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

Dominique Anderson¹, Eloy Ferreras², Marla Tuffin¹ and Don Cowan¹,²

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 Hypersensitivity 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°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-redundant 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 AAAACATATGTCTACCTGGCTAC and oWNdeIΔPSDir AAAACATATGTGCGTCATCTGGTA respectively as forwards, with a NdeI site (underlined) and, as reverse primer, oWHindIIIRevHisTag AAAAAGCTTTTAATGGTGGTGTTG with a HindIII 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 OD₆₀₀ 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 -80°C 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 8°C 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
Alignment of the dWHy1 protein with a Water Stress and Hypersensitive response protein 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-joining method. 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 Arabidopsis 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 8°C.
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 dWHyAP1 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


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


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


