A novel bacterial Water Hypersensitivity-like protein shows *in vivo* protection against cold and freeze damage

Dominique Anderson[1], Eloy Ferreras[2], Marla Tuffin[1] and Don Cowan[1,2]

[1] Institute for Microbial Biotechnology and Metagenomics, Department of Biotechnology, University of the Western Cape, Bellville 7535, Cape Town, South Africa

[2] Centre for Microbial Ecology and Genomics, Department of Genetics, University of Pretoria, Hatfield 0028, Pretoria, South Africa

Corresponding author:

Professor DA Cowan, Centre for Microbial Ecology and Genomics, NWII, Room 3-12, Private Bag X20, University of Pretoria, Hatfield 0028, Pretoria, Republic of South Africa. Email address: don.cowan@up.ac.za

Tel: Number: +27 12420 5873

Fax: Number: +27 12 420 6870

Keywords: Antarctic soil, Functional metagenomics, Freeze-thaw, Desiccation tolerance, Water Hypersensitivity protein, Late Embryogenesis Abundant protein.

Abstract

Metagenomic library screening, by functional or sequence analysis, has become an established method for the identification of novel genes and gene products, including genetic elements implicated in microbial stress response and adaptation. We have identified, using a sequence based approach, a fosmid clone from an Antarctic desert soil metagenome library containing a novel gene which codes for a protein homologous to a Water Hypersensitive domain (WHy). The WHy domain is typically found as a component of specific LEA (Late Embryogenesis Abundant) proteins, particularly the LEA-14 (LEA-8) variants, which occur widely in plants, nematodes, bacteria and archaea and which are typically induced by exposure to stress conditions. The novel Why-like protein, (165 amino acid, 18.6 kDa) exhibits a largely invariant NPN motif at the N-terminus and has high sequence identity to genes identified in *Pseudomonas* genomes. Expression of this protein in *E. coli* significantly protected the recombinant host against cold and freeze stress.

Introduction

Low water potential is considered to be the most life threatening abiotic stress, and is known to negatively affect all biological functions (Krisko *et al.*, 2010). Water confers structural order to cells, stabilises proteins, lipids and nucleic acids, and maintains a cellular microenvironment in which vital metabolic systems and chemical reactions occur (Billi and Potts, 2002). In prokaryotes, desiccation tolerance is generally attributed to the ability of the microorganisms to efficiently repair DNA damage, scavenge free radicals and accumulate high levels of compatible solutes (Krisko *et al.*, 2010). Various proteins are up-regulated in response to drought stress, including those involved in direct pathways for combating stress (Heat Shock Proteins, transporters and osmoprotectants) and those involved in regulatory processes (transcription factors, signalling proteins and kinases) (Roelofs *et al.*, 2008).

In microorganisms, turgor pressure across the semi-permeable membrane is maintained by controlling the amount of osmotically active solutes in the cytoplasm and is vital for coping with fluctuating osmolarity and water availability in the external milieu (Potts, 1994; Kempf and Bremer, 1998; Mahajan and Tuteja, 2005). For example, potassium ions (K^+) serve as a major intracellular osmolyte for the maintenance of turgor and accumulate to high levels in halophilic microorganisms in order to cope with extreme extracellular salt concentration and osmotic pressure (Csonka, 1989; Billi and Potts, 2002). Similarly, intracellular levels of Mn(II) have been shown to increase during radiation and desiccation stress in *Deinococcus radiodurans* and, while insufficient to provide complete radiation and desiccation tolerance, clearly contribute to the organisms' survival by scavenging toxic oxygen species (Potts *et al.*, 2005). Growth in hypertonic conditions requires the synthesis, or import of osmoprotectant solutes like sugars, free amino acids, polyols, quaternary amines, sulphate esters, inositol phosphates or manosylglyceramides (Imhoff and Rodriguez-Valera, 1984; Csonka, 1989; Potts, 1994; Leslie *et al.*, 1995; Kempf and Bremer, 1998). Several osmoprotectant transport systems are utilised in microorganisms, and the uptake and/ or synthesis of these substances is a tightly regulated and controlled process (Sutherland *et al.*, 1986; Kempf and Bremer, 1998; Frossard *et al.*, 2012).

The recovery of vegetative prokaryotic cells from water deficit is a multifunctional process that not only relies in the accumulation of compatible solutes. Due to the varied affects that water deficit has on cellular processes, it is reasonable to assume that no single gene, or protein could offer complete protection from this stress, and that a number of synergistic mechanisms would be employed. Studies in *D. radiodurans* have shown that 33 of the 72 genes upregulated during radiation stress were also induced in cultures recovering from desiccation stress (Mattimore and Battista, 1996). One particular group of proteins which are consistently upregulated during salt- and osmotic-stress, and are suggested to play an important role in desiccation tolerance are the Late Embryogenesis Abundant (LEA) proteins (Krisko *et al.*, 2010; Tolleter *et al.*, 2010).

LEA proteins were first described over 30 years ago, associated with the late stages of cotton seed development (Dure *et al.*, 1981; Hundertmark *et al.*, 2011), and are primarily found in seeds, pollens and anhydrobiotic plants (Wise, 2002; Singh *et al.*, 2005). Homologous genes have been found in nematodes (*Caenorhabditis elegans, Aphelenchus avenae*), archaea (Campos *et al*., 2013) and bacteria (*Bacillus subtilis, D. radiodurans* and *Haemophilus influenzae*) (Tunnacliffe and Wise, 2007). Although many of these proteins have no proven function, those that have been characterised are generally associated with cellular recovery processes (Kriško *et al.*, 2010; Cuevas-Velazquez *et al.*, 2014). In 2005, a novel domain known as the Water Hypersensitity domain (WHy) was reported (Ciccarelli and Bork, 2005), and provided a link to *Hin1* genes (induced in plants in response to bacterial infection and part of the general stress response pathway) and the plant *Lea14* genes. This domain has also been observed in a number of uncharacterised bacterial and archaeal proteins (Ciccarelli and Bork, 2005). In this study, next generation sequencing and *in silico* data-mining of Antarctic Dry Valley soil metagenomic library clones led to the discovery of a full-length gene encoding a putative bacterial Water Hypersensitivity response protein. The desert soils of the McMurdo Dry Valleys of Eastern Antarctica (Fitzsimons *et al.*, 2001; Hopkins *et al.*, 2006) are considered to be one of the most extreme desert habitats on Earth, where microbial populations are subject to multiple simultaneous abiotic stresses (Wynn-Williams, 1996; Hogg *et al.*, 2006). Low atmospheric humidity in combination with very low levels of precipitation [< 10 mm/ annum] (Cary *et al.*, 2010) result in low water input to the upper soil horizons (Balks and Campbell, 2001). In these sandy gravels, the water potential may be further reduced by the high levels of mineral salts, effectively transforming the region into a cold, hyper-arid desert (Potts, 1994; Mahajan and Tuteja, 2005).

Here, we present the first *in vivo* characterisation of a bacterial Why-like protein. Further functional analysis of this protein may identify novel mechanisms for desiccation survival employed by microorganisms inhabiting desert soil environments.

Materials and Methods

Sample acquisition

Mineral soil samples were recovered aseptically from the McKelvey Valley, Eastern Antarctica (S 78 04; E 163 51) during the 2011 austral summer, stored frozen during transport to the laboratory, and thereafter at -80^{o}C.

Fosmid library construction

Metagenomic DNA was extracted from the soil samples using the Zhou methodology (Zhou *et al.*, 1996). The high molecular weight DNA fraction was cloned into the Epicentre Biotechnologies® CopyControl™ Fosmid Library Production Kit according to manufacturer's specifications.

Next-generation DNA sequencing

A number of fosmid clones, selected on the basis of functional enzymatic properties (data not shown), were sequenced using the Illumina Solexa Next Generation sequencing platform at the University of the Western Cape, Cape Town, South Africa. Short sequence reads (36bp) were generated and assembled using a combination of *de novo* tools (CLC Genomics workbench and Velvet), primer walking and vector end-sequencing.

Bioinformatics analysis

Open reading frames were identified using the open access online tool, FgenesB from SoftBerry (Solovyev and Salamov, 2011). The translated nucleotide sequences for each ORF were used for homology searches in the Uniprot, InterPro and NCBI protein databases. Candidate ORFs encoding genes which could confer tolerance to abiotic stress factors were searched for and 13ORF6 (named *dwhy1* as it contains a domain homologous to the Water Hypersensitive domain described by Ciccarelli and Bork in

2005) was selected. Signal peptide prediction was performed using Phobius (Kall *et al.*, 2004) PrediSi (Hiller *et al.*, 2004) SignalP4.1 (Petersen *et al.*, 2011) and Signal-3L (Chou, 2001; Shen and Chou, 2006; Chou and Shen, 2007; Shen and Chou, 2007). For the transmembrane fragments prediction TMHMM (Krogh *et al.*, 2001), DAS TM prediction server (Cserzo *et al.*, 1997) and TMpred (Hofmann, 1993) were used. General protein characteristics such as amino acid content, isoelectric point (pI), molecular weight and protein Grand Average Hydropathy (GRAVY) were predicted using the ProtParam tool (Expasy: http://web.expasy.org/protparam/). PSIPRED was used to predict the secondary structure of the protein (McGuffin *et al.*, 2000) and regions of protein disorder were predicted by using the gensilico metadisorder service (http://iimcb.genesilico.pl/metadisorder) (Kozlowski and Bujnicki, 2012), Cspritz (http://protein.bio.unipd.it/cspritz/) (Walsh *et al.*, 2011), Disopred3 & Disopred2 (http://bioinf.cs.ucl.ac.uk/psipred/) and Kyte and Doolittle hydrophilicity plots were generated (www.vivo.colostate.edu/molkit). Putative phosphorylation sites were predicted using NetPhosBac 1.0 (Miller *et al.*, 2009). Multiple sequence alignments using Clustal Omega (Sievers *et al.*, 2011) were used to determine conserved nucleotides. Amino acid sequences harbouring the Water Hypersensitive domain described by Ciccarelli and Bork (Ciccarelli and Bork, 2005) were downloaded from the SMART (Simple Modular Architecture Research Tool) non-reduntant database (http://smart.embl-heidelberg.de/) and used to construct neighbour joining phylogenetic trees in Mega (Tamura *et al.*, 2013). The stability of the relationships was assessed by performing bootstrap analysis based on 1000 resamplings. Statistical significance values were calculated in R using the Student's t-test (R Core Team, 2014).

Sub-cloning of dwhy1

The original nucleotide sequence of the *dwhy1* gene was engineered and synthesized by GenScript USA Inc. to avoid codon bias during its expression in *E. coli*. The complete DNA sequence of the optimized *dwhy1* gene and of the mutant without the predicted signal peptide (*dwhy1Δsp*) were amplified using the primers oWNdeIDir AAAACATATGTCCTACCTGGCTAC and oWNdeIΔPSDir AAAACATATGTGTGCGTCATCTGGTA respectively as forwards, with a *Nde*I site (underlined) and, as reverse primer, oWHindIIIRevHisTag AAAAAAGCTTTTAATGGTGGTGGTG with a *Hind*III site (underlined) respectively. PCR products were digested and cloned into pET21a and pET28a vectors (Novagen). Recombinant plasmids were verified by sequencing. *E. coli* BL21 (DE3) strain was used for protein expression. The *dwhy1* gene sequence has been deposited in the NCBI database with the accession number KM111254. The protein sequence is available in Genbank with the accession number: AIS22443.1.

Protein expression

Recombinant clones transformed with either the parental vector or the vector containing *dwhy1* or *dwhy1Δsp* were grown in LB broth supplemented with Ampicillin (100 μg/ml) at 37°C until an OD_{600} of 0.5-0.6 was obtained. Cultures were then induced with 0.5 mM IPTG and grown for 16-24 hours at 30°C. Cultures were centrifuged at 6000 × *g* for 15 minutes. Pellet fractions were resuspended in sonication buffer (50 mM Phosphate buffer pH 7.0, 300 mM NaCl) and sonicated for 4 cycles of 30 seconds each at 40% amplitude on ice. The soluble and insoluble fractions were separated by centrifugation at 36000 × *g* for 60 minutes. Aliquots of all fractions were mixed with Laemmli buffer and analysed by SDS-PAGE.

In vivo assays for freeze tolerance

Cultures transformed with pET21a harbouring the engineered *dwhy1* and *dwhy1Δsp* genes were induced as above. The OD_{600} of each culture was measured spectrophotometrically and corrected with LB broth to 0.6. Serial dilutions (up to 10^{-6}) were prepared in quarter strength Ringer's solution. Each one were frozen 3 times at -80ºC for 20 minutes and thawed at room temperature for 20 minutes. For survival rate analysis, 100 µl of each dilution was spread-plated onto LB agar plates, together with a no treatment control. Following overnight incubation at 37 °C, CFU's were quantified and the percentage survival rates were calculated. All assays were performed three times, in triplicate, and control cultures containing parental vector in the expression host were routinely included.

In vivo cold stress tolerance

Cultures transformed with pET28a harbouring the engineered *dwhy1* and *dwhy1Δsp* were induced as above. After the induction, the OD_{600} was measured and the culture was adjusted to an OD_{600} = 0.1 with ice cold LB supplemented with ampicillin and 0.5 mM IPTG, growth at 8ºC up to 16 days and the OD_{600} monitored. All assays were performed three times, in triplicate, and control cultures containing parental vector in the expression host were routinely included.

Results

Bioinformatic analysis

Clones selected from a fosmid library, constructed from metagenomic DNA from Antarctic soil samples (Anderson, 2012), were sequenced using NGS technology. Sequence assembly, annotation and analysis of open reading frames encoded on Fosmid LD13 (average insert size: 26420 pb, GC content: 47 %; comprising 29 ORFs) indicated sequence identities of 72-99% to proteins from *Pseudomonas caeni*, a bacterium isolated from an anaerobic ammonium-oxidizing bioreactor (Xiao *et al.*, 2009) (NCBI Reference Sequence: NZ_ATXQ00000000.1). We conclude that the cloned DNA fragment of this Fosmid originated from a species of the genus *Pseudomonas* and possibly closely related to *Pseudomonas caeni*.

An open reading frame (13ORF6: also designated *dwhy1*) of 498 bp encoding a 165 amino acid protein with a predicted molecular mass of 18.6 kDa, was identified during sequence analysis. BlastP of the amino acid sequence against the NCBI database revealed that dWHy1 showed high sequence identity to a lipoprotein (NCBI ref number: WP_033421079.1) from *Pseudomonas caeni* (78% identity, E-value $2e^{-86}$). It also showed significant identity to Water Hypersensitivity proteins from *Pseudomonas mendocina* (51 % identity, E-value $2e^{-44}$) and *Pseudomonas pseudoalcaligenes* (50%, E-value $2e^{-58}$) (Figure 1). InterproScan analysis showed that 13ORF6 exhibited the classical NPN motif characteristic of Water Hypersensitivity domains (referred to as the WHy domain) and the atypical LEA-14 protein family [PF03168], (characterized with higher content of hydrophobic residues than typical LEA proteins). Further bioinformatic analysis using signal peptide, transmembrane segment and cleavage site prediction programs suggested the presence of a signal peptide of 25-27 amino acids with a cleavage site for a Signal peptidase II upstream the Cys^{26}. Using the NetPhosBac 1.0 server (Miller *et al.*, 2009), putative phosphorylation sites were predicted, with high probability scores, for Ser^{11}, Ser^{92}, and Thr^{112}. The GRAVY value for dWHy1 was predicted to be -0.087, suggesting a moderate level of overall hydrophilicity.

The results for protein disorder prediction do not suggest the presence of disordered regions within the dWHy1 protein. Only the N-terminal sequence corresponding to the putative signal peptide and the C-terminal domain showed a significant degree of

```
P.mendocina              MLSQAQMIRIISLLMFFGLFSGLTGCSTWMTGSFKDPDVQLIKVDVVKARLLEQEFRLRF 60
P.pseudoalcaligenes      -----------------------------MTGGFKDPDVQLIKVDVVKARLLEQEFRLRF 31
dWHy1                    MSYLATIKKTSAYLIFITLLSGVGGCASSGSSGFKDPDVQLVDVELIHAKLLEQQFMLHF 60
P.caeni                  -------------MVLLALLGSVSGCSTWLTGNFKKPDLQLVDVELVHAKLLEQQFVLHF 47
                                                      :..**.**:**:.*::::*:****:* *:*

P.mendocina              RIDNPNGVNLPVRGLDYNVHLNGMLLAEGHSNEWFTVPAHGHHVFEVPVRTNLWRHVRQI 120
P.pseudoalcaligenes      RIDNPNGVSLPVRGLDYNVHLNGIQLAEGHSNEWFTVPAHGHHTFEVPVRTNLWRHVRQI 91
dWHy1                    RVDNPNSKSLPMRGMDYRILLNNTPLATGSNSQWLTVPAHDYAYFKIPVHTNLWRHMKVV 120
P.caeni                  RVDNPNSKSLPMRSINYRLLLNDTPLATGSNNQWLTVPANGHAYFKIPVHTNLWRHMKVV 107
                       *:****. .**:*.::*.: **.   ** * ..:*:****:.:  *::**:******:: :

P.mendocina              VKALEKPDVPIRYSLKGEVKTGMLFGRRVHMARNGEIIPGDFIPE 165
P.pseudoalcaligenes      VKALEKPNEPIRYSLKGQVKTGMLFGRSVHMARNGEIIPGDFIPE 136
dWHy1                    LRMLENPDQPIHYALHAEVKTGLMFSKKINILRHGDIIPGDYIRE 165
P.caeni                  VRMLEKPDQPIHYALHADVKTGLLFSKKINIVRHGDIIPGDYIRE 152
                         :: **:*: **:*:*:.:****::*.: ::: *:*:*****:* *
```

Figure 1. Clustal Omega alignments. Alignment of the dWHy1 protein with a Water Stress and Hypersensitive response from *Pseudomonas mendocina* ymp (Sequence ID: gb|ABP84182.1|) , a water stress/hypersensitive response protein from *Pseudomonas pseudoalcaligenes* CECT 5344 (Sequence ID: emb|CDM41519.1|) and a lipoprotein from *Pseudomonas caeni* (Sequence ID: ref|WP_033421079.1).

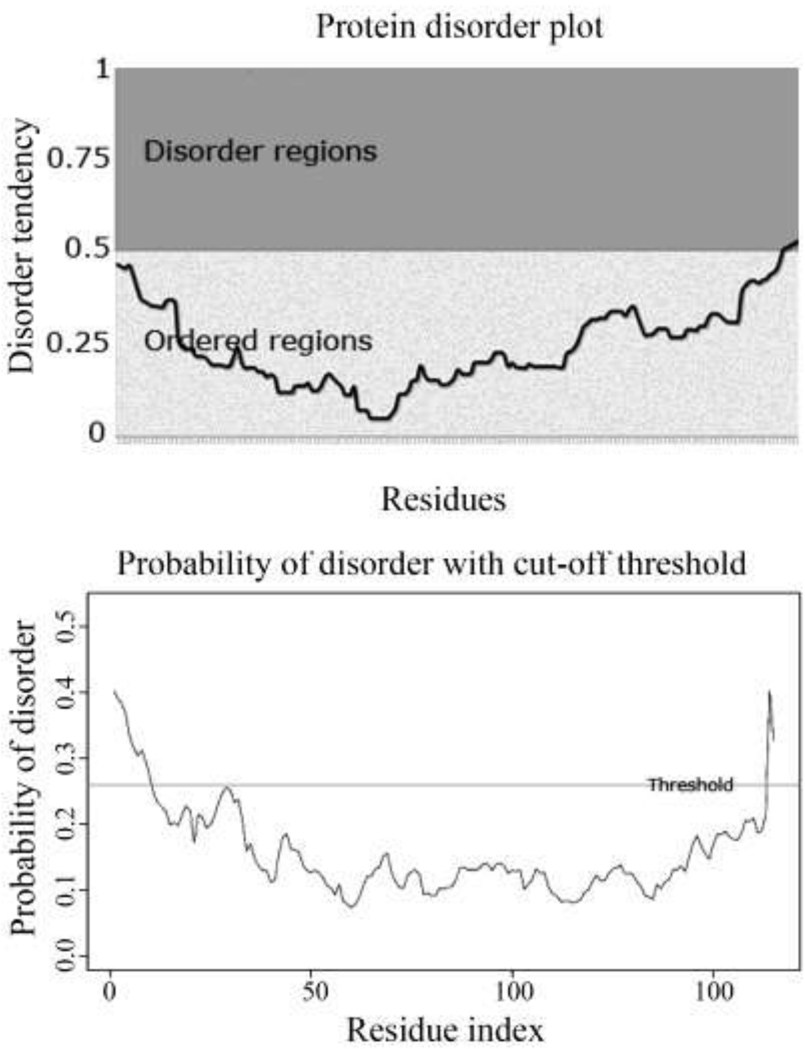

Figure 2. *In silico* prediction of protein disorder. Prediction of dWHy1 disordered regions using Gensilico Meta-disorder (up) and Cspritz (down).

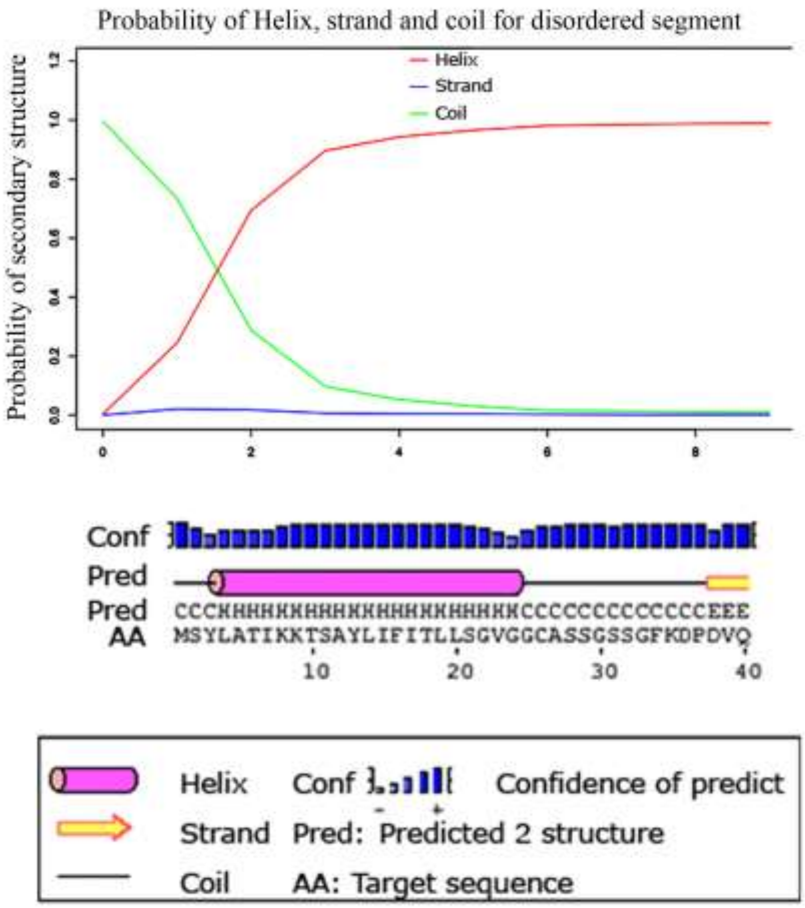

Figure 3. Secondary structure prediction of the N-Terminal domain of dWHy1 using Cspritz (up) and Psipred (down).

disorder (Figure 2). Moreover, the prediction by Cspritz (Walsh *et al.*, 2011) and Psipred (McGuffin *et al.*, 2000) for α-helical secondary structure in the N-terminal 'unstructured' domain (Figure 3), together with the hydrophilic character, its prediction as a transmembrane segment and the possible cleavage site for a peptidase II (data not shown), are all consistent with the presence of a signal peptide.

To establish the phylogenetic relationship of dWHy1 to other WHy domain-containing sequences from plants, archaea and bacteria, a phylogenetic tree was constructed using the neighbour-joining algorithm. Of the 850 WHy domains in 709 proteins in the SMART nrdb database, 70 belong to Archaea, 462 to bacteria and 177 to Eukaryota. The sequences of the non-redundant bacterial protein, plus three from *Arabidopsis thaliana*, grouped into distinct domain-specific clusters (Figure 4), where the positioning of dWHy1 with the bacterial proteins, within the Pseudomonadales and with *Pseudomonas caeni* as the closest relative, supports the putative bacterial origin of the sequence.

Protein expression and partial purification

SDS-PAGE analysis of the recombinant dWHy1 expressed in *E. coli* BL21(DE3) showed a strong band of the expected molecular weight for the His-tagged protein, although both dWHy1 and dWHy1ΔSP (the truncated construct lacking the signal peptide) were present mostly in the insoluble fraction. The truncated form of the protein was used in expression, in order to obtain higher amounts of soluble and active protein and to test the effect of this signal peptide on the protein activity. The presence of the His-tag, either at the N or C-terminus, did not change the expression pattern (data not shown).

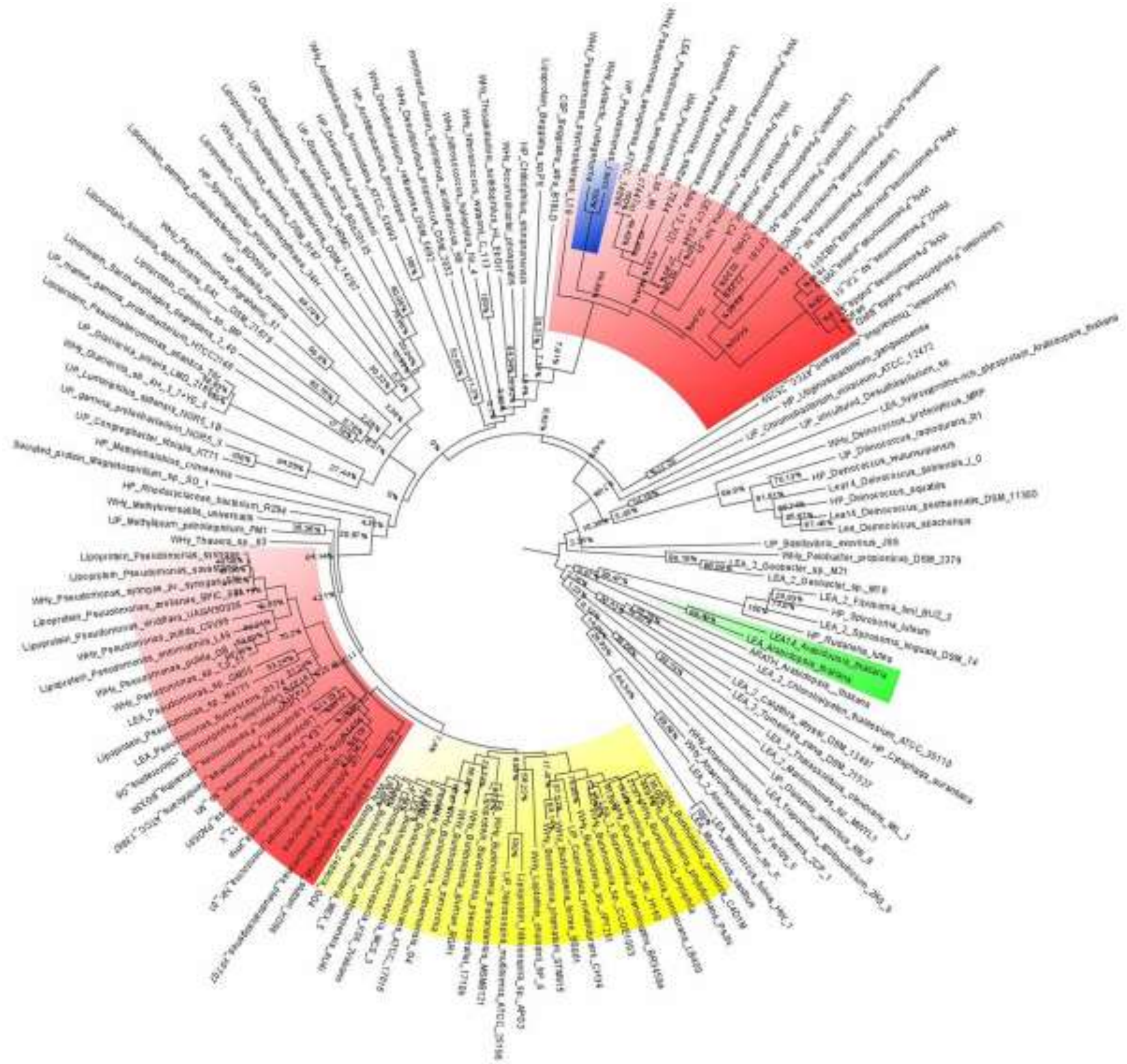

Figure 4. Phylogenetic tree of dWHy1 and related protein sequences. The evolutionary history of dWHy1 and related amino acid sequences was inferred by using the neighbour-joiningmethod. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Evolutionary analyses were conducted in MEGA6 (Tamura *et al.* 2013). Pseudomonadales are highlighted in red, *Burkholderia* spp. in yellow and *Arabidospis* spp. in green. dWHy1 protein and *P. caeni* lipoprotein (NCBI reference number: WP 033421079.1) are highlighted in blue.

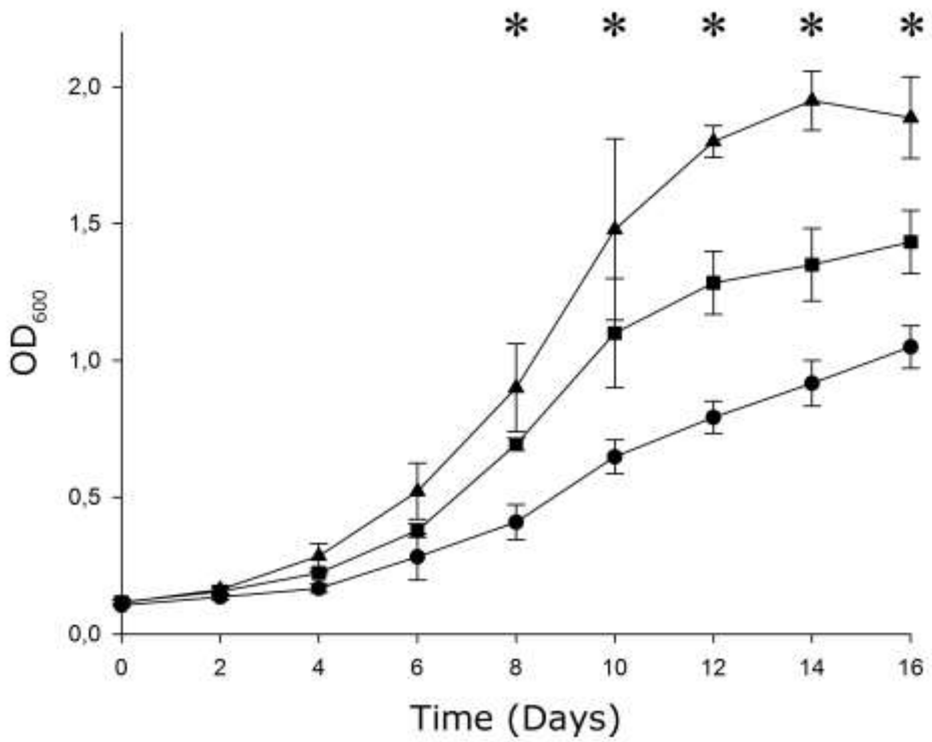

Figure 5. In vivo cold stress assay: growth of *E. coli* BL21 (DE3) at 8ºC. Circles, Negative control (*E. coli* harbouring the empty pET28a plasmid); Squares, *E. coli* expressing the entire dWHy1 protein; Triangles, *E. coli* expressing the mutant dWHy1ΔSP. (*) show significant statistical differences (p-value < 0.05).

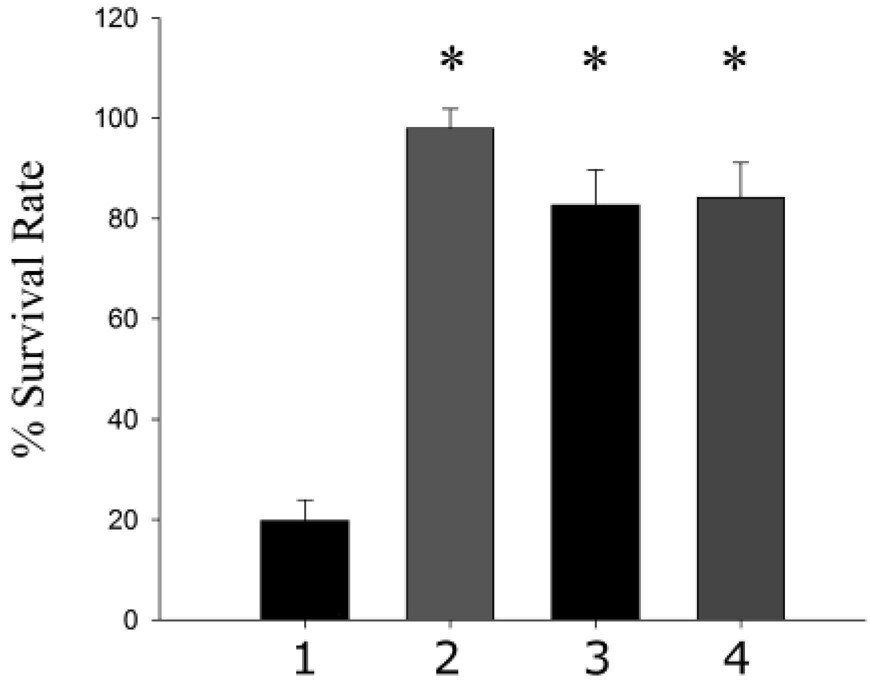

Figure 6. *E. coli* BL21 (DE3) freeze/thaw survival assay. Growth of *E. coli* on LB plates after 3 freeze/thaw cycles. 1, Control (cells transformed with the empty pET21a vector); 2, *E. coli* expressing the dWHyΔSP1 protein, with a His-Tag in C-Terminus; 3, *E. coli* expressing the entire dWHy1 protein with a His-Tag in C-Terminus; 4, *E. coli* expressing the entire dWHy1 protein without a His-Tag. (*) show significant statistical differences (p-value < 0.05).

In vivo phenotype assays

Since dWHy1 showed sequence identity to LEA homologs associated with desiccation stress and cold tolerance (Tunnacliffe and Wise, 2007), we hypothesised that the bacterial dWHy1 protein may exhibit a similar function. To test *in vivo* cold and freeze tolerance, growth assays at low temperatures and after freeze/thaw treatments were performed in *E. coli* BL21(DE3) expressing the recombinant dWHy1 and dWHy1ΔSP. Percentage survival rates and generation times were significantly higher under stress conditions for the heterologous organism expressing dWHy1 compared to controls (p-value < 0.05: Student's t-test), and notably higher when the dWHy1ΔSP protein was expressed (Figures 5 and 6). These results strongly suggest that the *in vivo* function of dWHy1 is related to cold and freeze stress tolerance.

Discussion

The screening of metagenomic DNA derived from environments in which specific functional characteristics are expected (such as protein thermostability or psychrophilicity) is a valid and un-biased approach for obtaining genetic elements which encode those properties (Sabree *et al.*, 2009; Berlemont *et al.*, 2011). It is therefore reasonable to expect that the microbial populations inhabiting the cold desert soils of the Antarctic Dry Valleys, characterised by very low levels of water bioavailability, low temperatures with extreme variations and frequent freeze-thaw events (Cary *et al.*, 2010) would exhibit both physiological and genomic stress-response adaptations. Recent 'omics' studies (Kriško *et al.*, 2010; Farrant and Moore, 2011; Tyson *et al.*, 2012) have identified a range of molecular mechanisms potentially implicated in desiccation or general stress response mechanisms.

In this study an ORF, designated *dwhy1*, was identified through a sequence based screen of an Antarctic desert soil metagenomic fosmid library. Several characteristics of the dWHy1 protein sequence are indicative of a possible *in vivo* function in desiccation-stress adaptation. Firstly, the LEA proteins, to which dWHy1 shows low but significant sequence homology, have consistently been identified as the most highly up-regulated and differentially-expressed proteins in desiccation-tolerant organisms (Dunaeva and Adamska, 2001; Tanaka *et al.*, 2004; Kovacs *et al.*, 2008; Shih *et al.*, 2008; Kriško *et al.*, 2010; Tolleter *et al.*, 2010). The low molecular mass of the dWHy1 protein product is also consistent with this concept, given that proteins of less than 25 kDa are commonly involved in major biochemical processes such as ribosome functioning, transcriptional regulation and stress response and/or adaptation (Müller *et al.*, 2010). In addition, analyses of the physicochemical properties of the WHy domain have suggested a role in dehydration tolerance, probably by interaction with water and small polar molecules (Jaspard and Hunault, 2014). A high confidence prediction of phosphorylation and lipoprotein signal peptide sites in dWHy1 may further support a stress related function. Phosphoproteins are involved in many aspects of cellular metabolism and include enzymes required for protein and DNA metabolism which confer protection during stress response (Soufi *et al.*, 2008).

Regions of disordered polypeptide, which enhance protein conformational flexibility and facilitate promiscuous specificity, are commonly found in LEA proteins (Bardwell and Jakob, 2012). None of our results suggest that the dWHy1 protein has substantially disordered regions (Figure 2), in line with other publications on the characteristics of the WHy domain (Jaspard and Hunault, 2014).

Survival data from *E. coli* growing at low temperatures or after freeze/thaw cycles demonstrate that dWHy1 actively confers cold and freeze tolerance *in vivo*. Possible mechanisms for the observed function include; 1) maintenance of cell membrane transport by regulating water at the boundary and influencing water dynamics, in a similar manner to glycerol, 2) involvement in sensing pathways and signal cascade regulated by phosphorylation events, or 3) chaperone-like activity, preventing unfolding and protein aggregation, or stabilising partially unfolded states.

LEA proteins such as LEA18, ERD10 and ERD14 have been shown to bind acidic phospholipid vesicles, possibly indicating a function in maintaining the structural or functional integrity of bacterial or plant cell membranes (Kovacs *et al.*, 2008; Thalhammer *et al.*, 2010; Tolleter *et al.*, 2010; Hundertmark *et al.*, 2011). ATP-independent chaperone-like activities have been reported for LEA proteins, particularly ERD10 (Kovacs *et al.*, 2008; Kovacs *et al.*, 2009; Hara, 2010).

Here we demonstrate that dWHy1 confers *in vivo* stress protection in bacteria, either when growing at low temperatures (Figure 5) or when they undergo repeated freeze/thaw cycles (Figure 6). To our knowledge, this is the first report of the functional characterisation of a bacterial WHy domain containing protein. Nevertheless, the exact cryoprotective functions of dWHy1 (and similar proteins) remain unknown.

Acknowledgements

The authors gratefully acknowledge support from the South African National Antarctic Program of the National Research Foundation. DAC and DEA wish to thank Professor Roberta Farrell (University of Waikato, New Zealand) for providing access to the Antarctic McMurdo Dry Valleys through the Antarctica-NZ 2011 K021 field expedition.

References

Anderson, D.E. Metagenome sequencing and in silico gene discovery: form genetic potential to function, *PhD Thesis*, University of the Western Cape, 2012, 249 pp.

Bardwell, J.C.A., Jakob, U., Conditional disorder in chaperone action, *Trends Biochem Sci*, 2012; **37**: 517-525.

Berlemont, R., Pipers, D., Delsaute, M., Angiono, F., Feller, G., Galleni, M., Power, P., Exploring the Antarctic soil metagenome as a source of novel cold-adapted enzymes and genetic mobile elements, *Rev Argent Microbiol*, 2011; **43**: 94-103.

Billi, D., Potts, M., Life and death of dried prokaryotes, *Res Microbiol*, 2002; **153**: 7-12.

Campos, F., Cuevas-Velazquez, C., Fares, M.A., Reyes, J.L., Covarrubias, A.A., Group 1 LEA proteins, an ancestral plant protein group, are also present in other eukaryotes, and in the archeae and bacteria domains, *Mol Genet Genomics*, 2013; **288**: 503-517.

Cary, S.C., McDonald, I.R., Barrett, J.E., Cowan, D.A., On the rocks: the microbiology of Antarctic Dry Valley soils, *Nat Rev Micro*, 2010; **8**: 129-138.

Ciccarelli, F.D., Bork, P., The WHy domain mediates the response to desiccation in plants and bacteria, *Bioinformatics*, 2005; **21**: 1304-1307.

Cserzo, M., Wallin, E., Simon, I., von Heijne, G., Elofsson, A., Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method, *Protein Eng*, 1997; **10**: 673-676.

Csonka, L.N., Physiological and genetic responses of bacteria to osmotic stress, *Microbiol Rev*, 1989; **53**: 121-147.

Chou, K.-C., Using subsite coupling to predict signal peptides, *Protein Eng*, 2001; **14**: 75-79.

Chou, K.C., Shen, H.B., Signal-CF: a subsite-coupled and window-fusing approach for predicting signal peptides, *Biochem Biophys Res Comm*, 2007; **357**: 633-640.

Cuevas-Velazquez, C.L., Rendón-Luna, D.F., Covarrubias, A.A., Dissecting the cryoprotection mechanisms for dehydrins, *Front Plant Sci*, 2014; **5**: 583.

Dunaeva, M., Adamska, I., Identification of genes expressed in response to light stress in leaves of Arabidopsis thaliana using RNA differential display, *Eur J Biochem*, 2001; **268**: 5521-5529.

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

Farrant, J.M., Moore, J.P., Programming desiccation-tolerance: from plants to seeds to resurrection plants, *Curr Opin Plant Biol*, 2011; **14**: 340-345.

Fitzsimons S., Campbell I., Balks M., Green T.G.A. & Hawes I. The state of the Ross Sea region terrestrial environment. In E.J. Waterhouse (ed.): *Ross Sea region 2001: a state of the environment report for the Ross Sea region of Antarctica*, 2001, Section 4.2–4.8. Christchurch: New Zealand Antarctic Institute.

Frossard, S.M., Khan, A.A., Warrick, E.C., Gately, J.M., Hanson, A.D., Oldham, M.L., Sanders, D.A., Csonka, L.N., Identification of a third osmoprotectant transport system, the osmU system, in Salmonella enterica, *J Bacteriol*, 2012; **194**: 3861-3871.

Hara, M., The multifunctionality of dehydrins: an overview, *Plant Signal Behav*, 2010; **5**: 503-508.

Hiller, K., Grote, A., Scheer, M., Münch, R., Jahn, D., PrediSi: prediction of signal peptides and their cleavage positions, *Nucleic Acids Res*, 2004; **32**: W375-W379.

Hofmann, TMbase - A database of membrane spanning proteins segments, *Biol. Chem. Hoppe-Seyler*, 1993; **374**:

Hogg, I.D., Craig Cary, S., Convey, P., Newsham, K.K., O'Donnell, A.G., Adams, B.J., Aislabie, J., Frati, F., Stevens, M.I., Wall, D.H., Biotic interactions in Antarctic terrestrial ecosystems: Are they a factor?, *Soil Biol Biochem*, 2006; **38**: 3035-3040.

Hopkins, D.W., Sparrow, A.D., Elberling, B., Gregorich, E.G., Novis, P.M., Greenfield, L.G., Tilston, E.L., Carbon, nitrogen and temperature controls on microbial activity in soils from an Antarctic dry valley, *Soil Biol Biochem*, 2006; **38**: 3130-3140.

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

Imhoff, J.F., Rodriguez-Valera, F., Betaine is the main compatible solute of halophilic eubacteria, *J Bacteriol*, 1984; **160**: 478-479.

Jaspard, E., Hunault, G., Comparison of Amino Acids Physico-Chemical Properties and Usage of Late Embryogenesis Abundant Proteins, Hydrophilins and WHy Domain, *PLoS ONE*, 2014; **9**: e109570.

Kall, L., Krogh, A., Sonnhammer, E.L., A combined transmembrane topology and signal peptide prediction method, *J Molec Biol*, 2004; **338**: 1027-1036.

Kempf, B., Bremer, E., Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments, *Arch Microbiol*, 1998; **170**: 319-330.

Kovacs, D., Kalmar, E., Torok, Z., Tompa, P., Chaperone Activity of ERD10 and ERD14, Two Disordered Stress-Related Plant Proteins, *Plant Physiol*, 2008; **147**: 381-390.

Kovacs, D., Rakacs, M., Agoston, B., Lenkey, K., Semrad, K., Schroeder, R., Tompa, P., Janus chaperones: assistance of both RNA- and protein-folding by ribosomal proteins, *FEBS Lett*, 2009; **583**: 88-92.

Kozlowski, L., Bujnicki, J., MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins, *BMC Bioinformatics*, 2012; **13**: 1-11.

Krisko, A., Smole, Z., Debret, G., Nikolic, N., Radman, M., Unstructured hydrophilic sequences in prokaryotic proteomes correlate with dehydration tolerance and host association, *J Molec Biol*, 2010; **402**: 775-782.

Kriško, A., Smole, Z., Debret, G., Nikolić, N., Radman, M., Unstructured Hydrophilic Sequences in Prokaryotic Proteomes Correlate with Dehydration Tolerance and Host Association, *J Mol Biol*, 2010; **402**: 775-782.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, *J Mol Biol*, 2001; **305**: 567-580.

Leslie, S.B., Israeli, E., Lighthart, B., Crowe, J.H., Crowe, L.M., Trehalose and sucrose protect both membranes and proteins in intact bacteria during drying, *Appl Environ Microbiol*, 1995; **61**: 3592-3597.

Mahajan, S., Tuteja, N., Cold, salinity and drought stresses: An overview, *Arch. Biochem. Biophys.*, 2005; **444**: 139-158.

Mattimore, V., Battista, J.R., Radioresistance of Deinococcus radiodurans: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation, *JBacteriol*, 1996; **178**: 633-637.

McGuffin, L.J., Bryson, K., Jones, D.T., The PSIPRED protein structure prediction server, *Bioinformatics*, 2000; **16**: 404-405.

Miller, M.L., Soufi, B., Jers, C., Blom, N., Macek, B., Mijakovic, I., NetPhosBac – A predictor for Ser/Thr phosphorylation sites in bacterial proteins, *Proteomics*, 2009; **9**: 116-125.

Müller, S.A., Kohajda, T., Findeiß, S., Stadler, P.F., Washietl, S., Kellis, M., von Bergen, M., Kalkhof, S., Optimization of parameters for coverage of low molecular weight proteins, *Anal Bioanal Chem*, 2010; **398**: 2867-2881.

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

Potts, M., Desiccation tolerance of prokaryotes, *Microbiol. Rev.*, 1994; **58**: 755-805.

Potts, M., Slaughter, S.M., Hunneke, F.U., Garst, J.F., Helm, R.F., Desiccation tolerance of prokaryotes: application of principles to human cells, *Integr Comp Biol*, 2005; **45**: 800-809.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2014). ISBN 3-900051-07-0, URL: http://www.R-project.org/.

Roelofs, D., Aarts, M.G.M., Schat, H., Van Straalen, N.M., Functional ecological genomics to demonstrate general and specific responses to abiotic stress, *Funct. Ecol.*, 2008; **22**: 8-18.

Sabree, Z.L., Kambhampati, S., Moran, N.A., Nitrogen recycling and nutritional provisioning by Blattabacterium, the cockroach endosymbiont, *Proc Natl Acad Sci U S A*, 2009; **106**: 19521-19526.

Shen, H.-B., Chou, K.-C., Ensemble classifier for protein fold pattern recognition, *Bioinformatics*, 2006; **22**: 1717-1722.

Shen, H.B., Chou, K.C., Signal-3L: A 3-layer approach for predicting signal peptides, *Biochem Biophys Res Commun*, 2007; **363**: 297-303.

Shih, M.-D., Hoekstra, F.A., Hsing, Y.-I.C., Late Embryogenesis Abundant Proteins, in Adv. Bot. Res., Jean-Claude, K., Michel, D., eds.; Academic Press, 2008.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.2011.

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

Solovyev, V., Salamov, A., Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies (Ed. R.W. Li), *Nova Science Publishers*, (2011); p. 61-78.

Soufi, B., Jers, C., Hansen, M.E., Petranovic, D., Mijakovic, I., Insights from site-specific phosphoproteomics in bacteria, *Biochim Biophys Acta*, 2008; **1784**: 186-192.

Sutherland, L., Cairney, J., Elmore, M.J., Booth, I.R., Higgins, C.F., Osmotic regulation of transcription: induction of the proU betaine transport gene is dependent on accumulation of intracellular potassium, *J Bacteriol*, 1986; **168**: 805-814.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0, *Mol Biol Evol*, 2013; **30**: 2725-2729.

Tanaka, S., Ikeda, K., Miyasaka, H., Isolation of a new member of group 3 late embryogenesis abundant protein gene from a halotolerant green alga by a functional expression screening with cyanobacterial cells, *FEMS Microbiol Lett*, 2004; **236**: 41-45.

Thalhammer, A., Hundertmark, M., Popova, A.V., Seckler, R., Hincha, D.K., Interaction of two intrinsically disordered plant stress proteins (COR15A and COR15B) with lipid membranes in the dry state, *Biochim Biophys Acta*, 2010; **1798**: 1812-1820.Tolleter, D., Hincha, D.K., Macherel, D., A mitochondrial late embryogenesis abundant protein stabilizes model membranes in the dry state, *Biochim Biophys Acta*, 2010; **1798**: 1926-1933.

Tunnacliffe, A., Wise, M., The continuing conundrum of the LEA proteins, *Naturwissenschaften*, 2007; **94**: 791-812.

Tyson, T., O'Mahony Zamora, G., Wong, S., Skelton, M., Daly, B., Jones, J., Mulvihill, E., Elsworth, B., Phillips, M., Blaxter, M., Burnell, A., A molecular analysis of desiccation tolerance mechanisms in the anhydrobiotic nematode Panagrolaimus superbus using expressed sequenced tags, *BMC Res Notes*, 2012; **5**: 1-24.

Walsh, I., Martin, A.J.M., Di Domenico, T., Vullo, A., Pollastri, G., Tosatto, S.C.E., CSpritz: accurate prediction of protein disorder segments with annotation for homology, secondary structure and linear motifs, *Nucleic Acids Res*, 2011; **39**: W190-W196.

Wise, M.J., The POPPs: clustering and searching using peptide probability profiles, *Bioinformatics*, 2002; **18**: S38-S45.

Wynn-Williams, D.D., Antarctic microbial diversity: the basis of polar ecosystem processes, *Biodivers Conserv*, 1996; **5**: 1271-1293.

Xiao, Y.P., Hui, W., Wang, Q., Roh, S.W., Shi, X.Q., Shi, J.H., Quan, Z.X., Pseudomonas caeni sp. nov., a denitrifying bacterium isolated from the sludge of an anaerobic ammonium-oxidizing bioreactor, *Int J Syst Evol Microbiol*, 2009; **59**: 2594-2598.

Zhou, J., Bruns, M.A., Tiedje, J.M., DNA recovery from soils of diverse composition, *Appl Environ Microbiol*, 1996; **62**: 316-322.