data it is difficult to provide conclusive evidence for HGT of LEA genes.

Pfam classification of ArLEA1A and ArLEA1B highlights the similarities and differences in the domain compositions of bdelloid LEA proteins and their homologs and provides further evidence for the independent evolution of these proteins. The discovery of the Pfam-B domain PB00621 in LEA proteins raises the possibility that these proteins might play a role in cell signalling mechanisms.

Recently, Boschetti et al. have reported that bdelloid LEA proteins are unusual compared with group 3 LEA proteins found in other species since they possess a greater degree of order and folding as evident by their high GRAVY and FoldIndex unfoldability scores (Boschetti et al., 2011). I-TASSER further corroborates these findings since it predicts the presence of higher order structure of ArLEA1A and ArLEA1B consisting of mainly α-helices. The fact that I-TASSER fails to yield any statistically significant results for AavLEA1 points towards the reliability of this program in distinguishing between unstructured and structured proteins. Interestingly, ArLEA1B is predicted to polymerise into a filamentous tropomyosin like structure. This prediction could explain the tendency of ArLEA1B to aggregate in solution (Pouchkina-Stantcheva et al., 2007). Deleting 44 amino acids in ArLEA1A corresponding to ArLEA1B, also results in the formation of a filamentous tropomyosin-like structure. These results highlight the key differences between ArLEA1A and ArLEA1B that contribute towards functional divergence of these proteins. Verifying the structures of ArLEA1A and ArLEA1B proteins by NMR or X-ray diffraction, might provide important insights regarding their mode of action during anhydrobiosis. For example, if ArLEA1B polymerises into a filamentous structure as predicted, it could provide structural integrity to membrane-enclosed organelles like the ER by forming a supportive scaffolding.

Computational predictions for intra-cellular localisation of ArLEA1A and ArLEA1B indicate that these proteins might be targeted to the ER via their N-terminal signal sequence. Additionally, a putative C-terminal ER retention signal is present that might retain these proteins within the ER (Pouchkina-Stantcheva et al., 2007). In

Chapter four, experimental evidence for the ER localisation of these proteins and the functions of their N-terminal and C-terminal signal sequences is provided.

Scansite further predicts the presence of numerous kinase binding and SH2 binding domains in these protein sequences. Overall, these findings suggest that bdelloid LEA proteins might participate in signal transduction mechanisms during anhydrobiosis, presumably within the ER environment to counter ER-stress (Kaufman et al., 2002; Ron, 2002). These predictions would have to be verified by screening for protein-protein interactions using ArLEA1A or ArLEA1B as 'bait' proteins and identifying potential interacting partners as 'prey' using two-hybrid assays. Alternatively tandem-affinity purification, followed by mass spectrometry could also be performed.

Chapter Four: Trafficking of Bdelloid LEA Proteins

4.1 Introduction

Previous work involving the study of ArLEA1A and ArLEA1B proteins during desiccation has looked at their activity in simple homogenous mixtures outside living cells (McGee, 2006; Pouchkina-Stantcheva et al., 2007). However, our knowledge of their functions in the context of the cellular environment remains obscure. Analysis of LEA protein distribution by immunohistochemistry and confocal microscopy in bdelloids demonstrated that LEA protein staining was absent from nuclei, and present in the cytoplasmic space, but it was difficult to resolve the signal further (McGee, 2006). Currently, it remains unclear whether ArLEA1A and ArLEA1B might act in a specific subcellular location like LEAM, which localises to plant mitochondria (Tolleter et al., 2007), or whether they might be more widely distributed in cells.

It is known that both ArLEA1A and ArLEA1B possess an N-terminal ER translocation signal (Pouchkina-Stantcheva et al., 2007). These signal sequences allow translated proteins to attach to signal-recognition particles (SRP) which enable the recruitment of translated peptides along with ribosomes to the ER membranes by binding to SRP receptors (Gilmore et al., 1982; Walter and Blobel, 1980, 1982). The crystal structure of the SRP protein shows that it consists of hydrophobic amino acid residues like methionines that allow it to bind to a large variety of hydrophobic sequences (Keenan et al., 1998). Once bound to the ER membrane, the translated peptide sequence is translocated across the membrane in a loop-like manner via a protein translocator called the Sec61 complex (Deshaies and Schekman, 1987, 1989; Goerlich et al., 1992; Rothblatt et al., 1989; Stirling et al., 1992). X-ray crystallography has shown that this pore is gated by a short helix that keeps the pore closed when no translocation occurs (Van den Berg et al., 2004). After the protein has been completely translocated into the ER lumen, the pore closes and the hydrophobic signal sequence is cleaved off by a signal peptidase (Milstein et al.,

1972). Certain chaperone proteins like BiP have been shown to facilitate the release of the cleaved protein into the ER lumen (Brodsky et al., 1995; Vogel et al., 1990). Günter Blobel was awarded the Nobel Prize in Physiology and Medicine in 1999 for his discovery of signal peptides and proposition of the 'signal hypothesis' which states that proteins are delivered to specific intra-cellular compartments via their signal sequences (Blobel and Dobberstein, 1975; Blobel and Sabatini, 1971). Figure 4.1 shows a model to depict the translocation of a protein across the ER membrane.

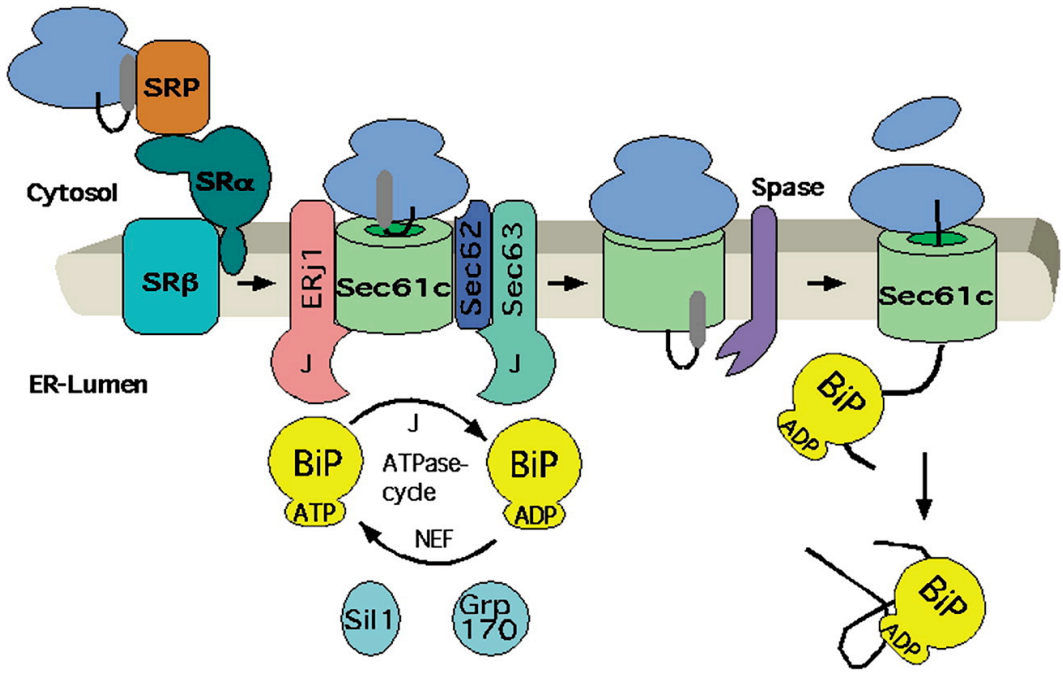


Figure 4.1: Model for co-translational SRP and SR dependent protein transport into the ER. Note that cleavable signal peptides within nascent precursor proteins insert into the Sec61 complex in a loop like fashion that orients the aminoterminus to the cytosol and the carboxyterminus plus the SPC cleavage site to the ER lumen. The signal peptide is subsequently cleaved by the signal peptidase (Spase). The protein is released in the ER lumen with the help of a chaperone protein, BiP. Taken from Zimmermann et al., 2011 with permission.

The signal peptides from various proteins usually consist of positively charged n-region, followed by a hydrophobic h-region and a neutral but polar c-region. The (–3,–1) rule states that the residues at positions -3 and -1 (relative to the cleavage site) must be small and neutral for cleavage to occur correctly (von Heijne, 1983, 1985,1986). The most widely used method for predicting the location of the cleavage site is a weight matrix which was published in 1986 (von Heijne, 1986). More

recently, a neural network based method for signal peptide detection in protein sequences has been devised (Emanuelsson et al., 2007; Petersen et al., 2011).

Once the newly synthesised proteins enter the ER, these proteins are subsequently transported to the Golgi or the cell surface for secretion or are retained within the ER with the help of ER retention signals (Munro and Pelham, 1987). These signals are located at the C-terminal end of proteins and directly bind to COPI coated vesicles to enable retrograde delivery of proteins to the ER that might escape into the Golgi. Such sequences consist of two lysines (KKXX) or KDEL-like amino acid residues. A more detailed analysis of efficiency of different ER retention signals has been recently performed by Raykhel et al. (Raykhel et al., 2008). Receptors that bind to these signals demonstrate high binding affinity in the Golgi compartments and low affinity in the ER, where transported cargo proteins are allowed to unload. This differential affinity might be regulated by pH (Wilson et al., 1993) or other factors that are currently not known. Bdelloid LEA proteins are also predicted to contain a putative ER retention signal ATEL at their C-terminal end (Pouchkina-Stantcheva et al., 2007).

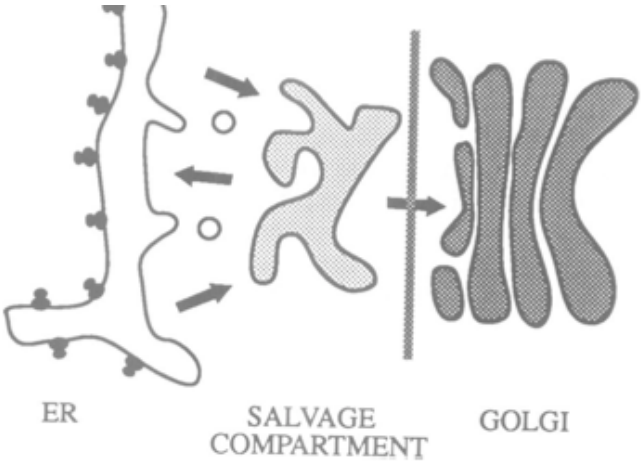


Figure 4.2: Sorting pathway for ER proteins. Both secretory proteins and ER proteins travel together to the salvage compartment lying in between ER and Golgi compartments. ER proteins are returned to the ER via COPI vesicles, while secretory proteins move to the Golgi. Taken from Pelham, 1989, National Library of Medicine.

In this chapter, the intracellular localisation of both full length ArLEA1A and ArLEA1B tagged with the fluorescent protein mCherry (Section 4.2.1), and of

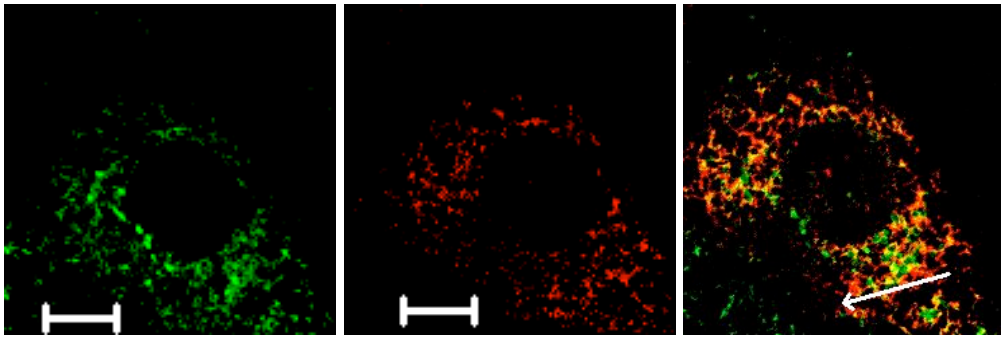
modified versions of both mCherry and EGFP containing bdelloid LEA protein N and C-terminal signal sequences (Sections 4.2.2,4.2.3,4.2.4) has been studied using confocal microscopy. Moreover, the ER-retention efficiency of ATEL has been assessed by measuring the amount of secreted proteins in cell supernatants (Section 4.2.5).

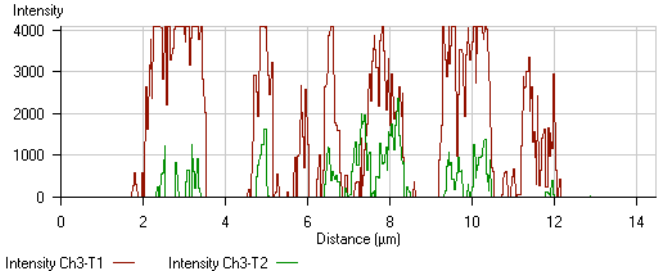
4.2 Results

4.2.1 ArLEA1A and ArLEA1B colocalise with ER-Tracker Green

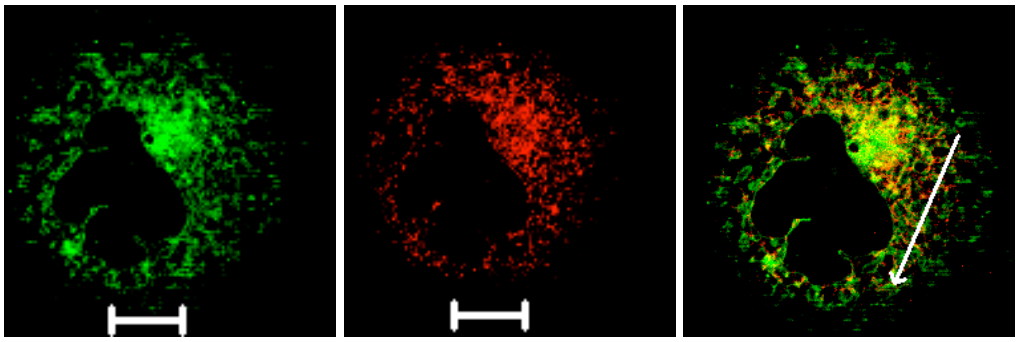
Since rotifer cell cultures were not available, intracellular LEA protein localisation was assessed after expression of tagged proteins in mammalian cells. COS-7 cells were transfected with pcDNA3.0 containing ArLEA1A-FLAG-mCherry, ArLEA1B-FLAG-mCherry or mCherry. Cells expressing ArLEA1A-FLAG-mCherry, ArLEA1B-FLAG-mCherry and mCherry were stained with ER-Tracker Green (BODIPY FL glibenclamide; Invitrogen), a cell permeant, live-cell stain that binds to sulfonylurea receptors of ATP-sensitive K^+ channels in the ER membrane, and colocalisation was assessed by confocal microscopy. Images obtained within the red and green channels were merged and the fluorescent intensity signals were measured in selected regions (Fig. 4.3a,b). For cells containing ArLEA1A-FLAG-mCherry and ArLEA1B-FLAG-mCherry, substantial overlap of the red and green fluorescence signals was observed, indicating colocalisation with ER-Tracker Green. The overlap is not complete, however, possibly due to the delivery of ArLEA1A and ArLEA1B to vesicles and compartments outside the ER. In contrast, mCherry alone is mostly distributed both in the cytoplasm and in the nucleus (Fig. 4.3c). Some mCherry is observed to colocalise with ER-Tracker Green possibly due to its hydrophobic nature that might favour its association with the ER membrane. Higher magnification might enable better resolution of the two compartments.

a. ER-Tracker Green ArLEA1A-FLAG-mCherry Merged





b. ER-Tracker Green ArLEA1B-FLAG-mCherry Merged



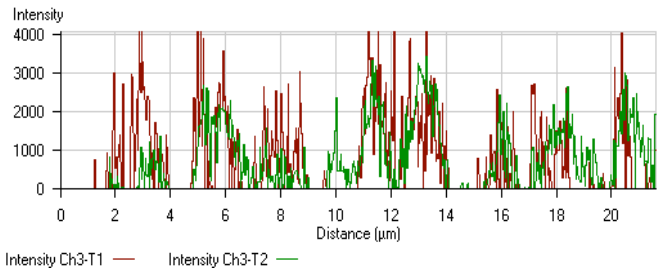
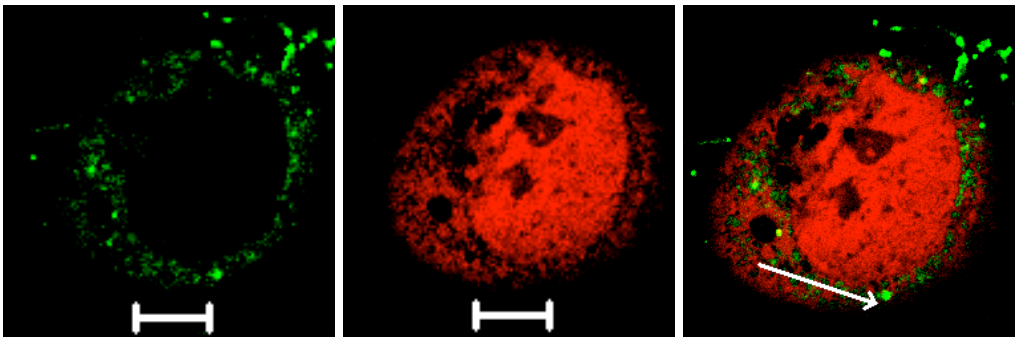


Figure 4.3: ER localisation of ArLEA1A and ArLEA1B. COS-7 cells were transfected with pcDNA3.0 vector containing mCherry tagged ArLEA1A and ArLEA1B. 24 h post-transfection, cells were stained with ER-Tracker Green. Live-cell imaging was performed with LSM510 confocal microscope. Colocalisation is observed between ER-Tracker Green and (a) ArLEA1A-FLAG-mCherry and (b) ArLEA1B-FLAG-mCherry. *Left*, green channel (ER-Tracker Green); *middle*, red channel (mCherry); *right*, red/green merged image. Plots depicting the red and green fluorescent intensities at selected region of interest (arrow) are also shown below. Scale bar: 10 µm.

c. ER-Tracker Green mCherry Merged



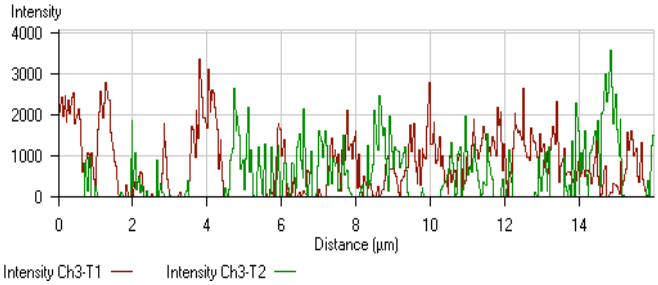
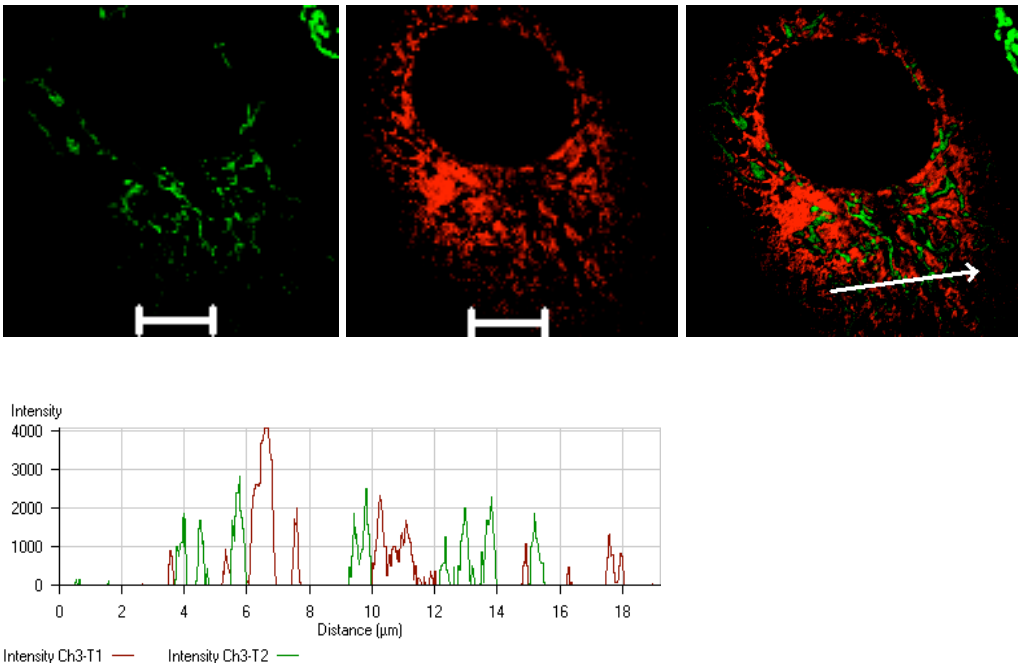


Figure 4.3: (continued) COS-7 cells were transfected with pcDNA3.0 vector containing mCherry. 24 h post transfection, cells were stained with ER-Tracker Green for 30 min. Live-cell imaging was performed with LSM510 confocal microscope. (c) No colocalisation is observed between ER-Tracker Green and mCherry. *Left*, green channel (ER-Tracker Green); *middle*, red channel (mCherry); *right*, red/green merged image. Plots depicting the red and green fluorescent intensities at selected region of interest (arrow) are also shown below. Scale bar: 10 µm.

As a control, ArLEA1A-FLAG-mCherry and ArLEA1B-FLAG-mCherry signal location was compared with MitoTracker Green, a mitochondrial stain (Fig. 4.4a,b). No overlap is observed between the red and green channels indicating absence of ArLEA1A and ArLEA1B in the mitochondria.

a. MitoTracker Green ArLEA1A-FLAG-mCherry Merged



b. MitoTracker Green ArLEA1B-FLAG-mCherry Merged

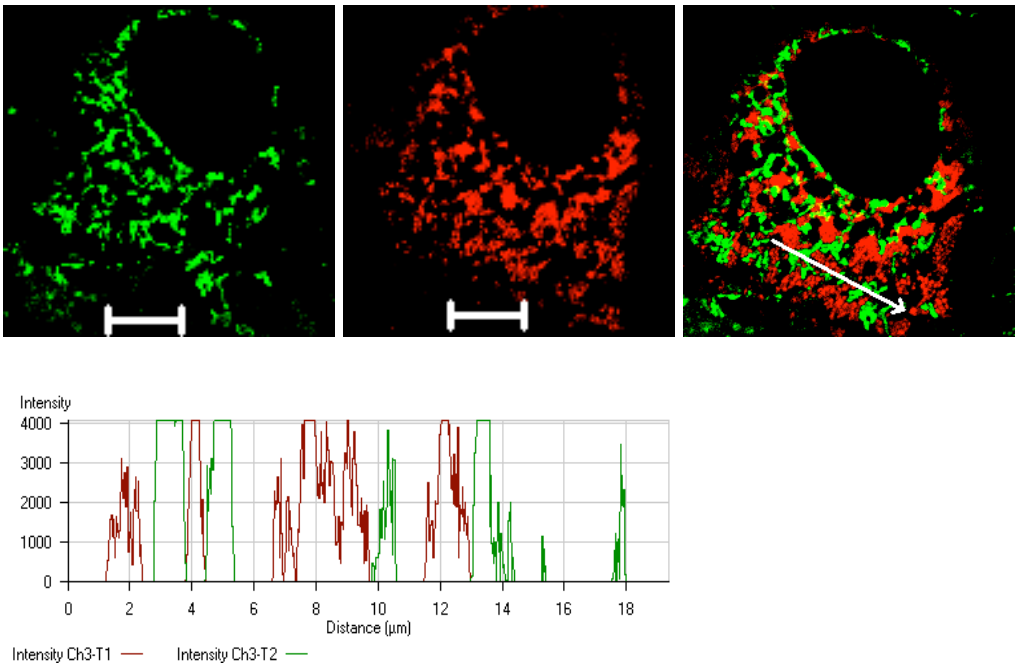


Figure 4.4: MitoTracker Green staining. 24 h post-transfection, cells were stained with MitoTracker Green for 15 min at 37°C. Live-cell imaging was performed with LSM510 confocal microscope. No colocalisation is observed between MitoTracker Green and (a) ArLEA1A-FLAG-mCherry and (b) ArLEA1B-FLAG-mCherry. *Left*, green channel (MitoTracker Green); *middle*, red channel (mCherry); *right*, red/green merged image. Plots depicting the red and green fluorescent intensities at selected region of interest (arrow) are also shown. Scale bar: 10 µm.

In order to confirm expression of ArLEA1A and ArLEA1B in COS-7 cells, western blot analysis was performed (Fig. 4.5) using COS-7 cell lysates. Cells expressing AavLEA1-FLAG-mCherry and untransfected cells were used as negative controls. The primary antibody used was specific for ArLEA1A and was raised in rabbit (McGee, 2006). The secondary antibody was HRP-linked anti-rabbit antibody. As can be observed in Fig. 4.5, the first two lanes correspond to full length ArLEA1A-FLAG-mCherry (72.6kDa) and ArLEA1B-FLAG-mCherry (67.9kDa). However, their sizes are slightly larger than expected indicating possible post-translational modifications of these proteins inside cells, or incomplete binding to SDS resulting in higher observed molecular weight, which is commonly observed for intrinsically disordered proteins (Chakrabortee et al., 2010). Moreover, there seem to be smaller sized bands observed in both lanes. These results suggest that these proteins might be truncated into smaller peptides inside cells.

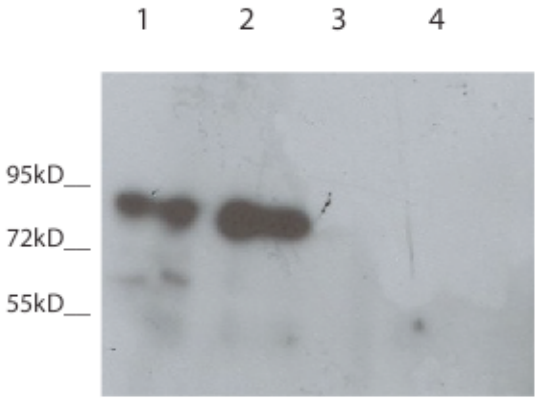


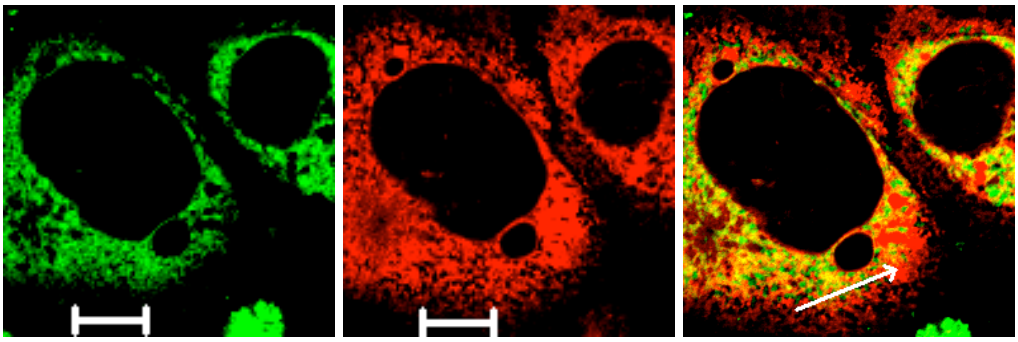
Figure 4.5: Western blot detection of ArLEA1A-FLAG-mCherry, (lane-1) and ArLEA1B-FLAG-mCherry (lane-2) in COS-7 cells. Primary antibody raised in rabbit against ArLEA1A and was used in 1:10000 dilution. HRP-linked-anti-rabbit secondary antibody was used in 1:4000 dilution. Protein extracts from cells expressing AavLEA1-FLAG-mCherry (lane-3) and untransfected cells (lane 4) were used as negative controls. The MW markers are shown on the left.

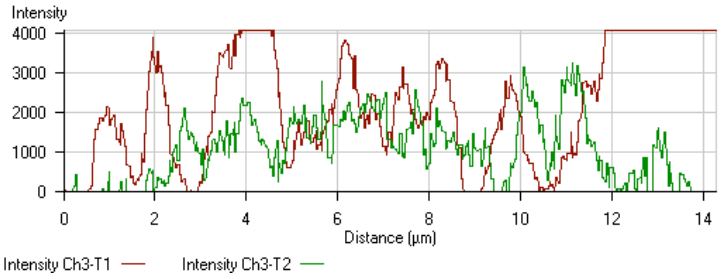
4.2.2 Functional verification of N- and C-terminal signals of ArLEA1A and ArLEA1B

In order to verify the functions of the N- and C-terminal signals of ArLEA1A and ArLEA1B, a new set of plasmids were made where the DNA sequence corresponding to the first 22 amino acids of ArLEA1A (termed 'NS'), which is almost identical to ArLEA1B except for a conserved leucine-> isoleucine amino-acid substitution at the N-terminal end, was cloned in frame with mCherry at the 5' end, and either the rotifer sequence ATEL or the conventional ER retention signal KDEL were engineered at the 3' end. NS-mCherry without a C-terminal signal was also produced as a control. COS-7 cells transfected with either NS-mCherry, NS-mCherry-KDEL or NS-mCherry-ATEL and stained with ER-Tracker Green show that NS-mCherry-KDEL colocalised substantially with ER-Tracker Green (Fig. 4.6a), but either ATEL or no retention signal at the C-terminal end of mCherry resulted in only partial overlap of red and green signals, suggesting the presence of mCherry in compartments other than the ER (Fig. 4.6 b,c).

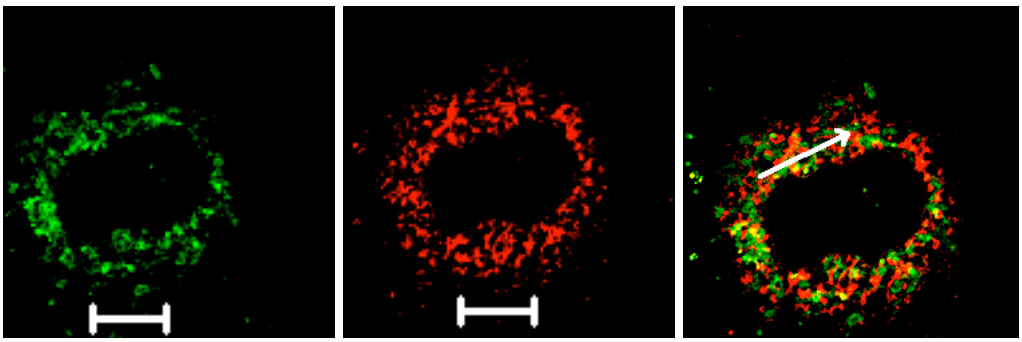
To test whether ATEL mediates differential localisation as compared to KDEL, constructs with EGFP containing N-terminal signal and either KDEL or ATEL at the C-terminal end were made. COS-7 cells were then cotransfected with either NS-mCherry-ATEL or NS-mCherry-KDEL, and either NS-EGFP-ATEL or NS-EGFP-KDEL. As shown in Fig. 4.7a and b, the red and green channel intensities show good overlap when cells are transfected with NS-mCherry-KDEL and NS-EGFP-KDEL or NS-mCherry-ATEL and NS-EGFP-ATEL. However, NS-EGFP-KDEL and NS-mCherry-KDEL seem to only partially colocalise with NS-mCherry-ATEL and NS-EGFP-ATEL, respectively (Fig. 4.7 c,d). Compartments enriched for ATEL-tagged mCherry or EGFP seem to lie close to the nucleus in a pattern reminiscent of the Golgi apparatus.

a. ER-Tracker Green NS-mCherry-KDEL Merged





b. ER-Tracker Green NS-mCherry-ATEL Merged



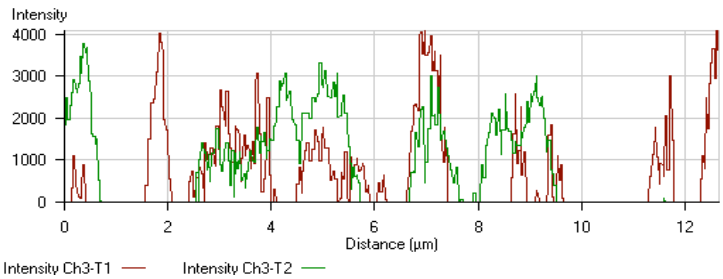
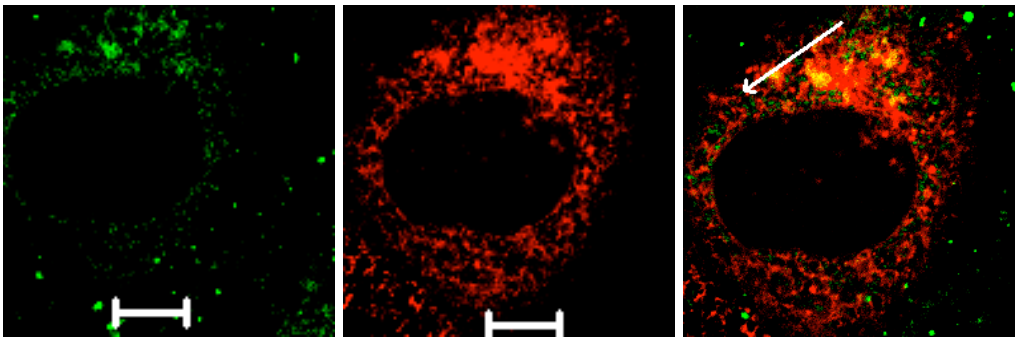


Figure 4.6: N-terminal signal peptide sequence mediates ER localisation. COS-7 cells were transfected with pcDNA3.3 vectors containing (a) NS-mCherry-KDEL and (b) NS-mCherry-ATEL, where NS corresponds to N-terminal sequence of ArLEA1A containing the first 22 amino acids. 24 h post-transfection, cells were stained with ER-Tracker Green. Colocalisation within the ER is observed for NS-mCherry-KDEL and NS-mCherry-ATEL. *Left*, green channel (ER-Tracker Green); *middle*, red channel (mCherry); *right*, red/green merged image. Plots showing red and green fluorescent intensities in the selected region of interest (arrow) are also shown. Scale bar: 10 µm.

c. ER-Tracker Green NS-mCherry Merged



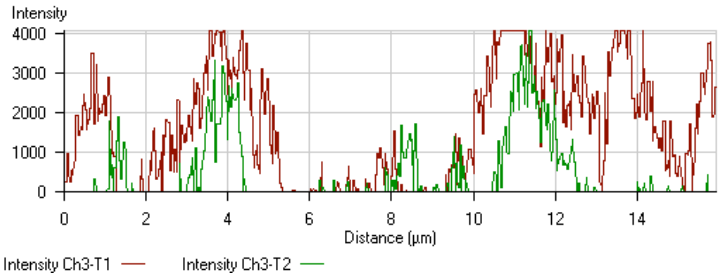


Figure 4.6 (continued). COS-7 cells transfected with pcDNA3.3 containing NS-mCherry were stained with ER-Tracker Green. (c) NS-mCherry is observed to colocalise with ER-Tracker Green. *Left,* green channel (ER-Tracker Green); *middle,* red channel (mCherry); *right,* red/green merged image. Plots showing red and green fluorescent intensities in the selected region of interest (arrow) are also shown. Scale bar: 10 µm.

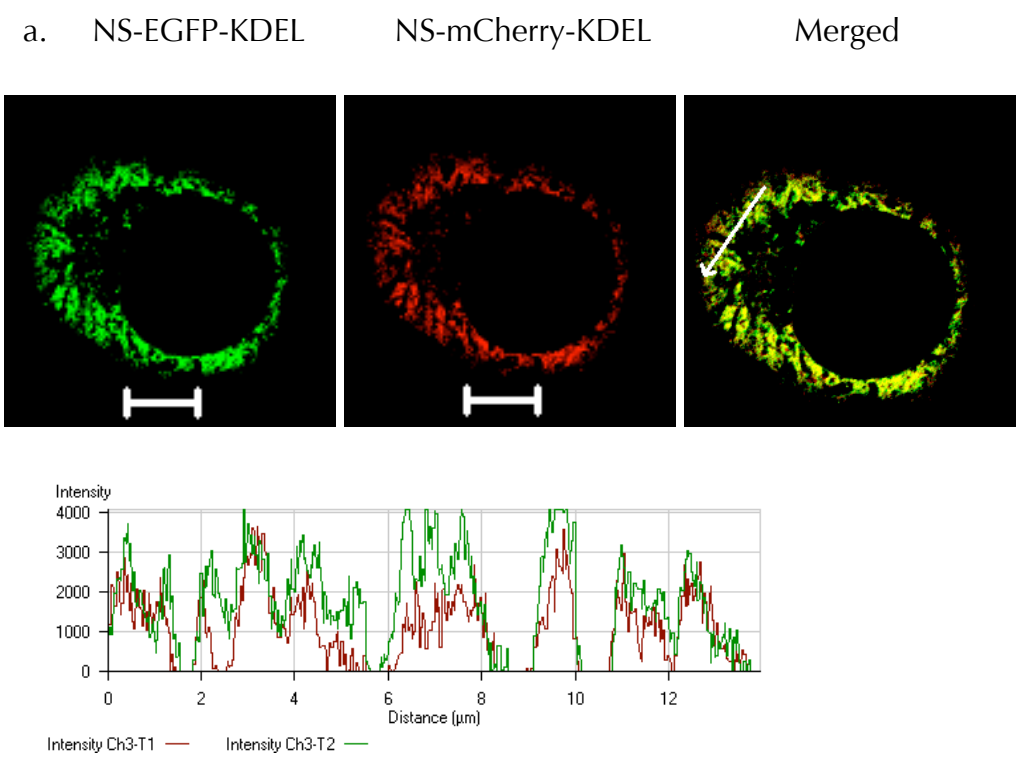
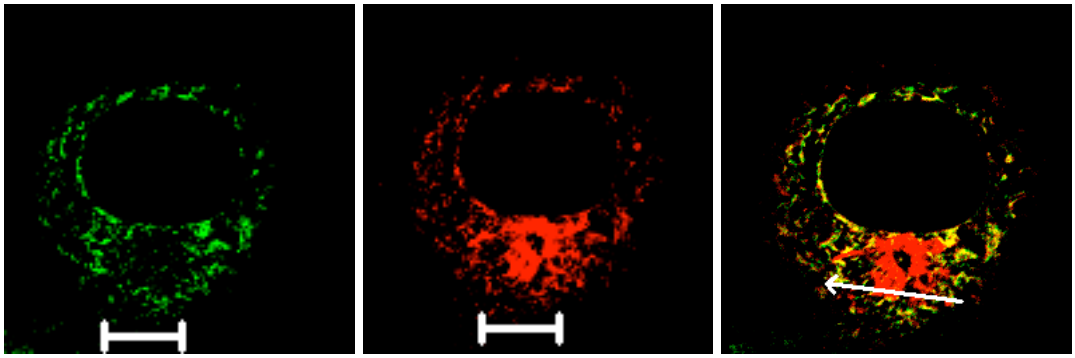
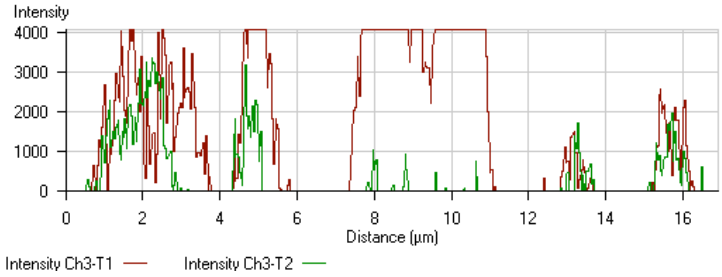


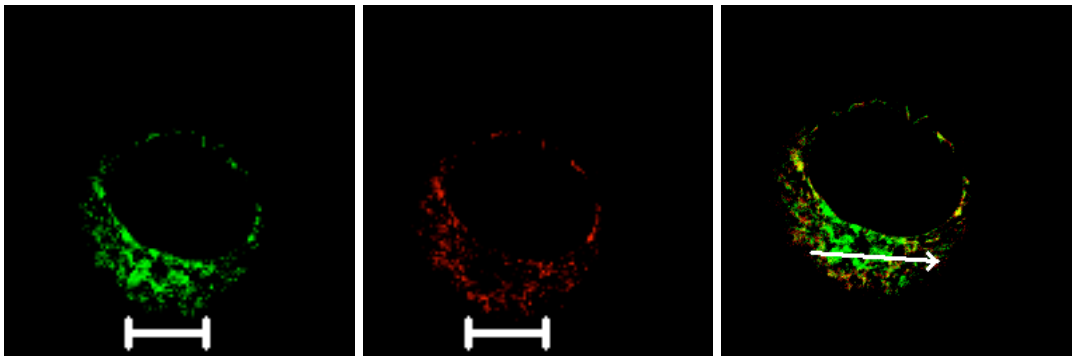
Figure 4.7: Differential localisation mediated by ATEL and KDEL. COS-7 cells were cotransfected with pcDNA3.3 containing the following constructs: (a) NS-mCherry-KDEL and NS-EGFP-KDEL; (b) NS-mCherry-ATEL and NS-EGFP-ATEL. Cells were observed under the confocal microscope, 24 h post-transfection. Similar localisation patterns for red and green channels are observed when NS-EGFP and NS-mCherry are fused with identical C-terminal sequences (a,b). *Left*, green channel (EGFP); *middle*, red channel (mCherry); *right*, red/green merged image. Plots showing red and green fluorescent intensities in the selected region of interest (arrow) are also shown. Scale bar: 10 µm.

c. NS-EGFP-KDEL NS-mCherry-ATEL Merged





d. NS-EGFP-ATEL NS-mCherry-KDEL Merged



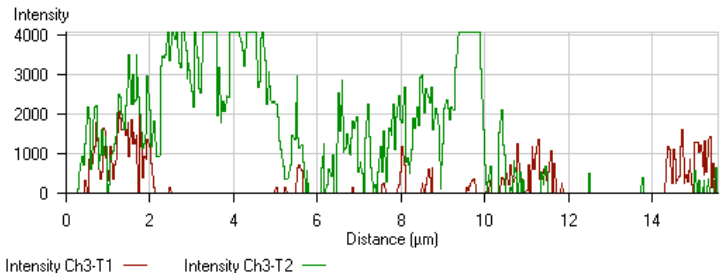
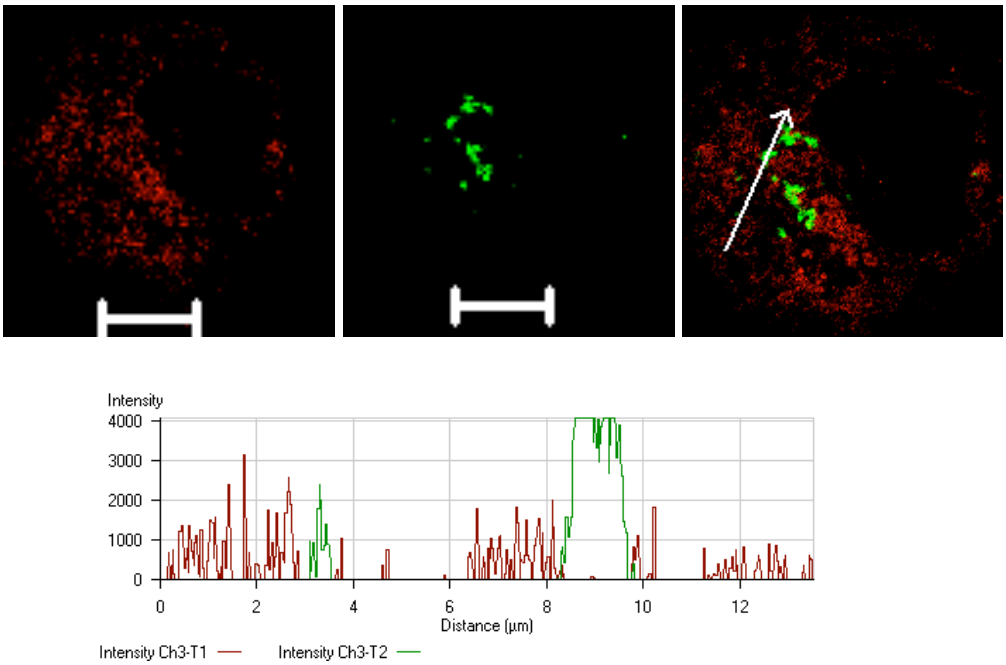


Figure 4.7 (continued): COS-7 cells were cotransfected with pcDNA3.3 containing the following constructs: (c) NS-mCherry-ATEL and NS-EGFP-KDEL and (d) NS-mCherry-KDEL and NS-EGFP-ATEL. Cells were observed under the confocal microscope, 24 h post-transfection. Differential localisation patterns are observed for EGFP and mCherry constructs. *Left*, green channel (EGFP); *middle*, red channel (mCherry); *right*, red/green merged image. Plots showing red and green fluorescent intensities in the selected region of interest (arrow) are also shown. Scale bar: 10 µm.

4.2.3 ATEL, but not KDEL, colocalises with Golgi stacks and vesicles

To confirm the presence of ATEL-tagged proteins in the Golgi, colocalisation analysis of NS-mCherry-KDEL or NS-mCherry-ATEL with GFP-tagged GMAP210 constructs was performed. GMAP210 is a long coiled-coil *cis*-Golgi associated protein that plays a role in maintaining Golgi ribbon integrity and position (Yadav et al., 2009). The N-terminal end of GMAP210 fused to GFP (NterGMAP210-GFP) has been found to concentrate in vesicles lying in the periphery of Golgi elements, while the C-terminal end of GMAP210 fused to GFP (GFP-CterGMAP210) appears to be distributed uniformly in the *cis*-cisternae of the Golgi apparatus (Cardenas et al., 2009). There seems to be poor colocalisation of NS-mCherry-KDEL with either NterGMAP210-GFP or GFP-CterGMAP210 (Fig. 4.8a,b), while NS-mCherry-ATEL is found to overlap partially with both NterGMAP210-GFP and GFP-CterGMAP210 (Fig. 4.8c,d).

a. NS-mCherry-KDEL GFP-CterGMAP210 Merged



b. NS-mCherry-KDEL NterGMAP210-GFP Merged

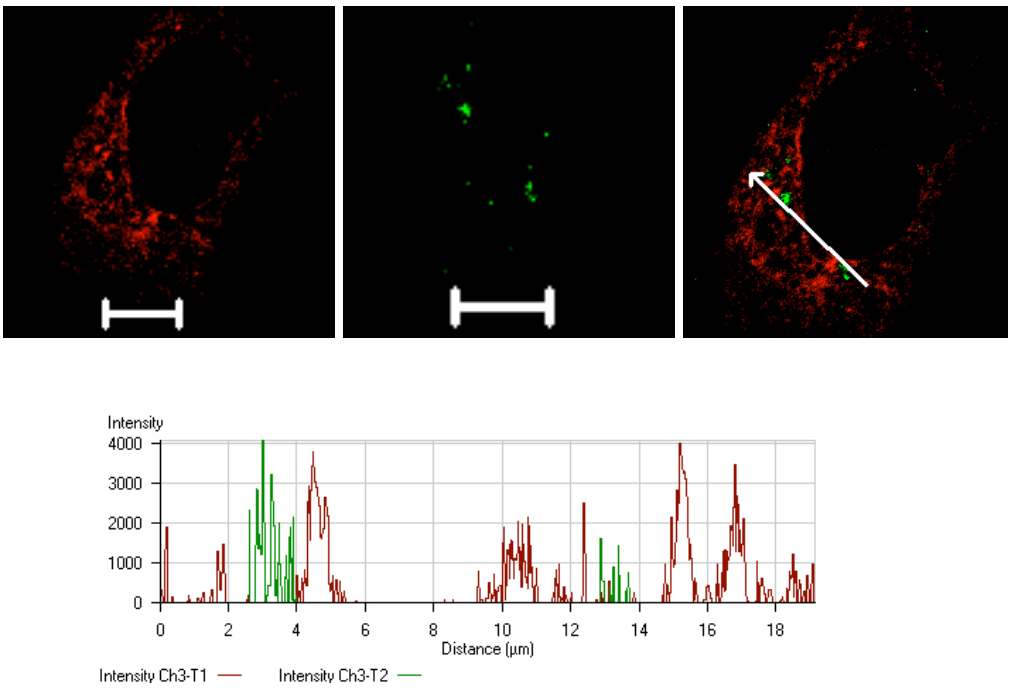
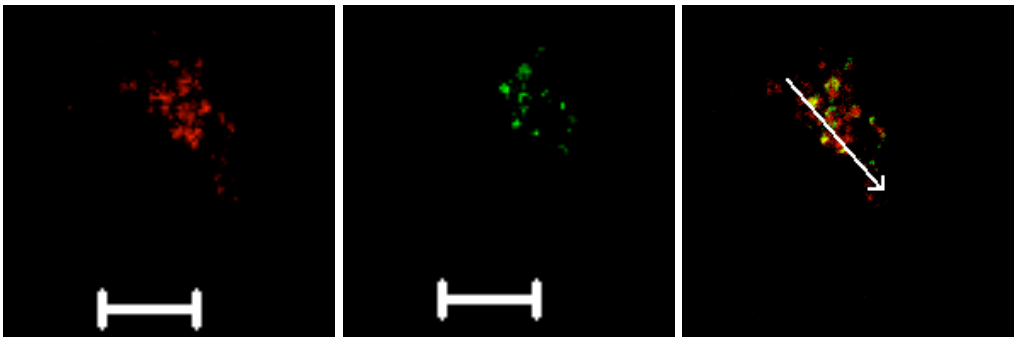
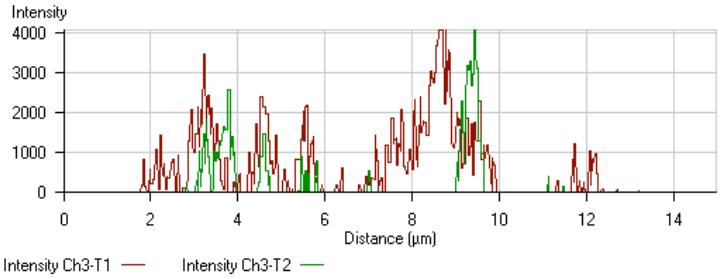


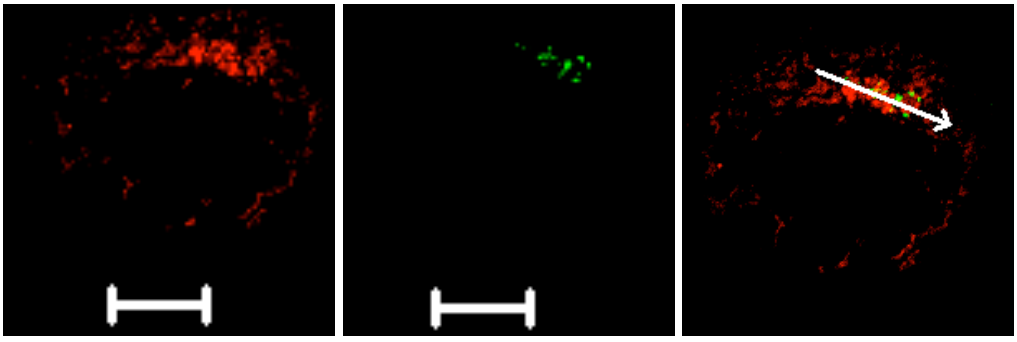
Figure 4.8: Golgi localisation. COS-7 cells were cotransfected with pcDNA3.3 containing NS-mCherry-KDEL and GFP tagged CterGMAP210/NterGMAP210 (Golgi markers). Cells were observed 24 h post-transfection. Scant colocalisation is observed between (a) NS-mCherry-KDEL and GFP-CterGMAP210 and (b) NterGMAP210-GFP. *Left*, red channel (mCherry); *middle*, green channel (GFP); *right*, red/green merged image. Plots showing red and green fluorescent intensities in the selected region of interest (arrow) are also shown. Scale bar: 10 µm.

c. NS-mCherry-ATEL GFP-CterGMAP210 Merged





d. NS-mCherry-ATEL NterGMAP210-GFP Merged



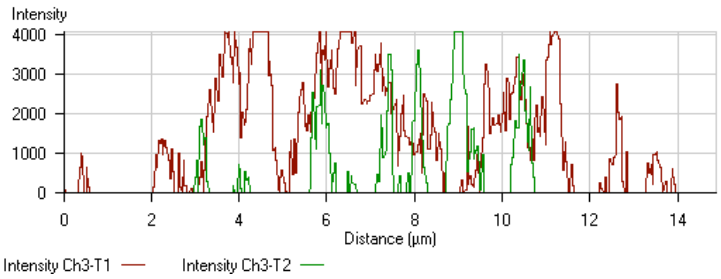


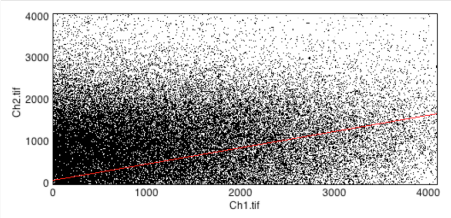
Figure 4.8 (continued): ATEL allows progression into Golgi. COS-7 cells were cotransfected with pcDNA3.3 containing NS-mCherry-ATEL and GFP tagged CterGMAP210/NterGMAP210 (Golgi markers). Cells were observed 24 h post-transfection. Colocalisation is observed between (c) NS-mCherry-ATEL and GFP-CterGMAP210 and (d) NterGMAP210-GFP. *Left*, red channel (mCherry); *middle*, green channel (GFP); *right*, red/green merged image. Plots showing red and green fluorescent intensities in the selected region of interest (arrow) are also shown. Scale bar: 10 µm.

4.2.4 Scatter plots of colocalised pixels

Fig. 4.9 and 4.10 show 2D-scatter plots of colocalised pixel intensities for the red and green channels and the corresponding correlation values for all the confocal images. Both ArLEA1A-FLAG-mCherry and ArLEA1B-FLAG-mCherry show higher Pearson's and Mander's correlation values with ER-Tracker Green (0.37,0.72,0.44 and 0.62,0.90,0.65 respectively) than mCherry (0.14,0.07,0.49) (Fig. 4.9). In contrast, ArLEA1A-FLAG-mCherry and ArLEA1B-FLAG-mCherry show low Pearson's and Mander's correlation values with MitoTracker Green (-0.03,0.02,0.09 and -0.02,0.09,0.13 respectively). NS-mCherry-KDEL shows higher Pearson's and Mander's correlation values with ER-Tracker Green (0.52,0.63,0.98) compared to NS-mCherry-ATEL (0.28,0.5,0.5) and NS-mCherry (0.26,0.28,0.62). The scatter plots also confirm differential localisation mediated by ATEL and KDEL sequences (Fig. 4.10) as evident by the higher Pearson's and Mander's correlation values for NS-mCherry-KDEL and NS-EGFP-KDEL (0.87,0.95,0.91); NS-mCherry-ATEL and NS-EGFP-ATEL (0.76,0.82,0.92); as compared to NS-mCherry-ATEL and NS-EGFP-KDEL (0.50,0.51,0.91) and NS-mCherry-KDEL and NS-EGFP-ATEL (0.43,0.79,0.54). Moreover, KDEL containing proteins are found to be excluded from the Golgi as indicated by the low Pearson's and Mander's correlation values for NS-mCherry-KDEL and GFP-CterGMAP210 (0.08,0.05,0.53) and NS-mCherry-KDEL and NterGMAP210-GFP (0.01,0.00,0.23). In contrast, NS-mCherry-ATEL shows higher correlation values with GFP-CterGMAP210 (0.48,0.33,0.94) and NterGMAP210-GFP (0.19,0.08,0.84) indicating Golgi translocation mediated by ATEL.

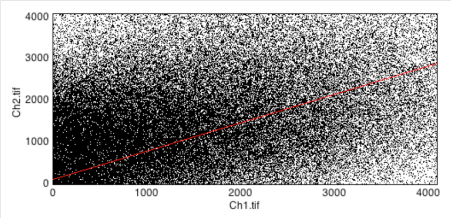
ArLEA1A-FLAG-mCherry (Ch1) vs ER-Tracker Green (Ch2)
(See Fig. 4.3a)

Pearson's correlation coefficient : 0.37, M1: 0.72, M2:0.44



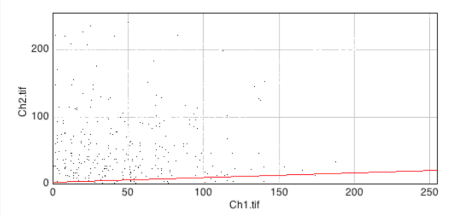
ArLEA1B-FLAG-mCherry (Ch1) vs ER-Tracker Green (Ch2)
(See Fig. 4.3b)

Pearson's correlation coefficient : 0.62, M1: 0.90, M2:0.65



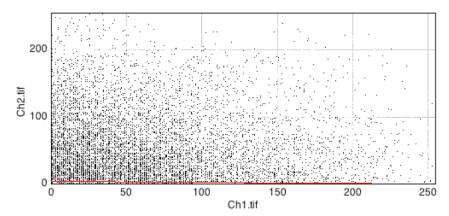
mCherry (Ch1) vs ER-Tracker Green (Ch2)
(See Fig. 4.3c)

Pearson's correlation coefficient : 0.14, M1: 0.07, M2:0.49



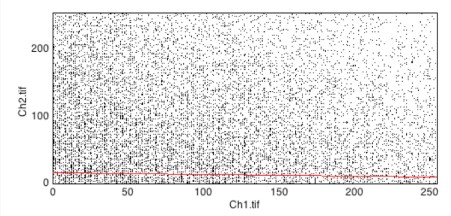
ArLEA1A-FLAG-mCherry(Ch1) vs MitoTracker Green (Ch2)
(See Fig. 4.4a)

Pearson's correlation coefficient :-0.03, M1: 0.02, M2:0.09



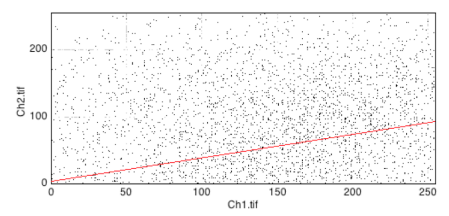
ArLEA1B-FLAG-mCherry (Ch1) vs MitoTracker Green (Ch2)
(See Fig. 4.4b)

Pearson's correlation coefficient : -0.02, M1:0.09, M2:0.13



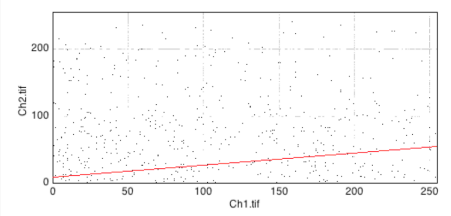
NS-mCherry-KDEL (Ch1) vs ER-Tracker Green (Ch2)
(See Fig. 4.6a)

Pearson's correlation coefficient : 0.52, M1: 0.63, M2:0.98



NS-mCherry-ATEL (Ch1) vs ER-Tracker Green (Ch2)
(See Fig. 4.6b)

Pearson's correlation coefficient : 0.28, M1: 0.5, M2: 0.5



NS-mCherry (Ch1) vs ER-Tracker Green (Ch2)
(See Fig. 4.6c)

Pearson's correlation coefficient : 0.26, M1: 0.28, M2: 0.62

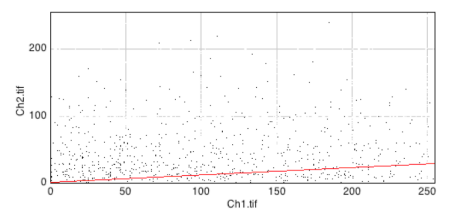
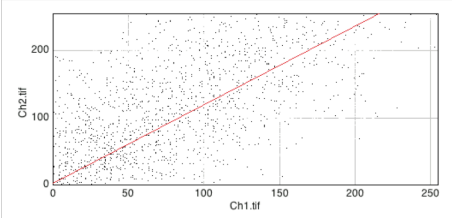


Fig. 4.9: 2D-pixel analysis of red and green fluorescent intensities. Channel 1 (Ch1) represents the red channel, while channel 2 (Ch2) represents the green channel. The linear regression line is shown in red. Pearson's correlation coefficient and Mander's correlation coefficients: M1 (ratio of the summed intensities of pixels from the green image for which the intensity in the red channel is above zero to the total intensity in the green channel) and M2 (ratio of the summed intensities of pixels from the red image for which the intensity in the green channel is above zero to the total intensity in the red channel) are depicted for each image.

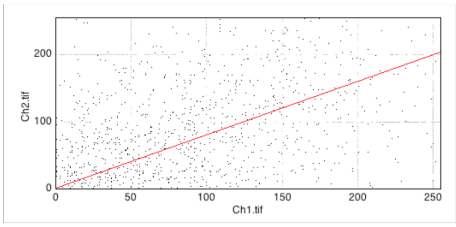
NS-mCherry-KDEL (Ch1) vs NS-EGFP-KDEL (Ch2)
See Fig. 4.7a

Pearson's correlation coefficient : 0.87, M1: 0.95, M2:0.91



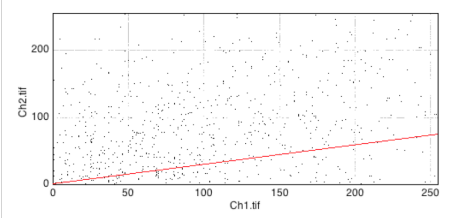
NS-mCherry-ATEL (Ch1) vs NS-EGFP-ATEL (Ch2)
See Fig. 4.7b

Pearson's correlation coefficient : 0.76, M1: 0.82, M2:0.92



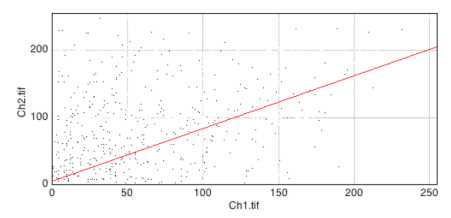
NS-mCherry-ATEL (Ch1) vs NS-EGFP-KDEL (Ch2)
See Fig. 4.7c

Pearson's correlation coefficient : 0.50, M1: 0.51, M2: 0.91



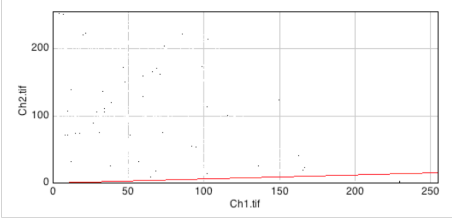
NS-mCherry-KDEL (Ch1) vs NS-EGFP-ATEL (Ch2)
See Fig. 4.7d

Pearson's correlation coefficient : 0.43, M1: 0.79, M2: 0.54



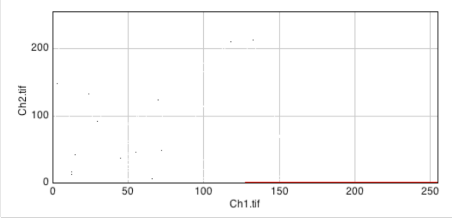
NS-mCherry-KDEL (Ch1) vs GFP-CterGMAP210 (Ch2)
See Fig. 4.8a

Pearson's correlation coefficient : 0.08, M1: 0.05, M2: 0.53



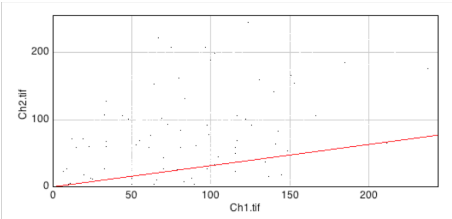
NS-mCherry-KDEL (Ch1) vs NterGMAP210-GFP (Ch2)
See Fig. 4.8b

Pearson's correlation coefficient : 0.01, M1: 0.00, M2: 0.23



NS-mCherry-ATEL (Ch1) vs GFP-CterGMAP210 (Ch2)
See Fig. 4.8c

Pearson's correlation coefficient : 0.48, M1: 0.33, M2: 0.94



NS-mCherry-ATEL (Ch2) vs NterGMAP210-GFP (Ch2)
See Fig. 4.8d

Pearson's correlation coefficient : 0.19, M1: 0.08, M2: 0.84

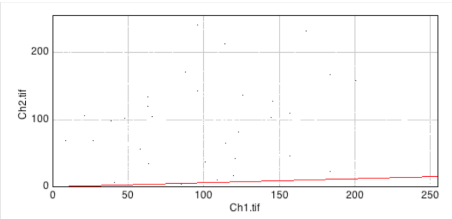


Fig. 4.10: 2D-pixel analysis of red and green fluorescent intensities. Channel 1 (Ch1) represents the red channel, while channel 2 (Ch2) represents the green channel. The linear regression line is shown in red. Pearson's correlation coefficient and Mander's correlation coefficients: M1 (ratio of the summed intensities of pixels from the green image for which the intensity in the red channel is above zero to the total intensity in the green channel) and M2 (ratio of the summed intensities of pixels from the red image for which the intensity in the green channel is above zero to the total intensity in the red channel) are depicted for each image.

4.2.5 ATEL retards secretion into the extra-cellular medium

The fact that ATEL-containing proteins are present in both ER and Golgi raises the question whether ATEL has any retention function at all; if not, we might expect ATEL-containing proteins to progress to the extracellular space to a similar extent as those without ATEL. Therefore, mCherry fluorescence was measured in supernatants of COS-7 cells expressing mCherry, NS-mCherry, NS-mCherry-KDEL or NS-mCherry-ATEL. Untransfected cells were used as controls. The means and standard deviations (Table 4.1) were plotted (Fig. 4.11). One-way ANOVA showed that statistical difference between various groups was highly significant ($F(4,10)=1671.5$, $p<0.0001$). Pairwise comparison between different groups performed using Tukey's post-hoc test (Table 4.2) revealed that mCherry fluorescence was significantly enhanced ($p<0.001$) in supernatants of cells expressing NS-mCherry compared to other experimental groups, suggesting efficient secretion of NS-mCherry to the cell exterior. The fluorescence values of both NS-mCherry-KDEL and NS-mCherry-ATEL were significantly reduced as compared to cells expressing NS-mCherry ($p<0.001$). However, secretion of NS-mCherry-ATEL was significantly higher ($p<0.001$) than NS-mCherry-KDEL. These results suggest that although both ATEL and KDEL mediate retention of mCherry within the secretory pathway, ATEL allows more leakage to the cell exterior and is not as effective as KDEL in retaining proteins within the ER.

Table 4.1: mCherry fluorescence in supernatants

	mCherry Fluorescence	Mean	Standard Deviation
NSmCherry	55728		
	54230		
	52782	54246.67	1473.07
NSmCherryATEL	28754		
	26922		
	27944	27873.33	918.04
NSmCherryKDEL	18453		
	18204		
	18140	18265.67	165.36
mCherry	11723		
	11740		
	10787	11416.67	545.37
Untransfected	9665		
	9964		
	9391	9673.33	286.59

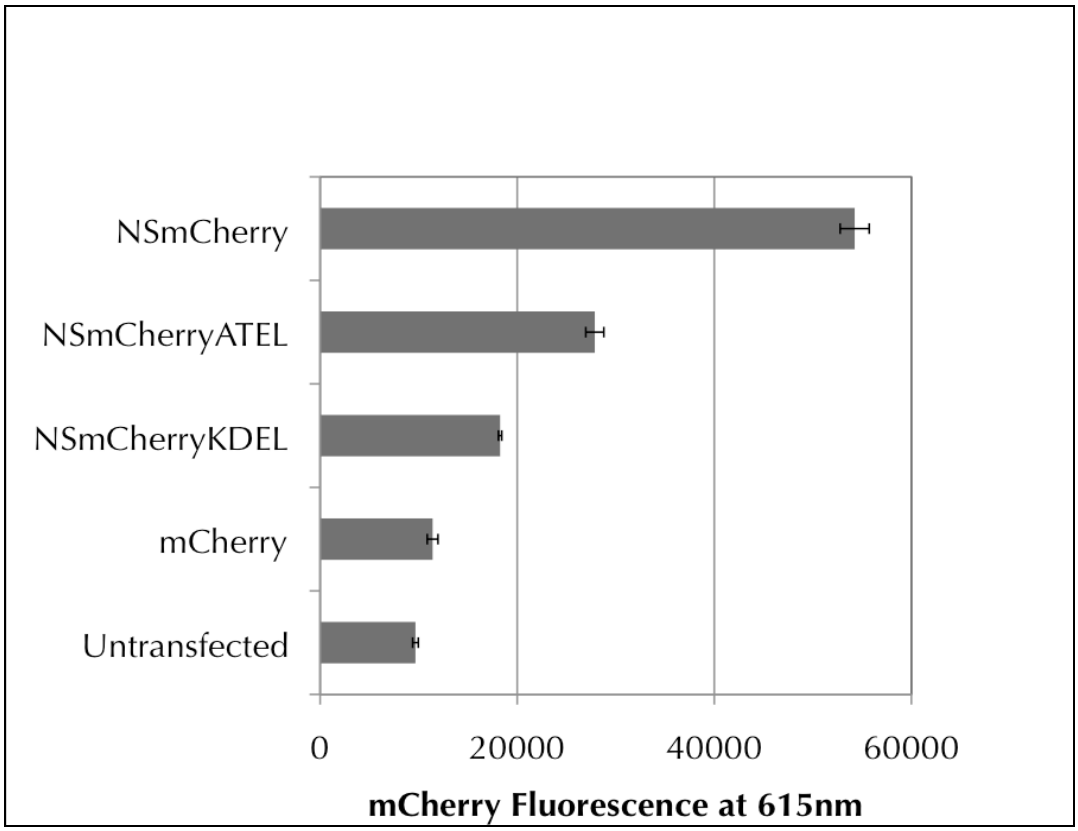


Figure 4.11: ATEL retards secretion into extracellular medium. mCherry fluorescence was measured in concentrated (20X) supernatants of COS-7 cells expressing NS-mCherry, NS-mCherry-ATEL, NS-mCherry-KDEL and mCherry, 24 h post-transfection. Untransfected cells were used as controls. Measurements were made using a PerkinElmer plate-reader at 615 nm. The mean and standard deviation are plotted.

Table 4.2: Tukey's post-hoc test

Comparison	Difference	P-value
NSmCherry vs NSmCherryATEL	97.74	<0.001
NSmCherry vs mCherry	126.07	<0.001
NSmCherryKDEL vs Untransfected	68.59	<0.001
NSmCherryKDEL vs NSmCherry	65.96	<0.001
NSmCherryKDEL vs NSmCherryATEL	31.79	<0.001
NSmCherryKDEL vs mCherry	60.11	<0.001
mCherry vs NSmCherryATEL	28.32	<0.001
untransfected vs NSmCherryATEL	36.8	<0.001
untransfected vs mCherry	8.48	<0.01
untransfected vs NSmCherry	134.55	<0.001

4.3 Discussion

Previous work on the functional characterisation of ArLEA1A and ArLEA1B suggested a role in preventing protein aggregation and in stabilising membranes respectively (Pouchkina-Stantcheva et al., 2007). It was speculated that these proteins were translocated into the ER via their N-terminal signals. In this chapter, the use of mammalian cells has allowed confirmation of the distribution of tagged bdelloid LEA proteins within the ER and the function of their N-terminal signal peptide sequences in delivering these proteins to this compartment. Further the function of the putative ER retention signal, ATEL, has been verified.

Proteins normally enter the ER cotranslationally and are either retained there or targeted to other destinations including the Golgi (Bannykh et al., 1998), lysosomes (Andrews, 2000) or the cell exterior (Burgess and Kelly, 1987). Their functions within the ER might include lipid biosynthesis (Black, 1972; Sisson and Fahrenbach, 1967), participation in 'quality control' processes of other proteins and lipids (Hammond and Helenius, 1995), mediating the unfolded protein response within the ER (Chakrabarti et al., 2011), calcium storage (Martone et al., 1993; Sammels et al., 2010) and secretion (Wiest et al., 1990). Protein retention within the ER is most commonly signal dependent, controlled by sequence-specific receptors (Munro and Pelham, 1987; Pelham, 1990).

It is known that the C-terminal sequence KDEL (or its variants) retains soluble proteins within the ER by retrograde transport from the early Golgi compartments by binding to KDEL receptors (Wilson et al., 1993) and trafficking them back to the ER via COPI-coated vesicles (Letourneur et al., 1994). However, it has also been shown that this system allows some leakage of ER-resident proteins into the extracellular medium, as observed for phytohemaglutinin in plant cells (Herman et al., 1990) and protein disulfide isomerase in a rat pancreatic exocrine cell line (Yoshimori et al., 1990). Hence the finding that a small amount of KDEL-tagged protein is detected in the supernatants is not surprising. Additionally, it is revealed that proteins containing ATEL are retained to a limited degree within the ER, and that unlike KDEL, greater population of other compartments of the secretory system, and indeed the extracellular space, also occurs. This is consistent with a study that examined the ER

retention capacity of 152 variants of KDEL in mammalian cells (Raykhel et al., 2007). Although the authors did not describe ATEL, they showed that the –3 position (i.e. T in ATEL) is not crucial (unless it is proline). Therefore, ATEL is probably similar in function to ADEL, for example, which was shown to allow at least 50% of associated protein to escape the ER.

Although extrapolating results from a mammalian cell to an invertebrate requires caution, trafficking mechanisms are highly conserved between yeast and humans (Dancourt and Barlowe, 2010), so it is likely that these results are also applicable to bdelloids. The presence of ArLEA1A and ArLEA1B throughout intracellular vesicular compartments, as well as the extracellular tissue fluid would allow a small number of protective proteins to protect a large physiological space in the animal, consistent with their role in stress tolerance.

The loss of structure and integrity of the ER has been previously reported in cortical parenchyma cells of mulberry trees during freezing and dehydration (Fujikawa and Takabe, 1996). For example, initiation of freezing at –5 °C resulted in the formation of multiplex lamellae (MPL) that completely covered the area beneath the plasma membrane. These MPL were produced by fusion of pre-existing ER vesicular network consisting of a parallel array of sheet-like ER cisternae. This structural reorganisation of the ER was observed within 10 min upon freezing at –5 °C and was quickly reversed upon thawing. The same structural reorganisation of the ER was produced by osmotic dehydration of the cortical tissues with a 2.7 osmol sorbitol solution at 20 °C. It is speculated that the formation of MPL with the initiation of freezing and drying might play a specific role in inhibiting the close apposition of membranes during water loss. In addition to MPL formation, cortical parenchyma cells have also been shown to accumulate group 3 LEA proteins in their vesicular and multiplex lamellae-form ER (Ukaji et al., 2001). It is hypothesised that conversion of the ER to MPL and accumulation of WAP27 in the ER during winter might minimize plasma membrane destabilization due to the close approach of membranes and consequently confer extremely high freezing tolerance to cortical parenchyma cells of mulberry tree. It is speculated that ArLEA1B might have similar protective effects in rotifers during water loss since it has been demonstrated to

interact with membranes (Pouchkina-Stantcheva et al., 2007). In the previous chapter, ArLEA1B has been predicted to polymerise into a filamentous structure, hence it can be imagined that this function might be crucial in its role in forming a supportive meshwork to stabilise intra-cellular membranes during anhydrobiosis.

Chapter Five: Evaluating the Protective Effects of ArLEA1A and ArLEA1B Within the Mammalian ER

5.1 Introduction

The ER is the site of synthesis for various secreted, plasma membrane and chaperone proteins. Analysis of the human genome has revealed that approximately a third of all open reading frames code for proteins that enter the ER, demonstrating the importance of this organelle for global protein maturation in living cells (Braakman and Bulleid, 2011). Desiccation in bdelloids is expected to result in perturbation of ER homeostasis due to accumulation of unfolded or misfolded proteins causing ER stress (Puthalakath et al., 2007; Ron, 2002; Rutkowski et al., 2006; Szegezdi et al., 2006; Urano et al., 2000). Since ArLEA1A has been shown to reduce aggregation of proteins *in vitro* (Pouchkina-Stantcheva et al., 2007), it was hypothesised that it might protect cells from the potential toxic effects of protein misfolding and aggregation in the physiological context of the ER-Golgi compartments where it has been shown to localise. Since, it was not possible to directly monitor protein misfolding and aggregation in bdelloids, it was decided to test this hypothesis using a spontaneously aggregating polyglutamine (polyQ) expansion protein fused with EGFP (Ravikumar et al., 2002) and bdelloid LEA protein N- and C-terminal signals (NS-EGFPHDQ74-ATEL) targeted to the ER-Golgi compartments in mammalian cells.

PolyQ proteins are often the causative agents of a number of neurodegenerative disorders (Bates, 2005; Gatchel and Zoghbi, 2005) including Huntington's disease, SBMA (spinobulbar muscular atrophy), SCA1 (spinocerebellar ataxia) etc. and are believed to disrupt neural functions by interfering with cellular processes like transcription (Sugars and Rubinsztein, 2003), proteasomal degradation (Finkbeiner and Mitra, 2008; Sakahira et al., 2002) or conductance of cell membranes (Jeub et al., 2006; Sanchez et al., 2008). Although inclusions are

observed in a number of diseased neurons, it is believed that these inclusions do not directly contribute towards disease pathogenesis (Arrasate et al., 2004) and may even provide protection to cells by sequestering toxic proteins. Instead a role of misfolded polyQ monomers and possibly small oligomers has been recently shown to contribute towards increased risk of cell death (Miller et al., 2011).

Previously, Rousseau et al. have claimed that targeting polyQ protein to the ER completely abolished its aggregation (Rousseau et al., 2004). These authors cited the presence of chaperone proteins within the ER that might prevent formation of aggregates. They also suggested that when these proteins escape from the ER to the cytoplasm by retrograde transport, formation of inclusions is accelerated.

There are many reasons to question the conclusions made by the authors in their study. (1) The microscopic imaging performed by the authors in their study might not be sensitive enough to detect the formation of small aggregates within the ER. (2) Although the filter retardation assay shows marked reduction in aggregates observed in cells expressing SP-HttQ73-KDEL in the ER, compared to cells expressing HttQ73 in the cytoplasm, one cannot rule out the possibility that small soluble aggregates might be formed within this compartment that might not be retained on the cellulose acetate filter. Indeed it has been previously shown that the cytotoxic polyQ proteins are not detected in filters of 0.2μm pore size (Miller et al., 2011). (3) If the environment in the ER is unfavourable for protein aggregation as claimed by the authors, then one should not observe the formation of any inclusions in this compartment using other aggregation-prone proteins. However, there have been many studies that have reported the formation of aggregates within this compartment (Gong et al., 2009; Rivera et al., 2000).

Therefore, it was decided to re-evaluate polyQ aggregation within the ER by fluorescence lifetime imaging microscopy (FLIM) (Section 5.2.1). FLIM is a useful tool for investigating the molecular environment of a fluorophore in living cells, including detection of conformational changes, protein-protein interactions and protein aggregation (Borst and Visser, 2010; Kaminski Schierle et al., 2011). A decrease in fluorescence lifetime during protein aggregation has been attributed to the formation supramolecular assemblies that result in disruption of fluorophore emission

properties due to Fluorescence Resonance Energy Transfer (FRET) between the donor fluorophore and the acceptor aggregate. Lifetime measurements are made using time and space-correlated single photon counting (TSCSPC) method that enables simultaneous acquisition of fluorescence lifetime information from each pixel at picosecond level resolution. Average fluorescence lifetimes can subsequently be computed for samples containing aggregated fluorophore and unaggregated controls. However, this method cannot be used to establish the intra-cellular localisation of aggregates. Hence confocal microscopy was performed to confirm the location of aggregates observed in cells (Section 5.2.2).

Additionally, MTS assay was carried out to assess the viability of cells expressing NS-EGFPHDQ74-ATEL alone or in the presence of ArLEA1A-FLAG-mCherry/ArLEA1B-FLAG-mCherry/MG132/TMAO (Section 5.2.3). MTS is a colorimetric method for determining the number of viable cells in a given cell population. The CellTiter 96® AQ_{ueous}One Solution Reagent contains a tetrazolium compound [3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium, inner salt; MTS] and an electron coupling reagent (phenazine ethosulfate; PES). PES has enhanced chemical stability, which allows it to be combined with MTS to form a stable solution. Phenazine methosulfate (PMS) is used as an electron coupling reagent in the assay. The quantity of formazan product is measured by the amount of absorbance at 490 nm which is indicative of mitochondrial activity which in turn is directly proportional to the number of living cells in culture. As misfolded proteins expressed within the ER might be targeted to the ubiquitin-proteasomal machinery in the cytoplasm by retrograde transport, inhibition of this complex by MG132, is expected to increase protein aggregation (Rousseau et al., 2004) and therefore decrease cell viability. Trimethylamine N-oxide (TMAO) is a chemical chaperone that has been recently demonstrated to reduce aggregation of mutant αA-crystallin within the ER (Gong et al., 2009). Adding TMAO might increase the viability of cells expressing NS-EGFPHDQ74-ATEL.

Lastly, the expression of NS-EGFPHDQ74-ATEL with or without ArLEA1A-FLAG-mCherry/ArLEA1B-FLAG-mCherry/MG132/TMAO was analysed by western blot (Section 5.2.4).

5.2 Results

5.2.1 Fluorescence lifetime imaging of cells expressing NS-EGFPHDQ74-ATEL

Fluorescence lifetime imaging (FLIM) was performed on COS-7 cells expressing NS-EGFPHDQ74-ATEL at 24 h and 48 h post-transfection. Cells expressing NS-EGFP-ATEL in the ER and EGFPHDQ74 in the cytoplasm and the nucleus were used as controls. As can be observed (Fig. 5.1a-i), there is no significant difference in the fluorescence lifetime observed in cells expressing either NS-EGFPHDQ74-ATEL or NS-EGFP-ATEL at 24 h post-transfection, indicating absence of amyloid formation. In contrast, distinct reduction in fluorescence lifetime is observed in cells expressing EGFPHDQ74 in the cytoplasm (Fig. 5.1j). However, one observes that the ER-morphology in cells expressing NS-EGFPHDQ74-ATEL is considerably fragmented compared to the uniform morphology of the ER observed in control cells expressing NS-EGFP-ATEL. These results suggest the occurrence of ER stress in cells expressing NS-EGFPHDQ74-ATEL within 24 h of transfection. Similar fragmentation of the ER network has been described in various conditions of ER stress (Howarth et al., 2012; Santi-Rocca et al., 2012; Sokka et al., 2007).

At 48 h, a drastic reduction in fluorescence lifetime is observed in five out of eight cells expressing NS-EGFPHDQ74-ATEL, potentially indicating the formation of amyloid (Fig. 5.2a-h). Fig. 5.3 compares the average fluorescence lifetime observed at 24 h and 48 h in cells expressing NS-EGFPHDQ74-ATEL. The average lifetime is significantly reduced at 48 h compared with 24 h ($t=2.235$, $P = 0.0261$, one-sided t-test with Welch correction).

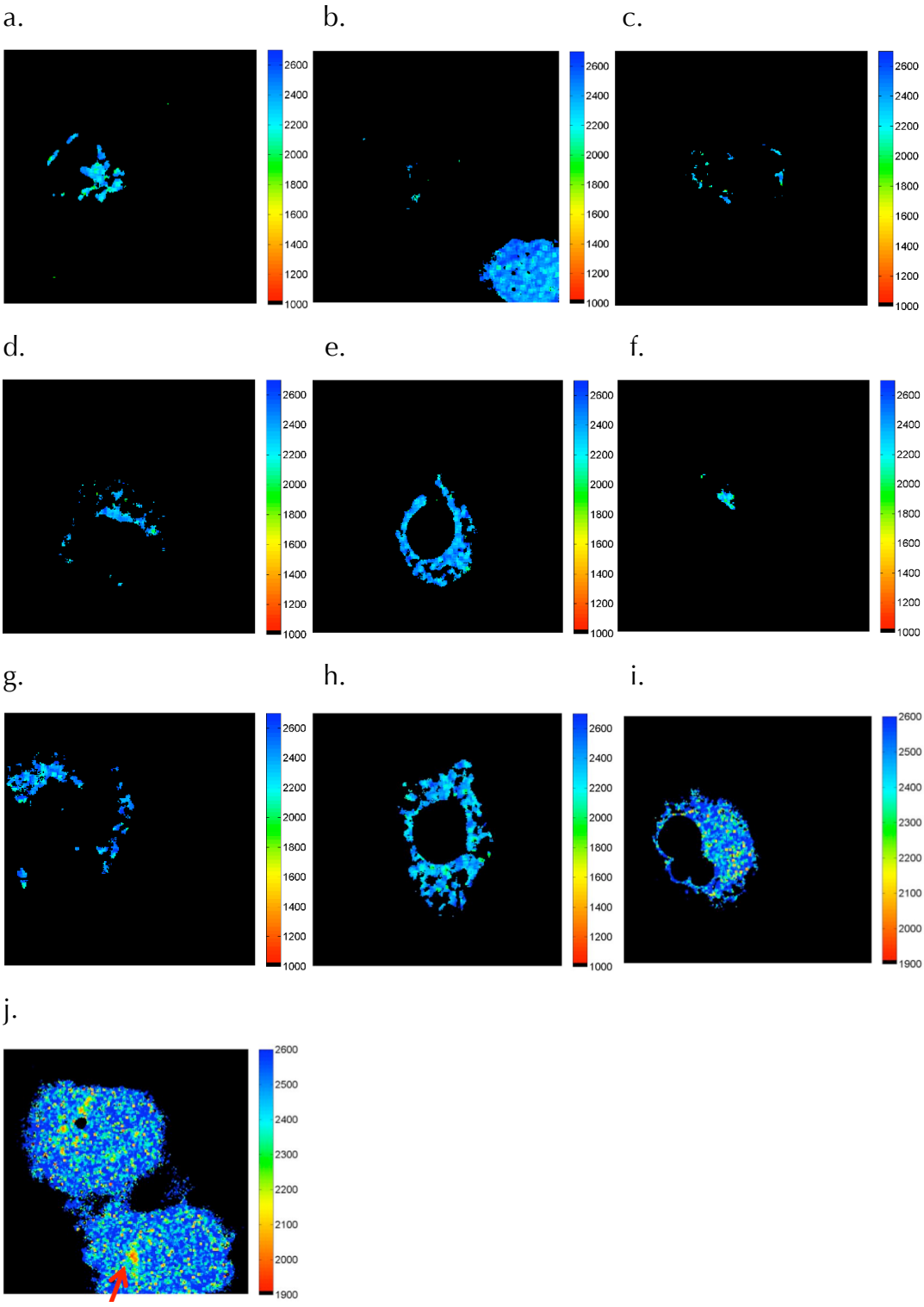


Figure 5.1: Fluorescence lifetime imaging of NS-EGFPHDQ74-ATEL in COS-7 cells at 24 h. (a-h) Fluorescence lifetime was measured in eight different cells. (i) NS-EGFP-ATEL was used as a control. (j) PolyQ expression observed in the cytoplasm. Lifetime was measured in picoseconds as shown in the colour-scale. This data was obtained in collaboration with Drs. Claire Michel and Gabriele S. Kaminski Schierle.

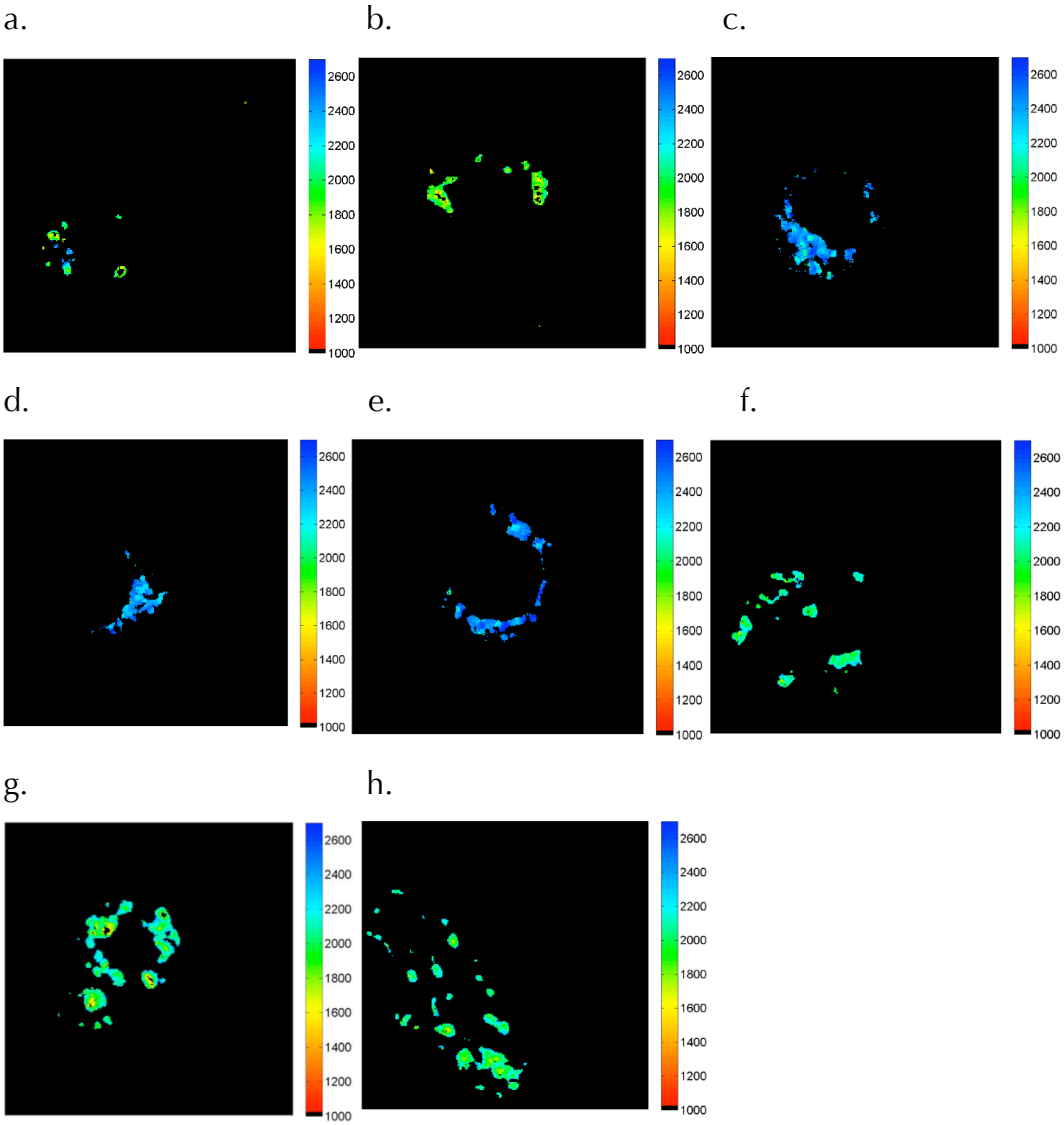


Figure 5.2: Fluorescence lifetime imaging of NS-EGFPHDQ74-ATEL in COS-7 cells at 48 h. (a-h) Fluorescence lifetime was measured in eight different cells. Lifetime was measured in picoseconds as shown in the colour-scale. This data was obtained in collaboration with Drs. Claire Michel and Gabriele S. Kaminski Schierle.

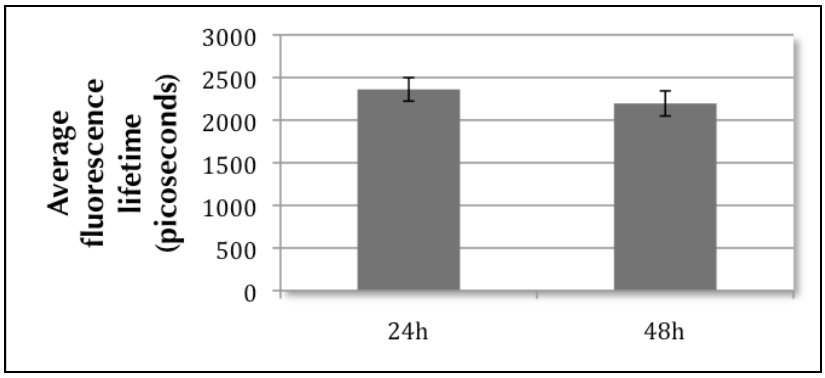


Figure 5.3: Average lifetime for NS-EGFPHDQ74-ATEL at 24 h and 48 h. Error bars indicate average SD for n=8 at 24 h and n=9 at 48 h.

5.2.2 Monitoring NS-EGFPHDQ74-ATEL expression in COS-7 cells by confocal microscopy

In order to confirm the intracellular localisation of aggregates, COS-7 cells were co-transfected with pcDNA3.3 containing NS-EGFPHDQ74-ATEL and NS-mCherry-ATEL (the latter was used as a marker for ER-Golgi compartments, as described in Chapter Four). Cells were subsequently fixed and stained with DAPI. Colocalisation analysis was performed as before. Around 70% of the cells observed at 24 h, showed colocalisation of NS-EGFPHDQ74-ATEL with NS-mCherry-ATEL, with little evidence of aggregate formation (Fig. 5.4a). However, in around 5% of the cells aggregate formation was detectable in the cytoplasm and in the nucleus (Fig. 5.4b). The ER network in these cells was found to be quite uneven. Moreover, the nuclei in these cells appeared to be fragmented indicating the initiation of apoptosis. These aggregates could have formed by retrograde transport of misfolded/aggregating proteins from the ER to the cytoplasm and nucleus. At 48 hours, as many as 50% of cells expressing NS-EGFPHDQ74-ATEL were observed to be in the later stages of apoptosis as evidenced by appearance of nuclear fragments (Fig. 5.4c). The uniform ER network in these cells was no longer visible and was replaced by large polyQ aggregates colocalising with NS-mCherry-ATEL. Moreover, nuclear fragments were also observed to co-cluster with NS-EGFPHDQ74-ATEL and NS-mCherry-ATEL indicating fusion of ER-Golgi compartments with these fragments. Such co-clustering of Golgi vesicles with other organelles has recently been described by Nozawa et al. in cells undergoing apoptosis (Nozawa et al., 2009). In a few cells, significant dilation and clumping within ER-Golgi vesicles was also observed (Fig. 5.4d). These results hint at the prevalence of aggregated polyQ proteins within the ER-Golgi compartments.

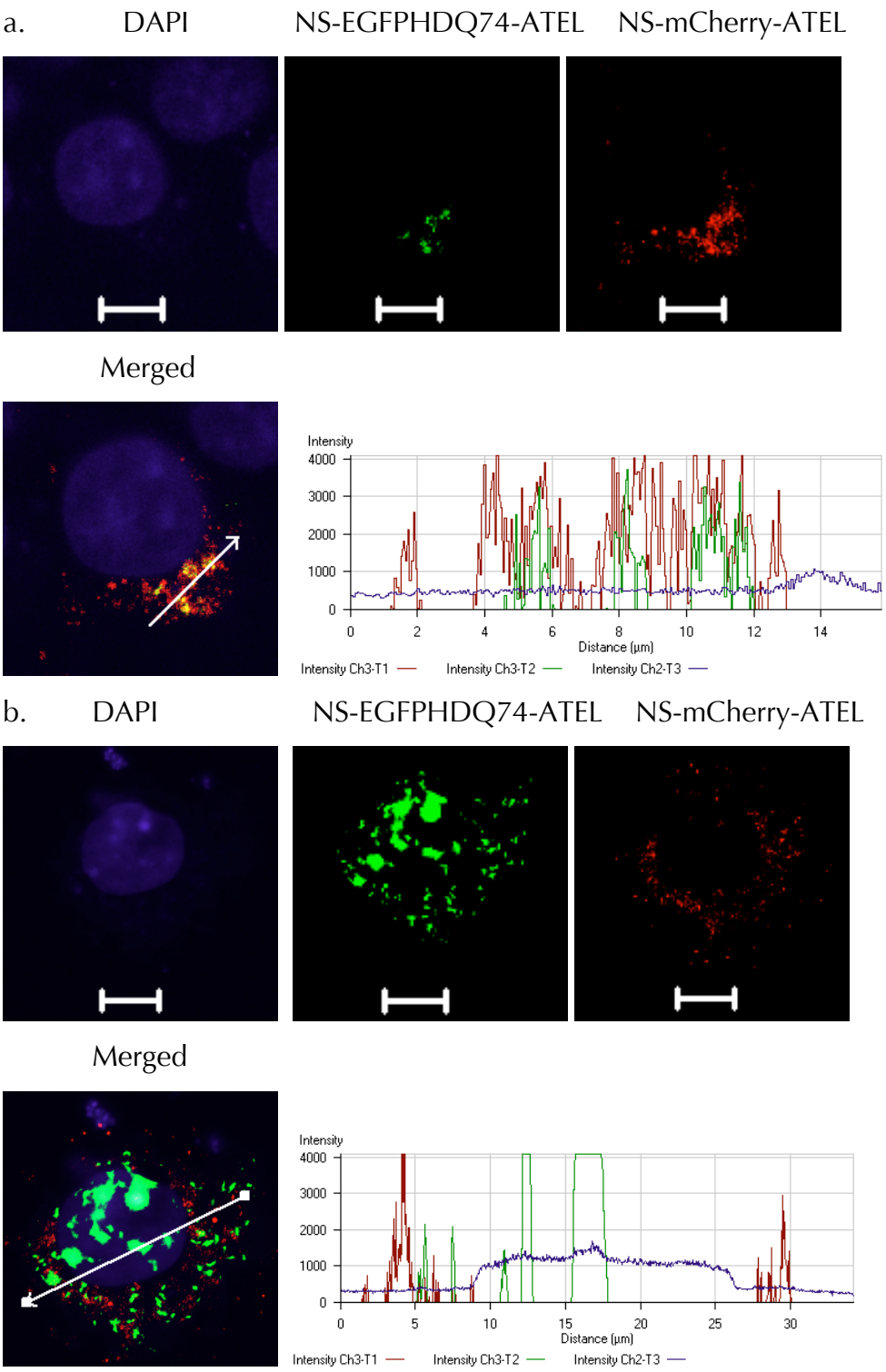
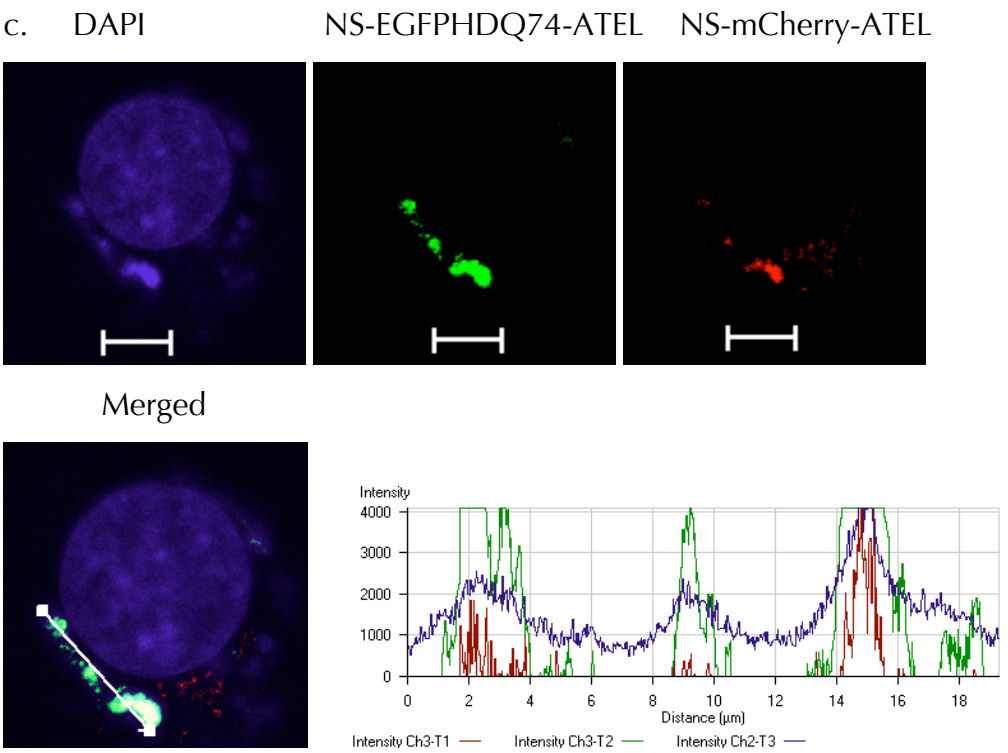


Figure 5.4: Confocal microscopy of COS-7 cells expressing (a,b) NS-EGFPHDQ74-ATEL and NS-mCherry-ATEL at 24 h. Distinct aggregates are observed in (b). Plots depict blue, green and red fluorescent signals in selected region of interest (arrow). Scale bar: 10 µm.



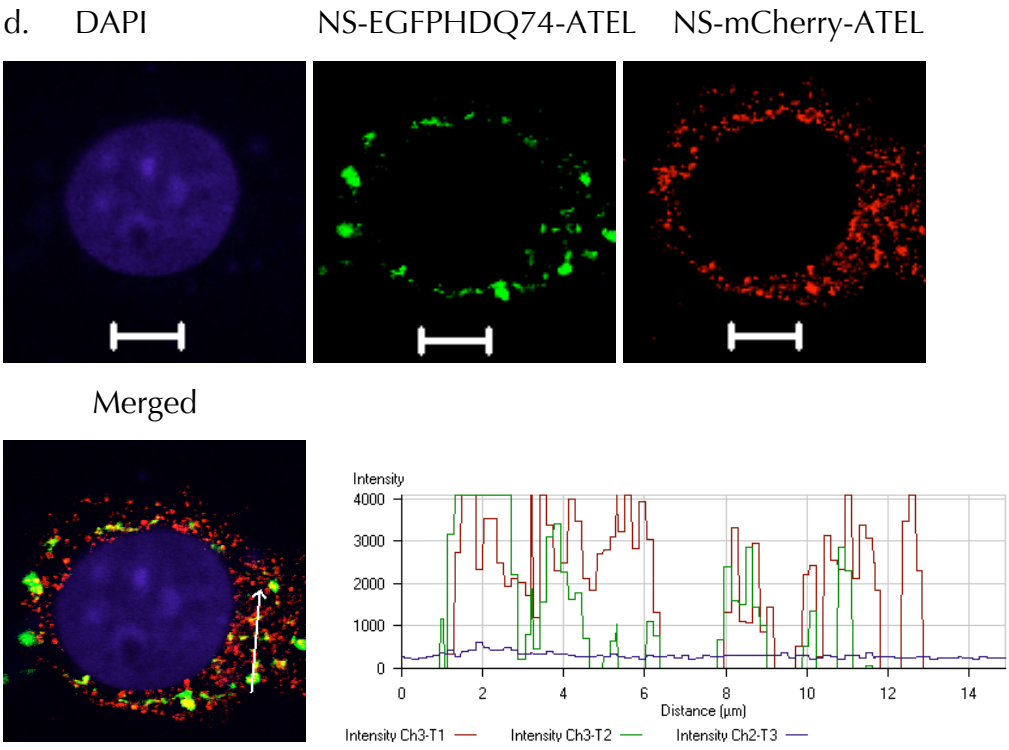


Figure 5.4 (continued): Formation of inclusions in the ER-Golgi compartments is observed at (c) 48 h and (d) 72 h. Plots depict blue, green and red fluorescent signals in selected region of interest (arrow). Scale bar: 10 µm.

5.2.3 Cell viability measurements

Cells expressing NS-EGFPHDQ74-ATEL alone, or with ArLEA1A-FLAG-mCherry/ArLEA1B-FLAG-mCherry/MG132/TMAO were treated with MTS reagent at 3, 12, 24 and 48 h post-transfection. Untransfected cells were used as controls. The absorbance was recorded in a spectrophotometer at 490nm. The means and standard deviations (n=3) were plotted (Fig. 5.5). Statistical significance of variation observed between different groups at 48 h was determined using One-way ANOVA, which demonstrated that the difference observed between different groups was highly significant ($F(5,12)=33.046$, $p<0.0001$). Pair-wise comparison between different groups performed using Tukey's post-hoc test revealed that although there was a decrease in viability of cells expressing NS-EGFPHDQ74-ATEL compared to untransfected controls the difference observed was not statistically significant. Adding TMAO to cells expressing NS-EGFPHDQ74-ATEL was observed to improve cell viability slightly however, the difference observed was also not statistically significant. Surprisingly, over-expressing either ArLEA1A-FLAG-mCherry or ArLEA1B-FLAG-mCherry was observed to reduce the viability of cells significantly ($p<0.001$). Adding proteasomal inhibitor MG132, results in the most dramatic decrease in cell viability ($p<0.001$), confirming previous results that have shown that failure to clear aggregates is detrimental for cell survival (Lauricella et al., 2003).

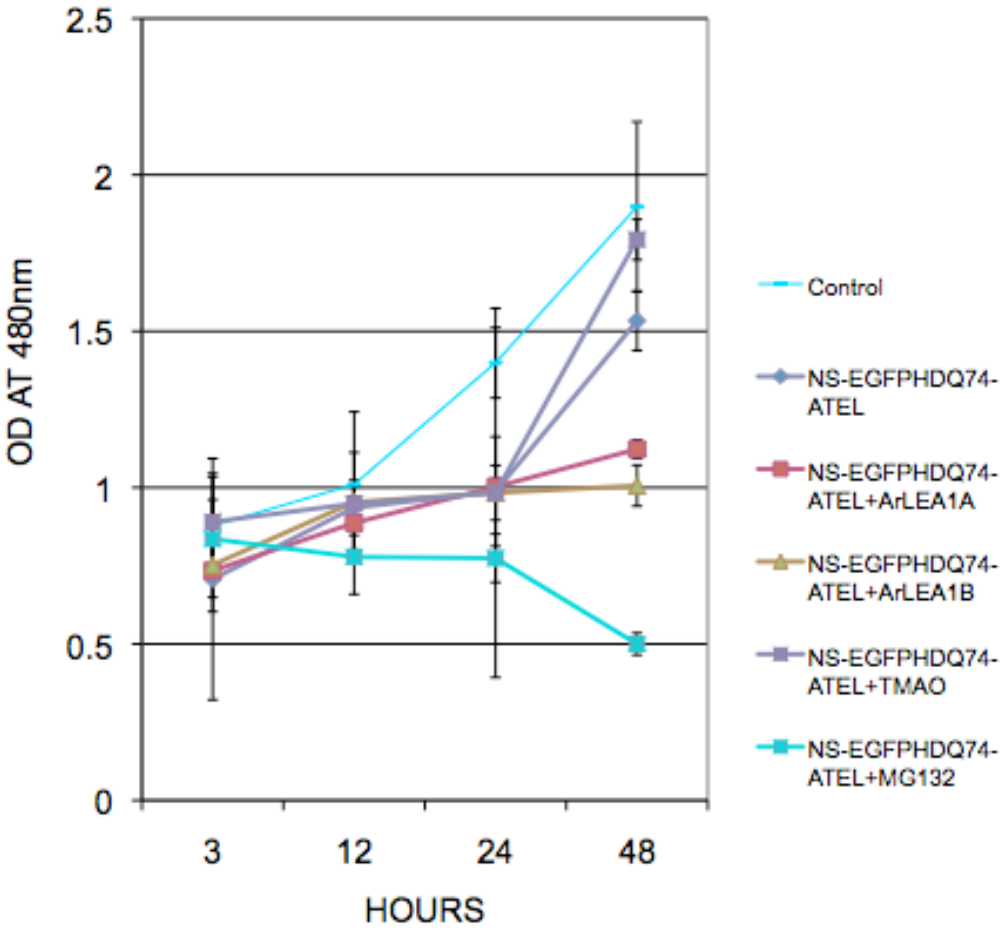


Figure 5.5: MTS assay for cell viability was performed on cells expressing NS-EGFPHDQ74-ATEL alone, or with ArLEA1A-FLAG-mCherry/ArLEA1B-FLAG-mCherry/TMAO/MG132. Untransfected cells were used as controls. Absorbance was measured at 490nm at 3, 12, 24 and 48 h post-transfection. Experiments were performed in triplicate. Error bars depict +-1SD.

5.2.4 Western blot detection of NS-EGFPHDQ74-ATEL

In order to evaluate NS-EGFPHDQ74-ATEL protein levels in COS-7 cells with or without ArLEA1A-FLAG-mCherry/ArLEA1B-FLAG-mCherry/MG132/TMAO, it was decided to focus on the soluble fraction of proteins, since it has been shown that it is the monomeric or small oligomeric polyQ protein that contributes towards cytotoxicity and not the protein present in insoluble aggregates (Miller et al., 2011). Western blot experiments were carried out 48 h post-transfection. As misfolded proteins expressed within the ER might be targeted to the ubiquitin-proteasomal

machinery in the cytoplasm by retrograde transport, inhibition of this complex by MG132, is expected to result in an increase in NS-EGFPHDQ74-ATEL (Rousseau et al., 2004). On the other hand, adding TMAO, which is a chemical chaperone, should have no effect on the amount of polyQ protein (Gong et al., 2009). As can be observed in Fig. 5.6, there is a drastic decrease in NS-EGFPHDQ74-ATEL in cells over-expressing ArLEA1A-FLAG-mCherry and ArLEA1B-FLAG-mCherry (lanes 4 and 5), as compared to cells expressing NS-EGFPHDQ74-ATEL alone (lane 2). Adding TMAO does not affect NS-EGFPQ74-ATEL expression (lane 3), while adding MG132 results in an increase in the amount of NS-EGFPHDQ74-ATEL as expected (lane 6). The lack of loading controls and input material controls is acknowledged in this data.

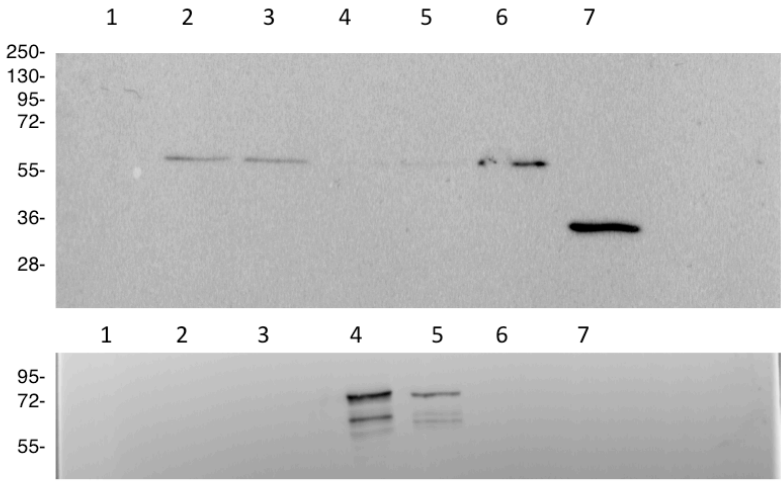


Figure 5.6: Western blot detection of NS-EGFPHDQ74-ATEL in COS-7 cells, 48 h post-transfection (top panel). 5μg of protein extract was loaded in each lane. The primary antibody used was anti-GFP, while the secondary body was HRP-linked anti-mouse antibody. ArLEA1A-FLAG-mCherry and ArLEA1B-FLAG-mCherry were detected (bottom panel) by using rabbit anti-ArLEA1A primary antibody and HRP-linked secondary anti-rabbit antibody. Lane 1: untransfected cells, lane 2: NS-EGFPHDQ74-ATEL, lane 3: NS-EGFPHDQ74-ATEL+TMAO, lane 4:NS-EGFPHDQ74-ATEL+ArLEA1A-FLAG-mCherry, lane 5: NS-EGFPHDQ74-ATEL+ArLEA1B-FLAG-mCherry, lane 6:NS-EGFPHDQ74-ATEL+MG132, lane 7: NS-EGFP-ATEL. MW markers are shown on the left.

5.3 Discussion

In order to evaluate the protective effects of ArLEA1A/ArLEA1B within the physiological context of the ER-Golgi compartments, it was important to devise a cell-based system where protein misfolding and aggregation could be induced and monitored within these organelles. Hence, expression of GFP-tagged polyQ protein was targeted to the ER and Golgi in COS-7 cells. At 24 h, there was little evidence of aggregation within these compartments, although the ER morphology was found to be considerably fragmented indicating ER stress. In a few cells, aggregate formation was detected in the cytoplasm and nucleus as reported previously (Rousseau et al., 2004). These aggregates are hypothesised to form as a result of retrograde transport of oligomerised/aggregated polyQ proteins from the ER to the cytoplasm. At later time points aggregates were also observed within the ER-Golgi compartments. Moreover, considerable nuclear fragmentation indicated the onset of apoptosis in these cells.

MTS assays revealed that cells expressing NS-EGFPHDQ74-ATEL experienced a reduction in viability compared to untransfected cells although the effect was not statistically significant. This could be due to the fact that the cells used in these experiments are dividing and although many cells might undergo apoptosis due to NS-EGFPHDQ74-ATEL expression, the reduction in metabolic activity might be neutralised by a net increase in the number of cells in the total population. However, co-expression of ArLEA1A/ArLEA1B with NS-EGFPHDQ74-ATEL did not protect cells from deleterious effects of NS-EGFPHDQ74-ATEL expression. On the contrary, expressing these proteins resulted in considerable loss of cell viability at the 48 h timepoint. The exact cause of this effect is not very well understood. It is possible that these proteins might provide protection exclusively in bdelloid cells under the influence of specific signalling factors. Alternatively, the possibility of 'ER overload' (Ewbank and Pujol, 2010) that might result in subsequent cell necrosis cannot be ignored since ArLEA1A and ArLEA1B have been expressed in 5:1 ratio to NS-EGFPHDQ74-ATEL. In the future, it might be useful to test the expression of ArLEA1A and ArLEA1B proteins in cells expressing NS-EGFPHDQ74-ATEL in 1:1 ratio without any fluorescent tags, since one cannot rule out the possibility that mCherry might contribute towards some amount of toxicity. One could also test the anti-aggregation

effects of AavLEA1 from *Aphelenchus avenae* in the ER by cloning the N and C-terminal ER translocation signals in frame with this protein.

It is interesting to note that adding TMAO, a chemical chaperone to cells expressing NS-EGFPHDQ74-ATEL, resulted in slight improvement of cell viability. These results are consistent with previous findings that have shown that TMAO alleviates aggregation of mutant αA-crystallin within the ER (Gong et al., 2009) by increasing the amount of ER chaperone proteins like heat shock protein 70 (HSP70).

As misfolded proteins are transported from the ER to the cytoplasm where they might be degraded by the ubiquitin-proteasomal complex, inhibition of proteasomal degradation by MG132 is expected to increase ER stress by inhibiting the clearance of improperly folded proteins from the ER, resulting in the activation of the unfolded protein response (UPR) and subsequent apoptosis (Bernales et al., 2006; Chakrabarti et al., 2011; Kaufman et al., 2002; Ron and Walter, 2007). Previously polyQ aggregation in the cytoplasm has been shown to trigger ER stress through proteasomal dysfunction (Nishitoh et al., 2002). Indeed, adding MG132 to cells expressing NS-EGFPHDQ74-ATEL is catastrophic for cell survival.

Western blot analysis showed that over-expression of ArLEA1A/ArLEA1B resulted in drastic reduction of NS-EGFPHDQ74-ATEL expression. It is speculated that this could be due to direct-translational inhibition of NS-EGFPHDQ74-ATEL due to ER overload (Ewbank and Pujol, 2010). Adding TMAO did not cause any change in NS-EGFPHDQ74-ATEL expression as expected, since it is not a regulator of protein translation or degradation mechanisms. In contrast, adding MG132, resulted in a net increase in NS-EGFPHDQ74-ATEL levels due to inhibition of proteasomal degradation.

Overall the results in this chapter highlight the complications involved in expressing heterologous proteins in the mammalian ER. Encouraging results for TMAO suggest possible therapeutic applications of this chemical chaperone in the treatment of various diseases caused by protein misfolding in the ER such as Alzhiemer's disease, diabetes mellitus, retinitis pigmentosa (Lin et al., 2008).

Chapter Six: Assessing Protective Effects of ArLEA1A and ArLEA1B During Desiccation-Induced Damage of mCherry Fluorescent Protein

6.1 Introduction

Fluorescent proteins are valuable biotechnological tools that have enabled the study of biological molecules and pathways in living cells, tissues, whole animals, plants and microorganisms. GFP was the first fluorescent protein to be characterised (Shimomura et al., 1962) as a companion protein to aequorin from jellyfish *Aequorea victoria*. In recognition of the importance of this protein in improving our understanding of life processes, Martin Chalfie, Osamu Shimomura and Roger Y. Tsien were awarded the Nobel Prize in Chemistry in 2008. Over the years since its discovery, GFP has been subject to several rounds of engineering to improve its brightness and fluorescent properties. Alteration of key residues around its fluorophore has resulted in the development of different colours such as blue, yellow and cyan, that has enabled multicolor tracking inside cells and FRET measurements involving energy transfer between different fluorophores. However, modification of GFP to derivatives emitting in the red or near-infrared region of the spectrum remained elusive for a long time. The quest for the discovery of fluorescent proteins with longer emission wavelengths ultimately led to the identification of red fluorescent protein drFP583 from *Discosoma* coral species, which later became known as DsRed (Matz et al., 1999). This protein has excitation and emission maxima at 558 and 583 nm respectively and shows stable fluorescence *in vivo*. The DsRed fluorophore is formed from the primary sequence Gln-Tyr-Gly (residues 66-68) arranged in a rectangular array in two approximately antiparallel pairs (Gross et al., 2000). However this protein is tetrameric in nature (Yarbrough et al., 2001) and

its fluorophore undergoes slow maturation kinetics involving post-translational modifications of Gln66, Tyr67 and Gly68 into an imidazolidinone heterocycle with p-hydroxybenzylidene and acylimine substituents (Baird et al., 2000; Matz et al., 1999).

In order to overcome problems of tetramerization and slow fluorophore maturation kinetics in DsRed, it was subject to multiple rounds of directed evolution that resulted in the formation of an improved monomeric version called mRFP (monomeric red fluorescent protein) (Bevis and Glick, 2002). Although mRFP proved to be widely useful, it was found to be sensitive to N-terminal fusions (Shaner et al., 2004) and its fluorophore maturation process had scope for further improvement. Therefore, Shaner et al. developed the next generation of monomers of mRFP through additional rounds of directed mutations in residues surrounding the fluorophore (Shaner et al., 2004). These variants (called 'mFruits') have overcome some of the previous drawbacks of mRFP and consist of seven members: mHoneydew, mBanana, mOrange, tdTomato, mTangerine, mStrawberry and mCherry.

mCherry contains an Q66M (glutamine->methionine) mutation that promotes a more complete maturation of the fluorophore and demonstrates an additional 5 nm red-shift in both the excitation and emission spectra relative to mRFP. It also demonstrates increased brightness, faster maturation and is less prone to photobleaching than mRFP. Furthermore, the N- and C-terminal ends of mRFP have been replaced with GFP residues to increase its tolerance to both N-terminal and C-terminal fusions. Its 3D structure is remarkably similar to GFP and DsRed with the typical 11-stranded beta barrel, a coaxial helix and a centrally located fluorophore (Shu et al., 2006). Fig. 6.1 compares the crystal structures of mCherry, DsRed and GFP.

The fluorophore environment of mCherry is substantially more hydrophobic than DsRed. The K163Q(lysine->glutamine) substitution results in loss of hydrogen bonding of the phenolate oxygen of the fluorophore leading to a possible shift of electron density towards the imidazolinone moiety. The K83M(lysine->methionine) mutation also results in minor side chain arrangements in the fluorophore. Fig. 6.2 depicts the fluorophore organisation in mCherry.

a.

b.

c.

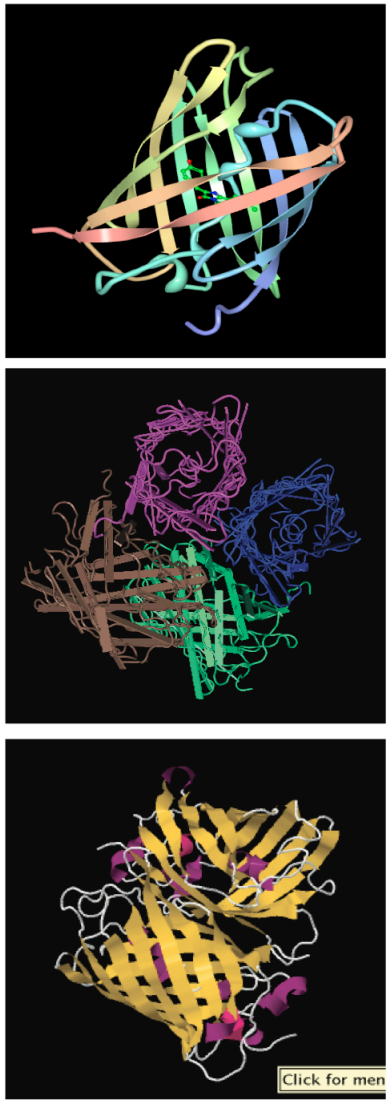


Figure 6.1: Crystal structures of (a) mCherry (Shu et al., 2006), (b) DsRed (Tubbs et al., 2005) and (d) GFP (Yang et al., 1996) downloaded from RCSB PDB. mCherry is observed to be monomeric, while DsRed is tetrameric and GFP is dimeric. These crystal structures were solved at a resolution of 1.36 Å for mCherry, 2.0 Å for DsRed and 1.9 Å for GFP.

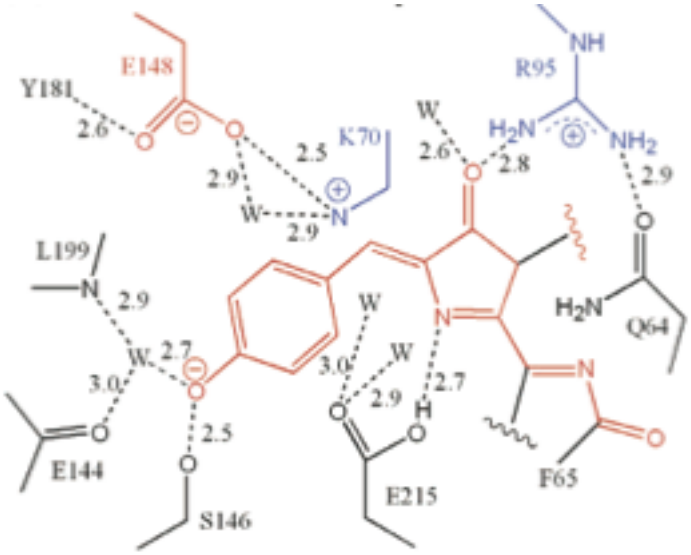


Figure 6.2: Fluorophore organisation of mCherry. The substitution K163Q (mCherry) leads to a loss of hydrogen-bonding opportunities for the phenolate oxygen of the fluorophore and presumably results in a net shift of electron density toward the imidazolinone moiety. Common to all of these structures is the mutation of K83L mutation in mCherry. This substitution results in the terminal N of conserved Lys 70 to move away by about 2.7 Å from its position over the methylene bridge of the fluorophore in DsRed. Its new position is stabilised through the interaction with Glu 148. Glu 215 appears to be protonated and forms a hydrogen bond (2.7 Å) with the imidazolinone ring nitrogen. Note the presence of water (W) molecules in determining the fluorophore structure. Taken from Shu et al., 2006, with permission.

It is important to recognise the role of water molecules in determining the quantum chemistry of the mCherry fluorophore. Topol et al. observed significant spectral tuning of the excitation and absorption spectra of mCherry by slightly modifying the orientation of the hydrogen bonded water network around the fluorophore (Topol et al., 2011). Moreover, the state of protonation of the conserved Glu215 residue in the fluorophore environment and its interaction with the neighboring water molecule might also affect its absorption properties (Topol et al., 2011).

It was hypothesised that drying mCherry might significantly alter its spectral properties presumably due to protein denaturation and loss of water molecules that are predicted to be crucial for the structural and functional integrity of the mCherry fluorophore. Since, ArLEA1A from bdelloid rotifers and AavLEA1 from *Aphelenchus avenae* have been demonstrated to function as 'molecular shields', it was decided to

test whether adding these proteins might protect mCherry from the deleterious effects of drying. Molecular shields are hypothesised to prevent intermolecular interactions between aggregating proteins by forming a physical barrier between them (Goyal et al., 2005b). Moreover, their functions have been found to be quite different from classical molecular chaperones. For example, molecular chaperones form complexes with their client proteins mostly through hydrophobic patches (Saibil, 2008; Tam et al., 2009) whereas both *in vivo* and *in vitro* experiments suggest that shield proteins act by slowing the collision rate of aggregating protein species (Liu et al., 2011). Moreover, the interaction of shield proteins with their clients has been found to be quite loose in comparison with chaperones that are more tightly associated (Chakrabortee et al., 2010; Chakrabortee et al., 2012). It can therefore be hypothesised that these proteins might be ineffective in preventing intra-molecular changes of protein structures caused by water loss.

In order to test this hypothesis, mCherry protein along with ArLEA1A, ArLEA1B, AavLEA1 and AavLEA1-FLAG-mCherry were expressed and purified as described in Sections 6.2.1 and 6.2.2. The effect of drying on the fluorescence and absorbance spectra of mCherry was analysed in Section 6.2.3. Subsequently the protective effects of ArLEA1A, ArLEA1B and AavLEA1 in rescuing mCherry fluorescence and absorbance properties were assessed in Section 6.2.4. Lastly, the protective effect of fusing AavLEA1 with mCherry *in cis* was analysed in Section 6.2.5.

6.2 Results

6.2.1 Expression and purification of mCherry

mCherry was cloned into pET-28a(+) vector, in frame with the His_6 tag at the N-terminal end. The plasmid was subsequently transformed into competent BL21DE3 *E. coli* cells. After screening transformants for optimal expression, single colonies were grown overnight in 5ml of LB. Overnight cultures were subsequently used to inoculate 400 ml LB and grown at 30°C until the OD_{600} reached 0.6. Protein expression was induced by adding 1mM IPTG. Cultures were further grown at three different temperatures (16°C, 30°C and 37°C) to optimise protein expression. Aliquots were withdrawn at 1, 2, 3 and 5 h post-induction and sample lysates were analyzed on SDS-PAGE gel (Fig. 6.3). One observes the presence of a truncated product, around 20 kDa in length within the first hour of expression at 37°C (Fig. 6.3 a), 2 h at 30°C (Fig. 6.3b) and within 3 h at 16°C (Fig. 6.3c). However, the amount of protein produced at 16°C was found to be drastically reduced. Hence it was decided to use cultures grown at 30°C for production of mCherry.

mCherry was purified using Nickel-NTA column (Fig. 6.4) and the His_6 tag was removed by thrombin cleavage (Fig. 6.5). The smaller truncated products have not been reported so far in the literature. However, other groups attempting to purify mCherry have also observed these products (personal communication by Dr. N.C. Shaner) and are believed to result from protease cleavage of the exposed helix-loop-helix region in the beta-barrel structure of mCherry.

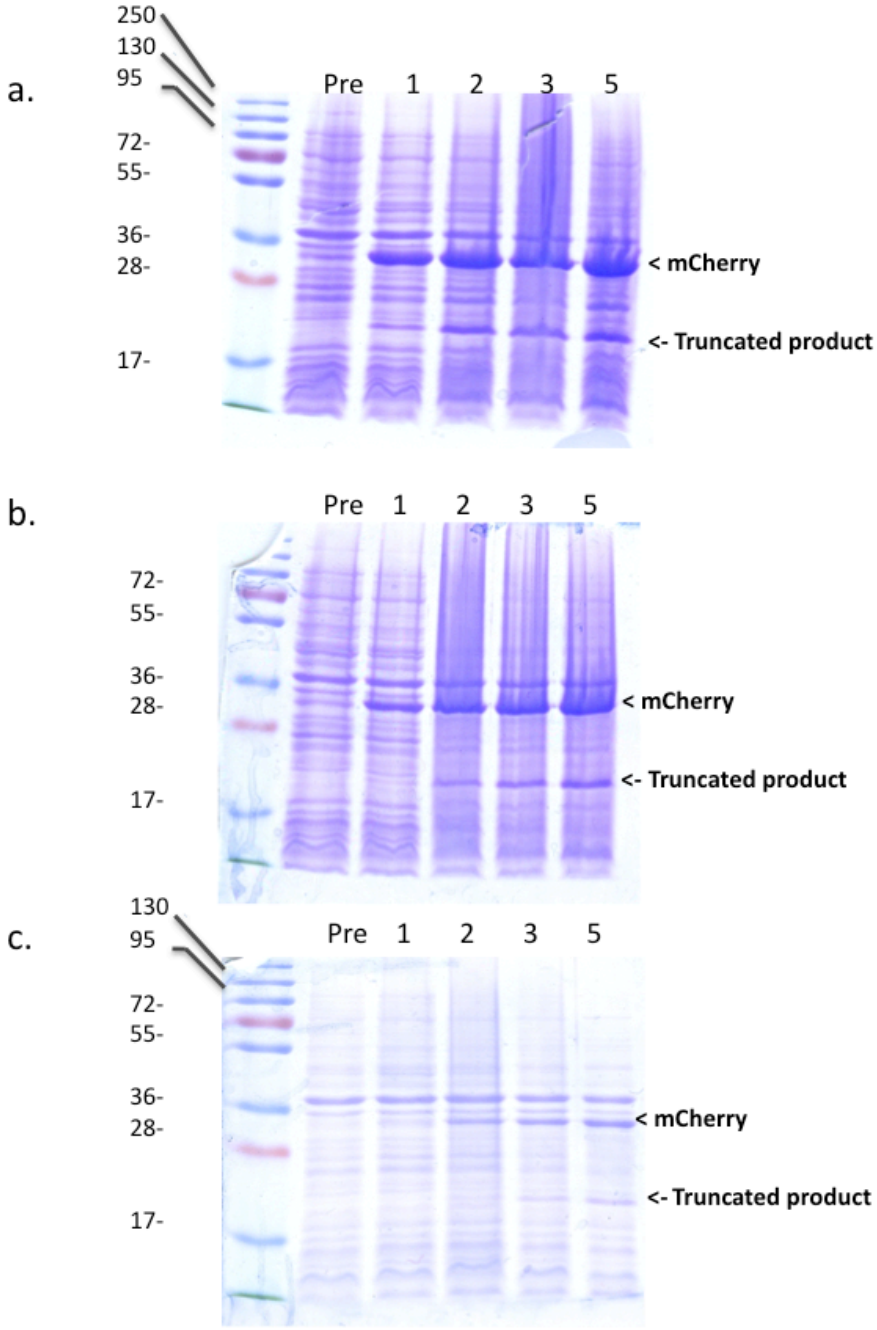


Figure 6.3: SDS-PAGE gels for monitoring mCherry expression at (a) 37°C, (b) 30°C and (c) 16°C. (10% SDS-PAGE gels were used). Pre: Pre-induced cell lysate. Cell lysates obtained at 1, 2, 3 and 5 h are in lanes labelled 1, 2, 3 and 5, respectively. Full length mCherry (29.3 kDa) and truncated mCherry (approximately 20 kDa) are observed. Molecular weight ladder (in kDa) is shown on the left. Performed in collaboration with Dr. Matthew Watson.

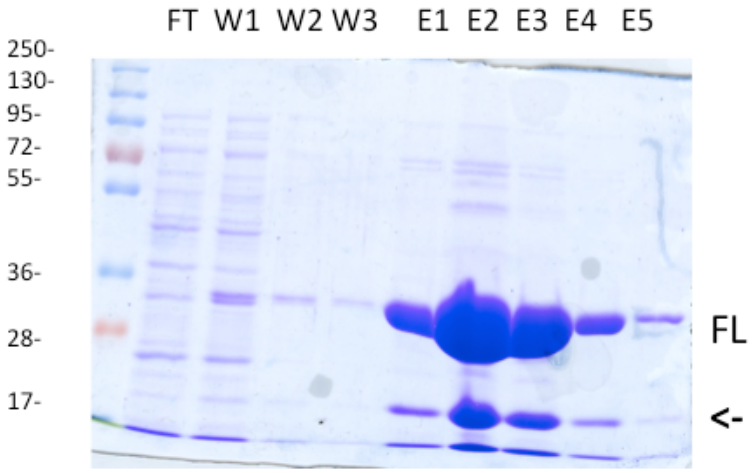


Figure 6.4: Purification of His_6-tagged mCherry. (10% SDS-PAGE gel was used). FT: Flowthrough. W1-3: Wash fractions. E1-5: Eluates. FL: Full length His_6-tagged mCherry (MW 29.3 kDa). Truncated product is marked by an arrow. MW ladder (in KDa) is shown on the left.

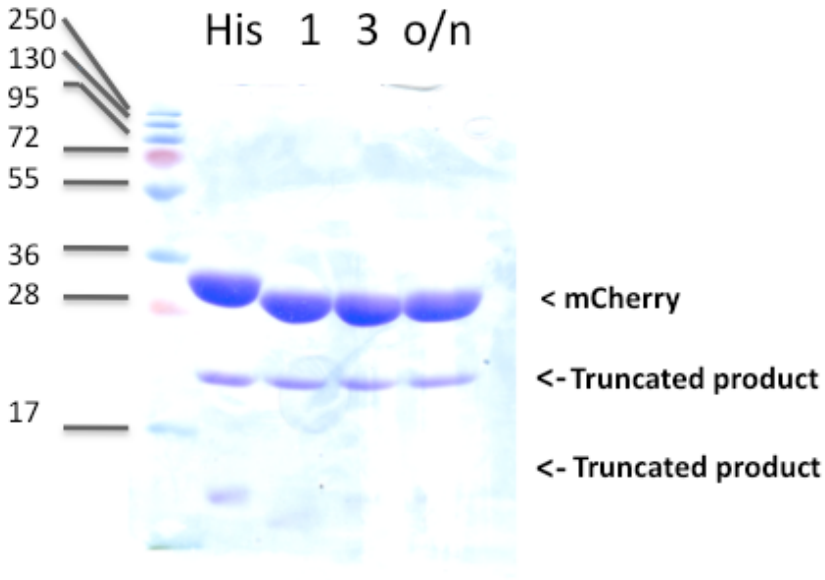


Figure 6.5: Thrombin cleavage of mCherry observed after 1h, 3h and overnight (o/n) incubation. (15% SDS-PAGE gel was used). Full-length His_6-tagged mCherry (His) is observed in the first lane (His). The molecular weight ladder is shown on the left. The arrows point towards truncated products.

Additional protein separation methods like gel-filtration and ion-exchange chromatography were carried out to enable the separation of the truncated product from the full length protein.

Gel filtration separates molecules according to differences in size as they pass through a gel filtration medium packed in a column. This technique also provides a means of determining the molecular weight of a given protein. This is usually done by comparing an elution volume parameter, such as K_{av} of the protein of interest with the values obtained for several known calibration standards.

$$K_{av} = V_e - V_o / V_t - V_o$$

where V_e = elution volume for the protein

V_o = column void volume = elution volume for Blue Dextran 2000

V_t = total bed volume

The void volume of the column was calculated by taking the elution volume of blue dextran (Fig. 6.6a). Fig. 6.6b demonstrates the elution curves for the protein calibration standards run on the HiLoad 16/60 Superdex 75 gel filtration column. The corresponding standard curve is plotted for the K_{av} values for individual protein standards (on the linear scale) against the corresponding molecular weight (on the logarithmic scale) (Fig. 6.6c). Fig. 6.6d demonstrates a single elution peak obtained for the mCherry protein sample. The estimated molecular weight of this complex is found to be approximately 50.1 kDa, which is close to the combined estimated weights of 27.1 kDa mCherry and the 20 kDa truncated product. These results indicate that the mCherry protein migrates as a tightly associated complex with the truncated product and does not resolve as a separate peak. The eluent when analyzed on the gel also resulted in two bands corresponding to the full-length protein and the truncated product (Fig. 6.6 e).

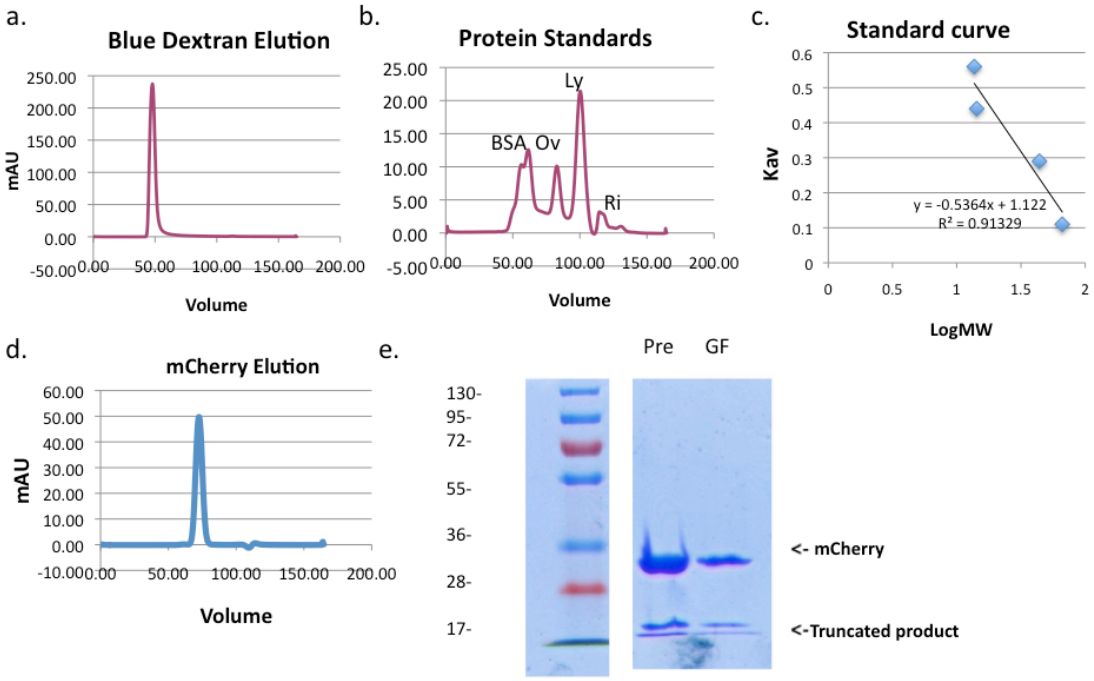


Figure 6.6: Gel-filtration of mCherry protein. (a) The void volume of the column was calculated according to the elution volume of blue dextran. (b and c) The column was calibrated using protein standards (BSA, ovalbumin (Ov), lysozyme (Ly) and Ribonuclease (Ri) for plotting the standard curve for molecular weight estimation of the mCherry complex. (d) mCherry was eluted as a single peak. (e) 10% SDS-PAGE gel was run to analyze the eluate. Pre: mCherry before gel filtration. GF: mCherry after gel filtration. Molecular weight marker is shown on the left. Top arrow points towards full length mCherry, while the bottom arrow points towards the truncated product.

Since gel filtration failed to separate the full-length mCherry protein and its truncated version, ion-exchange chromatography was tried as an alternative option. Ion exchange chromatography (IEX) is a form of an adsorption chromatography, which separates molecules on the basis of charge. After dialysing mCherry overnight in ion-exchange buffer (see Materials and Methods), mCherry was applied on a negatively charged MonoQ column. After washing off unbound material, elution was carried out by increasing the concentration of NaCl over time. As can be observed (Fig. 6.7 a), mCherry elutes as a single peak, along with its truncated version. Eluted fractions along with the flow-through were analyzed by SDS-PAGE (Fig. 6.7 b). These

results confirm the strong association between full length mCherry and its truncated version.

a. b.

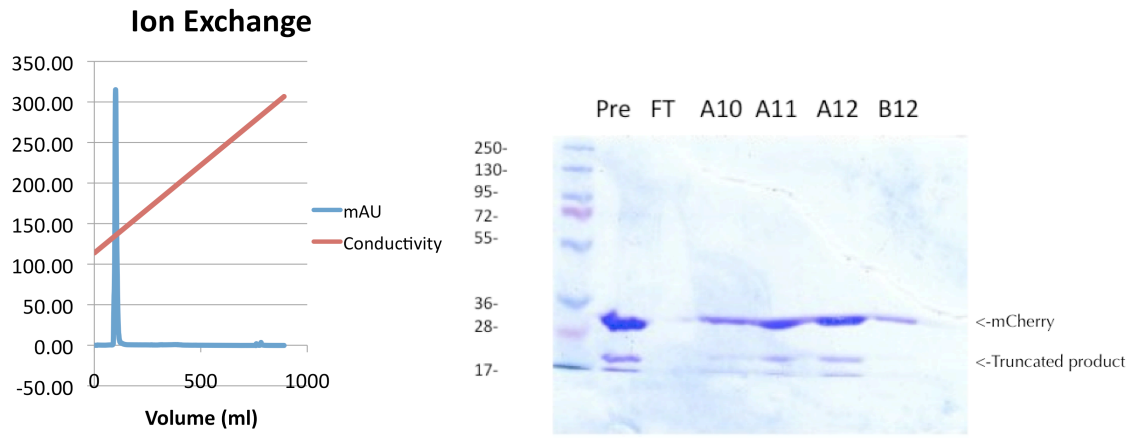


Figure 6.7: Ion-exchange chromatography of mCherry. Thrombin-cleaved mCherry was loaded onto MonoQ column. The flow-through was collected. (a) After washing off unbound material, mCherry was eluted by an increasing gradient of salt denoted by a rise in the conductivity. (b) 10% SDS-PAGE gel was run to analyze the protein fractions. Pre: Sample before ion-exchange chromatography. FT:Flow-through. A10-12, B12: Elution fractions. Molecular weight ladder is shown on the left. Performed in collaboration with Dr. Matthew Watson.

In order to verify the identity of the full length and the major truncated product, MALDI (Matrix-assisted laser desorption/ionization) mass fingerprinting of the tryptic digests of these proteins was carried out by isolating the protein bands from the SDS-PAGE gel (Fig. 6.8 a). The molecular masses of the peptide fragments of the full-length protein, match up with the mCherry protein sequence almost completely (Fig. 6.8b). There are a few fragments that are not detected as they might be too small in size. However, the truncated protein has several bands missing that are present in the full-length protein (Fig. 6.8c).

a.

b.

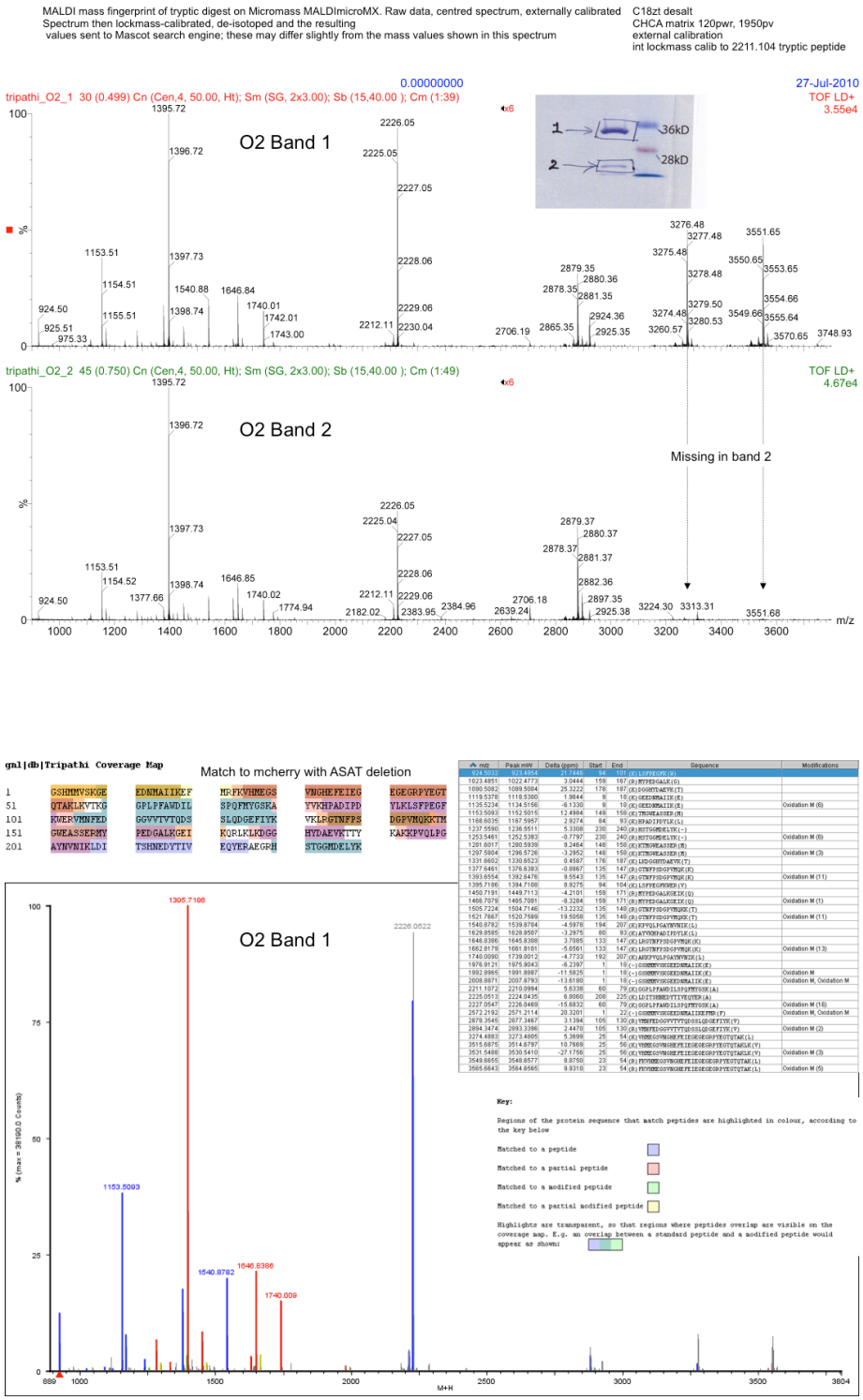


Figure 6.8: MALDI mass-fingerprinting analysis of full length and truncated mCherry protein. The tryptic digestion results in a series of peptide fragments whose masses are shown for (a) full-length mCherry (O2 Band 1) and (b) truncated mCherry (O2 band 2). The peptide fragments match up almost entirely with full length mCherry sequence. Peptide sequences that are detected are marked in colour, while those residues that are absent are uncoloured. (Performed by Protein Analysis Services, Department of Biochemistry).

c.

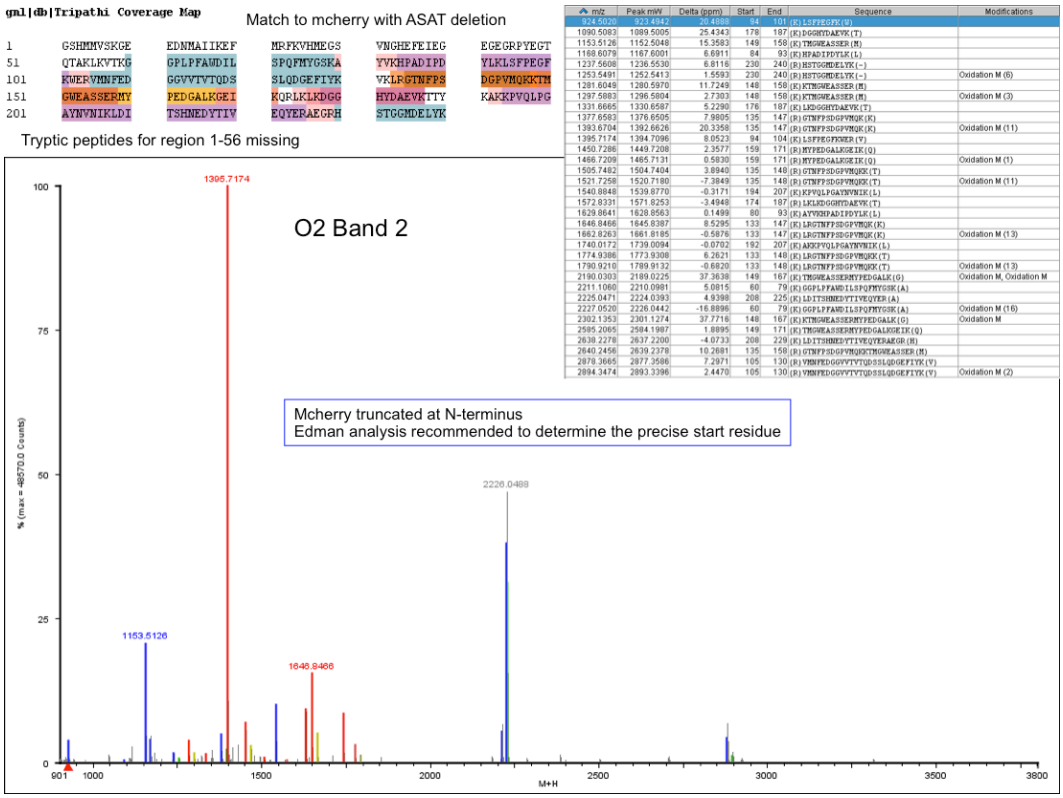


Figure 6.8: (Continued) (c) Peptide fragments corresponding to the truncated product (O2 Band 2). The N-terminal end of mCherry is missing. Matching peptide fragments are colored, while unmatched residues are uncolored. (Performed by Protein Analysis Services, Department of Biochemistry).

The truncated protein on the other hand lacks the first 59 amino acids at the N-terminal end (Fig. 6.8c). The exact position of the truncation cannot be determined in this analysis. Edman sequencing might help in determining the cleavage site. From the translated protein sequence of thrombin cleaved mCherry, its mass should be 27134.6Da. However, the observed mass of mCherry by mass spectrometry is 27113.8+/-1Da which is 20 Da less than expected. The current MALDI analysis is accurate to the level of 1Da, so the difference observed might be real. The fluorophore maturation process of mCherry involves the formation of an 'acylimine' by an oxidation step that results in the loss an H_2O_2 moiety and an additional proton whose approximate mass is in the range of 20 Da (Shu et al., 2006). It is possible that these modifications might result in the observed minor differences in the molecular

mass of mCherry. Similar differences have also been reported in the DsRed fluorophore (Gross et al., 2000).

6.2.2 Expression and purification of ArLEA1A, ArLEA1B, AavLEA1 and AavLEA1-FLAG-mCherry

pET-28a(+) vector containing N-terminally truncated versions of ArLEA1A, ArLEA1B (Pouchkina-Stantcheva et al., 2007) and AavLEA1-FLAG-mCherry (see Materials and Methods) were transformed into BL21DE3 *E. coli* strain for expression of recombinant proteins. Similarly, pET-15b vector containing AavLEA1 (Goyal et al., 2003), was also used to transform BL21DE3 strain for protein expression. Single colonies were inoculated in 5 ml of LB. Overnight cultures were subsequently used to inoculate 400 ml LB and grown at 30°C, until OD_{600} 0.6. Protein expression was then induced with the addition of 1 mM IPTG. Cultures were further grown overnight and harvested the next day. The bacteria were pelleted and lysed by sonication at 4°C. Proteins were subject to purification using the QIAGEN Ni-NTA columns (Fig. 6.9 a,b,c,d) and subsequent thrombin cleavage (Fig. 6.10 a,b,c,d).

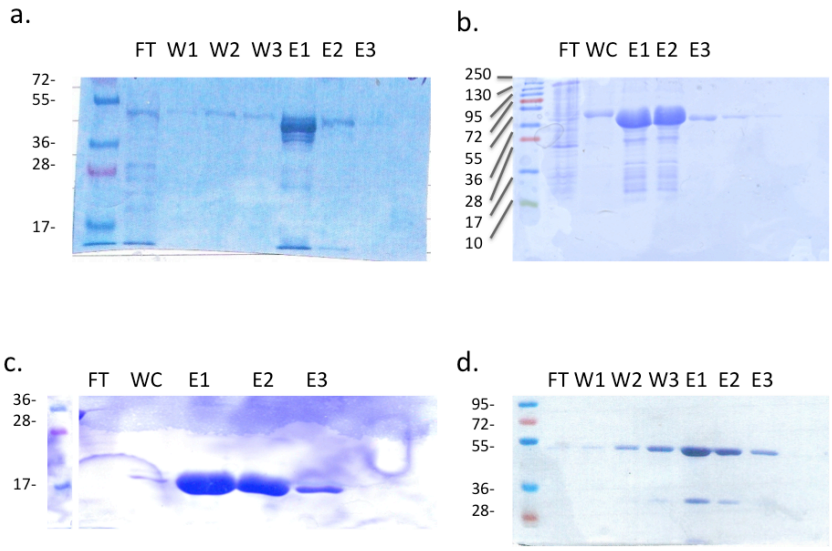


Figure 6.9: Purification of His_6-tagged (a) ArLEA1A, (b) ArLEA1B, (c) AavLEA1 and (d) AavLEA1-FLAG-mCherry. 10% SDS-PAGE gels were used. FT: Flowthrough, W1-W3: wash fractions, WC: wash fractions combined, E1-E3: Elution fractions.

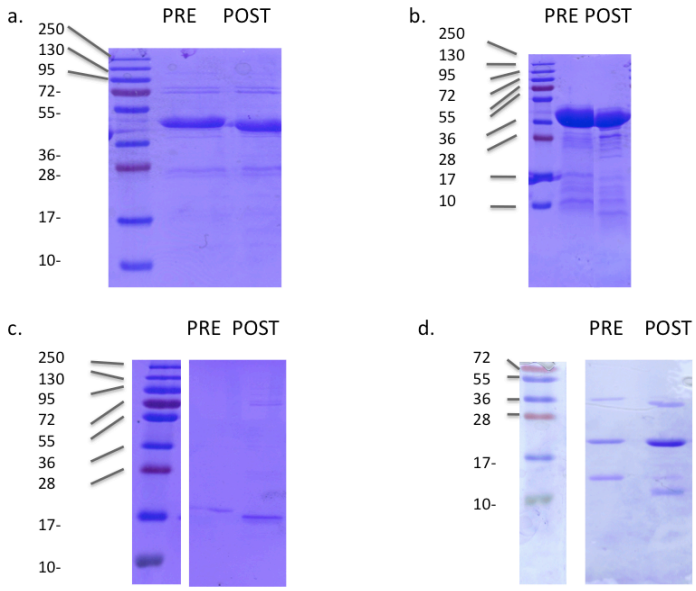
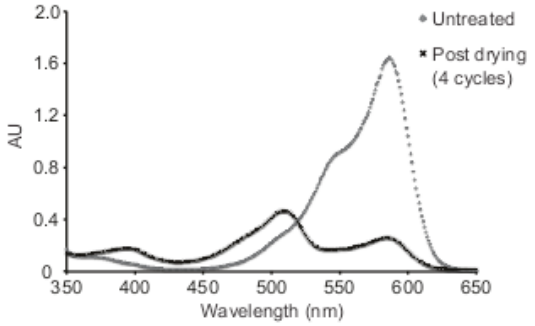


Figure 6.10: Thrombin cleavage of (a) ArLEA1A, (b) ArLEA1B, (c) AavLEA1 and (d) AavLEA1-FLAG-mCherry. 15% SDS-PAGE gels were used. Pre: Pre-thrombin cleavage. Post: Post-thrombin cleavage. Molecular weight ladder is shown on the left.

6.2.3 Monitoring mCherry absorbance and fluorescence upon drying

mCherry was subjected to four cycles of drying and the absorbance and fluorescence spectra were recorded before and after drying. There is a gradual loss of fluorescence and absorbance after one, two and four drying cycles, however for the sake of simplicity, data for only before and after four drying cycles is shown. Both spectra are observed to be substantially altered, indicating damage to the protein's fluorophore, possibly induced by loss of water molecules (Fig. 6.11a,b). In the absorption spectrum, the maximum at 587 nm was found to decrease with the number of drying cycles. Moreover, new absorption peaks were identified at 395 and 509 nm after drying. The fluorescence emission spectrum also decreased dramatically in intensity upon desiccation, consistent with the absorption data. Further, excitation at 395 and 509 nm did not give rise to appreciable levels of fluorescence, suggesting that the properties of the fluorophore had been lost entirely.

a.



b.

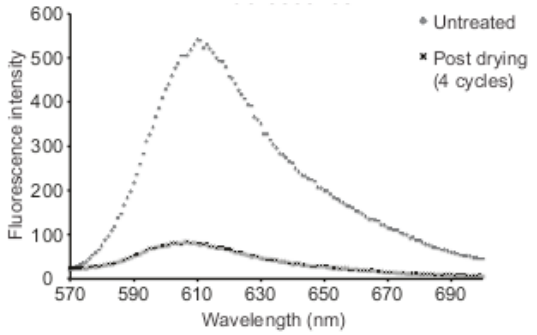


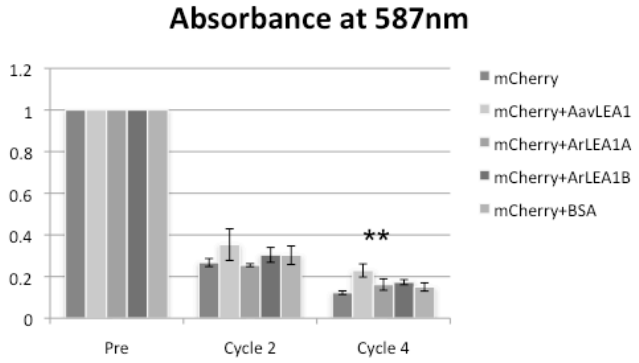
Figure 6.11: (a) Absorbance and (b) fluorescence emission spectra of mCherry, before and after four cycles of drying and rehydration. (Data obtained in collaboration with Dr. Matthew Watson).

6.2.4 ArLEA1A, ArLEA1B and AavLEA1 partially protect mCherry absorbance and fluorescence properties upon desiccation

In order to test the protective effects of LEA proteins on mCherry spectral properties during drying, bdelloid LEA proteins ArLEA1A and ArLEA1B, along with AavLEA1 from *A. avenae*, were mixed with mCherry in 5:1 molar ratio. Bovine serum albumin (BSA) was used as a control. Up to four drying cycles were performed and the changes in mCherry absorbance and fluorescence were recorded. Fig. 6.12a shows changes in mCherry absorbance values at a wavelength of 587 nm. As can be observed, there is gradual reduction in A_{587} of mCherry after two and four drying cycles in all cases tested. Intriguingly, only AavLEA1 resulted in moderate level of protection of mCherry at a 5:1 molar ratio after four cycles of desiccation, with

p<0.01 (Fig. 6.12 a). In contrast, ArLEA1A, ArLEA1B and BSA offered very limited protection to mCherry, both after two and four drying cycles.

a.



b.

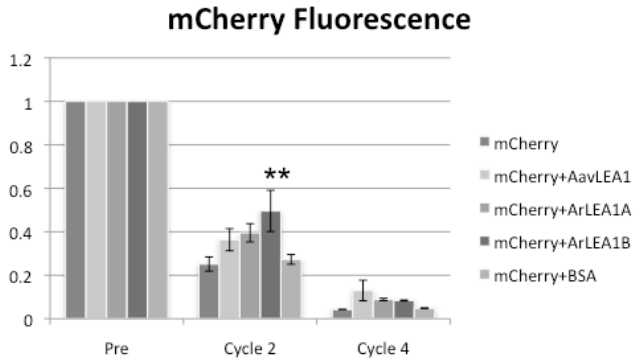


Figure 6.12: LEA proteins provide partial protection to mCherry upon drying. (a) Effect of drying on mCherry absorbance at 587nm in the absence or presence of AavLEA1, ArLEA1A, ArLEA1B or BSA at molar ratios of 1:5. (b) Effect of drying on mCherry fluorescence emission at 615nm in the absence or presence of AavLEA1, ArLEA1A, ArLEA1B or BSA at molar ratios of 1:5. Measurements after two and four cycles of drying and rehydration are shown. Data were normalised, with the absorbance and fluorescence values of the untreated sample represented as 1. All experiments were carried out in triplicate. Error bars indicate +-1SD. **:p<0.01.

Fig. 6.12b shows the changes in mCherry fluorescence emission at 615nm after two and four cycles of drying. The trends observed in fluorescence measurements after two drying cycles are slightly different compared to the changes observed in the absorbance values as one observes that ArLEA1B offered the best protection to mCherry fluorescence (p<0.01), followed by ArLEA1A, AavLEA1 and BSA in decreasing order. However, after four drying cycles no protein was found to significantly protect mCherry fluorescence, although AavLEA1 resulted in some amount of protection.

6.2.5 Fusion to AavLEA1 does not improve protection of mCherry against desiccation damage

If drying mCherry in the presence of AavLEA1 provides some protection of its absorbance properties, one might expect that a covalent linkage of the two proteins would increase the effective local concentration of AavLEA1 with respect to mCherry and afford better protection. Therefore, an AavLEA1-mCherry fusion protein was tested for protection of mCherry after four cycles of drying. However, as can be observed, this proved to be ineffective, since drying of the fusion protein (Fig. 6.13) gave similar results to mCherry alone.

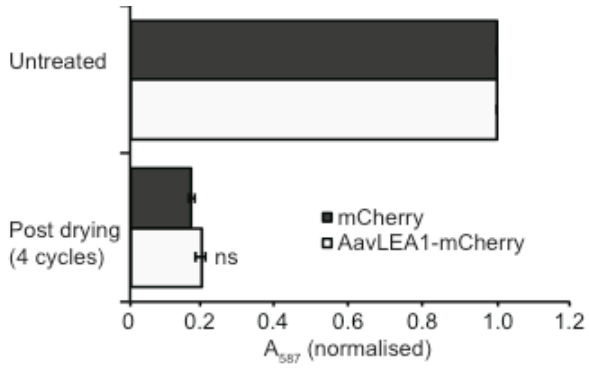


Figure 6.13: Effect of desiccation on mCherry and an AavLEA1-mCherry fusion protein. Absorbance at 587 nm was measured before and after four cycles of drying and rehydration. Data were normalised, with the absorbance of the untreated sample represented as 1. All experiments were carried out in triplicate; error bars indicate +-1 SD. As can be observed, the difference between AavLEA1-mCherry after four cycles of drying is statistically not significant (ns). Data was obtained in collaboration with Dr. Matthew Watson.

6.3 Discussion

Water is crucial for guiding the folding of protein molecules and determining their secondary, tertiary and quarternary structures. Moreover, water molecules have also been shown to be important in determining the quantum chemistry of fluorophores of fluorescent proteins like mCherry (Topol et al., 2011). In this chapter, I have analysed the effects of water loss on the absorbance and fluorescence properties of mCherry and assessed the functions of LEA proteins in preserving these properties upon desiccation.

mCherry protein was expressed and purified in *E. coli*, revealing its unique properties that have not been so far discussed in the literature. Firstly, mCherry is prone to truncation at the N-terminal end, although the exact position of this truncation is currently not determined. It is not known whether this truncation might affect mCherry absorbance and fluorescence, however despite the presence of the truncated protein, my mCherry preparation exhibits the expected properties. Secondly, the estimated mass of mCherry by MALDI mass fingerprinting analysis falls short of the theoretical value by 20 Da. The current MALDI analysis is accurate to 1 Da, so the difference observed might be real. It is possible that these differences might be caused by the fluorophore maturation process of mCherry that involves the removal of H_2O_2 and deprotonation of the -X-Y-G- amino acids into an acylimine residue (Shu et al., 2006). This modification delocalises the fluorophore electron density over the polypeptide backbone and results in the observed red-shift in the excitation and emission spectra of the molecule. Similar differences have also been reported in the DsRed fluorophore (Gross et al., 2000).

Subjecting mCherry to four drying cycles is observed to greatly reduce the fluorescence and absorbance of this protein. Loss of water is predicted to affect both the structural integrity of the 11-stranded beta-barrel structure of mCherry and its internal fluorophore. The beta sheets in the barrel consist of beta strands connected laterally by at least two or three hydrogen bonds, forming a twisted, pleated sheet. A number of salt bridges and hydrogen bonds are formed in this structure with the aid of surrounding water molecules. Moreover, the fluorophore structure of mCherry is also found to be dependent on the surrounding hydrogen bonded water network (Topol et al., 2011). Hence it is hardly surprising that removal of water molecules surrounding mCherry significantly impacts its absorbance and fluorescence properties. Moreover this damage is irreversible and does not restore mCherry properties upon rehydration.

Further, I have assessed the effects of LEA proteins in providing some amount of protection to mCherry during water loss. LEA proteins ArLEA1A and AavLEA1 have been classically proposed to function as molecular shields, whereby they have been shown to prevent the inter-molecular interactions between aggregating proteins

during desiccation. It is not known whether these proteins offer any intramolecular protection to protein structures upon water loss. Drying mCherry in the presence of these proteins should help validate this hypothesis. It is observed that mCherry suffers marked decrease in absorbance and emission properties after four drying cycles regardless of whether other proteins, like ArLEA1A, ArLEA1B, AavLEA1 or the control protein BSA, are present or not. Hence these results support the hypothesis that LEA proteins mostly act as shields and are quite ineffective in preventing intramolecular changes in protein structures caused by water loss.

However, limited protection of mCherry fluorescence is mediated by ArLEA1B after two drying cycles. ArLEA1B has mean Kyte-Doolittle hydrophobicity -0.58 with alternating hydrophilic and hydrophobic patches that might allow it to partially protect both the hydrophilic and hydrophobic regions of mCherry from denaturation. However, after four drying cycles, only AavLEA1 is able to provide very slight protection of mCherry fluorescence and absorbance. These results suggest AavLEA1 might be able to associate loosely with mCherry and partially replace the water network surrounding the beta-barrel structure and the fluorophore due to its highly hydrophilic nature. It is also possible that it might sequester water molecules, thereby reducing water loss. Alternatively it has been proposed that that disordered regions of LEA proteins might gain structure on interaction with misfolded client proteins, thus allowing the client to partially unfold through an entropy transfer effect, and then follow the correct folding pathway leading to its native conformation (Tompa et al., 2008).

If AavLEA1 is able to offer moderate protection to mCherry, it can be imagined that fusing this protein to mCherry *in cis* might enable greater protection of mCherry during desiccation. However, the data shows that the fusion protein AavLEA1-mCherry is as sensitive to desiccation damage as mCherry by itself. These results are contrary to recent reports that have suggested that the presence of an unstructured regions in proteins correlates with desiccation tolerance of proteomes in organisms (Krisko et al., 2010). Singh et al. have also shown that covalently linked LEA proteins can decrease the potential for aggregation under some circumstances (Singh et al., 2005). It is possible that a high molar ratio of AavLEA1 might be more desirable to

mediate sufficient protection of the mCherry 3D structure than using fusion proteins containing AavLEA1 and mCherry in 1:1 ratio.

Overall, these results provide evidence regarding the importance of water molecules in determining the spectral properties of mCherry fluorescent protein, and how LEA proteins like ArLEA1A, ArLEA1B and AavLEA1 might provide only very limited protection to these properties in the absence of water. Engineering key amino acid residues surrounding the mCherry fluorophore might enhance its tolerance to desiccation. Adopting principles of the state of disorder and hydrophilicity of LEA proteins in fluorescent protein structures might enable the derivation of brighter and more photo-stable molecules. Imaging dried biological structures using such desiccation-resistant fluorescent proteins might enable better understanding of the principles of anhydrobiosis and bio-stability *in vivo*.

Chapter Seven: Conclusions and Future Work

Nature demonstrates tremendous flexibility that allows both unicellular and multi-cellular organisms to thrive in harsh geochemical or physical conditions that might seem detrimental to most life forms. The survival of bdelloid rotifers *A. ricciae* in the absence of water is one such fine example of robustness observed in organisms to perpetuate in extreme conditions. This phenomenon has been termed anhydrobiosis (Giard, 1894) or 'life without water'. The biochemical and cellular processes that form the basis of the adaptive response observed in these animals remain largely unknown. Various other anhydrobiotic organisms like brine shrimps (Clegg, 1965, 1967, 2001; Crowe and Clegg, 1978), nematodes (Erkut et al., 2011; Goyal et al., 2005a), plants (Quillet and Soulet, 1964) and tardigrades (Westh and Ramlov, 1991) have been shown to rely on the presence of non-reducing disaccharides like trehalose to protect their tissues in the absence of water. However, the absence of these sugars in bdelloids (Tunnacliffe and Lapinski, 2003) has led to the search for other factors that might be involved in regulating this phenomenon.

Recently, the discovery of two late embryogenesis abundant proteins, ArLEA1A and ArLEA1B, has raised the possibility that these proteins might play a central role during anhydrobiosis in bdelloids (Pouchkina-Stantcheva et al., 2007). LEA proteins are a broad family implicated in desiccation tolerance in a variety of organisms like plants, insects, worms and bacteria (Tunnacliffe and Wise, 2007). ArLEA1A has been shown to prevent aggregation of proteins while ArLEA1B has been found to associate with membranes, possibly playing a role in membrane stabilisation during anhydrobiosis. However, these functions have been deduced *in vitro* and it is possible that these proteins might play a more multifaceted role in the physiological environment of cells. Hence, it was decided to examine the functional and structural properties of these proteins in greater detail in this current study.

Preliminary findings indicate that bdelloid LEA proteins have evolved as an independent evolutionary clade raising the possibility that the functions of these proteins might be unique compared to LEA proteins found in other anhydrobiotic organisms. Moreover, it is speculated that LEA genes might have been acquired by HGT in rotifers from other organisms. It has been recently shown that bdelloid rotifers have acquired many foreign genes of bacterial, plant and fungal origin through this mechanism (Gladyshev et al., 2008). This genetic transfer is thought to be facilitated by membrane disruption and DNA fragmentation and repair occurring during anhydrobiosis and is presumably an adaptive response to desiccation in these animals. HGT has also been implicated in the acquisition of novel metabolic functions in a variety of prokaryotic and eukaryotic species. For example nematodes have presumably acquired genes encoding cell wall degrading enzymes from bacteria and/or fungi through this mechanism as a key adaptation towards parasitism and pathogenicity (Smant, 1998). However, it is hard to be completely confident of the occurrence of HGT events, as the topology of a given phylogenetic tree can never be 100% accurate (Bapteste et al., 2004; Doyon et al., 2011).

Moreover, ArLEA1A and ArLEA1B have been shown to demonstrate homology with PvLEA1 and ankyrin. These results hint towards their role in vitrification and membrane anchorage respectively. Verifying these properties might provide additional clues related to their functions during anhydrobiosis.

The fact that ArLEA1B is predicted to polymerise into a filamentous structure lends further support to the hypothesis that it might a play a role in stabilising membranes and other integral membrane proteins by forming a supportive scaffolding. Structural characterisation of ArLEA1B by NMR or electron microscopy might provide additional insights regarding its conformation.

So far there has been little evidence available regarding the role of LEA proteins in interacting with cell signalling molecules. Scansite predicts numerous kinase-binding, serine-threonine kinase binding and SH2-interacting motifs in both ArLEA1A and ArLEA1B. Verifying these predictions by two-hybrid assays or tandem affinity purification followed by mass spectrometry might identify the target signalling molecules of ArLEA1A and ArLEA1B during anhydrobiosis. It can be speculated that

bdelloids might respond to water loss by triggering various stress response pathways and bdelloid LEA proteins might acquire novel functions in collaboration with various signalling kinases to protect cells during anhydrobiosis. This possibility remains to be tested.

Current data seems to suggest that bdelloid LEA proteins are present in both the hydrated (McGee, 2006) and desiccating animals (Pouchkina-Stantcheva et al., 2007). One might speculate that these proteins might prepare hydrated animals to respond quickly to dehydrating conditions. One cannot rule out the possibility that these proteins might acquire special functions upon water-loss. Moreover, the observation that these proteins might be truncated to smaller peptides (McGee, 2006) suggests that these proteins might be processed to smaller active molecules within cells whose functions currently remain unknown. Characterisation of truncated peptides by mass spectrometry might elucidate their properties. Recently, several LEA proteins have been identified in *Arabidopsis* seeds as components of the phosphoproteome (Irar et al., 2006). Similar analysis of the proteome in bdelloids might result in the identification of physiologically relevant post-translational modifications in LEA proteins.

Previously it has been predicted that bdelloid LEA proteins might localise to the ER inside cells. Confocal microscopy has enabled the visualisation of these proteins within the ER in living cells for the first time. This localisation has been shown to be dependent on the N-terminal signal sequences in these proteins. A novel function of the putative ER retention signal ATEL in regulating the distribution of proteins in both the ER and Golgi compartments has been elucidated. This is in contrast with other conventional ER retention signals that normally restrict protein distribution in the ER compartments. ATEL has also been shown to regulate the secretion of proteins to the cell exterior. These results provide novel insights regarding how LEA protein concentration might be modulated in the entire secretory system both inside and outside cells, a property that might be crucial to their functions in protecting bdelloid tissues during anhydrobiosis.

The endoplasmic reticulum is the site of synthesis and folding of a large fraction of the total proteins present in a cell. Therefore, it can be envisaged that the

protein flux through the ER must be carefully monitored for abnormalities including the buildup of misfolded proteins during anhydrobiosis. Since ArLEA1A has been demonstrated to prevent protein aggregation *in vitro* (Pouchkina-Stantcheva et al., 2007), it was hypothesised that it might be able to protect cells from the deleterious effects of protein misfolding and aggregation within the mammalian ER. Thus, a cell-based system consisting of aggregating polyQ proteins in the ER-Golgi compartments was devised. However, it was demonstrated that instead of providing protection, bdelloid LEA protein overexpression in the mammalian ER reduced cell viability. These results alert us to the possibility of heterologous proteins being toxic in mammalian cells and hence they should be used with caution for any future therapeutic purposes in diseases caused by protein aggregation within the ER like neurodegeneration and diabetes (Lin et al., 2008). Alternatively it is possible that this effect observed might be solely due to 'ER overload' (Ewbank and Pujol, 2010) that might disturb ER homeostasis by inhibiting translation of other ER resident proteins. It is also possible that mCherry fusions might result in the observed cytotoxicity. Optimising expression within this compartment might be an option to observe any beneficial effects of these proteins. Previously Chakrabortee et al. have looked at the effects of overexpression of cytoplasmic LEA proteins in cells expressing polyQ proteins and have found that LEA proteins reduced the number of aggregates observed in mammalian cells (Chakrabortee et al., 2007). Moreover these proteins have been shown to act as kinetic stabilisers of aggregating polyQ proteins, reducing the number of observed aggregates in cells over time (Liu et al., 2011). Although these results hint at their possible functions as molecular shields, it is not clear whether LEA expression in eukaryotic cells is toxic. It might be also useful to test the cytotoxicity of these proteins in mammalian cells.

The protective ability of bdelloid LEA proteins could also be tested using other cell-based systems expressing ER-aggregating proteins like α-crystallin (Gong et al., 2009), or proinsulin and human growth hormone fused with conditional aggregation domain (Rivera et al., 2000). This might enable the development of strategies that could stabilise secreted proteins within the ER and enhance ER quality processes resulting in production of bioactive proteins like human hormones and antibodies in cell expression systems. TMAO or Trimethylamine N-oxide might be one such

compound that could be applied to enhance ER quality control since it has been shown to alleviate ER stress caused by the aggregation of α-crystallin in the ER by increasing the level of chaperone proteins like HSP70 (Gong et al., 2009). Indeed, this molecule is also found to slightly improve the viability of cells expressing polyQ proteins within the ER.

Lastly, an elegant experiment involving the use of mCherry whose fluorophore structure has been shown to be dependent on the presence of surrounding water molecules has shown that bdelloid LEA proteins and AavLEA1 from *A. avenae* species are only able to provide partial protection of its fluorescence and absorbance upon drying. These results demonstrate how LEA proteins might be ineffective in preserving structures of surrounding protein molecules during anhydrobiosis. Bdelloids and other anhydrobiotic organisms might rely on other factors such as chaperone proteins to mediate this protection that might be crucial for cell survival. Identification of these factors in the desiccation-induced gene set of bdelloids might enable the use of such molecules to stabilise therapeutic proteins such as vaccines, monoclonal antibodies, anti-coagulation factors and enzymes that are critically dependent on their structural integrity for their biological functions. Discovery of such stabilising proteins might enable the storage of such therapeutic agents at ambient temperatures, enabling their delivery to areas with little access to modern storage facilities.

Hence, although our understanding of ArLEA1A and ArLEA1B functions has improved, more investigations regarding their mode of action are warranted. It remains to be seen how their expression might be regulated inside cells. Tools like RNA interference and gene-knock outs in bdelloids might help us understand LEA protein functions in these animals. Such efforts are currently underway in the Tunnacliffe laboratory. Recently, cross-species RNAi has been applied to silence genes in anhydrobiotic worms *Panagrolaimus superbus* using DNA from *Aphelenchus avenae* species since *A.avenae* was found to be recalcitrant to gene silencing due to technical difficulties (Reardon et al., 2010). This resulted in the identification of two genes, one novel gene of unknown function and the other encoding glutathioine peroxidase which might function as an anti-oxidant during

anhydrobiosis. A similar strategy could be applied in bdelloids to identify other components of the anhydrobiotic toolkit.

The rapid pace of development of novel technologies like genomics, transcriptomics, proteomics and systems biology should enable the deciphering of molecular mechanisms of these proteins in greater detail and result in the identification of additional factors involved in the phenomenon of anhydrobiosis. The Tunnacliffe laboratory has recently generated both genome and transcriptome assemblies (www.bdelloid.org) which are currently being analysed for the presence of novel genes and pathways that might play a role during anhydrobiosis in bdelloid rotifers. The molecular circuitry of anhydrobiosis is expected to include chaperone proteins, proteins that protect cells against DNA damage, signalling kinases and molecules that protect cells against reactive oxygen species. Some of these factors have indeed been identified by gene-expression profiling of human cells subject to desiccation (Huang et al., 2010; Huang and Tunnacliffe, 2005). However, understanding the actual interplay between different factors in the cellular environment will always remain a challenge due to the lack of sufficient diversity in fluorescently coloured proteins. Future expansion of the colour palette of fluorescent proteins might result in the better understanding of physiological processes like anhydrobiosis inside cells.

Bibliography

Abu-Abied, M., Golomb, L., Belausov, E., Huang, S., Geiger, B., Kam, Z., Staiger, C.J., and Sadot, E. (2006). Identification of plant cytoskeleton-interacting proteins by screening for actin stress fiber association in mammalian fibroblasts. Plant Journal *48*, 367-379.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research *25*, 3389-3402.

Andrews, N.W. (2000). Regulated secretion of conventional lysosomes. Trends in Cell Biology *10*, 316-321.

Arakawa, T., and Timasheff, S.N. (1982). Stabilization of protein-structure by sugars. Biochemistry *21*, 6536-6544.

Arrasate, M., Mitra, S., Schweitzer, E.S., Segal, M.R., and Finkbeiner, S. (2004). Inclusion body formation reduces levels of mutant Huntingtin and the risk of neuronal death. Nature *431*, 805-810.

Baird, G.S., Zacharias, D.A., and Tsien, R.Y. (2000). Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. Proceedings of the National Academy of Sciences of the United States of America *97*, 11984-11989.

Baker, E.H., Bradford, K.J., Bryant, J.A., and Rost, T.L. (1995). A comparison of desiccation-related proteins (dehydrin and QP47) in peas (*Pisum sativum)*. Seed Science Research *5*, 185-193.

Bannykh, S.I., Nishimura, N., and Balch, W.E. (1998). Getting into the Golgi. Trends in Cell Biology *8*, 21-25.

Baker, J., Steele, C., and Dure, L. (1988). Sequence and characterization of 6 LEA proteins and their genes from cotton. Plant Molecular Biology *11*, 277-291.

Bapteste, E., Boucher, Y., Leigh, J., and Doolittle, W.F. (2004). Phylogenetic reconstruction and lateral gene transfer. Trends in Microbiology *12*, 406-411.

Bates, G.P. (2005). History of genetic disease - The molecular genetics of Huntington disease - a history. Nature Reviews Genetics *6*, 766-773.

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

Battista, J.R., Park, M.J., and McLemore, A.E. (2001). Inactivation of two homologues of proteins presumed to be involved in the desiccation tolerance of plants sensitizes *Deinococcus radiodurans* R1 to desiccation. Cryobiology *43*, 133-139.

Bennett, V., and Baines, A.J. (2001). Spectrin and ankyrin-based pathways: Metazoan inventions for integrating cells into tissues. Physiological Reviews *81*, 1353-1392.

Bernales, S., Papa, F.R., and Walter, P. (2006). Intracellular signalling by the unfolded protein response. Annual Review of Cell and Developmental Biology, 487-508.

Bevis, B.J., and Glick, B.S. (2002). Rapidly maturing variants of the *Discosoma* red fluorescent protein (DsRed). Nature Biotechnology *20*, 83-87.

Black, V.H. (1972). Development of smooth-surfaced endoplasmic reticulum in adrenal cortical cells of fetal guinea-pigs. American Journal of Anatomy *135*, 381-417.

Bies-Etheve, N., Gaubier-Comella, P., Debures, A., Lasserre, E., Jobet, E., Raynal, M., Cooke, R., and 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*, 107-124.

Birky, C.W. Jr. (2004). Bdelloid rotifers revisited. Proceedings of the National Academy of Sciences of the United States of America *101*, 2651-2652.

Blobel, G., and Dobberstein, B. (1975). Transfer of proteins across membranes .1. Presence of proteolytically processed and unprocessed nascent immunoglobulin light-chains on membrane-bound ribosomes of murine myeloma. Journal of Cell Biology *67*, 835-851.

Blobel, G., and Sabatini, D.D. (1971). Ribosome-membrane interaction in eukaryotic cells. In Biomembranes. (ed. Manson, L.A.) 193-195.

Bolte, S., and Cordelières, F.P. (2006). A guided tour into subcellular colocalization analysis in light microscopy. Journal of Microscopy *224*, 213-232

Bogdanov, M., and Dowhan, W. (1999). Lipid-assisted protein folding. Journal of Biological Chemistry *274*, 36827-36830.

Bokor, M., Csizmók, V., Kovács, D., Bánki, P., Friedrich, P., Tompa, P., and Tompa, K. (2005). NMR relaxation studies on the hydrate layer of intrinsically unstructured proteins. Biophysical Journal *88*, 2030-2037.

Borovskii, G.B., Stupnikova, I.V., Antipina, A.I., Downs, C.A., and Voinikov, V.K. (2000). Accumulation of dehydrin-like-proteins in the mitochondria of cold-treated plants. Journal of Plant Physiology *156*, 797-800.

Borst, J.W., and Visser, A.J.W.G. (2010). Fluorescence lifetime imaging microscopy in life sciences. Measurement Science & Technology *21*.

Boschetti, C., Pouchkina-Stantcheva, N., Hoffmann, P., and Tunnacliffe, A. (2011). Foreign genes and novel hydrophilic protein genes participate in the desiccation response of the bdelloid rotifer *Adineta ricciae*. Journal of Experimental Biology *214*, 59-68.

Boucher, Y., Douady, C.J., Papke, R.T., Walsh, D.A., Boudreau, M.E.R., Nesbo, C.L., Case, R.J., and Doolittle, W.F. (2003). Lateral gene transfer and the origins of prokaryotic groups. Annual Review of Genetics *37*, 283-328.

Bousse, L., and Parce, W. (1994). Applying silicon micromachining to cellular-metabolism. IEEE Eng Med. Biol. Mag. *13*, 396-401.

Braakman, I., and Bulleid, N.J. (2011). Protein folding and modification in the mammalian endoplasmic reticulum. Annual Review of Biochemistry *80*, 71-99.

Brodsky, J.L., Goeckeler, J., and Schekman, R. (1995). BiP and Sec63p are required for both co- and posttranslational protein translocation into the yeast endoplasmic-reticulum. Proceedings of the National Academy of Sciences of the United States of America *92*, 9643-9646.

Browne, J., Tunnacliffe, A., and Burnell, A. (2002). Anhydrobiosis - Plant desiccation gene found in a nematode. Nature *416*, 38-38.

Browne, J.A., Dolan, K.M., Tyson, T., Goyal, K., Tunnacliffe, A., and Burnell, A.M. (2004). Dehydration-specific induction of hydrophilic protein genes in the anhydrobiotic nematode *Aphelenchus avenae*. Eukaryotic Cell *3*, 966-975.

Burgess, T. L., and Kelly, R. B. (1987). Constitutive and regulated secretion of proteins. Annual Review of Cell Biology. 243-294.

Burke, M.J. (1986). The vitreuos state and survival of anhydrous biological systems. In Membranes, metabolism and dry organisms. (ed. Burke,M.J.) 358-364.

Butlin, R. (2002). Evolution of sex: The costs and benefits of sex: new insights from old asexual lineages. Nature Review Genetics *3*, 311-317.

Campbell, S.A., and Close, T.J. (1997). Dehydrins: genes, proteins, and associations with phenotypic traits. New Phytologist *137*, 61-74.

Caprioli, M., Katholm, A.K., Melone, G., Ramlov, H., Ricci, C., and Santo, N. (2004). Trehalose in desiccated rotifers: a comparison between a bdelloid and a monogonont species. Comparative Biochemistry and Physiology A-Molecular & Integrative Physiology *139*, 527-532.

Chakrabarti, A., Chen, A.W., and Varner, J.D. (2011). A review of the mammalian unfolded protein response. Biotechnology and Bioengineering *108*, 2777-2793.

Chakrabortee, S., Boschetti, C., Walton, L.J., Sarkar, S., Rubinsztein, D.C., and 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*, 18073-18078.

Chakrabortee, S., Meersman, F., Schierle, G.S.K., Bertoncini, C.W., McGee, B., Kaminski, C.F., and Tunnacliffe, A. (2010). Catalytic and chaperone-like functions in an intrinsically disordered protein associated with desiccation tolerance. Proceedings of the National Academy of Sciences of the United States of America *107*, 16084-16089.

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

Chevenet, F., Brun, C., Banuls, A.L., Jacq, B., and Christen, R. (2006). TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics *7*, 439-447.

Chothia, C. (1992). Proteins - 1000 families for the molecular biologist. Nature *357*, 543-544.

Clegg, J.S. (1965). Origin of trehalose and its significance during formation of encysted dormant embryos of *Artemia salina*. Comparative Biochemistry and Physiology *14*, 135-143.

Clegg, J.S. (1967). Metabolic studies of cryptobiosis in encysted embryos of *Artemia salina*. Comparative Biochemistry and Physiology *20*, 801.

Clegg, J.S. (1974). Biochemical adaptations associated with embryonic dormancy of *Artemia salina*. Transactions of the American Microscopical Society *93*, 481-490.

Clegg, J.S. (1978). Hydration dependent metabolic transitions and the state of water in *Artemia* cysts. In Dry Biological Systems. (ed. Crowe, J.H. and Clegg, J.S.) 117-154.

Clegg, J.S. (1986). The physical properties and metabolic status of *Artemia* cysts at low water contents the water replacement hypothesis. In Membranes, Metabolism and Dry Organisms. (ed. Leopold, C.)

Clegg, J.S. (2001). Cryptobiosis - a peculiar state of biological organization. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology *128*, 613-624.

Clegg, J.S., and Evans, D.R. (1962). Free glycerol in dormant cysts of the brine shrimp, *Artemia salina*. American Zoologist *2*, 399-399.

Clegg, J.S., Seitz, P., Seitz, W., and Hazlewood, C.F. (1982). Cellular-responses to extreme water-loss - the water-replacement hypothesis. Cryobiology *19*, 306-316.

Close, T.J. (1996). Dehydrins: Emergence of a biochemical role of a family of plant dehydration proteins. Physiologia Plantarum *97*, 795-803.

Close, T.J., Fenton, R.D., Yang, A., Asghar, R., DeMason, D.A., Crione, D., Meyer, N.C., and Moonan, F. (1993). Dehydrin: the protein. In Plant Responses to Cellular Dehydration During Environmental Stress. (ed. Close, T.J., Bray, E.A.). American Society of Plant Physiologists, Rockville, MD, 104–118

Close, T.J., Kortt, A.A., and Chandler, P.M. (1989). A cDNA-based comparison of dehydration-induced proteins (dehydrins) in barley and corn. Plant Molecular Biology *13*, 95-108.

Crowe, J.H., Carpenter, J.F., and Crowe, L.M. (1998). The role of vitrification in anhydrobiosis. Annual Review of Physiology *60*, 73-103.

Crowe, J.H., and Clegg, J.S. (1973). In Anhydrobiosis, Editor's comments. 226–231.

Crowe, J.H., and Clegg, J.S. (1978). In Dry Biological Systems. (ed. Crowe, J.H. and Clegg, J.S.).

Crowe, J.H., Hoekstra, F.A., and Crowe, L.M. (1992). Anhydrobiosis. Annual Review of Physiology *54*, 579-599.

Crowe, J.H., and Madin, K.A. (1974). Anhydrobiosis in tardigrades and nematodes. Transactions of the American Microscopical Society *93*, 513-524.

Cuming, A.C. (1999). LEA Proteins. In Seed Proteins. (ed. Shewry, P.R. and Casey, R.) 753-780.

Curry, J., and Walker-Simmons, M.K. (1993). Unusual sequence of group-3 LEA (ii) messenger-RNA inducible by dehydration stress in wheat. Plant Molecular Biology *21*, 907-912.

Dancourt, J., and Barlowe, C. (2010). Protein sorting receptors in the early secretory pathway. Annual Review of Biochemistry *79*, 777-802.

Danyluk, J., Perron, A., Houde, M., Limin, A., Fowler, B., Benhamou, N., and Sarhan, F. (1998). Accumulation of an acidic dehydrin in the vicinity of the plasma membrane during cold acclimation of wheat. Plant Cell *10*, 623-638.

Dereeper, A., Audic, S., Claverie, J.M., and Blanc, G. (2010). BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evolutionary Biology *10*, 8-13.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M., and Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Research *36*, 465-469.

Deshaies, R.J., and Schekman, R. (1987). A yeast mutant defective at an early stage in import of secretory protein precursors into the endoplasmic reticulum. Journal of Cell Biology *105*, 633-645.

Deshaies, R.J., and Schekman, R. (1989). SEC62 encodes a putative membrane-protein required for protein translocation into the yeast endoplasmic-reticulum. Journal of Cell Biology *109*, 2653-2664.

Do, C.B. and Katoh, K. (2008). Protein multiple sequence alignment. Functional Proteomics: Methods and Protocols *484*, 379-413.

Douglas, E.S., Hsiao, S.C., Onoe, H., Bertozzi, C.R., Francis, M.B., and Mathies, R.A. (2009). DNA-barcode directed capture and electrochemical metabolic analysis of single mammalian cells on a microelectrode array. Lab Chip *9*, 2010-2015.

Doyon, J.P., Ranwez, V., Daubin, V., and Berry, V. (2011). Models, algorithms and programs for phylogeny reconciliation. Briefings in Bioinformatics *12*, 392-400.

Dure, L. (1993). A repeating 11-mer amino-acid motif and plant desiccation. Plant Journal *3*, 363-369.

Dure, L. (2001). Occurrence of a repeating 11-mer amino acid sequence motif in diverse organisms. Protein Pept. Lett. *8*, 115-122.

Dure, L., Crouch, M., Harada, J., Ho, T.H.D., Mundy, J., Quatrano, R., Thomas, T., and Sung, Z.R. (1989). Common amino-acid sequence domains among the LEA proteins of higher-plants. Plant Molecular Biology *12*, 475-486.

Dure, L., Greenway, S.C., and Galau, G.A. (1981). Developmental biochemistry of cottonseed embryogenesis and germination - changing messenger ribonucleic-acid populations as shown by *in vitro* and *in vivo* protein-synthesis .14. Biochemistry *20*, 4162-4168.

Edgar, R.C. (2004a). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.

Edgar, R.C. (2004b). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research *32*, 1792-1797.

Elder, A.D., Domin, A., Schierle, G.S.K., Lindon, C., Pines, J., Esposito, A., and Kaminski, C.F. (2009). A quantitative protocol for dynamic measurements of protein interactions by Forster resonance energy transfer-sensitized fluorescence emission. Journal of the Royal Society Interface *6*, 59-81.

Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protocols *2*, 953-971.

Eom, J.W., Baker, W.R., Kintanar, A., and Wurtele, E.S. (1996). The embryo-specific EMB-1 protein of *Daucus carota* is flexible and unstructured in solution. Plant Science *115*, 17-24.

Erkut, C., Penkov, S., Khesbak, H., Vorkel, D., Verbavatz, J.M., Fahmy, K., and Kurzchalia, T.V. (2011). Trehalose renders the dauer larva of *Caenorhabditis elegans* resistant to extreme desiccation. Current Biology *21*, 1331-1336.

Esposito, A., Choimet, J.B., Skepper, J.N., Mauritz, J.M.A., Lew, V.L., Kaminski, C.F., and Tiffert, T. (2010). Quantitative Imaging of human red blood cells infected with *Plasmodium falciparum*. Biophysical Journal *99*, 953-960.

Ewbank, J.J., and Pujol, N. (2010). Cellular homeostasis: Coping with ER overload during an immune response. Current Biology *20*, 452-455.

Felsenstein, J. (1996). Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. Computer Methods for Macromolecular Sequence Analysis *266*, 418-427.

Finkbeiner, S., and Mitra, S. (2008). The ubiquitin-proteasome pathway in Huntington's disease. The Scientific World Journal *8*, 421-433.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E.L., Eddy, S.R., and Bateman, A. (2010). The Pfam protein families database. Nucleic Acids Research *38*, 211-222.

Fitch, W.M. (1970). Distinguishing homologous from analogous proteins. Systematic Zoology *19*, 99-113.

Fontaneto, D., Herniou, E.A., Boschetti, C., Caprioli, M., Melone, G., Ricci, C., and Barraclough, T.G. (2007). Independently evolving species in asexual bdelloid rotifers. PLoS Biology *5*, 914-921.

Frank, J.H., Elder, A.D., Swartling, J., Venkitaraman, A.R., Jeyasekharan, A.D., and Kaminski, C.F. (2007). A white light confocal microscope for spectrally resolved multidimensional imaging. Journal of Microscopy *227*.

Frankel, N., Carrari, F., Hasson, E., and Iusem, N.D. (2006). Evolutionary history of the Asr gene family. Gene *378*, 74-83.

Fujikawa, S., and Takabe, K. (1996). Formation of multiplex lamellae by equilibrium slow freezing of cortical parenchyma cells of mulberry and its possible relationship to freezing tolerance. Protoplasma *190*, 189-203.

Gabaldon, T. (2005). Evolution of proteins and proteomes: a phylogenetics approach. Evolutionary Bioinformatics *1*, 51-61.

Gaff, D.F. (1989). Responses of desiccation-tolerant 'resurrection' plants to water stress. In Structural and functional responses to environmental stresses (ed. Krebb, K.H., Richter,H. and Hinkley, T.M.). 1602-1606.

Gal, T.Z., Glazer, I., and Koltai, H. (2004). An LEA group 3 family member is involved in survival of *C.elegans* during exposure to stress. FEBS Letters *577*, 21-26.

Galau, G.A., and Dure, L. (1981). Developmental biochemistry of cottonseed embryogenesis and germination - changing messenger ribonucleic-acid populations as shown by reciprocal heterologous complementary deoxyribonucleic-acid messenger ribonucleic-acid hybridization. Biochemistry *20*, 4169-4178.

Galau, G.A., Wang, H.Y.C., and Hughes, D.W. (1992). Cotton Lea4 (D19) and LeaA2 (D132) group-1 lea genes encoding water stress-related proteins containing a 20-amino acid motif. Plant Physiology *99*, 783-788.

Galau, G.A., Wang, H.Y.C., and Hughes, D.W. (1993). Cotton LEA5 and LEA14 encode atypical late embryogenesis-abundant proteins. Plant Physiology *101*, 695-696.

Garay-Arroyo, A., Colmenero-Flores, J.M., Garciarrubio, A., and Covarrubias, A.A. (2000). Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. Journal of Biological Chemistry *275*, 5668-5674.

Gatchel, J.R., and Zoghbi, H.Y. (2005). Diseases of unstable repeat expansion: Mechanisms and common principles. Nature Review Genetics *6*, 743-755.

Giard, A. (1894). L'anhydrobiose ou ralentissement des phe´no- me`nes vitaux. . C R Soc Biol Paris *46*, 497–500.

Gilles, G.J., Hines, K.M., Manfre, A.J., and Marcotte, W.R. (2007). A predicted N-terminal helical domain of a group 1 LEA protein is required for protection of enzyme activity from drying. Plant Physiology and Biochemistry *45*, 389-399.

Gilmore, R., Walter, P., and Blobel, G. (1982). Protein translocation across the endoplasmic-reticulum 2. Isolation and characterization of the signal recognition particle receptor. Journal of Cell Biology *95*, 470-477.

Gladyshev, E., and Meselson, M. (2008). Extreme resistance of bdelloid rotifers to ionizing radiation. Proceedings of the National Academy of Sciences of the United States of America *105*, 5139-5144.

Gladyshev, E.A., Meselson, M., and Arkhipova, I.R. (2008). Massive horizontal gene transfer in bdelloid rotifers. Science *320*, 1210-1213.

Goerlich, D., Prehn, S., Hartmann, E., Kalies, K.U., and Rapoport, T.A. (1992). A mammalian homolog of SEC61p and SECYp is associated with ribosomes and nascent polypeptides during translocation. Cell *71*, 489-503.

Goldgur, Y., Rom, S., Ghirlando, R., Shkolnik, D., Shadrin, N., Konrad, Z., and Bar-Zvi, D. (2007). Desiccation and zinc binding induce transition of tomato abscisic acid stress ripening 1, a water stress- and salt stress-regulated plant-specific protein, from unfolded to folded state. Plant Physiology *143*, 617-628.

Gong, B., Zhang, L.Y., Pang, C.P., Lam, D.S.C., and Yam, G.H.F. (2009). Trimethylamine N-oxide alleviates the severe aggregation and ER stress caused by G98R alpha A-crystallin. Molecular Vision *15*, 2829-2840.

Goyal, K., Browne, J.A., Burnell, A.M., and Tunnacliffe, A. (2005a). Dehydration-induced tps gene transcripts from an anhydrobiotic nematode contain novel spliced leaders and encode atypical GT-20 family proteins. Biochimie *87*, 565-574.

Goyal, K., Tisi, L., Basran, A., Browne, J., Burnell, A., Zurdo, J., and Tunnacliffe, A. (2003). Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein. Journal of Biological Chemistry *278*, 12977-12984.

Goyal, K., Walton, L.J., and Tunnacliffe, A. (2005b). LEA proteins prevent protein aggregation due to water stress. Biochemical Journal *388*, 151-157.

Grelet, J., Benamar, A., Teyssier, E., Avelange-Macherel, M.H., Grunwald, D., and Macherel, D. (2005). Identification in pea seed mitochondria of a late-embryogenesis abundant protein able to protect enzymes from drying. Plant Physiology *137*, 157-167.

Gross, L.A., Baird, G.S., Hoffman, R.C., Baldridge, K.K., and Tsien, R.Y. (2000). The structure of the chromophore within DsRed, a red fluorescent protein from coral.

Proceedings of the National Academy of Sciences of the United States of America *97*, 11990-11995.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology *59*, 307-321.

Hammond, C., and Helenius, A. (1995). Quality-control in the secretory pathway. Current Opinion in Cell Biology *7*, 523-529.

Hand, S.C., Menze, M.A., Toner, M., Boswell, L., and Moore, D. (2011). LEA proteins during water stress: Not just for plants anymore. Annual Review of Physiology *73*, 115-134.

Hara, M., Fujinaga, M., and Kuboi, T. (2004). Radical scavenging activity and oxidative modification of citrus dehydrin. Plant Physiology and Biochemistry *42*, 657-662.

Herman, E. M., Tague, B. W., Hoffman, L. M., Kjemtrup, S. E., and Chrispeels, M. J. (1990). Retention of phytohemagglutinin with carboxyterminal tetrapeptide KDEL in the nuclear-envelope and the endoplasmic-reticulum. Planta *182*, 305-312.

Hesper, B., and Hogeweg, P. (1970). Bioinformatica: een werkconcept. Leiden: Leidse Biologen Club. Kameleon *1*, 28-29.

Howarth, D.L., Vacaru, A.M., Tsedensodnom, O., Mormone, E., Nieto, N., Costantini, L.M., Snapp, E.L., and Sadler, K.C. (2012). Alcohol disrupts endoplasmic reticulum function and protein secretion in hepatocytes. Alcoholism-Clinical and Experimental Research *36*, 14-23.

Hsing, Y.C., Chen, Z.Y., Shih, M.D., Hsieh, J.S., and Chow, T.Y. (1995). Unusual sequences of group 3 LEA mRNA inducible by maturation or drying in soybean seeds. Plant Molecular Biology *29*, 863-868.

Huang, Z., Banton, M.C., and Tunnacliffe, A. (2010). Modeling anhydrobiosis: activation of the mitogen-activated protein kinase ERK by dehydration in both human cells and nematodes. Journal of Experimental Zoology Part a-Ecological Genetics and Physiology *313A*, 660-670.

Huang, Z., and Tunnacliffe, A. (2005). Gene induction by desiccation stress in human cell cultures. FEBS Letters *579*, 4973-4977.

Hundertmark, M., Dimova, R., Lengefeld, J., Seckler, R., and Hincha, D.K. (2011). The intrinsically disordered late embryogenesis abundant protein LEA18 from *Arabidopsis thaliana* modulates membrane stability through binding and folding. Biochimica et Biophysica Acta- Biomembranes *1808*, 446-453.

Hur, J.H., Van Doninck, K., Mandigo, M.L., and Meselson, M. (2009). Degenerate tetraploidy was established before bdelloid rotifer families diverged. Molecular Biology and Evolution *26*, 375-383.

Irar, S., Oliveira, E., Pages, M., and Goday, A. (2006). Towards the identification of late-embryogenic-abundant phosphoproteome in Arabidopsis by 2-DE and MS. Proteomics *6*, 175-185.

Ismail, A.M., Hall, A.E., and Close, T.J. (1999). Allelic variation of a dehydrin gene cosegregates with chilling tolerance during seedling emergence. Proceedings of the National Academy of Sciences of the United States of America *96*, 13566-13570.

Jeub, M., Herbst, M., Spauschus, A., Fleischer, H., Klockgether, T., Wuellner, U., and Evert, B.O. (2006). Potassium channel dysfunction and depolarized resting membrane potential in a cell model of SCA3. Experimental Neurology *201*, 182-192.

Joh, T., Honjoh, K., Yoshimoto, M., Funabashi, J., Miyamoto, T., and Hatano, S. (1995). Molecular-cloning and expression of hardening-induced genes in *Chlorella*-

vulgaris C-27 - the most abundant clone encodes a late embryogenesis abundant protein. Plant and Cell Physiology *36*, 85-93.

Kalifa, Y., Gilad, A., Konrad, Z., Zaccai, M., Scolnik, P.A., and Bar-Zvi, D. (2004). The water- and salt-stress-regulated Asr1 (abscisic acid stress ripening) gene encodes a zinc-dependent DNA-binding protein. Biochemical Journal *381*, 373-378.

Kaminski Schierle, G.S., Bertoncini, C.W., Chan, F.T.S., Van der Goot, A.T., Schwedler, S., Skepper, J., Schlachter, S., van Ham, T., Esposito, A., Kumita, J.R., Nollen, E.A., Dobson, C.M., and Kaminski, C.F. (2011). A FRET sensor for non-invasive imaging of amyloid formation *in vivo*. Chemphyschem : a European Journal of Chemical Physics and Physical Chemistry *12*, 673-680.

Kaufman, R.J., Scheuner, D., Schröder, M., Shen, X.H., Lee, K., Liu, C.Y., and Arnold, S.M. (2002). The unfolded protein response in nutrient sensing and differentiation. Nature Reviews Molecular Cell Biology *3*, 411-421.

Keenan, R.J., Freymann, D.M., Walter, P., and Stroud, R.M. (1998). Crystal structure of the signal sequence binding subunit of the signal recognition particle. Cell *94*, 181-191.

Keilin, D. (1959). The Leeuwenhoek lecture - the problem of anabiosis or latent life - history and current concept. Proceedings of the Royal Society of London Series B-Biological Sciences *150*, 149-191.

Kikawada, T., Nakahara, Y., Kanamori, Y., Iwata, K.I., Watanabe, M., McGee, B., Tunnacliffe, A., and Okuda, T. (2006). Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid. Biochemical and Biophysical Research Communications *348*, 56-61.

Kim, H.S., Lee, J.H., Kim, J.J., Kim, C.H., Jun, S.S., and Hong, Y.N. (2005). Molecular and functional characterization of CaLEA6, the gene for a hydrophobic LEA protein from *Capsicum annuum*. Gene *344*, 115-123.

Kiyosue, T., Yamaguchi-Shinozaki, K., Shinozaki, K., Higashi, K., Satoh, S., Kamada, H., and Harada, H. (1992). Isolation and characterization of a cDNA that encodes ecp31, an embryogenic-cell protein from carrot. Plant Molecular Biology *19*, 239-249.

Koag, M.C., Fenton, R.D., Wilkens, S., and Close, T.J. (2003). The binding of maize DHN1 to lipid vesicles. Gain of structure and lipid specificity. Plant Physiology *131*, 309-316.

Krisko, A., Smole, Z., Debret, G., Nikolic, N., and Radman, M. (2010). Unstructured hydrophilic sequences in prokaryotic proteomes correlate with dehydration tolerance and host association. Journal of Molecular Biology *402*, 775-782.

Ladle, R.J., Johnstone, R.A., and Judson, O.P. (1993). Coevolutionary dynamics of sex in a metapopulation - escaping the red queen. Proceedings of the Royal Society of London Series B-Biological Sciences *253*, 155-160.

Lapinski, J., and Tunnacliffe, A. (2003). Anhydrobiosis without trehalose in bdelloid rotifers. FEBS Letters *553*, 387-390.

Lauricella, M., D'Anneo, A., Giuliano, M., Calvaruso, G., Emanuele, S., Vento, R. Tesoriere, G. (2003). Induction of apoptosis in human osteosarcoma Saos-2 cells by the proteasome inhibitor MG132 and the protective effect of pRb. Cell Death and Differentiation *10*, 930-932.

Letourneur, F. Gaynor, E. C., Hennecke, S., Demolliere, C., Duden, R., Emr, S. D., Riezman, H., and Cosson, P. (1994). Coatomer is essential for retrieval of dilysine-tagged proteins to the endoplasmic reticulum. Cell *79*, 1199-1207.

Li, D., and He, X. (2009). Desiccation induced structural alterations in a 66-amino acid fragment of an anhydrobiotic nematode late embryogenesis abundant (LEA) protein. Biomacromolecules *10*, 1469-1477.

Lin, J.H., Walter, P., Yen, T.S. Benedict. (2008). Endoplasmic reticulum stress in disease pathogenesis. Annual Review of Pathology-Mechanisms of Disease *3*, 399-425.

Lisse, T., Bartels, D., Kalbitzer, H.R., and Jaenicke, R. (1996). The recombinant dehydrin-like desiccation stress protein from the resurrection plant *Craterostigma plantagineum* displays no defined three-dimensional structure in its native state. Biological Chemistry *377*, 555-561.

Liu, Y., Chakrabortee, S., Li, R., Zheng, Y., and Tunnacliffe, A. (2011). Both plant and animal LEA proteins act as kinetic stabilisers of polyglutamine-dependent protein aggregation. Febs Letters *585*, 630-634.

Lomberk, G.A., Imoto, I., Gebelein, B., Urrutia, R., and Cook, T.A. (2010). Conservation of the TGFbeta/labial homeobox signalling loop in endoderm-derived cells between *Drosophila* and mammals. Pancreatology *10*, 74-84.

Loomis, S.H., Odell, S.J., and Crowe, J.H. (1979). Anhydrobiosis in nematodes - inhibition of the browning reaction of reducing sugars with dry proteins. Journal of Experimental Zoology *208*, 355-360.

Maitra, N., and Cushman, J.C. (1994). Isolation and characterisation of a drought-induced soybean cDNA-encoding a D95 family late-embryogenesis-abundant protein. Plant Physiology *106*, 805-806.

Makarova, K.S., Aravind, L., Wolf, Y.I., Tatusov, R.L., Minton, K.W., Koonin, E.V., and Daly, M.J. (2001). Genome of the extremely radiation-resistant bacterium

Deinococcus radiodurans viewed from the perspective of comparative genomics. Microbiology and Molecular Biology Reviews *65*, 44.

Manning, C.M., Mathews, W.R., Fico, L.P., and Thackeray, J.R. (2003). Phospholipase C-gamma contains introns shared by SRC homology 2 domains in many unrelated proteins. Genetics *164*, 433-442.

Mark Welch, D.B., Cummings, M.P., Hillis, D.M., and Meselson, M.S. (2004). Divergent gene copies in the asexual class Bdelloidea (Rotifera) separated before the bdelloid radiation or within bdelloid families. Proceedings of the National Academy of Sciences of the United States of America *101*, 1622-1625.

Mark Welch, D.B., and Meselson, M.S. (2000). Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science *288*, 1211-1215.

Mark Welch, D.B., Mark Welch, J.L., and Meselson, M. (2008). Evidence for degenerate tetraploidy in bdelloid rotifers. Proceedings of the National Academy of Sciences of the United States of America *105*, 5145-5149.

Marotta, R., Leas, F., Uggetti, A., Ricci, C., and Melone, G. (2010). Dry and survive: Morphological changes during anhydrobiosis in a bdelloid rotifer. Journal of Structural Biology *171*, 11-17.

Martone, M. E., Zhang, Y., Simpliciano, V. M., Carragher, B. O., and Ellisman, M. H. (1993). 3-dimensional visualization of the smooth endoplasmic reticulum in purkinje-cell dendrites. Journal of Neuroscience *13*, 4636-4646.

Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L., and Lukyanov, S.A. (1999). Fluorescent proteins from nonbioluminescent Anthozoa species. Nature Biotechnology *17*, 969-973.

McGee, B. (2006). Hydrophilic proteins in the anhydrobiosis of bdelloid rotifers. PhD Thesis Cambridge: University of Cambridge.

Miller, J., Arrasate, M., Brooks, E., Libeu, C.P., Legleiter, J., Hatters, D., Curtis, J., Cheung, K., Krishnan, P., Mitra, S., Widjaja, K., Shaby, B.A., Lotz, G.P., Newhouse,Y., Mitchell, E.J., Osmand, A., Gray, M., Thulasiramin, V., Saudou, F., Segal, M., Yang,X. W., Masliah, E., Thompson, L.M., Muchowski, P.J., Weisgraber, K. H., and Finkbeiner, S. (2011). Identifying polyglutamine protein species in situ that best predict neurodegeneration. Nature Chemical Biology *7*, 925-934.

Milstein, C., Brownlee, G.G., Harrison, T.M., and Mathews, M.B. (1972). Possible precursor of immunoglobulin light chains. Nature New Biology *239*, 117-120.

Mouillon, J.M., Gustafsson, P., and Harryson, P. (2006). Structural investigation of disordered stress proteins. Comparison of full-length dehydrins with isolated peptides of their conserved segments. Plant Physiology *141*, 638-650.

Mundy, J., and Chua, N.H. (1988). Abscisic-acid and water-stress induce the expression of a novel rice gene. EMBO Journal *7*, 2279-2286.

Munro, S., and Pelham, H. R. B. (1987). A C-terminal signal prevents secretion of luminal ER proteins. Cell *48*, 899-907.

Nierman, W.C., Feldblyum, T.V., Laub, M.T., Paulsen, I.T., Nelson, K.E., Eisen, J., Heidelberg, J.F., Alley, M.R.K., Ohta, N., Maddock, J.R., et al. (2001). Complete genome sequence of *Caulobacter crescentus*. Proceedings of the National Academy of Sciences of the United States of America *98*, 4136-4141.

Nishitoh, H., Matsuzawa, A., Tobiume, K., Saegusa, K., Takeda, K., Inoue, K., Hori, S., Kakizuka, A., and Ichijo, H. (2002). ASK1 is essential for endoplasmic reticulum stress-induced neuronal cell death triggered by expanded polyglutamine repeats. Genes & Development *16*, 1345-1355.

Nozawa, K., Fritzler, M.J., Takasaki, Y., Wood, M.R., and Chan, E.K.L. (2009). Co-clustering of Golgi complex and other cytoplasmic organelles to crescentic region of half-moon nuclei during apoptosis. Cell Biology International *33*, 148-157.

Obenauer, J.C., Cantley, L.C., and Yaffe, M.B. (2003). Scansite 2.0: proteome-wide prediction of cell signalling interactions using short sequence motifs. Nucleic Acids Research *31*, 3635-3641.

Park, J.A., Cho, S.K., Kim, J.E., Chung, H.S., Hong, J.P., Hwang, B., Hong, C.B., and Kim, W.T. (2003). Isolation of cDNAs differentially expressed in response to drought stress and characterization of the Ca-LEAL1 gene encoding a new family of atypical LEA-like protein homologue in hot pepper (*Capsicum annuum* L. cv. Pukang). Plant Science *165*, 471-481.

Pelham, H. R. B. (1989). Heat-shock and the sorting of luminal ER proteins. EMBO Journal *8*, 3171-3176.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods *8*, 785-786.

Plana, M., Itarte, E., Eritja, R., Goday, A., Pages, M., and Martinez, M.C. (1991). Phosphorylation of maize rab-17 protein by casein kinase-2. Journal of Biological Chemistry *266*, 22510-22514.

Potts, M. (1994). Desiccation tolerance of prokaryotes. Microbiological Reviews *58*, 755-805.

Pouchkina-Stantcheva, N.N., McGee, B.M., Boschetti, C., Tolleter, D., Chakrabortee, S., Popova, A.V., Meersman, F., Macherel, D., Hincha, D.K., and Tunnacliffe, A.

(2007). Functional divergence of former alleles in an ancient asexual invertebrate. Science *318*, 268-271.

Prágai, Z., and Harwood, C.R. (2002). Regulatory interactions between the Pho and sigma(B)-dependent general stress regulons of *Bacillus subtilis*. Microbiology *148*, 1593-1602.

Puthalakath, H., O'Reilly, L.A., Gunn, P., Lee, L., Kelly, P.N., Huntington, N.D., Hughes, P.D., Michalak, E.M., McKimm-Breschkin, J., Motoyama, N., Gotoh, T., Akira, S., Bouillet, P. and Strasser, A. (2007). ER stress triggers apoptosis by activating BH3-only protein Bim. Cell *129*, 1337-1349.

Quillet, M., and Soulet, M. (1964). Sur l'accumulation concomitante du saccharose et du tre´halose chez plusieurs espe`ces de Se´laginelles indige`nes et exotiques. C R Acad Sci Paris *259*, 635-637.

Ravikumar, B., Duden, R., and Rubinsztein, D.C. (2002). Aggregate-prone proteins with polyglutamine and polyalanine expansions are degraded by autophagy. Human Molecular Genetics *11*, 1107-1117.

Raykhel, I., Alanen, H., Solo, K., Jurvansuu, J., Nguyen, V.D., Latva-Ranta, M., and Ruddock, L. (2008). A molecular specificity code for the three mammalian KDEL receptors. Journal of Cell Biology *180*, 1193-1204.

Reardon, W., Chakrabortee, S., Pereira, T.C., Tyson, T., Banton, M.C., Dolan, K.M., Culleton, B.A., Wise, M.J., Burnell, A.M., and Tunnacliffe, A. (2010). Expression profiling and cross-species RNA interference (RNAi) of desiccation-induced transcripts in the anhydrobiotic nematode *Aphelenchus avenae*. BMC Molecular Biology *11*.

Ricci, C., and Fontaneto, D. (2009). The importance of being a bdelloid: Ecological and evolutionary consequences of dormancy. Italian Journal of Zoology *76*, 240-249.

Ricci, C., Melone, G., Santo, N., and Caprioli, M. (2003). Morphological response of a bdelloid rotifer to desiccation. Journal of Morphology *257*, 246-253.

Riera, M., Figueras, M., Lopez, C., Goday, A., and Pages, M. (2004). Protein kinase CK2 modulates developmental functions of the abscisic acid responsive protein Rab17 from maize. Proceedings of the National Academy of Sciences of the United States of America *101*, 9879-9884.

Rivera, V.M., Wang, X.R., Wardwell, S., Courage, N.L., Volchuk, A., Keenan, T., Holt, D.A., Gilman, M., Orci, L., Cerasoli, F.*, et al.* (2000). Regulation of protein secretion through controlled aggregation in the endoplasmic reticulum. Science *287*, 826-830.

Ron, D. (2002). Translational control in the endoplasmic reticulum stress response. Journal of Clinical Investigation *110*, 1383-1388.

Ron, D., and Walter, P. (2007). Signal integration in the endoplasmic reticulum unfolded protein response. Nature Reviews Molecular Cell Biology *8*, 519-529.

Rothblatt, J.A., Deshaies, R.J., Sanders, S.L., Daum, G., and Schekman, R. (1989). Multiple genes are required for proper insertion of secretory proteins into the endoplasmic-reticulum in yeast. Journal of Cell Biology *109*, 2641-2652.

Rousseau, E., Dehay, B., Ben-Haiem, A., Trottier, Y., Morange, M., and Bertolotti, A. (2004). Targeting expression of expanded polyglutamine proteins to the endoplasmic reticulum or mitochondria prevents their aggregation. Proceedings of the National Academy of Sciences of the United States of America *101*, 9648-9653.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols *5*, 725-738.

Rutkowski, D.T., Arnold, S.M., Miller, C.N., Wu, J., Li, J., Gunnison, K.M., Mori, K., Akha, A.A.S., Raden, D., and Kaufman, R.J. (2006). Adaptation to ER stress is mediated by differential stabilities of pro-survival and pro-apoptotic mRNAs and proteins. PLoS Biology *4*, 2024-2041.

Saibil, H.R. (2008). Chaperone machines in action. Current Opinion in Structural Biology *18*, 35-42.

Saitou, N., and Nei, M. (1987). The neighbor-joining method - a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution *4*, 406-425.

Sakahira, H., Breuer, P., Hayer-Hartl, M.K., and Hartl, F.U. (2002). Molecular chaperones as modulators of polyglutamine protein aggregation and toxicity. Proceedings of the National Academy of Sciences of the United States of America *99*, 16412-16418.

Sammels, E., Parys, J.B., Missiaen, L., De Smedt, H., and Bultynck, G. (2010). Intracellular Ca(2+) storage in health and disease: A dynamic equilibrium. Cell Calcium *47*, 297-314.

Sammut, S.J., Finn, R.D., and Bateman, A. (2008). Pfam 10 years on: 10,000 families and still growing. Briefings in Bioinformatics *9*, 210-219.

Sanchez, A.M.E., Mejia-Toiber, J., and Massieu, L. (2008). Excitotoxic neuronal death and the pathogenesis of Huntington's disease. Arch Med Res *39*, 265-276.

Santi-Rocca, J., Smith, S., Weber, C., Pineda, E., Hon, C.C., Saavedra, E., Olivos-Garcia, A., Rousseau, S., Dillies, M.A., Coppee, J.Y., and Guillen, N. (2012). Endoplasmic reticulum stress-sensing mechanism is activated in *Entamoeba histolytica* upon treatment with nitric oxide. PLoS One *7*, 31777.

Smant, G., Stokkermans, J.P., Yan, Y. T., de Boer, J. M., Baum, T. J., Wang, X.H.,

Hussey, R. S., Gommers, F. J., Henrissat, B., Davis, E. L., Helder, J., Schots, A., Bakker, J. (1998). Endogenous cellulases in animals: Isolation of beta-1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes. Proceedings of the National Academy of Sciences of the United States of America *95*, 4906-4911.

Schadt, E.E., Sinsheimer, J.S., and Lange, K. (1998). Computational advances in maximum likelihood methods for molecular phylogeny. Genome Research *8*, 222-233.

Schlachter, S., Schwedler, S., Esposito, A., Kaminski Schierle, G.S., Moggridge, G.D., and Kaminski, C.F. (2009). A method to unmix multiple fluorophores in microscopy images with minimal a priori information. Optics Express *17*, 22747-22760.

Scriven, D. R. L., Lynch, R. M., Moore, E.D. (2008). Image acquisition for colocalization using optical microscopy. American Journal of Physiology-Cell Physiology *294*, 1119-1122.

Segers, H., and Shiel, R.J. (2005). Tale of a sleeping beauty: a new and easily cultured model organism for experimental studies on bdelloid rotifers. Hydrobiologia *546*, 141-145.

Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N.G., Palmer, A.E., and Tsien, R.Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma sp* red fluorescent protein. Nature Biotechnology *22*, 1567-1572.

Shimizu, T., Kanamori, Y., Furuki, T., Kikawada, T., Okuda, T., Takahashi, T., Mihara, H., and Sakurai, M. (2010). Desiccation-induced structuralization and glass formation of group 3 late embryogenesis abundant protein model peptides. Biochemistry *49*, 1093-1104.

Shimomura, O., Johnson, F.H., and Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. Journal of Cellular and Comparative Physiology *59*, 223-239.

Shu, X.K., Shaner, N.C., Yarbrough, C.A., Tsien, R.Y., and Remington, S.J. (2006). Novel chromophores and buried charges control color in mFruits. Biochemistry *45*, 9639-9647.

Silhavy, D., Hutvagner, G., Barta, E., and Bánfalvi, Z. (1995). Isolation and characterization of a water-stress-inducible cDNA clone from *Solanum chacoense*. Plant Molecular Biology *27*, 587-595.

Singh, S., Cornilescu, C.C., Tyler, R.C., Cornilescu, G., Tonelli, M., Lee, M.S., and Markley, J.L. (2005). Solution structure of a late embryogenesis abundant protein (LEA14) from *Arabidopsis thaliana*, a cellular stress-related protein. Protein Science *14*, 2601-2609.

Sisson, J.K., and Fahrenbach, W. (1967). Fine structure of steroidogenic cells of a primate cutaneous organ. American Journal of Anatomy *121*, 337-367.

Smith, T.F., and Waterman, M.S. (1981). Identification of common molecular subsequences. Journal of Molecular Biology *147*, 195-197.

Sokka, A.L., Putkonen, N., Mudo, G., Pryazhnikov, E., Reijonen, S., Khiroug, L., Belluardo, N., Lindholm, D., and Korhonen, L. (2007). Endoplasmic reticulum stress inhibition protects against excitotoxic neuronal injury in the rat brain. Journal of Neuroscience *27*, 901-908.

Sonnhammer, E.L., Eddy, S.R., and Durbin, R. (1997). Pfam: A comprehensive database of protein domain families based on seed alignments. Proteins *28*, 405-420.

Soulages, J.L., Kim, K., Arrese, E.L., Walters, C., and Cushman, J.C. (2003). Conformation of a group 2 late embryogenesis abundant protein from soybean. Evidence of poly (L-proline)-type II structure. Plant Physiology *131*, 963-975.

Soulages, J.L., Kim, K., Walters, C., and Cushman, J.C. (2002). Temperature-induced extended helix/random coil transitions in a group 1 late embryogenesis-abundant protein from soybean. Plant Physiology *128*, 822-832.

Stacy, R.A., Nordeng, T.W., Culiánez-Macia, F.A., and Aalen, R.B. (1999). The dormancy-related peroxiredoxin anti-oxidant, PER1, is localized to the nucleus of barley embryo and aleurone cells. Plant Journal *19*, 1-8.

Stirling, C.J., Rothblatt, J., Hosobuchi, M., Deshaies, R., and Schekman, R. (1992). Protein translocation mutants defective in the insertion of integral membrane-proteins into the endoplasmic-reticulum. Molecular Biology of the Cell *3*, 129-142.

Stout, E.P., La Clair, J.J., Snell, T.W., Shearer, T.L., and Kubanek, J. (2010). Conservation of progesterone hormone function in invertebrate reproduction. Proceedings of the National Academy of Sciences of the United States of America *107*, 11859-11864.

Sugars, K.L., and Rubinsztein, D.C. (2003). Transcriptional abnormalities in Huntington disease. Trends in Genetics *19*, 233-238.

Sun, W.Q., and Leopold, A.C. (1997). Cytoplasmic vitrification acid survival of anhydrobiotic organisms. Comparative Biochemistry and Physiology A-Physiology *117*, 327-333.

Svensson, J., Palva, E.T., and Welin, B. (2000). Purification of recombinant *Arabidopsis thaliana* dehydrins by metal ion affinity chromatography. Protein Expression and Purification *20*, 169-178.

Szegezdi, E., Logue, S.E., Gorman, A.M., and Samali, A. (2006). Mediators of endoplasmic reticulum stress-induced apoptosis. EMBO Reports *7*, 880-885.

Tam, S., Spiess, C., Auyeung, W., Joachimiak, L., Chen, B., Poirier, M.A., and Frydman, J. (2009). The chaperonin TRiC blocks a huntingtin sequence element that promotes the conformational switch to aggregation. Nature Structural and Molecular Biology *16*, 1279-1298.

Tolleter, D., Jaquinod, M., Mangavel, C., Passirani, C., Saulnier, P., Manon, S., Teyssier, E., Payet, N., Avelange-Macherel, M.H., and Macherel, D. (2007). Structure and function of a mitochondrial late embryogenesis abundant protein are revealed by desiccation. Plant Cell *19*, 1580-1589.

Tompa, P., and Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. FASEB Journal *18*, 1169-1175.

Tompa, P., and Kovacs, D. (2010). Intrinsically disordered chaperones in plants and animals. Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire *88*, 167-174.

Topol, I., Collins, J., Savitsky, A., and Nemukhin, A. (2011). Computational strategy for tuning spectral properties of red fluorescent proteins. Biophysical Chemistry *158*, 91-95.

Trucksis, M., Michalski, J., Deng, Y.K., and Kaper, J.B. (1998). The *Vibrio cholerae* genome contains two unique circular chromosomes. Proceedings of the National Academy of Sciences of the United States of America *95*, 14464-14469.

Tubbs, J.L., Tainer, J.A., Getzoff, E.D. (2005). Crystallographic structures of Discosoma red fluorescent protein with immature and mature chromophores: Linking peptide bond trans-cis isomerization and acylimine formation in chromophore maturation. Biochemistry *44*, 9833-9840.

Tunnacliffe, A., and Lapinski, J. (2003). Resurrecting van Leeuwenhoek's rotifers: a reappraisal of the role of disaccharides in anhydrobiosis. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences *358*, 1755-1771.

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

Ukaji, N., Kuwabara, C., Takezawa, D., Arakawa, K., and Fujikawa, S. (2001). Cold acclimation-induced WAP27 localized in endoplasmic reticulum in cortical parenchyma cells of mulberry tree was homologous to group 3 late-embryogenesis abundant proteins. Plant Physiology *126*, 1588-1597.

Urano, F., Wang, X.Z., Bertolotti, A., Zhang, Y.H., Chung, P., Harding, H.P., and Ron, D. (2000). Coupling of stress in the ER to activation of JNK protein kinases by transmembrane protein kinase IRE1. Science *287*, 664-666.

Uversky, V.N., and Dunker, A.K. (2010). Understanding protein non-folding. Biochimica et Biophysica Acta-Proteins and Proteomics *1804*, 1231-1264.

Van den Berg, B., Clemons, W.M. Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., and Rapoport, T.A. (2004). X-ray structure of a protein-conducting channel. Nature *427*, 36-44.

Vilardell, J., Goday, A., Freire, M.A., Torrent, M., Martinez, M.C., Torne, J.M., and Pages, M. (1990). Gene sequence, developmental expression, and protein-phosphorylation of RAB-17 in maize. Plant Molecular Biology *14*, 423-432.

Vogel, J.P., Misra, L.M., and Rose, M.D. (1990). Loss of BiP/GRP78 function blocks translocation of secretory proteins in yeast. Journal of Cell Biology *110*, 1885-1895.

von Heijne, G. (1983). Patterns of amino-acids near signal-sequence cleavage sites. European Journal of Biochemistry *133*, 17-21.

von Heijne, G. (1985) Signal sequences. The limits of variation. Journal of Molecular Biology *184*, 99-105.

von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. Nucleic Acids Research *14*, 4683-4690.

Wallace, R.L. (2002). Rotifers: Exquisite metazoans. Integr. Comp. Biol. *42*, 660-667.

Walter, P., and Blobel, G. (1980). Purification of a membrane-associated protein complex required for protein translocation across the endoplasmic reticulum. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences *77*, 7112-7116.

Walter, P., and Blobel, G. (1982). Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. Nature *299*, 691-698.

Wang, H.J., Jauh, G.Y., Hsu, Y.H., and Wang, C.S. (2003). The nuclear localization signal of a pollen-specific, desiccation-associated protein of lily is necessary and sufficient for nuclear targeting. Botanical Bulletin of Academia Sinica *44*, 123-128.

Wang, W., Meng, B., Chen, W., Ge, X., Liu, S., and Yu, J. (2007a). A proteomic study on postdiapaused embryonic development of brine shrimp (*Artemia franciscana*). Proteomics *7*, 3580-3591.

West, S.A., Lively, C.M., and Read, A.F. (1999). A pluralist approach to sex and recombination. Journal of Evolutionary Biology *12*, 1003-1012.

Westh, P., and Ramlov, H. (1991). Trehalose accumulation in the tardigrade *Adorybiotus coronifer* during anhydrobiosis. Journal of Experimental Zoology *258*, 303-311.

Whitby, F.G., Kent, H., Stewart, F., Stewart, M., Xie, X., Hatch, V., Cohen, C., and Phillips, G.N. Jr. (1992). Structure of tropomyosin at 9-angstroms resolution. Journal of Molecular Biology *227*, 441-452.

Whitby, F.G., and Phillips, G.N. (2000). Crystal structure of tropomyosin at *7* Angstroms resolution. Proteins *38*, 49-59.

Wilson, C.G., and Sherman, P.W. (2010). Anciently asexual bdelloid rotifers escape lethal fungal parasites by drying up and blowing away. Science *327*, 574-576.

Wiest, D. L., Burkhardt, J. K., Hester, S., Hortsch, M., Meyer, D. I., and Argon, Y. (1990). Membrane biogenesis during B-cell differentiation - most endoplasmic-reticulum proteins are expressed coordinately. Journal of Cell Biology *110*, 1501-1511.

Wilson, D.W., Lewis, M.J., and Pelham, H.R. (1993). pH-dependent binding of KDEL to its receptor invitro. Journal of Biological Chemistry *268*, 7465-7468.

Wise, M.J. (2002). The POPPs: clustering and searching using peptide probability profiles. Bioinformatics *18*, S38-45.

Wise, M.J., and Tunnacliffe, A. (2004). POPP the question: what do LEA proteins do? Trends in Plant Science *9*, 13-17.

Wolkers, W.F., McCready, S., Brandt, W.F., Lindsey, G.G., and Hoekstra, F.A. (2001). Isolation and characterization of a D-*7* LEA protein from pollen that stabilises glasses in vitro. Biochimica et Biophysica Acta-Protein Structure and Molecular Enzymology *1544*, 196-206.

Wolkers, W.F., van Kilsdonk, M.G., and Hoekstra, F.A. (1998). Dehydration-induced conformational changes of poly-L-lysine as influenced by drying rate and carbohydrates. Biochimica et Biophysica Acta-General Subjects *1425*, 127-136.

Yadav, S., Puri, S., and Linstedt, A.D. (2009). A primary role for Golgi positioning in directed secretion, cell polarity, and wound healing. Molecular Biology of the Cell *20*, 1728-1736.

Yamaguchi-Shinozaki, K., and Shinozaki, K. (1993). The plant hormone abscisic-acid mediates the drought-induced expression but not the seed-specific expression of RD22, a gene responsive to dehydration stress in *Arabidopsis thaliana*. Molecular & General Genetics *238*, 17-25.

Yancey, P.H., Clark, M.E., Hand, S.C., Bowlus, R.D., and Somero, G.N. (1982). Living with water-stress - evolution of osmolyte systems. Science *217*, 1214-1222.

Yang, F., Moss, L.G., Phillips, G.N. Jr. (1996). The molecular structure of green fluorescent protein. Nature Biotechnology *14*, 1246-1251.

Yarbrough, D., Wachter, R.M., Kallio, K., Matz, M.V., and Remington, S.J. (2001). Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-angstrom resolution. Proceedings of the National Academy of Sciences of the United States of America *98*, 462-467.

Yoshimori, T., Semba, T., Takemoto, H., Akagi, S., Yamamoto, A., and Tashiro, Y. (1990). Protein disulfide-isomerase in rat exocrine pancreatic-cells is exported from the endoplasmic-reticulum despite possessing the retention signal. Journal of Biological Chemistry *265*, 15984-15990.

Zegzouti, H., Jones, B., Marty, C., Lelièvre, J.M., Latché, A., Pech, J.C., and Bouzayen, M. (1997). ER5, a tomato cDNA encoding an ethylene-responsive LEA-

like protein: characterization and expression in response to drought, ABA and wounding. Plant Molecular Biology *35*, 847-854.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. BMC Bioinformatics *9*. 40.

Zimmermann, R., Eyrisch, S., Ahmad, M., and Helms, V. (2011). Protein translocation across the ER membrane. Biochimica et Biophysica Acta-Biomembranes *1808*, 912-924.