2008

Composition, conservation, evolution, and function of the cold shock domain proteins in plants

Kari Beth Thompson
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Thompson, Kari Beth, "Composition, conservation, evolution, and function of the cold shock domain proteins in plants" (2008). *Graduate Theses, Dissertations, and Problem Reports*. 4427.
<https://researchrepository.wvu.edu/etd/4427>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

**COMPOSITION, CONSERVATION, EVOLUTION, AND FUNCTION OF THE COLD
SHOCK DOMAIN PROTEINS IN PLANTS**

by

Kari Beth Thompson

**Thesis submitted to the Davis College of Agriculture and Forestry and Consumer
Sciences
at West Virginia University
in partial fulfillment of the requirements
for the degree of**

**Master of Science
in
Genetics and Developmental Biology**

Approved by

**Dale Thomas Karlson, Ph.D., Chair
Jed Doelling, Ph.D.
Carina Barth, Ph.D.**

Division of Plant and Soil Sciences

**Morgantown, West Virginia
2008**

**Keywords: Cold induction, CBF, ABA, localization, phylogeny, *Physcomitrella patens*,
*Arabidopsis thaliana***

Copyright 2008 Kari Beth Thompson

Abstract

COMPOSITION, CONSERVATION, EVOLUTION, AND FUNCTION OF THE COLD SHOCK DOMAIN PROTEINS IN PLANTS

by Kari Beth Thompson

Cold shock domain proteins, or CSPs, are among the most ancient and conserved nucleic acid binding proteins. They have been well characterized in prokaryotic systems but little is known about their function in plants. The overall objective of this study was to understand the evolution and function of the cold shock domain in plant model systems. To pinpoint conserved areas within CSD an alignment was made of the cold shock domain protein sequences based on nucleotide sequence information found within EST databases. A description of domains found in the plant CSPs was prepared and a descriptive nomenclature system was developed. The alignment was also used to analyze the phylogeny of CSPs by making a neighbor-joining tree and a maximum likelihood tree. Prediction software was used to predict sub-cellular localization and post-translation modifications by phosphorylation and sumoylation. To characterize the CSPs on a functional level, their expression in the cold response pathways of ICE1, CBF 1, 2, and 3, and ABA were analyzed by qRT-PCR. The characteristics of nucleic acid binding, cold induction, and localization of the CSPs in bryophytes, the earliest land plants, were also determined. *Physcomitrella patens* CSPs are capable of binding to nucleic acids and are cold inducible. Expression lines for two CSPs from *Physcomitrella patens* were created in order to view sub-cellular and tissue specific localization patterns over the course of the moss life cycle. It is hoped that this foundation of knowledge of the cold shock domain will lead to a better understanding of its functional importance in higher plants of agronomic importance.

This work is dedicated to
my family
and all of the people who
have helped and supported me.

ACKNOWLEDGMENTS

First and foremost I wish to thank Dr. Dale Karlson for allowing me work toward my master's degree in his lab. I quickly felt at ease working with him and without his excellent guidance, advice, understanding, and kind words this experience may not have been so rewarding for me. I truly grew both as a researcher and a person while working with him and I cannot find enough words to express my thanks. I would also like to thank my committee members Dr. Jed Doelling and Dr. Carina Barth for all of their suggestions, advice, and help. Next, I would like to thank all of my fellow lab members, Yongil Yang, Prashanthi Mulinti, Swamy Krishna Tripurani, Vijaya Vardhana Reddy Chaikam, and Dr. Kentaro Nakaminami. They are all wonderful people to work with, each of them with their own special qualities and good advice. I hope they all know how much each of them has meant to me and how they have all helped me to feel at home. I will always think of them as my scientific brothers and sisters. I would also like to thank my family. I want to thank my father Larry for always being proud of me no matter what I do or where my life takes me. My mother Brenda, who is ever my best friend, I want to thank her for always being there and giving me good guidance whether I am ready to hear it or not. Finally, I want to thank my brother Scott and my sister Kathryn also for being there and listening to me when I needed to talk.

I would like to thank Dr. Michael Thomashow for kindly donating the CBF over-expression seeds and for helpful discussions with Sarah Gilmour. I would also like to thank Dr. Jian-Kang Zhu for graciously donating the seeds for *ice1* and *ICE-OX* as well as their vector control. I would like to acknowledge the ABRC for donating ABA mutant seeds and Dr. Eiji Nambara (RIKEN Plant Science Center, Japan) for helpful advice regarding the ABA mutant lines. I want to thank Tsuyoshi Aoyama from the National Institute for Basic Biology (NIBB) in Okazaki Japan, for teaching me how to use the PHYLIP and MOLPHY programs. I also want to thank Dr. Yuji Hiwatashi and all the assistants in the Hasebe laboratory for teaching me how to work with moss. Finally, I would also like to send my deepest thanks to Dr. Mitsuyasu Hasebe also from NIBB for allowing me to come and work in his lab and his great advice during the summers of 2006 and 2007. My work was

funded by the National Science Foundation grant #IBN-0416945 (D.K) and the NSF/JSPS EAPSI (K.T.) summer program.

TABLE OF CONTENTS

Abstract	ii
Dedication	iii
Acknowledgements	iv
Table of contents	vi
List of Figures	vii
List of Tables	viii
List of Symbols / Nomenclature	ix
Literature Review	1
Chapter 1: Conservation, Composition, Evolution, and Function of Cold Shock Proteins...	6
1. Abstract	6
2. Introduction	7
3. Materials and Methods	10
4. Results	13
5. Discussion	19
Chapter 2: Function of CSPs in <i>Physcomitrella patens</i>	37
1. Abstract	37
2. Introduction	38
3. Materials and Methods	39
4. Results	44
5. Discussion	46
Summary and Conclusions	56
Bibliography	58
Curriculum Vitae	63

LIST OF FIGURES

Multiple Sequence Alignment	27
Highly Conserved Regions of the CSD	28
Neighbor-Joining Tree	29
Maximum Likelihood Tree	30
Monocot and Dicot Subgroup Sequence Comparison.....	31
CSP Motif arrangement and Nomenclature	32
<i>At</i> CSPs in well-known cold regulons.....	36
PpCSPs compared to well-known CSPs.....	49
PpCSPs bind nucleic acids	50
Cold induced Up-regulation of the PpCSPs.....	51
Expression Construct Design and Southern Blot confirming transformant lines	52
Localization of PpCSP1 and 2, protonema, gametophore, gametangia	53
PpCSP localization in embryo and developing sporophyte	54
PpCSP Nucleolar localization	54
PpCSP dot-like pattern in the cytoplasm	55

LIST OF TABLES

Amino Acid Composition	33
Sumoylation Prediction	34-35

LIST OF SYMBOLS / NOMENCLATURE

1. ABA	Absciscic acid
2. AtCSP	<i>Arabidopsis thaliana</i> cold shock protein
3. CBF	Cold Binding Factor
4. CSD	Cold Shock Domain
5. CSP	Cold Shock Protein
6. COR	Cold responsive gene
7. DREB	Dehydration response element bind factor
8. EST	Expressed Sequence Tag
9. GUS	β -glucuronidase
10. ICE1	Inducer of CBF Expression 1
11. NJ	Neighbor-Joining Tree
12. ML	Maximum likelihood Tree
13. PpCSP	<i>Physcomitrella patens</i> Cold shock proteins
14. RNP	Ribonucleoprotein motif
15. uidA	GUS gene
16. YFP	Yellow fluorescent protein

LITERATURE REVIEW

The cold shock domain (CSD) is one of the most evolutionarily conserved nucleic acid binding domains known to date (Graumann and Marahiel 1998). The CSD is thought to have been present before the beginning of single-cell evolution (Graumann and Marahiel 1998) and has been found in organisms ranging from *Archaea* into higher eukaryotes including humans (Saunders et al. 2003; Sommerville 1999). The bacterial cold shock proteins differ from those found in eukaryotes in that bacterial cold shock proteins consist solely of a Cold Shock Domain (CSD), whereas eukaryotic counterparts contain additional auxiliary C-terminal domains.

CspA, from *Escherichia coli* was the first protein found to contain a CSD (Jones et al. 1987). It is one member of a nine-member family comprised of CspA-CspI, four of which are cold induced; CspA is the predominant member accumulating to up to 10% of total cellular proteins (Phadtare and Inouye 1999). The cold induced Csps include CspA (Jones et al. 1987; Lee et al. 1994), CspB (Lee et al. 1994), CspG (Nakashima et al. 1996), and CspI (Bae et al. 1999; Wang et al. 1999). The other five members are involved in various physiological activities such as cell division and response to low nutrient levels (as reviewed by Graumann and Marahiel 1998). In 1997, Jiang et al. determined that CspA functions as an RNA chaperone inhibiting RNA secondary structures at low temperatures. At low temperatures, RNA tends toward secondary structures because they are more thermodynamically favored (Graumann and Marahiel 1996). The chaperone activity of bacterial CSPs serves to anti-terminate the transcription that would otherwise be hindered by these secondary structures (Jiang et al. 1997). Bae et al. showed that CspC and CspE also display transcription anti-termination ability (2000). The expression of CspB and CspC from *Bacillus subtilis*, two *E. coli* Csp orthologs, also respond to a temperature downshift (Graumann et al. 1997). Other Csp orthologs have been identified in more than 50 gram positive and negative bacterial species (Graumann et al. 1998), and they are often present in multiple copies comprising families (Phadtare and Inouye 1999).

The solution structures of CspA from *E.coli*, CspB from *B. subtilis*, as well as the CSD from

the human Y-box 1 protein have been solved. All three proteins have a β -barrel structure made by 5 anti-parallel β -strands (Newkirk et al. 1994; Schindelin et al. 1993), however, Y-box 1 contains a large loop between β -3 and β -4 (Kloks et al. 2002). Two RNA recognition motifs, RNP-1 and RNP-2, are present on adjacent β -sheets (Manival et al. 2001). The sequences of RNP-1 and RNP-2 within CspB from *B. subtilis* are important for ssDNA binding (Schroder et al. 1995). This binding occurs via base stacking between the aromatic side chains of amino acids contained within the RNPs (Graumann and Marahiel 1998). In 1995, N. Sato identified the RNA binding domain (RBD) that contains a RNA recognition motif (RRM) and proposed that it may share a common function with the CSD. Like the CSD, the RBD contains RNP-1 and 2, but it adopts a $\beta\alpha\beta\beta\alpha\beta$ structure (Sommerville 1999). The similarity between the structures of these proteins has led scientists to believe that they may be the product of convergent evolution (Graumann and Marahiel 1996). RNA-binding proteins are associated with many aspects of RNA metabolism such as processing, transport, and the translation of RNA molecules (Dreyfuss et al. 2002).

In addition to the extensive study of cold shock domain proteins in prokaryotes, eukaryotic animal orthologues (Y-box proteins) have been heavily studied. The *Xenopus laevis* Y-box protein, FRGY2, packages mRNA into RNPs (Matsumoto et al. 2003) and masks maternal mRNA as a precise means to control developmental timing via repression of translation (Richter and Smith, 1984; Ranjan et al. 1993; Bouvet and Wolffe 1994). Not surprisingly, it localizes to the nucleolus and cytoplasm (Matsumoto et al. 1996). Human Y-box proteins are involved in many steps of RNA metabolism. Specifically, Y-box proteins regulate transcription (Ohga et al. 1998; Lodomery and Sommerville 1995), enhance pre-mRNA splicing (Stickeler 2001), control translation (Sommerville 1999; Evdokimova et al. 1998), are involved in mRNA stabilization (Edokimova et al. 2001) and have been implicated in detecting RNA damage (Hayakawa et al. 2002). The Y-box structure is characterized as having a variable N-terminal domain, a CSD, and a C-terminal tail domain with interspersed basic and aromatic islands (as reviewed by Kohno et al. 2003). LIN-28 is another well-studied cold shock domain protein from *Caenorhabditis elegans* that is involved with post-transcriptional regulation and is important for developmental timing (Moss and Tang 2003). This

protein is composed of a variable N-terminus followed by a CSD, and two C-terminal retroviral-like zinc fingers (Moss et al. 1997). Another protein containing five CSDs (UNR) was also found in human; it specifically binds RNA and is potentially involved in cytoplasmic mRNA metabolism (Triqueneaux et al. 1999). The majority of animal Y-box proteins are not related to low temperature stress, studies highlight their functional roles relating to RNA metabolism. To my knowledge, only one study has linked an animal CSD protein to cold stress: chicken YB-1 has been shown to reinstate the growth of cells at low temperature (Matsumoto et al. 2005).

Plant CSDs are widely distributed and show high sequence similarity with each other and to those found in bacteria. Plant CSPs, like bacterial Csps, are often present in small gene families (Karlson and Imai 2003). A major difference between plant CSPs and bacterial Csps, however, is that the CSD is present in the amino-terminus of plant cold shock domain proteins while various auxiliary domains are present in the carboxyl-terminus. The domains present in the C-terminus include glycine-rich regions, retroviral-like zinc fingers, RGG-like repeats, RRM domains (Mussgung et al. 2005), and additional CSDs. Unlike animal cold shock domain proteins, many plant CSPs appear to respond to cold stress. In 2002, Karlson et al. were the first to show the low temperature induction of WCSP1, one of at least three CSPs present in winter wheat. Like most plant CSPs, WCSP1 has an N-terminal CSD and three zinc fingers that are interspersed between glycine-rich regions (Karlson et al. 2002). WCSP1 exhibits nucleic acid binding activity (Karlson et al. 2002), which was found to be heat stable in that its binding activity to mRNA, ssDNA, and dsDNA after is maintained after boiling (Nakaminami et al. 2005). The quadruple KO mutant (*cspA*, *cspB*, *cspE*, and *cspG*) BX04 line of *E.coli* fails to grow at low temperatures and serves as a useful *in vivo* system to assess the function of CSPs during cold stress. WCSP1 complemented the cold sensitive phenotype of these mutant cells and exhibited transcription anti-termination activity (Nakaminami et al. 2006). Nakaminami et al. also performed transient expression analysis of WCSP1 and found that it localized to the endoplasmic reticulum and the nucleus of wheat cells (2006). The expression of *Arabidopsis thaliana* CSPs is also induced by temperature downshift (Karlson et al. 2003; Fusaro et al. 2006; Kim et al. 2006). Like WCSP1, AtCSP2/AtGRP2 is also

capable of binding nucleic acids (Fusaro et al. 2006). Kim et al. found that Arabidopsis AtCSP1/CSDP1 (At4g36020) can improve the growth of BX04 *E. coli* cells at cold temperatures and that it exhibits RNA chaperone activity (2006). In 2007, Sasaki et al. determined the same activities for AtCSP2/AtGRP2. Fusaro et al. fused AtCSP2/AtGRP2 to GFP (green fluorescent protein) and found it to be present in the cytoplasm and the nucleus of *Arabidopsis* cells (2006). Sasaki et al. further confirmed this localization and determined that AtCSP2/AtGRP2 is also localized in the nucleolus (2007). Nucleolar targeting required the last zinc finger and glycine-rich region (Saskai et al. 2007). Using *in situ* hybridization they also determined that AtCSP2/AtGRP2 proteins accumulate in the apical meristem and at the tips of the roots of *Arabidopsis* plants (Fusaro et al. 2006). Preferential accumulation was further seen in reproductive tissues such as ovules and pollen, and finally in the embryos of *Arabidopsis* seeds (Fusaro et al. 2006). Fusaro et al. also used RNAi lines to confirm that AtCSP2/AtGRP2 has roles in flowering time as well as flower and seed development (2006). NAB1, a CSP from *Chlamydomonas reinhardtii*, localizes to the cytoplasm and masks RNA similar to FRGY2 (Mussnug et al. 2005).

Our present work shows, in more detail, the conservation of CSPs among a wide range of plant species across many different phyla. Furthermore, our evolutionary analysis of CSPs shows a grouping of similar organisms which indicates that CSPs from similar organisms are more related to each other than CSPs from more distant species. Our data also illustrates the presence of more than one type of CSP in monocots and dicots. While these proteins vary in length, and even motif composition, they still share many characteristics such as solubility and predicted localization. As determined by prediction software, our data convincingly support the hypothesis that plant CSPs are targets for post-translational modifications such as phosphorylation and sumoylation. This hypothesis is also supported by our *in vitro* experimental data for both modifications (Nakaminami and Chaikam, unpublished). The cold inducible expression of AtCSPs was further analyzed by attempting to pinpoint their presence within well-known cold regulons such as ABA and CBF/DREB1A. This study also shows that all *Physcomitrella patens* CSPs (PpCSPs) are cold induced and capable of binding nucleic acids. These observations correlate well to those obtained

from CSPs of higher plants. Finally, we propose hypotheses which link CSPs to RNA metabolism based upon the *in vivo* sub-cellular localization patterns that we observed for PpCSPs.

CHAPTER 1: CONSERVATION, EVOLUTION, AND FUNCTION OF THE PLANT COLD SHOCK DOMAIN PROTEINS, AND ATCSPS IN WELL-KNOWN COLD REGULONS

1. Abstract

CSPs are predicted to have evolved prior to the divergence of prokaryotes and eukaryotes. Although they have been well characterized in both prokaryotes and eukaryotes, little is known about their evolutionary history and function in plants. CSPs can be identified in plants ranging from various algal species through angiosperms. In order to demonstrate their highly conserved nature, a recent publication aligned 25 orthologues of plant CSPs from various members within the plant kingdom (Karlson and Imai 2003). We expanded upon the previous study by gathering additional ESTs (expressed sequence tags) with the BLAST (Basic Local Alignment Search Tool) tool from numerous public databases. Eighty-one additional unique CSP sequences were found from plants spanning a wider evolutionary history. These sequences were aligned and analyzed with ClustalX and a nomenclature formula was developed for naming the various motifs contained within plant CSPs. To gain insight into the basic evolutionary history of the cold shock domain, these CSP sequences were used to generate phylogenetic trees using the PHYLIP and MOLPHY programs. Interestingly, the monocot and dicot CSPs diverge in the phylogenetic trees and can be further broken into subgroups. Monocots have two major types and dicots contain three major types of CSPs. In ancestral plants, such as gymnosperms, their CSP sequences occupy their own branches. All plant CSPs show sequence similarity in their N-terminal domains to CspA, the major cold related CSP from *E.coli*. The kingdom *Plantae* demonstrates a striking conservation of CSPs throughout their evolutionary history. Prediction software indicated that CSPs were largely cytosolic and commonly targets for phosphorylation and/or sumoylation. Finally, the presence of *Arabidopsis thaliana* CSPs (*AtCSPs*) within ABA and well-known cold signal transduction networks was determined using mutants lines and quantitative real-time PCR analysis.

2. Introduction

E.coli CspA was the first cold shock domain protein to be identified (Jones et al. 1987). Four of the nine Csps in *E.coli*, are cold inducible: CspA (Jones et al. 1987; Lee et al. 1994), CspB (Lee et al. 1994), CspG (Nakashima et al. 1996), and CspI (Bae et al. 1999; Wang et al. 1999). During low temperature stress, CspA, the predominant cold shock protein, accumulates to over 10% of total cellular proteins (as reviewed by Phadtare and Inouye 1999). CspA functions as an RNA chaperone (Jiang et al. 1997), and CspA, CspC, and CspE have also been linked to transcription anti-termination activity (Bae et al. 2000). Both of these functions are necessary due to the formation of stable secondary structure in mRNA upon exposure to cold (as reviewed by Graumann and Marahiel 1996).

From an evolutionary standpoint, the CSD is the most conserved nucleic acid binding domain identified and was hypothesized to have been present at the onset of single-cell evolution (as reviewed by Graumann and Marahiel 1998). Therefore, it is not surprising that the CSD has been extensively characterized from *Archaea* (Saunders et al. 2003) through higher eukaryotes. The solution structure of CspA from *E.coli* was solved and found to contain 5 anti-parallel β -strands that form a β -barrel structure (Newkirk et al. 1994). Within this structure are two RNA recognition motifs (RNP-1 and RNP-2) on adjacent β -sheets (reviewed by Manival et al. 2001). The sequences of RNP-1 and RNP-2 within CspB from *B. subtilis* are important for ssDNA binding (Schroder et al. 1995). The CSD shares a functional similarity with the RNA recognition motif (RRM) found in the RNA binding domain (RBD) of cyanobacterium *Anabaena variabilis* M3 (Sato 1995). The RBD also contains RNPs 1 and 2 responsible for RNA-binding, but it adopts a $\beta\alpha\beta\beta\alpha\beta$ structure (as reviewed by Sommerville 1999). The aromatic side chains of the RNPs contained in CSDs and RBDs interact with nucleic acids via base stacking (Graumann and Marahiel 1998). Because of these similarities, these two protein domains are thought to have arisen via convergent evolution (Graumann and Marahiel 1996).

Similar to bacteria cold shock proteins, plant CSPs are often present in multiple copies comprising small gene families (Karlson and Imai 2003). Unlike bacterial CSPs, which only contain

a CSD, plant CSPs usually contain a CSD in the amino-terminus and variable domains in the carboxyl-terminus. These variable domains include glycine-rich regions, RGG-like repeats, retroviral-like zinc fingers, RRM domains (Mussgung et al. 2005), and additional CSDs. Most of these domains are implicated in nucleic acid binding and/or mediating protein-protein interactions (Sommerville 1999). Thus far, very few plant CSPs have been studied on a functional level. Plant CSP expression is induced by temperature stress, which was documented for the first time with winter wheat WCSP1 (Karlson et al. 2002). Consistent with the function of previously characterized proteins containing a cold shock domain, WCSP1 exhibits nucleic acid binding activity (Karlson et al. 2002). WCSP1 also complements the cold sensitive phenotype of BX04 *E. coli* cells and exhibits transcription anti-termination activity (Nakaminami et al. 2006). *Arabidopsis thaliana* CSP homologues were subsequently shown to respond to cold temperature as well by RT-PCR and qRT-PCR (Karlson et al. 2003; Fusaro et al. 2006; Kim et al. 2006). Like WCSP1, Arabidopsis AtCSP1/CSDP1 (At4g36020) is capable of complementing the cold sensitive phenotype of BX04 *E. coli* cells and exhibits RNA chaperone activity (Kim et al. 2006). Recently, AtCSP2/AtGRP2 expression was found to be very high in meristematic tissues and transgenic studies confirm that it plays a role in flower and seed development (Fusaro et al. 2006; Sasaki et al. 2007).

Since the expression of *AtCSPs* is responsive to cold treatment, we were interested to determine whether *AtCSPs* are under regulation by well-characterized cold regulons. The CRT/DRE binding factors (CBFs) or dehydration response element binding factors class 1 (DREB1) regulon is perhaps the most well-studied cold regulated pathway to date. CBF1 (DREB1B), CBF2 (DREB1C), and CBF3 (DREB1A) bind to the DRE (dehydration, high-salt, or low-temperature responsive element) element that was discovered in the promoters of many cold regulated genes (Yamaguchi-Shinozaki and Shinozaki 1994; Stockinger et al. 1997; Jaglo-Ottosen et al. 1998; Gilmour et al. 1998; Liu et al. 1998). These transcription factors function as transcriptional activators for a battery of cold responsive, or *COR*, genes (Stockinger et al. 1997; Jaglo-Ottosen et al. 1998; Gilmour et al. 1998), and induce freezing tolerance (Jaglo-Ottosen et al. 1998; Liu et al. 1998). The ICE1 (inducer of CBF expression) protein is an activator of *CBF3* expression and

therefore affects both *CBF3* transcription and its downstream targets (Chinnusamy et al. 2003). Mutants of *ICE1* are unable to properly cold acclimate, so *ICE1* has a very important role in cold stress tolerance (Chinnusamy et al. 2003). By monitoring the expression of *AtCSPs* in the *ice1* mutant background, I aimed to determine if the *AtCSPs* function upstream of *CBF3* in the *CBF* regulon.

The function of abscisic acid (ABA) in relation to the adaptation to abiotic stress is well studied (reviewed by Leung and Giraudat, 1998). During cold stress, ABA levels increase as well as the expression of stress-related genes responsive to both cold stress and ABA treatment (reviewed by Shinozaki and Yamaguchi-Shinozaki 1996). These up-regulated genes include well known *COR* (cold responsive genes) genes (Thomashow 1998). In 1996, *COR15a* was the first *COR* gene found to be both up-regulated under cold stress and to have a role in freezing tolerance (Artus et al. 1996). Thus, ABA has been implicated to play an important role in plant cold stress signal transduction (Shinozaki and Yamaguchi-Shinozaki 1996). ABA deficient mutants in the major ABA biosynthesis pathway (*aba*) and mutants that are insensitive to ABA (*abi*) have been identified. *aba* mutants have difficulty cold acclimating, therefore, ABA signal transduction has been implicated as an important component for acquiring optimum freezing tolerance (Mantyla et al. 1995).

In 2003, it was reported the high conservation of cold shock domain proteins within the plant kingdom (Karlson and Imai, 2003). This study documents the presence of CSPs across an extensive range of photosynthetic organisms and represents a more robust and complete evolutionary history. Neighbor-joining and maximum likelihood trees illustrate groupings of plant shock domain proteins within related plant species. I predicted their sub-cellular localization and potential sites for phosphorylation and sumoylation using publicly available software. Furthermore, I propose a nomenclature system to describe categories of plant cold shock domain proteins based on the presence of various motifs and their arrangements within cold shock domain proteins. Lastly, I studied the four *Arabidopsis thaliana* CSPs in terms of their relationship to the well-characterized

CBF/DREB1 and ICE1 low temperature signal transduction regulons as well as their relationship to ABA.

3. Materials and Methods

Cold Shock Domain Sequence Data

Known cold shock domain protein sequences were used as queries in BLAST (tBlastn basic local alignment sequence tool) searches of expressed sequence tag (EST) databases. Databases searched include the National Center for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov/BLAST/), The Institute for Genomic Research (TIGR) (tigrblast.tigr.org/tgi/), The Joint Genome Institute (JGI) (www.jgi.doe.gov/), The Diatom EST Database (avesthagen.sznbowler.com/) (these sequences contain PTMM or xzd in their identification numbers), The Castor Bean Genome Database (castorbean.tigr.org/) from which the *R.communis* sequences were obtained, Blast Assemble Data Submission available from NIBB (the National Institute for Basic Biology) (moss.nibb.ac.jp/cgi-bin/blast-assemble) which provided the *P.patens* PpCSP3b scaffold 64 sequence and the *S. moellendorffii* sequences, and Physcobase (moss.nibb.ac.jp/) provided the *P.patens* PpCSP3a contig 4 sequence.

Sequence Alignments and Phylogenetic Trees

The Clustal X v1.83.1 program was used to align the N-terminal domain of the cold shock domain protein sequences using the following changes to the default parameter settings: Pairwise Alignment Parameter: gap opening penalty = 35.00, gap extension penalty = 0.75, and Multiple Alignment Parameter: gap opening penalty = 15.00 and gap extension penalty = 0.30. The initial start codon (methionine) of the protein sequences was not used for the alignment. The alignment was then visually aligned and adjusted manually with MacClade v. 4.08 OSX according to the protocol described by Maddison and Maddison, 2000. Using MacClade, regions with gaps were excluded and redundant sequences were combined.

A neighbor-joining (NJ) tree was generated from the MacClade sequence alignment using the PHYLIP 3.65 package. The programs used from PHYLIP include: seqboot, protdist, neighbor, and consensus with bootstrap data generated for 1000 replicates. The NJ tree was used to construct a maximum likelihood (ML) tree using the MOLPHY program, version 2.3b3, by a local rearrangement search (Adachi and Hasegawa, 1996). The probabilities of the trees were calculated using the ProtML program under the JTT model (Jones et al. 1992), and finally the trees were sorted by their Akaike information criterion (AIC) values (Adachi and Hasegawa, 1996). The bootstrap values were calculated for 100 replicates using the resampling-of-estimated-log-likelihood (RELL) method (Kishino et al. 1990; Hasegawa and Kishino, 1994).

Protein Characteristics and Predictions

The solubility of the full-length proteins was predicted using the SOSUI server (http://bp.nuap.nagoya-u.ac.jp/sosui/sosui_submit.html) (Hirokawa et al. 1998). The TargetP v1.1 server available at <http://www.cbs.dtu.dk/services/TargetP/> was used to predict the localization of CSPs with the organism group set to plants and the default of no cutoffs (Emanuelsson et al. 2000). We used the Protein Calculator v3.3 program to calculate the isoelectric points and molecular weights of the full-length proteins (<http://www.scripps.edu/~cdputnam/protcalc.html>). The amino acid compositions of both the CSD alone and full-length CSPs were compiled using the ExPASy ProtParam tool available at <http://expasy.org/tools/protparam.html> (Gasteiger et al. 2005). The NetPhos 2.0 server was used with the default settings for the prediction of putative phosphorylation sites within the CSD (<http://www.cbs.dtu.dk/services/NetPhos/>) (Blom et al. 1999). Sumoylation site predictions were performed using the SUMOplot™ Analysis Program (<http://www.abgent.com.cn/doc/sumoplot/login.asp>).

CSD Gene Expression Analysis

Arabidopsis thaliana ecotype Col-0 was grown on Metromix soil for three weeks under long day conditions (16 hours light, 8 hours dark) at 20°C. Leaf tissue was collected at time point 0 hours and the plants were subsequently transferred to 4°C with dim light. Leaf tissue was collected

from plants at 12, 24, and 48 hours after the transfer to low temperature. For assessment of recovery from low temperature stress, plants were transferred back to 20°C under dim light and leaf tissue was collected 24 and 48 hours after their transfer. Immediately upon all harvests, samples were frozen in liquid nitrogen and stored at -80° C until further use for RNA extraction.

I investigated the relationship of *AtCSPs* to cold stress in the CBF/DREB1 and ICE1 low temperature signal transduction pathways. In addition, we also evaluated the role of abscisic acid in mediating their expression under low temperature stress. Seeds of CBF1, CBF2, CBF3 transgenic over-expression lines, as well as an appropriate vector control and wild-type WS-2 background, were generously donated by Dr. Michael Thomashow (Michigan State University). For my analysis of the ICE1 regulon, Dr. Jian-Kang Zhu (Univ. of Cal. Riverside) generously provided seeds for ICE-WT-CBF3-LUC (vector), *ice1*, and *ICE-OX*, and their wild-type background of Col-gl1. Mutant seeds of Arabidopsis that were either deficient in endogenous ABA (*aba1-4*) or insensitive to ABA (*abi1-1*, and *abi2-1*) were obtained from the ABRC stock center (Columbus, OH). *Landsberg erecta* was used as the wild-type control for comparison to ABA mutants. Seeds were surface sterilized according to standard protocols and approximately twenty seeds of each type were plated on MS (Murashige and Skoog) salts media (Caisson Laboratories Inc.; North Logan, UT, U.S.A.) and stratified at 4°C in the dark for 3 days. Plates were then transferred to a Percival growth chamber and grown for an additional 11 days under 16 hour long days with day/night temperatures of 23°C and 22°C, respectively. The seedlings were then cold treated at 4°C under long day conditions. Whole seedlings were removed from plates at 0, 12 and 24 hours after their transfer to 4°C and were immediately frozen in liquid nitrogen. The samples were stored at -80°C until further use for RNA extraction. RNA was extracted using Invitrogen's Trizol Reagent (Carlsbad, CA, USA) according to the manufacturer's instructions. The Qiagen QuantiTect Reverse Transcription Kit (Valencia, CA, U.S.A.) was used to generate cDNA from 500 ng of isolated total RNA from each sample. The cDNA stocks were then diluted 100x for subsequent use as DNA templates for quantitative Real-Time PCR analysis.

Quantitative Real-Time PCR

Quantitative real-time PCR was performed with pre-developed gene-specific Taq-man assays (Applied Biosystems (Foster City, CA, USA)) for the following genes (TAIR accession numbers in parentheses): (*ACT2*: AtACTIN2) NM_180280.1, (*AtCSP1*: At4g36020) NM_119769.2, (*AtCSP2*: At4g38680) NM_120029.2/S47408.1, (*AtCSP3*: At2g17870) NM_127341.3, (*AtCSP4*: At2g21060) NM_127676.2. The amplification of cDNA template was accomplished by using the Universal Taqman PCR Master Mix according to supplier's instructions with an Applied Biosystems 7500 Real-Time PCR System (Branchburg, NJ, USA). All qRT-PCR reactions were carried out for 40 cycles in triplicate for each gene. The context sequences and amplicon length for the genes analyzed are as follows: *ACT2*, (5'- CTGGATCGGTGGTTCCATTCTTGCT-3'), 85bps, *AtCSP1*, (5'-TTCTGGAAATTTTGACCTCATCATA-3'), 85bps, *AtCSP2*, (5'-TATTTATGGTACTTGTGATATGGGT-3'), 108bps, *AtCSP3*, (5'-ATGATATGTGTGCTTGCTTATTAC-3'), 91bps, *AtCSP4*, (5'-TAATTAGGACTTGTGGTTGTTATGG-3'), 142bps. The relative fold changes were calculated by the $2^{-\Delta\Delta CT}$ method (Livak et al. 2001).

4. Results

EST databases were searched to identify novel cold shock domain sequences for multi-sequence alignment. Due to the variability of C-terminal regions and incomplete sequence information, only N-terminal regions for cold shock domain sequences were aligned with the Clustal-X program (Figure 1). Four sequences including, *F.cylindrous* DR026829, *T.pseudonana* xzd241, *T.pseudonana* xzd3738x1, and *S.moellendorffii* contig6, had a variable N-terminal domain followed by a CSD; only their CSD sequences were used for the alignment. Figure 1 contains 121 unique CSP sequences, including the 40 sequences from 19 different genera published previously (Karlson and Imai, 2003). The previously published *T.monococcum* BG607909 sequence was excluded because it is incomplete and *S.cereale* BE495434 and *H.vulgare* BE060365 were removed because they caused the MOLPHY program to oscillate when making the maximum

likelihood tree. In some cases, sequences from the earlier publication were replaced with different EST accessions that contained more sequence information and are denoted by a single asterisk after their accession number. The multi-sequence alignment contains 74 different photosynthetic organisms, representing 56 genera, some of which possess multiple CSP genes. This multi-sequence alignment clearly demonstrates that cold shock domain proteins are highly conserved in plants. This conservation is especially evident within the β -sheet regions 1, 2, 3, and 5, but is also apparent between them (Figure 2). For example, between β -sheets 3 and 4 the sequence (F/Y) R (S/T) L occurs in every sequence with only two exceptions.

The multiple sequence alignment from Figure 1 was used to make a neighbor-joining (NJ) tree with the PHYLIP package (Figure 3). The NJ tree was used in turn to make a maximum likelihood (ML) (Figure 4) tree using the MOLPHY program. Neither tree shows outstanding bootstrap values, which is likely due to the short sequence of the CSD used in the alignment. Only 65 amino acid residues were alignable due to the considerable amount of variation in C-terminal regions and incomplete EST sequences. Despite this fact, closely related organisms group together in both trees, similar to the neighbor-joining tree that was published in 2003 (Karlson and Imai 2003). In the current trees, monocot CSPs were placed within two subgroups suggesting that monocots possess two types of CSPs. Sequence comparisons show that monocot CSPs in subgroup 1 are more conserved in amino acid sequences within the CSD than monocot CSPs in subgroup 2 (Figure 5). According to the trees, dicot CSPs fall into three different subgroups. The dicot CSPs within subgroup 1 are more related to the monocot subgroups than those in subgroups 2 or 3; they show the least conservation of amino acids (Figure 5). Unexpectedly monocot *A. cepa* CF448014 and diatom *T. pseudonana* xzd1827.x1 group with dicot subgroup I, in the ML tree (Figure 4). Sequence comparisons show that the CSPs within dicot subgroup 2 have less sequence conservation than those within dicot subgroup 3 (Figure 5). As denoted by a yellow box, a major discrepancy occurred between the NJ and ML trees for seven dicot CSPs (Figure 3). On the NJ tree, these dicot CSPs were placed within subgroup 2; however, they were placed within subgroup 3 in the ML tree. For further characterization, these protein sequences were placed within

subgroup 3 because the ML tree showed better bootstrap support. Also, *P.armeniaca* CV049731, *R.communis* rcal.assembly.29908, *G.raimondii* CO108408, and *G.max* CO980000, four dicot CSPs that were outliers on the NJ tree (Figure 3), group in dicot subgroup 2 on the ML tree with bootstrap support; they were in dicot subgroup 2 for further sequence analysis. In both phylogenies, species representing the early land plants (Bryophyta, Lycopphyta, Pteridophyta, and Coniferophyta) group together. Species of autotrophic protists, representing the Haptophyta, Bacillariophyta (Diatoms), and *Mesostigma viride*, also group together. An exception to this is the Charophyte *Closterium peracerosum-strigosum-littorale complex* which grouped with the Bryophytes. As would be expected, both of these groupings of CSPs indicate more relation to each other than any of the other higher plants. Oddly, the representative members of the Chlorophyta, *Chlamydomonas* species *reinhardtii* and *incerta*, group with the *Medicago truncatula* *BG451829 sequence in proximity to the other autotrophic protists in both phylogenetic trees. Almost all plant CSPs include an N-terminal CSD (note the four aforementioned sequences with variable N-termini) that is succeeded by other variable conserved motifs. As plant CSP sequences were compiled and analyzed, common domains in the C-terminal regions became more apparent. For the sake of simplification and classification, I propose a nomenclature system for the domains that are commonly present within cold shock domain proteins from photosynthetic organisms (Figure 6). The simplest CSP type, which I name C₁, is composed of only a CSD (C); it can be found in bacteria. CSPs from the Haptophyceae representatives *Prymnesium parvum* and *Isochrysis galbana* contain two tandem CSDs, and are thus denoted as C₂. To my knowledge, the observation of tandem cold shock domains (C₂) has never been previously described for a photosynthetic organism. Interestingly, the two *Chlamydomonas* species, *Chlamydomonas reinhardtii* and *Chlamydomonas incerta*, each encode CSPs containing both a CSD in the N-terminus and a RRM (M) in the C-terminus. These two domains are separated by a short stretch of RGRG (R) repeats; thus this type is named C₁R₁M₁. These RGRG, or in some cases RGG repeats are also known as RGG Boxes or GAR domains (Ginisty et al. 1999). CSPs in representative Bacillariophyta (diatoms), commonly contain a N-terminal CSD and highly variable C-termini. Thus,

I have not yet developed a complete nomenclature for their CSPs. The lack of complete sequence information for a Charophyta representative (*Closterium peracerosum*) prohibits the development of a complete nomenclature. In this study, retroviral-like CCHC zinc fingers first appeared in the C-terminus of the CSP from *Mesostigma viride*. This organism is currently debated to be the closest living relative of land plants (Lemieux et al. 2007). To the best of my knowledge, C-terminal zinc fingers remain present in all land plant CSPs including Bryophytes, Lycophytes, Pterophytes, and higher plants such as monocots and dicots. These C-terminal zinc fingers are commonly separated from the CSD by either a glycine-rich region (G), as in the case of *Mesostigma*, Lycophyta, and Higher plants, or a RGRG repeat (R) as seen in Bryophytes and Pterophytes. The number of zinc fingers is highly variable ranging from as few as one to as many as seven. It is of interest to note that conserved sequences such as DRGY, DRY, or RY are commonly interspersed in the glycine-rich regions between the zinc fingers. Additionally, the glycine-rich regions of at least five dicot proteins, for example *P.tremula* CK097750, are also asparagine-rich.

To gain a better understanding of the function of these proteins in plants, I collected predictive data relating sub-cellular localization patterns and trends for putative phosphorylation and sumoylation target sites. Of the 121 distinct CSP sequences included in my multi-sequence and phylogenetic analyses, only 70 were full-length and ranged in size from 68 to 301 amino acids. The SOSUI server (http://bp.nuap.nagoya-u.ac.jp/sosui/sosui_submit.html) predicted that all 70 of these full-length CSP proteins are soluble; it is unlikely, therefore, that they are membrane proteins. The TargetP v1.1 program did not predict any of these CSPs to localize to the chloroplast or mitochondria, or to be secreted. These data support the notion that CSPs are predominately localized in the nucleus and/or cytoplasm. The majority of these predictions are based on high reliability class (RC) values ranging from one to three. The RC value interpretation is as follows: RC value 1 indicates that difference between the highest score and the second highest score is greater than 0.800, RC value 2 follows the same idea but with a difference between 0.800 and 0.600, RC 3 indicates a difference between 0.600 and 0.400 and etc. Approximately 65.71% of the CSP protein predictions are at an RC value of 1 or 2. Ninety-percent of the proteins have an RC

value of 3 or above. In the case of an RC value of a 4 or 5, the prediction for mitochondrial localization was the next highest prediction. An example of this occurred with the *O.sativa* AK119758 CSP.

The Protein Calculator v3.3 program (<http://www.scripps.edu/~cdputnam/protcalc.html>) was used to calculate the isoelectric point and molecular weight for all full-length proteins. The molecular weights of the CSPs are highly variable, the lowest being 7,335.3 for the bacterial-type *H.vulgare* (BF253683) and the highest of 26,503.27 for *R.communis* (rcal.assembly.29405). The isoelectric points range from 4.52 for *P.tricornatum* (PTMM06075) to 9.44 for *Chlamydomonas* NAB1. More than 50% of the isoelectric points fall below pH 7.0, but approximately 27% are neutral and 14% are basic. Table 1 shows the amino acid compositions of both full-length proteins and full-length CSPs as calculated using the ExPASy ProtParam tool. Full-length CSPs, as well as CSDs on their own, both have glycine as the most common amino acid. Methionine and tryptophan are the least common amino acids in full-length CSPs and cysteine and tyrosine the least common in the CSD. NetPhos 2.0 identified three major, putative phosphorylation sites within the CSD itself. These include the serine/threonine (S/T) commonly present at the end of RNP-1, the S/T just after RNP-2, and the S/T located in the consensus (F/Y) R (S/T) L sequence between β -sheets 3 and 4 (indicated by shading in Figure 2). NetPhos 2.0 predicted that the S/T residue at the end of RNP-1 was phosphorylated with a score greater than 0.700 in 69% of all full-length plant CSPs. The software did not support phosphorylation of the S/T appearing just after RNP-2. However, the adjacent serine residue, appearing in many of the CSPs, (not shown in Figure 2) had a phosphorylation prediction with a score greater than 0.700 in 45% of the proteins. Phosphorylation at the S/T present in the consensus (F/Y) R (S/T) L sequence had the highest support with 79% of the proteins showing a prediction score of >0.700.

Sumoylation is another post-translational modification of proteins that has been of great interest recently. Since Miura et al. 2007 found that sumoylation of ICE1 was critical for its activation of CBF3, I was interested to determine if plant CSPs might be sumoylated. The sumoylation consensus motif is Ψ KX E/D, where Ψ denotes a large hydrophobic residue, K is the

lysine site where the Sumo protein is conjugated, X could be any amino acid, and D (aspartate) or E (glutamate) (as reviewed by Miura et al. 2007). The SUMOplot™ Analysis Program (<http://www.abgent.com.cn/doc/sumoplot/login.asp>) predicts that this consensus motif was present in 62 out of the 121 CSP proteins reported in this paper (Table 2) at one to four sites. The most conserved potential sumoylation site is indicated by an asterisk in Table 2. Subsequent *in vitro* experiments confirmed that plant CSPs are indeed targets of sumoylation (Kenji Miura, Center for Plant Environmental Stress Physiology, Purdue University).

Previous studies have shown that *AtCSP* expression is up-regulated by cold treatment (Karlson and Imai 2003; Fusaro et al. 2006; Kim et al. 2006). I first characterized the wild-type expression patterns of all four *Arabidopsis* CSP members in response to low temperature. Three-week old *Arabidopsis* plants were treated at 4°C and quantitative real-time PCR analysis deduced *AtCSP* transcript levels (Figure 7). *AtCSP2* and *AtCSP4* transcripts were found to be much more abundant than those of *AtCSP3* or *AtCSP1* (Fig 7a). Thus, the relative expression patterns for *AtCSP3* and *AtCSP1* are presented separately (Fig 7b). The expression of all four *AtCSPs* was up-regulated in response to low temperature. Interestingly, *AtCSP2* and *AtCSP4* transcript levels continue to increase after being shifted back to normal growing temperatures for forty-eight hours. Transcription of both *AtCSP3* and *AtCSP1* were induced by low temperature treatment with *AtCSP1* exhibiting the greatest response of any *AtCSPs*. Unlike *AtCSP2* and *AtCSP4* transcript levels, those of *AtCSP1* returned to basal levels within forty-eight hours of being shifted to normal growing conditions.

Various pathways in *Arabidopsis* have been linked to cold stress. These include, but are not limited to, ABA-dependent, CBF/DREB1 regulon, and the ICE1 regulon pathways. Since the function of *AtCSPs* is induced by cold treatment I was interested to determine whether any of these three signal transduction networks were involved in cold regulation of plant cold shock domain proteins. I first characterized *AtCSP* transcript levels in the ABA mutants *aba1-4*, *abi1-1* and *abi2-1* in the *Landsberg* ecotype. The *Landsberg* ecotype shows little to no expression of *AtCSP4* (Fig 7c). Comparative expression patterns of the *AtCSPs* in the *aba1-4* mutant revealed that the cold

regulation of *AtCSP2* and *AtCSP3* was altered (Figs 7c and 7d). For *AtCSP3*, the lack of endogenous ABA resulted in failure to respond to low temperature treatment. *AtCSP2* expression, which was unaffected by cold treatment in WT, decreased in response to cold treatment in *aba1-4*. Thus, the cold-related expression of both *AtCSP2* and *AtCSP3* appears to be ABA-dependent. I next examined the expression patterns of the *AtCSPs* in *CBF1*, 2, and 3 over-expression lines in the *Wassilewskija-2* (WS-2) ecotype. The transcription of *AtCSP2*, *AtCSP4*, and *AtCSP1* did not appear to be affected by the ectopic over-expression of any CBF transcription factor. Interestingly, *AtCSP3* exhibited an up-regulation in response to cold treatment in *35s::CBF3* over-expression background (Figs 7e and 7f). These data documenting constitutive up-regulation of *AtCSP3* in the CBF3 over-expression background suggest that the CBF3 transcription factor is important for the transcription of *AtCSP3*. The effect of *ICE1* on *AtCSP* transcription was the least clear of the cold pathway components analyzed. While in neither the *ice1* or the *ICE-OX* mutants the expression of *AtCSP1*, *AtCSP3* or *AtCSP4* appeared affected, *AtCSP2* expression showed no change in the *ice1* mutant but an increased expression in the *ICE-OX* mutant with cold treatment.

Thus, we conclude that *AtCSP2* and *AtCSP3* expression are related to ABA. The *AtCSP3* gene expression is up-regulated, at least in part, by the expression of CBF3. Finally, the expression of *AtCSP2* may be related to ICE1.

5. Discussion

Cold shock domain proteins were first reported to be widely distributed among plants by Karlson and Imai (2003). In their report, the conservation of cold shock domain proteins was noted from the Chlorophyte *Chlamydomonas reinhardtii* through ferns, gymnosperms, and higher plants such as monocots and dicots. In the present study, multiple gaps of plant evolutionary history have been filled by adding members from Bacillariophyta, Charophyta, Haptophyta, Lycopphyta, Bryophyta, and Angiosperms, as well as additional sequences from the groups previously represented. The number of known unique CSP sequences has been increased by four times.

This paper supports Karlson and Imai's conclusion that the CSD is highly conserved in plants (2003).

I had hoped that sequences from a more complete representation of the plant kingdom would provide a clear picture of the evolutionary history of CSPs in plants. This, however, was not the case. We suspect the reason for this is that the length of alignable CSD sequence is comprised of only of 65 amino acid residues, not providing a strong basis for comparison. Full-length CSP sequence could not be used in this analysis because the C-terminal motifs are highly variable between plant types and full-length sequence data is lacking in the databases for these proteins. Thus, instead of a clear evolutionary picture of CSPs in Figures 3 and 4, we are only able to observe the grouping of similar organisms. Earlier land plants group together and the single-celled autotrophs cluster together. The same is true of the neighbor-joining tree published by Karlson and Imai in 2003. They chose to divide their tree into two groups of CSP proteins, Type-I with N-terminal sequence and Type-II that are bacterial-like (Karlson and Imai, 2003). Clearly, these proteins are related to bacterial CSPs and remain labeled Type-II. I further divided the Type-I group into subgroups, two subgroups in monocots and three in dicots. The monocot I subgroup shows higher conservation than that of monocot II as can be seen in Figure 5. The major differences between the monocot sub-groups occur at 5 positions as indicated in orange (Figure 5). The dicots show much more variation between the subgroups than the monocots and these variations are indicated in blue (Figure 5). There are two positions (indicated in green) at which variation may not affect the activities of the proteins due to the conservation of amino acids with similar side-group properties at these sites (Figure 5). Interestingly, four CSPs in dicot subgroup II show conspicuous variation at the last two positions of β -5. Despite the variations between the major groupings of CSDs all amino acid sequences possess the key hydrophobic residues responsible for the three-dimensional structure of bacterial Csp's (columns indicated with an asterisk beneath them) (Yamanaka et al. 1998) (Figure 5). The conclusion made by Karlson and Imai in 2003 that plant CSPs may assume a similar 3-D structure as CspA from *E. coli* is further supported by this work.

The overall structure of plant CSPs is variable due to the presence of auxiliary C-terminal domains. However, the length of the CSD is very consistent and averages 73.5 ± 3 residues. At the time of this analysis, only 70 full-length CSPs had been identified with sizes ranging from 68 to 301 residues. As more genomic resources become available in the future, it is likely that larger CSP proteins will be identified. A common feature of higher plant CSPs is that they usually either possess RGRG repeats or have glycine-rich repeats in the C-terminal domain. According to the classification of Fusaro et al. in 2001 plant CSPs were classified as subclass III type glycine-rich proteins because their glycine content is between 20-70%. The full-length plant CSPs identified thus far range from 21.4 to 42.3 percent glycine (Table 1). Another consistent property these proteins share is that all full-length proteins were predicted to be soluble. This strongly supports the protein localization predictions which suggested that the majority of CSPs localize in the cytoplasm or the nucleus. Considering that this protein type is associated with various aspects of RNA processing in many living systems, these predictions are not unexpected. Experimental evidence has been reported in the literature that also support these predictions. Winter wheat WCSP1 was found to localize in the nucleus and the endoplasmic reticulum (Nakaminami et al. 2006). Fusaro et al. found that AtCSP2 (AtGRP2) localized in the cytoplasm and the nucleus of *Arabidopsis* cells (2006). NAB1 localized in the cytoplasm of *Chlamydomonas reinhardtii* (Mussnug et al. 2005). We also see similar localization patterns in the moss *Physcomitrella patens*. An intriguing aspect of these findings is that plant CSPs show similar localization patterns despite their motif differences. AtCSP2 is C₁R₂Z₂, WCSP1 is C₁R₃Z₃, and NAB1 is C₁R₁M₁. It will be interesting to see, if this is true for all plant CSP proteins.

Through evolutionary time, as organisms became more complex, plant CSPs became more diverse. This can be seen through the continual additions to and variations of the C-terminal region. The first discovered CSP, CspA of *Escherichia coli*, and many other bacterial types solely consist of a single CSD. Notable exceptions include the C₂ CSPs of *Caulobacter crescentus*, *Agrobacterium tumefaciens*, and other α -proteobacteria (Lang and Marques, 2004). The addition of a RRM or another CSD to the CSP proteins of the Chlorophytes and Haptophytes, respectively, suggests a

more potent activity of these proteins. A potential milestone in the evolution of CSPs was the appearance of a zinc-finger, or zinc knuckle, in *Mesostigma viride*. *Mesostigma viride* is debated to be one of the earliest streptophyte species and therefore represents one of the earliest relatives of land plants currently known (Lemieux et al. 2007). Almost all CSPs of plants after *M.viride* possess at least one zinc-finger. In fact, there is a longer read of the *C.richardii* BE641524 sequence, BQ086954, which shows that fern CSPs also can have zinc-fingers. Zinc-fingers are known to facilitate protein-RNA, protein-DNA, and protein-protein interactions. The conserved regions of these zinc-fingers are identical to those found in retroviruses and have a CarXCX₃GHhyd/arX₃C (ar for aromatic and hyd for hydrophobic) consensus sequence (Dannull et al. 1994). These zinc-fingers have been shown to increase ssDNA binding and are linked to specific RNA recognition of the HIV-1 virus (Dannull et al. 1994). Consequently, it is reasonable to hypothesize that the zinc-fingers, as well as the CSD, of CSPs could be responsible for nucleic acid binding, possibly inferring some specificity. In the case of WCSP1, the glycine-rich regions and zinc fingers were required for dsDNA binding (Karlson et al. 2002). For *Arabidopsis* AtCSP1/CSDP1, the zinc-finger and glycine-rich regions were found to be important for RNA chaperone activity and nucleic acid-binding (Kim et al. 2006). The first glycine-rich region and zinc-finger were necessary for WCSP1 to localize to the nucleus and endoplasmic reticulum (Nakaminami et al. 2006). Other potential activities of zinc-fingers, such as facilitating protein-protein interaction, also remain a possibility for affecting function of plant CSPs.

CSPs of eukaryotes were found to have additional motifs in their C-terminal regions, such as RG repeats, basic aromatic islands, zinc-fingers, RRM, and additional CSDs (reviewed by Sommerville 1999; Musgnug et al. 2005; reviewed by Graumann and Marahiel 1998). In this study, we have not identified clear basic/aromatic islands in plant CSPs, but the other aforementioned domains were found. Many of the motifs present in the C-terminal region are also implicated in binding nucleic acids and/or mediating protein-protein interaction, for example the glycine-rich region of AtCSP2 mentioned above. The glycine-rich regions in Chlorophyta, Bryophyta, and Pterophyta most closely resemble the RGG boxes, but the glycine-rich regions of the other plant

CSPs have been compared to RGG boxes and could be a variation on this motif (Nakaminami et al. 2005). It is very interesting to note that the positively charged arginines of the RGRG, or RGG boxes have been found to increase the affinity for RNA by RNA-binding proteins by 10-fold (Burd and Dreyfuss, 1994). In the case of WCSP1, Nakaminami et al. showed that the CSD as well as the first glycine-rich region are required for ssDNA and mRNA binding in 2005. It is possible that modification of these arginine residues via methylation aids in mediating RNA-binding and protein-protein interactions (Gendra et al. 2004). The RGG domain has also been found to be the crucial sequence required for nuclear and nucleolar targeting of proteins (Gendra et al. 2004). It is also possible that the glycine-rich regions serve another purpose. In 1991, Steinert et al. proposed a glycine-loop velcro model for glycine-rich proteins. In their model, the amino acids interspersed in the glycine-rich regions pack together to form glycine-loops that could interact like velcro with other proteins or direct subcellular localization (Steinert et al. 1991). This could be the reason that some proteins have DRGY or DRY sequences interspersed in the glycine-rich regions. Another reason for these sequences could be for mediating protein-RNA interactions.

Another interesting prediction for the plant CSPs is regarding phosphorylation. There are three obvious potential phosphorylation sites in the CSD; these are shaded in Figure 2. Based on the fact that of almost 80% of plant CSPs obtained a prediction score greater than 0.700 for phosphorylation at this site, the best candidate for phosphorylation is the S/T in the consensus (F/Y) R (S/T) L. Normally, if a protein can be phosphorylated, this modification is critical for protein function. For this reason, there would be a very strong selective pressure against mutations at the phosphorylation site, or essential or important proteins. This site is also the most conserved; only two plant CSPs identified contain something besides serine or threonine at this position. In concert with this argument, the human Y-box 1 protein is phosphorylated at an analogous site (Sutherland et al. in 2005).

Recently, it was found that the ability of ICE1 (inducer of CBF expression 1) to activate the CBF3/DREB1A cold signaling regulon is dependent on its sumoylation (Miura et al. 2007). Sumoylation occurs when a SUMO (small ubiquitin-related modifier) protein is attached to a target

protein, similar in a manner to ubiquitination (Melchior 2000). Unlike ubiquitin, however, the SUMO protein does not mark its target for degradation; instead it may promote other activities such as protein-protein interaction, stress response, nuclear transport, and cell cycle (Melchior 2000; reviewed by Miura et al. 2007). About half of the CSPs identified in this paper were predicted to contain one or more consensus sumoylation sites (Table 2). Since sumoylation is involved in stress response (Miura et al. 2007), it was of interest to determine if CSPs are actually sumoylated. Plant CSPs were indeed found to be sumoylated (Kenji Miura, Center for Plant Environmental Stress Physiology, Purdue University). In the future, it will be of great interest to determine what role this modification plays with regards to the activity of these proteins.

All four of the *AtCSPs* are induced by cold treatment (Figure 7a and 7b) (Karlson and Imai 2003; Fusaro et al. 2006; Kim et al. 2006). Various pathways have been identified that are induced by cold-stress. I investigated major pathways of cold stress induction: the ABA-dependent pathway and the CBF/DREB1 regulon. Figures 7c and 7d show that the cold induction of both *AtCSP2* and *AtCSP3* are ABA-dependent, because their expression was not induced by cold in the *aba1-4* background. The ABA-insensitive mutants, *abi1-1* and *abi2-1*, did not appreciably affect the cold induction of *AtCSP* transcription. The *aba1-4* mutant contains a premature stop codon in the gene coding for the zeaxanthin epoxidase enzyme (Barrero et al. 2005). For this reason, the ABA precursors zeaxanthin and antheraxanthin are not synthesized into violaxanthin and neoxanthin. This epoxidation represents an important step in the ABA biosynthesis pathway (Duckham et al. 1991; Rock and Zeevaart 1991). It is interesting to note that even though *aba1-4* exhibits one of the greatest ABA deficiencies among the *aba* mutants, it still possesses ABA levels 30% of those found in wild-type plants (Gilmour and Thomashow 1991; Barrero et al. 2005). The *ABI1* and *ABI2* genes were found to be important members of the ABA cold signal transduction pathway (reviewed by Shinozaki and Yamaguchi-Shinozaki 1996). They code for homologous type 2C protein phosphatases, in fact the *abi1* and *abi2* mutants have the same mutation (Leung et al. 1997). Thus, if the *AtCSP2* and *AtCSP3* genes are part of the ABA cold signal transduction pathway then it is reasonable to consider that their expression would be affected in these mutant backgrounds as well.

A potential reason why this effect is not seen is that ABI1 and ABI2 have been shown to have some overlapping function (Leung et al. 1997); therefore in either mutant, the functioning ABI could be acting in a redundant fashion in relation to the *AtCSP2* and *AtCSP3* expression patterns.

Over-expression of *CBF3/DREB1A* confers freezing tolerance (Liu et al. 1998) and leads to the induction of molecules related to freezing tolerance (Gilmour et al. 2000). *AtCSP3* expression was already induced at time zero in the *35S:CBF3* mutant background (Figure 7f) indicating that *AtCSP3* is located somewhere downstream in the CBF3 signal transduction pathway. Recently, Maruyama et al. identified genes downstream of CBF3/DREB1A using a cDNA microarray and an Affymetrix GeneChip, but *AtCSP3* was not among them (2004). Due to its very low expression levels, it is possible the *AtCSP3* gene did not satisfy the criteria set for identifying genes of interest from this study. The study did identify two categories of up-regulated genes, genes that directly function in cold stress tolerance and genes responsible for further regulation (Maruyama et al. 2004). It is likely that *AtCSP3* falls into the latter category because plant CSPs are RNA-binding proteins and may in turn affect the subsequent expression of other genes. Liu et al. found that *CBF3/DREB1A* was not induced by ABA (1998), leading to the idea that the CBF/DREB1 family of transcription factors belonged to an ABA-independent cold signaling pathway. Narusaka et al. also confirmed this idea in 2003. More recently, however, members of the CBF/DREB1 family, like *CBF3/DREB1A*, showed induction after treatment with ABA (Knight et al. 2004). Interestingly, the DRE/CRT promoter element target of the CBF/DREB1 TFs, which was once also thought to be ABA-independent (Yamaguchi-Shinozaki and Shinozaki 1994), has also been shown to be a target of ABA (Knight et al. 2004). With this information, the apparent link to induction of *AtCSP3* by both ABA and CBF3 is not unlikely. It is possible that the ABA induction of *AtCSP3* could occur downstream in the CBF3/DREB1A regulon or that it is a member of both pathways exclusive of each other. Further experiments are needed to pinpoint the location of the *AtCSPs* in both of these signal transduction pathways.

In 2003, Chinnusamy et al. found that ICE1 is an upstream activator of the *CBF3/DREB1A* gene, and therefore has a drastic effect on the CBF regulon via the CBF3/DREB1A pathway. The

ice1 mutation abolished the transcription of *CBF3* and in turn produced plants that could not effectively cold acclimate (Chinnusamy et al. 2003). As expected, the effect of the *ice1* mutation is carried to genes downstream of *CBF3* sometimes even without cold treatment (Lee et al. 2005). In some cases, cold-related genes are up-regulated by the *ice1* mutation (Lee et al. 2005). The results of our study are not conclusive. However, it is possible that *ICE-OX* up-regulates *AtCSP2* expression with cold treatment as compared to the wild-type. These data suggest that *AtCSP2* may function downstream of the ICE1 regulon. The observation that *AtCSP2* was not up-regulated in *35S::CBF3* suggests that may not be a member of the ICE1-CBF3 pathway, but part of another pathway up-regulated by ICE1. Further analysis is necessary to re-confirm this pattern.

Almost all plants with available EST data contain one or more CSP genes. The cold shock domain, first discovered in bacteria, is an important and common RNA-binding domain, which facilitates adaptation to low temperature stress. In addition to an N-terminal cold shock domain, nearly all plant CSPs also contain additional motifs in their C-terminal region. Interestingly, as plants progress in “complexity” during the course of evolution, these C-terminal regions become more complex. Since these C-terminal motifs are linked to nucleic acid-binding as well as protein-protein interaction, it is tempting to speculate that this diversification of C-terminal regions of plant CSPs possibly led to more specialized *in vivo* functions or preferential binding of nucleic acids as these proteins evolved. Plant CSP function has been recently linked to similar action as their bacterial counterparts, such as cold induction, aiding in cold tolerance, and RNA-chaperone activity. Fusaro et al. extended the functional analysis of plant CSPs in 2006 and linked the activity of *AtCSP2* with floral and embryo development. So, what is the role of these proteins in plants? We hope that this paper will serve as a solid foundation for future explorations designed to increase our understanding of plant CSPs and their *in vivo* functional roles.

Figure 1: Multiple Sequence Alignment. This is an alignment of the deduced amino acid sequence of the N-terminal CSD of plant CSPs generated using Clustal-X. Completely conserved sites are indicated by an asterisk above the column, and sites with conserved substitutions are indicated by either a colon or a period above the column. The amino acids with similar properties share the same color. The organism names are given as an abbreviated genus name followed by the species name and they are followed by their accession number from their respective databases (see materials and methods for more detail). An asterisk between the organism name and accession number indicates a sequence that was previously published by Karlson and Imai in 2003. Number symbols above the accession number indicate an organism name that has been abbreviated, their real names are as follows: *Citrus* is *Citrus x paradisi*, *Malus* is *Malus x domestica*, *Aquilegia* is *Aquilegia formosa* x *Aquilegia pubesens*, and *Closterium* is *Closterium peracerosum-strigosum-littorale-complex*. The bar graph along the bottom of the alignment aids in the visualization of highly conserved sites versus sites with low conservation.

R ₈₃	V ₂₀	T ₅₅	G ₁₂₀	T ₅₆	V ₁₂₀	K ₁₀₄	W ₁₁₈	F ₁₂₁	N ₇₈		K ₁₁₉	G ₁₁₉	F ₁₀₃	G ₁₂₁	F ₁₂₁	I ₁₂₀	T ₇₇	P ₁₂₀		L ₁₀₇	F ₁₂₁	V ₁₂₀	H ₁₂₁	Q ₁₁₂	S ₈₈
K ₂₆	L ₁₉	K ₂₆	R ₁	K ₄₈	A ₁	N ₄	R ₂		S ₂₂		R ₂	C ₂	Y ₁₇			L ₁	S ₂₅	T ₁		V ₁₀		I ₁		F ₆	T ₃₀
A ₃	S ₁₈	S ₁₉		V ₅		R ₃	K ₁		D ₂₀				S ₁				K ₆				I ₃			H ₃	K ₁
P ₃	K ₁₄	R ₈		S ₃		S ₃			A ₁								A ₆				F ₁				L ₁
V ₃	Q ₁₁	Q ₅		I ₂		Q ₂											V ₂								A ₁
Q ₂	M ₁₀	I ₃		E ₂		V ₂											H ₂								
L ₁	R ₉	L ₂		Q ₁		L ₂											R ₁								
	E ₆	N ₁		C ₁		T ₁											E ₁								
	I ₅	V ₁		M ₁													M ₁								
	A ₄	F ₁		N ₁																					
	F ₂			R ₁																					
	C ₁																								
	H ₁																								
	T ₁																								

F ₉₂	R ₁₁₆	S ₉₉	L ₁₁₈		G ₁₀₆	R ₉₃	T ₇₅	K ₁₀₅	A ₁₂₁	V ₆₃	D ₈₁	V ₁₂₁	T ₁₁₀	G ₈₈	P ₁₀₉
Y ₂₉	K ₅	T ₂₀	V ₃		K ₅	K ₂₂	S ₁₀	Q ₆		L ₁₇	N ₁₈		S ₄	A ₂₉	I ₅
		C ₁			D ₃	G ₅	P ₁₀	S ₅		I ₁₅	E ₁₂		K ₄	P ₃	L ₃
		V ₁			N ₃	Q ₁	A ₉	R ₄		A ₁₂	A ₅		R ₁	S ₁	S ₁
					S ₂		L ₅	M ₁		S ₄	S ₂		V ₁		A ₁
					V ₁		Y ₄			T ₃	Q ₂		I ₁		R ₁
					Q ₁		M ₃			C ₂	K ₁				M ₁
							K ₂			G ₁					
							V ₂			Q ₁					
							W ₁			H ₁					
										N ₁					
										M ₁					

Figure 2: Regions of high amino acid sequence similarity with the CSD. The subscripts indicate the number of times (out of 121) each particular amino acid was present at the given position. The shaded position highlights a highly conserved predicted phosphorylation site.

Figure 3: Neighbor-Joining Tree for phylogenetic analysis of plant CSDs. This neighbor-joining tree was made using the PHYLIP program using the multiple sequence alignment shown in figure 1. The tree was edited in the TreeViewPPC program in order to root it to CspA from *E. coli* and the names were further adjusted and colors and labels were added using Adobe Illustrator. For confidence the bootstrap value was set to 1000 and the branch lengths indicate evolutionary distance. All proteins not labeled Type II fall into the Type I category of plant CSPs. The monocot sequences are indicated in blue and group into two subgroups of the Type I CSPs. The dicot sequences are indicated in red and group into three subgroups of the Type I CSPs. The yellow highlighted area indicates CSPs that grouped with the dicot subgroup II in this tree but grouped with subgroup III in the maximum likelihood tree. An asterisk between the organism name and accession number indicates that this sequence was previously reported by Karlson and Imai 2003.

Figure 4: Maximum likelihood tree showing phylogenetic relationships between plant CSDs. This maximum likelihood tree was made using the neighbor-joining tree in figure 3 in the MOLPHY programs. The names were adjusted and colors and labels were added using Adobe Illustrator. For confidence, the bootstrap value was set to 100 and the branch lengths indicate evolutionary distance. All of the organisms that are not labeled Type II are in the Type I category of plant CSPs. Like in the NJ tree, monocot sequences are indicated in blue and group into two subgroups of the Type I CSPs. The dicot sequences are indicated in red and group into three subgroups of the Type I CSPs as they did in the NJ tree. An asterisk between the accession number and organism name indicate that this sequence was published by Karlson and Imai in 2003.

		Beta Sheet 1										Beta Sheet 2									
Monocots 1	R	X	+	G	p/n	V	K	W	F	p	p/-	p/n	p/n	K	G	F	G	F	I	p	P
Monocots 2	R	br	X	G	T	V	K	W	F	p/-	p/n	p/-	K	G	F	G	F	I	p	P	
Dicots 1	R	p/n/ar	p/n/+	G	p/n/+	V	p/n/+	ar/+	F	p/-	p/n/-	p/n/+	K	G	ar	G	F	I	p/+	P	
Dicots 2	+	p/n/+	p/n/+	G	p/n	V	K	W	F	p/n/-	p/n/-	p/+/-	K	p/n	F	G	F	I	p	P	
Dicots 3	+	n/+/-	p/+	n/+	p/+	V	p/+	W	F	p/-	n/-	p	K	G	p/ar	G	F	I	p/+	p/n	
		*										*									
		Beta Sheet 3																			
Monocots 1	-	D	G	p/n		E	D	L	F	V	H	Q	S	S	I	p/+	p/n	-	G		
Monocots 2	-	p/-	n	p/n		E	D	L	F	V	H	Q	S	p/n	br	K	p/n	D	G		
Dicots 1	D	p/n/-	p/n/-	p/n/-	p/bl	n/+/-	-	L	F	V	H	p/+	p/+	p/n/-	I	br/+	S	-	G	bl/n/+	
Dicots 2	X	D	G	p/n		p/-	D	L	F	V	H	p/ar	p	p/n	I	p/n/+	S	p/-	G		
Dicots 3	p/-	p/n/-	p/n	p/n		-	-	br/ar	F	br	H	Q	p	p/-	I	p/+	p/n	-	G		
		*										*									
		Beta Sheet 4																			
Monocots 1	ar	R	p	L	n	E	G	E	X	V	E	F	p/n	br	p/n	p/-	p/n	-	D		
Monocots 2	ar	R	S	L	p/+/-	-	p/n	n/-	p/n	V	E	ar	p/n/-	br	n/-	pol	G	p/-	D		
Dicots 1	ar	R	p	br	p/n/ar	br/-	p/n/-	p/-	p/n/+	V	E	ar	X	br,n	p/n/-	p/n/ar	X	p/n/-	n/-		
Dicots 2	ar	R	p	L	p/n	n/-	p/n/-	p/-	p/n/-	V	-	ar	p/n/-	br	p/-	p/n/ar	n/-	p/n/-	p/-	p/bl	
Dicots 3	ar	R	p	L	n	br/-	n/-	-	p/n	V	E	ar	p/n	br	n/-	p/+	p/-	p/n	n/-		
		*										*									
		Beta Sheet 5																			
Monocots 1	G	R	T	K	A	V	-	V	T	G	P										
Monocots 2	G	R	p/n	K	A	p/n/+	D	V	T	n	P										
Dicots 1	p/n	K	p/n/ar	p/+	A	br	p/-	V	T	n	n										
Dicots 2	p/n	R	p/n	n/+	A	p/br	p/-	V	p/br	n/p	p/n										
Dicots 3	p/n	R	p/n	K	A	n	p/-	V	p	G	P										

p = polar

n = nonpolar

ar = aromatic

br = branched

bl = blank site

X = any

(+) = positively charged

(-) = negatively charged

p = polar ar = aromatic bl = blank site (+) = positively charged
n = nonpolar br = branched X = any (-) = negatively charged

Figure 5: Sequence comparison of CSDs between the monocot subgroups and the dicot subgroups. Yellow highlights the more major differences between the two monocot subgroups and blue indicates the significant differences between the dicot subgroups. Green highlights indicate protein sequence differences within the dicot subgroups that normally do not vary among CSDs, but may not have an effect on the overall protein activity. An asterisk below a column denotes the hydrophobic core residues of the CSD. Capital numbers are standard amino acid abbreviated names, lower case letters indicate the character of amino acid side groups. p stands for polar, uncharged amino acids, n stands for nonpolar aliphatic side groups, ar indicates a position occupied by an amino acid with an aromatic side group, br indicates an amino acid with a branched side group, bl indicates a position at which some proteins have a blank site, X indicates any amino acid, + stands for a positively charged amino acid and - stands for a negatively charged amino acid.

	Motifs			Nomenclature
Bacterial	CSD			C ₁
Haptophyta	CSD	CSD		C ₂
Bacillariophyta	CSD	Variable		C ₁ ?
Chlorophyta	CSD	RGRG	RRM	C ₁ R ₁ M ₁
Charophyta	CSD	Glycine	?	C ₁ G ₁ ?
Mesostigma	CSD	Glycine	Zn	C ₁ G ₁ Z ₁
Bryophyta	CSD	RGRG	Zn Glycine Zn	C ₁ R ₁ G ₁ Z ₂
Lycophyta	CSD	Glycine	Zn Glycine Zn	C ₁ G ₁ Z ₂
Pterophyta	CSD	RGRG	Zn ?	C ₁ R ₁ Z ₁ ?
Higher Plants	CSD	(Glycine Zn) _n		C ₁ G _n C _n

Figure 6: Motifs found in CSPs from representatives of the indicated phyla and proposed CSP nomenclature. The nomenclature system is designed as follows: C stands for CSD, R stands for a region of RGRG repeats, M stands for an RRM domain, Glycine indicates a glycine-rich region and is represented by a G, and Zn represents retroviral-like zinc fingers. In higher plants the glycine-rich region and zinc fingers are present in multiple copies as designated by the n subscript. The question marks indicate where sequence information was not sufficient to determine the C-terminal motifs.

Table 1: This table shows the average percent amino acid composition of both full-length CSPs and the CSD only. These percentage compositions were gathered using the ExPASy Protparam tool available at <http://expasy.org/tools/protparam.html>.

Amino Acid	Full Length	CSD Only
Alanine (A)	5.08 ± 2.50	5.61 ± 2.28
Arginine (R)	6.53 ± 2.00	4.41 ± 1.32
Asparagine (N)	2.89 ± 1.82	2.32 ± 1.45
Aspartate (D)	5.64 ± 1.41	8.83 ± 2.34
Cysteine (C)	3.16 ± 1.69	0.16 ± 0.57
Glutamine (Q)	2.30 ± 1.27	3.20 ± 2.03
Glutamate (E)	5.48 ± 1.62	8.82 ± 2.26
Glycine (G)	31.85 ± 10.42	13.04 ± 2.04
Histidine (H)	1.94 ± 0.59	1.74 ± 0.72
Isoleucine (I)	2.15 ± 1.34	3.90 ± 1.18
Leucine (L)	1.88 ± 1.04	3.50 ± 1.08
Lysine (K)	3.84 ± 2.20	6.29 ± 1.44
Methionine (M)	1.07 ± 0.75	1.61 ± 0.63
Phenylalanine (F)	4.51 ± 1.91	7.45 ± 1.23
Proline (P)	2.61 ± 1.28	2.99 ± 0.72
Serine (S)	6.85 ± 1.80	8.61 ± 2.40
Threonine (T)	3.51 ± 1.57	6.62 ± 2.04
Tryptophan (W)	0.74 ± 0.44	1.35 ± 0.26
Tyrosine (Y)	3.34 ± 1.68	1.02 ± 1.19
Valine (V)	4.55 ± 1.94	8.57 ± 1.56
Total Proteins	70	117

Table 2: This table shows the predicted sumoylated lysine residues (K) within target sequences and their prediction scores. The position indicates the position of the lysine with regards to the entire sequence of the protein and some proteins have more than one predicted site. A number symbol after the accession number indicates a full-length protein sequence. An asterisk next to the sequence indicates a highly conserved region within plant CSPs.

Organism	Accession	Plant Type	Sumoylation Site Number	Position	Sequence	Score
<i>I.galbana</i>	EC146548#	Haptophyta	1	K77	VKGD	0.93
			2	K135	GKED	0.67
<i>P.parvum</i>	DV100020#	Haptophyta	1	K76	VKGD	0.93
<i>F.cylindrus</i>	DR026829#	Bacillariophyta	1	K42	VKKG	0.76
<i>T.pseudonana</i>	xzd1827.x1#	Bacillariophyta	1	K155	VKLG	0.76
<i>T.pseudonana</i>	xzd3738.x1#	Bacillariophyta	1	K43	VKKG	0.76
			2	K76	IKVE*	0.94
<i>Closterium%</i>	BW646729	Charophyta	1	K46	IKAE*	0.94
<i>M.viride</i>	DN255890#	Mesostigma	1	K146	FKCG	0.68
<i>S.moellendorffii</i>	contig12#	Lycophyta	1	K49	LKDG	0.73
			2	K142	FKCG	0.68
<i>S.moellendorffii</i>	contig2#	Lycophyta	1	K49	LKDG	0.73
			2	K141	FKCG	0.68
<i>C.richardii</i>	BE643472	Pterophyta	1	K79	VKGG	0.76
<i>A.speltoides</i>	BQ840577#	Monocot	1	K38	IKSD*	0.94
<i>A.cepa</i>	CF448014#	Monocot	1	K37	IKSE*	0.94
			2	K254	GKEE	0.67
<i>A.officinalis</i>	CV292210#	Monocot	1	K80	VKGD	0.93
<i>B.distachyon</i>	DV479026	Monocot	1	K40	IKAD*	0.94
			2	K142	FKCG	0.68
<i>B.distachyon</i>	DV471044#	Monocot	1	K38	IKSD*	0.94
<i>B.distachyon</i>	DV479359#	Monocot	1	K38	IKSD*	0.94
<i>B.distachyon</i>	DV477485	Monocot	1	K38	IKSD*	0.94
<i>E.coracana</i>	CX265498	Monocot	1	K39	IKSE*	0.94
			2	K78	VKGG	0.76
<i>E.tef</i>	DN481300	Monocot	1	K39	IKSD*	0.94
<i>H.vulgare</i>	CV063400#	Monocot	1	K40	IKAD*	0.94
			2	K146	FKCG	0.68
<i>H.vulgare</i>	CV054578#	Monocot	1	K38	IKSD*	0.94
<i>H.vulgare</i>	AL501518#	Monocot	1	K58	AKGP	0.69
<i>H.vulgare</i>	BF253683#	Monocot	1	K58	AKGP	0.69
<i>O.sativa</i>	AK119758#	Monocot	1	K40	IKAD*	0.94
			2	K79	VKGG	0.76
			3	K128	FKCG	0.68
			4	K163	FKCG	0.68
<i>O.sativa</i>	AK101577#	Monocot	1	K39	LKSD	0.91
<i>P.virgatum</i>	DN143581	Monocot	1	K40	IKSE*	0.94
			2	K79	VKGG	0.76
<i>S.officinarum</i>	CA142059#	Monocot	1	K39	IKSD*	0.94
			2	K47	LKEG	0.73
<i>S.officinarum</i>	CA124251	Monocot	1	K40	IKSE*	0.94
			2	K79	VKGG	0.76
<i>S.bicolor</i>	CD224347	Monocot	1	K40	IKSE*	0.94
			2	K79	VKGG	0.76
			3	K165	FKCG	0.68

<i>S.bicolor</i>	BE917858#	Monocot	1	K39	IKCD*	0.94
			2	K47	LKEG	0.73
			3	K160	FKCG	0.68
<i>T.aestivum</i>	BQ578846#	Monocot	1	K59	AKGP	0.69
<i>T.aestivumWCSP3</i>	CV768436#	Monocot	1	K38	IKSD*	0.94
<i>T.aestivumWCSP2</i>	AB158409#	Monocot	1	K38	IKAD*	0.94
<i>T.aestivumWCSP1</i>	AB066265#	Monocot	1	K38	IKSD*	0.94
<i>T.turgidum</i>	BE428632	Monocot	1	K38	IKAD*	0.94
<i>Z.aethiopica</i>	AJ701877	Monocot	1	K24	IKPD	0.94
<i>Z.mays</i>	CD438932#	Monocot	1	K40	IKSE*	0.94
			2	K164	FKCG	0.68
<i>Z.mays</i>	CK371153#	Monocot	1	K39	IKSD*	0.94
			2	K161	FKCG	0.68
<i>Z.officinale</i>	DY377474	Monocot	1	K41	IKSD*	0.94
<i>AtCSP4</i>	At2g21060#	Dicot	1	K140	FKCG	0.68
<i>AtCSP1</i>	At4g36020#	Dicot	1	K83	LKKE	0.91
			2	K194	VKGG	0.76
<i>Citrus x paradisi</i>	DN958776	Dicot	1	K40	IKSD*	0.94
<i>F.vesca</i>	DY669397#	Dicot	1	K39	IKSD*	0.94
			2	K80	GKKE	0.67
<i>G.max</i>	CO979596#	Dicot	1	K38	IKSD*	0.94
<i>G.arboreum</i>	BG440883	Dicot	1	K46	IKSE*	0.94
<i>G.raimondii</i>	CO097977#	Dicot	1	K46	IKSE*	0.94
<i>L.perennis</i>	DW087293	Dicot	1	K152	FKCG	0.68
<i>L.sativa</i>	BQ854588#	Dicot	1	K144	FKCG	0.68
<i>L.japonicus</i>	BI418821	Dicot	1	K23	IKPD	0.94
<i>L.japonicus</i>	AW164158	Dicot	1	K39	IKSE*	0.94
<i>L.esculentum</i>	BW692471	Dicot	1	K81	VKGG	0.76
			2	K152	FKCG	0.68
<i>L.hirsutum</i>	DN169821#	Dicot	1	K81	VKGG	0.76
			2	K152	FKCG	0.68
<i>Malus x domestica</i>	CN943096	Dicot	1	K149	FKCG	0.68
<i>M.truncatula</i>	BG586674	Dicot	1	K23	IKPD	0.94
<i>M.truncatula</i>	BG451829	Dicot	1	K38	IKTD*	0.94
<i>NsGRP2</i>	NSGRP2MR#	Dicot	1	K161	FKCG	0.68
<i>N.tabacum</i>	EB428117#	Dicot	1	K161	FKCG	0.68
<i>P.coccineus</i>	CA898605	Dicot	1	K38	IKSD*	0.94
<i>P.tremula</i>	CK097750#	Dicot	1	K34	IKPD	0.94
			2	K50	IKSD*	0.94
<i>R.communis</i>	rca1.assembly.29405#	Dicot	1	K27	IKPD	0.94
			2	K43	IKSD*	0.94
			3	K155	FKCG	0.68
<i>S.tuberosum</i>	CK246632#	Dicot	1	K81	VKGG	0.76
			2	K151	FKCG	0.68
<i>S.tuberosum</i>	CK279018#	Dicot	1	K25	IKPD	0.94
			2	K41	IKSD*	0.94
			3	K61	VKGD	0.93

Figure 7: Quantitative Real-time PCR data for the *AtCSPs* within cold regulons. a. and b. indicate the cold regulation of the *AtCSPs*. c. and d. show the cold regulation of the *AtCSPs* in the *Landsberg erecta* and ABA mutants, *aba1*, a null mutant and *abi1-1* and *abi2-1* ABA-insensitive mutants. e. and f. graphs show the regulation of the *AtCSPs* in relation to cold in the CBF over-expression lines (*35S::CBF1*, *35S::CBF2*, and *35S::CBF3*) and their respective vector and wild-type background. g. and h. show the cold regulation of the *AtCSPs* in the ICE1 over-expression line (*ICE-OX*), *ice1* null mutant line, and their respective empty vector and wild-type line.

CHAPTER 2: PHYSCOMITRELLA PATENS COLD SHOCK DOMAIN PROTEINS

1. Abstract

Physcomitrella patens ssp. patens is a bryophyte that has been gaining popularity as a model system for molecular biology studies. Its evolutionary position, as well as its capability for efficient gene targeting, makes it a very unique system for the *in vivo* analysis of gene function. Cold Shock Domain Proteins (CSPs) are among the most evolutionarily conserved nucleic acid binding proteins. They are present in bacteria, plants, and animals. Functions of Cold Shock Domain Proteins (CSPs) from higher plants, such as winter wheat and *Arabidopsis thaliana*, are currently under investigation. This study identifies four CSPs in *Physcomitrella patens* and begins to determine CSP function in this basal land plant lineage. *PpCSP* gene expression is cold inducible and PpCSPs are capable of binding nucleic acids *in vitro*. Expression of PpCSPs with C-terminal YFP and GUS fusions were used to determine their sub-cellular localization during *P.patens* development as controlled by their respective native promoters. Signals were detected most highly in new and developing tissues of protonema, gametophores, gametangia, embryos, and developing sporophytes. PpCSP1 and PpCSP2 localize mainly in the cytoplasm with additional localization in the nucleolus after a recent cell division. Both PpCSP1 and PpCSP2 localized to cytoplasmic foci, resembling cytosolic processing bodies (P-bodies). Collectively, these data strongly suggest a role for the PpCSPs in the processes of RNA metabolism.

2. Introduction

The Cold Shock Domain (CSD) is hypothesized to have appeared before single-celled evolution, placing it among the most ancient domains known (as reviewed by Graumann and Marahiel 1998). The first cold shock protein to be identified, CspA, was characterized in *Escherichia coli* (Jones et al. 1987). CspA is composed of only a CSD while in plants and animals the CSD is often paired with various auxiliary motifs. Within the CSD, two consensus RNA binding motifs, RNP1 and RNP2, reside on adjacent β -sheets of a 5 anti-parallel β -barrel structure (Manival et al. 2001; Newkirk et al. 1994; Schindelin et al. 1993). CspA, and three other members of the nine-member family of the *E. coli* cold shock domain family, CspB, CspG, and CspI, are induced following cold stress (Jones et al. 1987; Lee et al. 1994; Nakashima et al. 1996; Bae et al. 1999; Wang et al. 1999). In 1997, Jiang et al. determined that CspA displays RNA chaperone activity. CspA, CspC and CspE all exhibit transcription anti-termination activity both *in vivo* and *in vitro* (Bae et al. 2000). A quadruple deletion of four *E. coli* Csps results in a cold-sensitive phenotype, confirming that Csps are necessary for cold adaptation (Xia et al. 2001).

Plant Cold Shock Domain Proteins (CSPs) have been shown to display similar function with their bacterial counterparts. Cold shock domain protein expression from *Arabidopsis thaliana* (AtCSPs), as well as winter wheat WCSP1, is induced by low temperature stress (Karlson and Imai 2003; Fusaro et al. 2006; Kim et al. 2007; Sasaki et al. 2007; Karlson et al. 2002). Furthermore, WCSP1, AtCSP4 (CSDP1), and AtCSP1 (AtGRP2, AtCSP2) also bind nucleic acids and exhibit *in vivo* chaperone activity (Karlson et al. 2002; Kim et al. 2006; Fusaro et al. 2006; Sasaki et al. 2007; Nakamami et al. 2006). Most interestingly, these three genes also complement the cold sensitive phenotype of the BX04 *E. coli* cells (Nakaminami et al. 2006; Kim et al. 2006; Sasaki et al. 2007). Although studies with plant CSPs have demonstrated functionality similar to bacterial Csps, little is known about their *in vivo* functional role *in planta*.

Approximately one decade ago, the moss *Physcomitrella patens* was the first plant identified with the capability of gene targeting with high efficiency (Schaefer and Zryd, 1997). This,

and other factors including, a short life cycle with a predominant haploid phase, simple body plan, and straightforward cultivation make this bryophyte an ideal plant model system for whole plant functional studies. Additionally, bryophytes are among the first land plants, appearing about 400 million years ago, which place them in a key position for understanding the evolution of plant molecular processes.

In this study we identified 4 cold shock domain proteins, PpCSPs (*Physcomitrella patens* **C**old **S**hock Domain **P**roteins), within the *P.patens* genome. We show that like other plant CSPs, moss PpCSPs are capable of nucleic acid binding and that their expression is cold inducible. In order to shed light on their physiological function, we examined their sub-cellular localization utilizing PpCSP-YFP and -GUS lines. PpCSP1 and PpCSP2 proteins were found in moss at all stages of development including protonema, caulonema, chloronema, gametophores, gametangia including antheridia and archegonia, early embryos, and developing sporophytes. They were mainly localized in cytoplasm but were also found in the nucleolus of recently divided cells. Cytoplasmic localization also occurred in discrete cytoplasmic foci that I hypothesize to be P-bodies. Thus, these data collectively support the involvement of the PpCSPs in RNA metabolism.

3. Materials and Methods

Identifying CSP proteins in Physcomitrella patens

The coding regions of the *PpCSP* genes, as well as the sequences of their extreme 5' and 3' ends, were found by BLAST searching in the PHYSCObase (<http://moss.nibb.ac.jp/> and <http://moss.nibb.ac.jp/cgi-bin/blast-assemble>) and JGI (Joint Genome Institute) (<http://genome.jgi-psf.org/>) databases. Southern blotting was performed to confirm *PpCSP* copy number (data not shown).

Moss culture conditions

The Gransden 2004 variety of *Physcomitrella patens* (Hedw.) Bruch & Schimp subsp. *patens* were

propagated as described by Nishiyama et al. 2000. For observation of gametophores and gametangia, these tissues were grown as previously described by Tanahashi et al. 2005. Protonema of the YFP lines were also treated with various abiotic stresses, such as heat at 30°C, cold at 4 and 0°C, 150 and 300 mM salt, and 250 and 500 mM sorbitol (for osmotic stress) with continuous light and subsequently visualized.

Generating recombinant proteins and nucleic acid binding assay

Primers were designed to amplify the full-length *PpCSP* genes from genomic DNA isolated via the CTAB (Hexadecyltrimethyl-ammonium bromide) method, as previously described (Murray and Thompson, 1980). For in-frame cloning, restriction enzyme sites were incorporated in the primer with a BamHI in the forward primers, and an XhoI in the reverse primers. The primer sequences are as follows: *PpCSP1* (5'-TCTGGATCCATGGAGGAAGTGGGAGCGAGC-3') and (5'-TCTCTCGAGGGAAGCGGCTGCAGGAGTGCAG-3'), *PpCSP2* (5'-TCTGGATCCATGACGGAAGAGGGGCACAGC-3') and (5'-TCTCTCGAGAGCGGCAGCCGCGGGAGCGCAG-3'), *PpCSP3a/b* (5'-TCTGGATCCATGGCGGAAGAGGGAGCGAGC-3') and (5'-TCTCTCGAGGGCAGCGGCTGCAGGAGTGCAG-3'). The *PpCSP* genes were amplified using iProof DNA polymerase (BioRad; Hercules, CA, U.S.A.) and the High-Fidelity buffer according to the manufacturers protocol with the cycling conditions as follows: 98.0°C for 10 sec, 53.0°C for 30 sec, and 72.0°C for 30 sec for a total of 35 cycles. The PCR products were gel purified using the Perfectprep Gel Clean-up Kit (Eppendorf; Westbury, NY, U.S.A.). Purified fragments were then digested overnight with BamHI and XhoI and ligated into the pET-43.1a vector that was previously digested with the same restriction enzymes. The plasmids were sequenced to confirm that the genes were intact. The pET-D43.1a vector incorporates a N-terminal NusA for increased solubility and a C-terminal 6X His-tag for purification. The plasmids were grown in BL21-DE3-pLys-S (Novagen; Madison, WI, U.S.A.) competent cells for recombinant protein expression. The cells were grown at 37°C in LB-Carbenicillin liquid media to an OD₆₀₀ of 0.6. Protein synthesis was then

induced by adding IPTG (isopropyl b-D-1-thiogalactopyranoside) to a final concentration of 1 mM and grown an additional 3 hours. Cells were pelleted by centrifugation to remove the LB-Carbenicillin. The cells were then re-suspended in BugBuster® Protein Extraction Reagent (Novagen; Madison, WI, U.S.A.) supplemented with benzonase nuclease (Novagen; Madison, WI, U.S.A.) according to the manufacturer's instructions. These suspensions were subsequently centrifuged at 14,000x g for 20 minutes at 4°C and soluble proteins were collected in the supernatant. The proteins were then purified by mixing the soluble fraction with Ni-NTA matrix beads and purified following the batch method and eluted according to the manufacturer's protocol (Qiagen; Valencia, CA, U.S.A.). Purified recombinant proteins were quantified using the Bradford assay reagent (BioRad; Hercules, CA, U.S.A.) and an SDS-page gel was run to determine protein quality.

The nucleic acid binding gel shift assay was carried out using phage M13 mp8 ssDNA (USB; Cleveland, OH, U.S.A) as previously described (Karlson et al. 2002). The PpCSP protein concentration ranged from 10 pmol to 100 pmol. Purified recombinant NusA-His protein alone was used at a concentration of 100 pmol to serve as a negative control.

RT-PCR PpCSP cold induction

Moss protonema tissue was grown on BCDAT media overlain with cellophane for 4 days at 25°C under continuous light. These pre-cultures were used for subsequent cold treatments and plates were transferred to 4°C for a time course of 0, 30 min, 1, 2, 6, 12, 24, and 48 hours under continuous light. At each respective time point, the moss was quickly collected from the cellophane with a spatula, blotted on Kim wipes, weighed, and subsequently frozen in liquid nitrogen. RNA was extracted using RNeasy Plant Mini-Kit (Qiagen; Valencia, CA, U.S.A.) according to the manufacturer's protocol. Reverse transcription was carried out using 0.5 µg of total RNA with the QuantiTect Reverse Transcription Kit (Qiagen; Valencia, CA, U.S.A.) at half-scale of the company's protocol, but otherwise according to their instructions. PCR was carried out using Go-Taq® Flexi DNA polymerase and the following gene specific primers: *PpCSP1* (5'-

GCTGTCGGTCTTCGCCTGCG-3') and (5'CCTCGACCTCCTCCATCGCC-3'), *PpCSP2* (5'- -3') and (5'- -3'), *PpCSP3a/b* (5'-CATTCTGCGCGGCTGTCTGA-3') and (5'-TCCCCCTACGCCACCGC-3'), and the *GapC* (glyceraldehyde-3-phosphate dehydrogenase) loading control gene (5'-GAGATAGGAGCATCTGTACCGCTTGTGC-3') and (5'-CATGGTGGGATCGGCTAAGATCAAGGTC-3'). The band intensities were analyzed using Image-J 1.33u software available at <http://rsb.info.nih.gov/ij/index.html>. The intensities of the *PpCSP* bands were divided by the corresponding intensity of the control gene, *GapC*, to correct any differing PCR efficiencies and loading differences.

Cloning of Expression Constructs

The coding regions of the *PpCSP* genes, as well as the sequences of their extreme 5' and 3' regions, were found by BLAST searching in the PHYSCObase (<http://moss.nibb.ac.jp/> and <http://moss.nibb.ac.jp/cgi-bin/blast-assemble>) and JGI (Joint Genome Institute) (<http://genome.jgi-psf.org/>) databases. Targeting constructs were made by C-terminally fusing YFP (yellow fluorescent protein) and GUS (*uidA*) genes to *PpCSP1* and *PpCSP2* just before their respective stop codons. The genomic fragment for the 5' end of the constructs included about 1,000 bp upstream of the stop codon including the *CSP* ORF (open reading frame) and the 5' UTR (un-translated region). The genomic fragment for the 3' end of the constructs included the stop codon and approximately 1,000 bp down-stream from the stop codon for the *PpCSP* genes. The 5' and 3' genomic sequences (including the UTRs) included in these constructs were present to facilitate homologous recombination. The primers used to amplify the 5' region for *PpCSP1* were forward (5'-TCTCTCGAGAGCTGAGCTGGAATCGGG-3') and reverse (5'-AGAATCGATGGAAGCGGCTGCAGGAGTG-3'), and the 3' region were forward (5'-TCTACTAGTTAGGTGCTTCCGAGTAG-3') and reverse (5'-AGAGCGGCCGCAAAAATCATCTACTCG-3'). The primers used to amplify the 5' region for *PpCSP2* were forward (5'-TCTCTCGAGAGACGAAAAGGGGGCAGG-3') and reverse (5'-AGAATCGATAGCGGCAGCCGCGGGAGC-3'), and the 3' region were forward (5'-

TCTACTAGTTAGTTGCTTCACTAGAG-3') and reverse (5'-AGAGCGGCCGCGCAGCAAGTAAGCCACAC-3'). These primers included restriction enzyme sites for cloning purposes. For both genes, the forward primer for the fragment upstream of the stop codon included a XhoI site, and the reverse a Clal site, and the forward primer for the fragment downstream of the stop codon included a SpeI site, and the reverse a NotI site.

Transformation and selection of positive transformant lines

The YFP and GUS targeting plasmids were introduced into *Physcomitrella patens* via the polyethylene glycol-mediated transformation method as previously described (Nishiyama et al. 2000). Using this method, the *PpCSP-YFP* and *-GUS* gene expression are driven by the *PpCSP1* and *PpCSP2* native promoters, respectively. Lines showing resistance to Geneticin (G418) (Invitrogen; Tokyo, Japan), at a concentration of 20 mg/L, through two cycles of selection were screened by PCR. The positive lines confirmed by PCR were further analyzed for single insertion by southern blot analysis.

Southern blot analysis

Genomic DNA (gDNA) of moss protonema and gametophore tissue was isolated using the CTAB (Hexadecyltrimethyl-ammonium bromide) method from the transformant lines, as previously described (Murray and Thompson 1980). Two µg of gDNA from each line was digested with a restriction enzyme and consequently blotted and hybridized as described previously (Shindo et al. 2001). The *PpCSP1* gDNA samples were digested with EcoRI and the *PpCSP2* gDNA samples were digested with EcoT14I. The probes used to identify positive lines were specific for either *PpCSP1* or *PpCSP2*.

Visualizing sub-cellular localization

The tissues were visualized using a Leica DMLB microscope. The bright-field images were taken with Normarski filters. A laser with a wavelength of 514 nm was used to excite the YFP

protein for fluorescence. Confocal images were taken using either an Olympus IX70 or a Leica SP2 system (and the YFP was excited at a wavelength of 514 nm).

4. Results

Database searching and genomic southern blot analysis (data not shown) identified four unique cold shock domain proteins in *Physcomitrella patens*. Each of the four proteins consist of an N-terminal cold shock domain, as is common for cold shock domain proteins, and two C-terminal retroviral-like zinc fingers interspersed in glycine-rich regions (Figure 1). None of the four genes contain introns and their coding regions range from 564 to 597 bp. PpCSP3a and PpCSP3b are identical in their coding regions, coding for 198 amino acid proteins, suggesting a more recent duplication event as compared to the other two PpCSPs. PpCSP1 shows the most similarity in the cold shock domain region to PpCSP3a and 3b, while the PpCSP2 protein is the most different, coding for a 195 amino acid protein.

We performed a ssDNA gel-shift binding assay to determine whether PpCSPs are capable of binding nucleic acids. Full-length PpCSP1, PpCSP2, and PpCSP3a/b were cloned into the pET-43.1a vector and expressed in BL21-DE3-pLysS cells and subsequently purified. PpCSP3a is not distinguishable from PpCSP3b since their coding regions are identical. Binding of M13 mp8 phage ssDNA was detected at the lowest protein concentration of 10 pmol for all three PpCSPs (Figure 2). As the protein concentrations were increased up to 100 pmols, the gel shift became increasingly apparent, while 100 pmol of recombinant NusA protein (negative control) showed no affinity for the ssDNA (Figure 2). Similar to bacterial and higher plant CSPs, these data clearly show that nucleic acid binding activity is conserved amongst moss CSPs.

Higher plant CSP genes, such as WCSP1 and AtCSP2/AtGRP2 (At4g38680), are up-regulated in response to low temperature stress (Karlson et al. 2002; Sasaki et al. 2007). To determine if moss CSP genes share this characteristic response, protonema cells were placed at 4°C for various time points and *PpCSP* transcript levels were determined semi-quantitatively with

RT-PCR analysis (Figure 3a and b). Both *PpCSP1* and *PpCSP2* transcripts show a clear and gradual increase subsequent to low temperature exposure. The expression of *PpCSP1* peaks at 24 hours while *PpCSP2* shows two peaks at 12 and 48 hours (Fig. 3b). Either one or both of the genes *PpCSP3a* and *PpCSP3b* also show a more gradual up-regulation. However, it is not possible to quantify them separately with this method due to their identical gene sequences. Clearly, the moss *PpCSPs* show temperature-induced transcription, a characteristic that is retained in higher plants.

The coding regions on *PpCSP1* and *PpCSP2* were C-terminally fused to both a gene for yellow fluorescent protein (YFP) and the *uidA* gene coding for GUS (β -glucuronidase). Both constructs contain an NPTII cassette conferring G418 (geneticin) resistance just before their stop codon (Figure 4A). Through this design, *PpCSP1*-YFP, *PpCSP1*-GUS, *PpCSP2*-YFP, and *PpCSP2*-GUS expression were all under the control of their respective native promoters. Once positive lines were confirmed by both antibiotic selection and PCR, southern blot analyses were conducted to confirm single-insertion targeted lines (Figure 4B). Southern blot analyses confirmed 5 positive single-insertion lines for both *PpCSP1*-YFP and *PpCSP2*-YFP lines as well as 4 lines for both *PpCSP1*-GUS and *PpCSP2*-GUS (Figure 4B).

The expression lines were used to observe the localization of *PpCSP1* and *PpCSP2* at various stages of *Physcomitrella patens* in development to gain insight into their putative *in vivo* functions. The YFP fusion lines were used primarily for sub-cellular localization studies while the GUS lines substantiated the YFP expression patterns. Upon broadly viewing *PpCSP1*-YFP and *PpCSP2*-YFP protonema tissue under 20x magnification, fluorescence was readily visible in the cytoplasm with the highest signals seen in apical cells and new branch cells (Figure 5A and B). The same expression pattern was confirmed in the *PpCSP1*-GUS and *PpCSP2*-GUS lines (Figure 5A and B). A high signal was also detected in phragmoplast regions. This discrete localization pattern is likely due to the tendency for cytoplasmic proteins to aggregate in the actively dividing region (Hiwatashi unpublished results). *PpCSP1*-YFP and *PpCSP2*-YFP signal was also very intense in the cytoplasm of new branching protonema cells as well as young gametophores (Figure 5B and C).

As the gametophores grow, signals became apparent in the apical regions of gametophores. Gametophores for both lines were shifted to 15°C to induce gametangia and sporophyte production. In the case of the gametangia, fluorescent signals were clearly observed in both the antheridia and the archegonia, as well as the unfertilized egg cells (Figure 5D, E, and F). After fertilization, fluorescence was detected in the cytoplasm of embryos, as well as the developing sporophytes for PpCSP1-YFP (Figure 6). Observation of dividing protonema cells with a confocal microscope revealed the localization of both PpCSP1-YFP and PpCSP2-YFP in the newly formed nucleoli of both the apical and sub-apical cells (Figure 7). YFP is however not detected in the nucleolus between cell division events (Figure 7). Also, upon careful examination, a dot-like fluorescent signal was detected within the cytoplasm for both fusion proteins (Figure 8). No changes in localization patterns could be detected under any of the introduced stresses of heat, cold, salt, and sorbitol (osmotic) (data not shown).

5. Discussion

All four members of the cold shock domain family in *P.patens* are composed of an N-terminal CSD and two retroviral-like zinc fingers interspersed within glycine-rich regions. This motif arrangement is common amongst higher plant CSP proteins (Karlson and Imai 2003). The glycine-rich regions resemble RGG-boxes that are capable of binding RNA and increasing RNA affinity 10-fold (Burd and Dreyfuss 1994). Like the CSPs of higher plants including WCSP1, AtCSP2/AtGRP2, and AtCSP4 (Karlson et al. 2002; Fusaro et al. 2006; Sasaki et al. 2007; Kim et al. 2007), moss PpCSPs are clearly capable of binding nucleic acids. Another common characteristic of the PpCSPs with both bacteria, such as *E. coli* CspA, and higher plant CSPs is that they are cold induced (Jones et al. 1987; Karlson et al. 2002; Karlson and Imai 2003). Secondary structures in mRNA are thermodynamically favored at low temperatures and are thus stabilized under these conditions and in turn hinder translation (Jiang et al. 1997). Bacterial CspA, and AtCSP4 and WCSP1 show RNA chaperone activity reducing secondary structures in mRNA. Thus, it is possible that these proteins may facilitate enhanced translation at low temperatures in *planta* (Nakaminami

et al. 2006; Kim et al. 2007). Considering that PpCSPs share with higher plants the characteristics of nucleic acid binding and cold induction, it is possible that they may also function as RNA chaperones under low temperature stress.

Physcomitrella patens PpCSP1 and PpCSP2 localize in the cytoplasm of protonema, gametophores, gametangia, egg cells, embryos, and in the sporophytes (Figures 5 and 6). The detected YFP signals were consistently higher in younger actively developing tissue. Specifically, expression was found to be higher in apical cells than in sub-apical cells. This pattern was also observed for MA16, a maize protein containing the RNA-recognition motif or RRM (Alba and Pages, 1998), which also contains the RNP-1 and RNP-2 sites. The moss CSPs were not the first found to localize in the cytoplasm; higher plant CSPs including *Arabidopsis* AtCSP2/AtGRP2 also localize in the cytoplasm (Fusaro et al. 2006; Sasaki et al. 2007). In dividing cells, both PpCSP1-YFP and PpCSP2-YFP were detected in the nucleolus of the newly formed nuclei (Figure 7). They were not detected in the nucleolus at the other stages of the cell cycle (Figure 7). However, it is possible that the concentrations were so low that either light or confocal microscopes could not detect the signal. This pattern is interesting considering that new cells are actively involved in translation requiring high quantities of rRNAs. If PpCSPs are involved in ribosomal assembly or are associated with rRNA, it is reasonable to consider that the intensity of YFP signals in the nucleolus would increase at this stage of development. The well-known *Xenopus* CSP, FRGY2, also shows nucleolar localization that requires basic and acidic islands of the tail-domain (Ranjan et al. 1993). This tail-domain is also reported to be important for the association of FRGY2 with mRNPs *in vivo* (Matsumoto et al. 1996). In the case of the RRM MA16, the C-terminal RGG domains necessary for nucleolar targeting (Gendra et al. 2004) and WCSP1 needs at least the first glycine-rich region and zinc-finger to localize to the endoplasmic reticulum (Nakaminami et al. 2006). Recently, Sasaki et al. reported that the last glycine-rich region and zinc finger of AtCSP2/AtGRP2 are also necessary for nucleolar targeting (2007). Based on the roles of other CSP and RRM C-terminal domains in relation to localization, it is possible that the glycine-rich region and zinc fingers of the PpCSPs are involved in sub-cellular localization and molecular interactions. Collectively, our

analysis of PpCSP sub-cellular localization and nucleic acid binding activity suggests an involvement in RNA metabolism and cell development, with potential importance in the cell cycle.

We were interested to explore the dot-like pattern observed in the cytoplasm of some cells for both PpCSP1 and PpCSP2 (Figure 8). Recently, a well-studied *C.elegans* cold shock domain protein (Lin28), containing both a CSD and retroviral zinc-fingers, showed a similar particulate localization (Balzer and Moss, 2007). Further experiments demonstrated that these particles were in fact P-bodies. P-bodies are cytoplasmic processing bodies that are the sites for multiple functions of mRNA regulation, including mRNA de-capping and degradation (Sheth and Parker 2003) and translation and translation suppression (Coller and Parker, 2005). In view of the evidence supporting PpCSP involvement with RNA and RNA metabolism, it is possible that the particles seen in the cytoplasm could in fact be P-bodies. Co-localization experiments with a Dcp (an mRNA de-capping protein known to localize in P-bodies) orthologue in *Physcomitrella patens*, are underway and will confirm if PpCSPs indeed localize to P-bodies. In addition, future functional analyses with knock-out lines for PpCSPs will confirm if plant cold shock proteins play a role in mRNA stability.

Cold shock domain protein function has been heavily studied and well characterized in both bacteria and animal systems. Functional roles of cold shock domain proteins in plants are just beginning to become better understood. In this study, we provided the first functional characterization of cold shock domain proteins from lower plants by using the bryophyte *Physcomitrella patens* model system. PpCSPs are capable of nucleic acid binding and are cold induced; both functions that are conserved in higher plant CSPs (see the previously mentioned citations). We confirmed that PpCSPs are expressed throughout all stages of the moss life cycle. Their prevalence in new developing tissue, nucleoli and putative P-bodies suggest their involvement in RNA metabolism and cell development. It is our hope that future experiments will determine their specific involvement and in turn determine the precise function of cold shock domain proteins in the kingdom *planta*.

Figure 1: The *Physcomitrella patens* PpCSPs as compared to other well-known CSD proteins. CspA is the first identified CSD protein in *E. coli*, Y-box1 is a human CSD protein, and WCSP1 comes from winter wheat. The box marked Var indicates a variable region, the boxes marked B indicate basic regions, and the boxes marked A indicate primarily acidic regions.

a.

b.

Figure 2: PpCSP protein purification and nucleic acid binding gel shift assay. a. purification of PpCSP proteins from the soluble fraction. The amount of each purified protein is 1 microgram. b. PpCSPs are able to bind ssDNA. The numbers above the gels indicate the amount of protein added in pmol. All lanes contain 100 ng of ssDNA. - indicates only ssDNA and NusA+His is the purified NusA protein with the 6x His tag without recombinant PpCSP proteins (empty vector) present at 100 pmol concentration. The empty vector is unable to bind to ssDNA, but the PpCSP proteins begin to bind the ssDNA at a concentration of only 10 pmol of protein.

a.

b.

	0 hr	0.5 hr	1 hr	2 hr	6 hr	12 hr	24 hr	48 hr
<i>PpCSP1</i>	27859	26687	24695	27938	29284	33458	36378	26227
<i>GapC</i>	34001	25409	31719	15415	13166	11440	9597	7569
Relative Expression	0.82	1.05	0.78	1.81	2.22	2.92	3.79	3.47
<i>PpCSP2</i>	27401	27350	32813	35433	35275	42273	34437	36955
<i>GapC</i>	29235	33893	37352	33843	30204	27069	28204	23626
Relative Expression	0.94	0.81	0.88	1.05	1.17	1.56	1.22	1.56
<i>PpCSP3a/b</i>	32334	34301	37501	37344	32869	32719	35506	29644
<i>GapC</i>	32124	31845	35330	33533	30720	18617	18616	9436
Relative Expression	1.01	1.08	1.06	1.11	1.07	1.76	1.91	3.14

Figure 3: Cold induced up-regulation of the *PpCSP* genes. *GapC* is present as the loading control for all of the present time points. a. All of the *PpCSPs* show some degree of up-regulation in response to treatment at 4°C. b. Band intensity analyzed using Image-J 1.33u. This table shows a clearer representation of the up-regulation of each of the genes relative to *GapC*.

a.

b.

Figure 4: a. shows the YFP and GUS expression construct design for both PpCSP1 and PpCSP2. The grey boxes indicate the PpCSP protein. b. Southern blot analysis of the GUS and YFP targeted moss lines. The expected band size for PpCSP1-GUS is about 6.6 kb and the expected size of PpCSP1-YFP is 2 kb. The expected band size for PpCSP2-GUS is 5 kb and the expected size for PpCSP2-YFP is about 3.6 kb. The PpCSP2-GUS line 338 has the expected band size but it may also have a band higher, so it was eliminated from further analysis.

Figure 5: Localization of PpCSP1 and PpCSP2 -YFP and -GUS proteins in protonema, gametophores, and gametangia. A. signal detected in the cytoplasm of protonema cells. B. very strong signal detected in new branch cells. C. localization of PpCSPs 1 and 2 in young gametophores. D. signal detected in antheridia. E. localization of the proteins in the archegonia. The GUS lines showed a different pattern with staining only at the archegonia tips while the YFP localized more strongly in the base. F. highlights the strong signals detected for both proteins in unfertilized eggs. Scale bar = 50 μ m.

Figure 6: Localization in egg cells, embryos and young sporophytes. a. shows the localization of PpCSP1-YFP in an embryo. b. shows that PpCSP1-YFP is still detected in the embryo after two cell divisions have occurred. c. shows the localization of PpCSP1-YFP in a young developing sporophyte and d. shows the continued detection in the sporophyte at a later developmental stage. e. shows the localization of PpCSP2-YFP in an egg cell.

Figure 7: Localization of PpCSP1-YFP and PpCSP2-YFP in the nucleolus of newly divided cells. a. shows a confocal image of a chloronema cell when it is not dividing. b. shows the nucleolar localization of PpCSP1-YFP in a cell finishing cell division. c. shows the nucleolar localization of PpCSP2-YFP just after a cellular division. The white arrows indicate the nucleolus.

PpCSP1

PpCSP2

Figure 8: Dot-like patterns detected in the cytoplasm of protonema for both PpCSP1 and PpCSP2 YFP lines. Scale bar = 50 μm .

SUMMARY AND CONCLUSIONS

The aim of the first part of this study was to investigate the diversity of cold shock domain proteins in plants. First, EST sequences were gathered for as many plant CSPs as possible and from as many species as possible. These protein sequences were then analyzed using multiple publicly available programs to predict potential localization, and potential phosphorylation and sumoylation sites, based on amino acid sequence. Because the ultimate goal is to understand CSP function. The cold-induced expression of *Arabidopsis thaliana* CSPs was analyzed in wild-type and various cold related regulons.

The second part of this study concentrated on the CSPs from one of the most basal plant lineages, the bryophytes. *Physcomitrella patens* CSPs were tested for their ability to bind nucleic acids and to determine if their expression is cold induced. To gain a better understanding of their function, expression constructs were made to determine their localization throughout the life cycle of moss.

The conclusions of this study are:

- ❖ Cold shock domain proteins are highly conserved in plants ranging from haptophytes to angiosperms.
- ❖ CSDs from similar organisms group together in the phylogenetic analyses showing that more closely related organisms have more similar CSDs.
- ❖ Monocot CSDs can be classified into two subgroups whereas dicot CSDs can be classified into three subgroups.
- ❖ It appears that as evolution proceeded the motifs found in CSPs changed and/or became more diverse.
- ❖ CSPs are soluble proteins predicted to localize in the cytoplasm.
- ❖ CSPs are predicted to be phosphorylated and the most likely site for this modification is the serine or threonine located in the (F/Y) R (S/T) L consensus sequence between β -sheets 3 and 4.

- ❖ Many CSPs are predicted to be sumoylated.
- ❖ The expression of *AtCSP3* and *AtCSP3* is affected by ABA.
- ❖ *AtCSP3* function is up-regulated by CBF3.
- ❖ *AtCSP1* and *AtCSP3* may be downstream targets in the ICE1 regulon.
- ❖ *Physcomitrella patens* CSPs are able to bind to nucleic acids.
- ❖ The transcription of *PpCSPs* are all induced by cold treatment.
- ❖ PpCSP1 and PpCSP2 proteins primarily localize in the cytoplasm and can be found at all stages of moss development mainly localizing in young and developing tissue.
- ❖ PpCSP1 and PpCSP2 localize in the nucleolus of newly divided cells.
- ❖ PpCSPs may localize in P-bodies and may have a role in RNA metabolism.

BIBLIOGRAPHY

1. Adachi, J., and Hasegawa, M. (1996) MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. Computer Science Monograph 28: 1-150.
2. Alba, M. and Pages, M. (1998) Plant proteins containing the RNA-recognition motif. Trends in Plant Science 3: 15-21.
3. Artus, N.N., Uemura, M., Steponkus, P.L., Gilmour, S.J., Lin, C.T., and Thomashow, M.F. (1996) Constitutive expression of the cold-regulated *Arabidopsis thaliana* *COR15a* gene affects both chloroplast and protoplast freezing tolerance. Proceedings of the National Academy of Science USA 93: 13404-13409.
4. Bae, W., Phadtare, S., Severinov, K., Inouye, M. (1999) Characterization of *Escherichia coli* *cspE*, whose product negatively regulates transcription of *cspA*, the gene for the major cold shock protein. Molecular Microbiology 31: 1429-1441.
5. Bae, W., Xia, B., Inouye, M., and Severinov, K. (2000) *Escherichia coli* CspA-family RNA chaperones are transcription antiterminators. Proceedings of the National Academy of Science USA 97: 7784-7789.
6. Balzer, E. and Moss, E.G. (2007) Localization of the developmental timing regulator Lin28 to mRNP complexes, P-bodies, and stress granules. RNA Biology 4(1): 16-25.
7. Barrero, J.M., Piqueras, P., Gonzalez-Guzman, M., Serrano, R., Rodriguez, P.L., Ponce, M.R., and Micol, J.L. (2005) A mutational analysis of the *ABA1* gene of *Arabidopsis thaliana* highlights the involvement of ABA in vegetative development. Journal of Experimental Botany 56(418): 2071-2083.
8. Blom, N., Gammeltoft, S., and Brunak, S. (1999) Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. Journal of Molecular Biology 294(5): 1351-1362.
9. Bouvet, P. and Wolffe, A.P. (1994) A role for transcription and FRGY2 in masking maternal mRNA within *Xenopus* oocytes. Cell 77(6): 931-941.
10. Burd, C.G., and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. Science 265: 615-621.
11. Chinnusamy, V., Ohta, M., Kanrar, S., Lee, B., Hong, X., Agarwal, M., and Zhu, J.K. (2003) ICE1: a regulator of cold-induced transcriptome and freezing tolerance in *Arabidopsis*. Genes and Development 17: 1043-1054.
12. Collier, J. and Parker, R. (2007) General translational repression by activators of mRNA decapping. Cell 122(6): 875-886.
13. Dannull, J., Surovoy, A., Jung, G., and Moelling, K. (1994) Specific binding of HIV-1 nucleocapsid protein to PSI RNA in vitro requires N-terminal zinc finger flanking basic amino acid residues. The EMBO Journal 13(7): 1525-1533.
14. Dreyfuss, G., Kim, V.N., and Kataoka, N. (2002) Messenger-RNA-binding proteins and the messages they carry. National Review of Molecular and Cellular Biology 3: 195-205.
15. Duckham, S.C., Linforth, R.S.T., and Taylor, I.B. (1991) Absciscic-acid-deficient mutants at the *aba* gene locus of *Arabidopsis thaliana* are impaired in the epoxidation of zeaxanthin. The Plant Cell and Environment 14: 631-636.
16. Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of Molecular Biology 300: 1005-1016.
17. Fusaro, A.F., Bocca, S.N., Ramos, R.L.B., Barroco, R.M., Magioli, C., Jorge, V.C., Coutinho, T.C., Rangel-Lima, C.M., De Rycke, R., Inze, D., Engler, G., and Sachetto-Martins, G. (2006) AtGRP2, a cold-induced nucleo-cytoplasmic RNA-binding protein, has a role in flower and seed development. Planta 225(6): 1339-1351.
18. Fusaro, A., Mangeon, A., Junqueira, R.M., Rocha, C.A.B., Coutinho, T.C., Margis, R., and Sachetto-Martins, G. (2001) Classification, expression pattern and comparative analysis of

- sugarcane expressed sequence tags (ESTs) encoding glycine-rich proteins (GRPs). *Genetics and Molecular Biology* 24(1-4): 263-273.
19. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. (2005) Protein identification and analysis tools on the ExPASy Server;_(In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press pp. 571-607.
 20. Gendra, E., Moreno, A., Alba, M.M., and Pages, M. (2004) Interaction of the plant glycine-rich RNA-binding protein MA16 with a novel nucleolar DEAD box RNA helicase protein from *Zea mays*. *The Plant Journal* 38: 875-886.
 21. Gilmour, S.J. and Thomashow, M.F. (1991) Cold acclimation and cold-regulated gene expressing in ABA mutants of *Arabidopsis thaliana*. *Plant Molecular Biology* 17: 1233-1240.
 22. Gilmour, S.J., Zarka, D.G., Stockinger, E.J., Salazar, M.P., Houghton, J.M., and Thomashow, M.F. (1998) Low Temperature regulation of the Arabidopsis CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. *The Plant Journal* 16(4): 433-442.
 23. Ginisty, H., Sicard, H., Roger, B., and Bouvet, P. (1999) Structure and functions of nucleolin. *Journal of Cell Science* 112: 761-772.
 24. Graumann, P. and Marahiel, M.A. (1996) A case of convergent evolution of nucleic acid binding modules. *Bioessays* 18(4): 309-315.
 25. Graumann, P. & Marahiel, M.A. (1996) Some like it cold: response of microorganisms to cold shock. *Archives of Microbiology* 166: 293-300.
 26. Graumann, P.L. & Marahiel, M.A. (1998) A superfamily of proteins that contain the cold-shock domain. *Trends Biochemical Science* 23: 286-290.
 27. Hasegawa, M. and Kishino, H. (1994) Accuracies of the simple methods for estimating the bootstrap probability of a maximum-likelihood tree. *Molecular Biology and Evolution*. 11: 142-145.
 28. Hirokawa, T., Boon-Chieng, S., and Mitaku, S. (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* 14: 378-379.
 29. Jaglo-Ottosen, K.R., Gilmour, S.J., Zarka, D.G., Schabenberger, O., and Thomashow, M.F. (1998) *Arabidopsis* CBF1 overexpression induces COR genes and enhances freezing tolerance. *Science* 280: 104-106.
 30. Jiang, W., Hou, Y., Inouye, M. (1997) CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone. *Journal of Biological Chemistry* 272: 196-202.
 31. Jones, D.T., Taylor, R.W., and Thornton, M.J. (1992) The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences*. 8: 275-282.
 32. Jones, P. G., Van Bogelen, R. A. and Neidhardt, F. C. (1987) Induction of proteins in response to low temperature in *Escherichia coli*. *Journal of Bacteriology* 169, 2092-2095.
 33. Karlson, D. and Imai, R. (2003) Conservation of the cold shock domain protein family in plants. *Plant Physiology*: 131: 12-15.
 34. Karlson, D., Nakaminami, K., Toyomasu, T., and Imai, R. (2002) A cold-regulated nucleic acid-binding protein of winter wheat shares a domain with bacterial cold-shock proteins. *Journal of Biological Chemistry* 20: 35248-35256.
 35. Kim, J.S., Park, S.J., Kwak, K.J., Kim, Y.O., Kim, J.Y., Song, J., Jang, B., Jung, C.H., and Kang, H. (2007) Cold shock domain proteins and glycine-rich RNA-binding proteins from *Arabidopsis thaliana* can promote the cold adaptation process in *Escherichia coli*. *Nucleic Acids Research* 35(2): 506-516.
 36. Kishino, H., Miyata, T., Hasegawa, M. (1990) Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution* 30: 151.
 37. Knight, H., Zarka, D.G., Okamoto, H., Thomashow, M.F., and Knight, M.R. (2004) Absciscic acid induces CBF gene transcription and subsequent induction of cold-regulated genes via the CRT promoter element. *Plant Physiology* 135: 1710-1717.
 38. Lang, E.A.S., and Marques, M.V. (2004) Identification and transcriptional control of *Caulobacter crescentus* genes encoding proteins containing a Cold Shock Domain.

- Journal of Bacteriology 186(17): 5603-5613.
39. Lee, B., Henderson, D.A., and Zhu, J.K. (2005) The *Arabidopsis* cold-responsive transcriptome and its regulation by ICE1. *The Plant Cell* 17: 3155-3175.
 40. Lee, S.J., Xie, A., Jiang, W., Etchegaray, J.P., Jones, P.G., Inouye, M. (1994) Family of the major cold-shock protein, CspA (CS7.4), of *Escherichia coli*, whose members show a high sequence similarity with the eukaryotic Y-box binding proteins. *Molecular Microbiology* 11: 833-839.
 41. Lemieux, C., Otis, C., and Turmel, M. (2007) A clade uniting the green algae *Mesostigma viride* and *Chlorokybus atmophyticus* represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. *BMC Biology* 5(2).
 42. Leung, J. and Giraudat, J. (1998) Absciscic acid signal transduction. *Annual Review of Plant Physiology and Plant Molecular Biology* 49: 199-222.
 43. Leung, J., Merlot, S. and Giraudat, J. (1997) The *Arabidopsis* Absciscic acid-insensitive2 (ABI2) and ABI1 genes encode homologous protein phosphatases 2C involved in Absciscic acid signal transduction. *The Plant Cell* 9: 759-771.
 44. Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998) Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. *The Plant Cell* 10: 1391-1406.
 45. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(- $\Delta\Delta C(T)$) Method. *Methods*: 25(4): 402-408.
 46. Maddison, D.R. and Maddison, W.P. (2000) MacClade 4. Analysis of phylogeny and character evolution. Sunderland, M.A., Sinaer Associates.
 47. Manival, X., Ghisolfi-Nieto, L., Joseph, G., Bouvet, P., and Erard, M. (2001) RNA-binding strategies common to cold-shock domain- and RNA recognition motif-containing proteins. *Nucleic Acids Research* 29(11): 2223-2233.
 48. Mantyla, E., Lang, V., and Palva, T. (1995) Role of absciscic acid in drought-induced freezing tolerance, cold acclimation, and accumulation of LTI78 and RAB18 proteins in *Arabidopsis thaliana*. *Plant Physiology* 107: 141-148.
 49. Maruyama, K., Sakuma, Y., Kasuga, M., Ito, Y., Seki, M., Goda, H., Shimada, Y., Yoshida, S., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2004) Identification of cold-inducible downstream genes of the *Arabidopsis* DREB1A/CBF3 transcriptional factor using two microarray systems. *The Plant Journal* 38: 982-993.
 50. Matsumoto, K., Meric, F., and Wolffe, A.P. (1996) Translational repression dependent on the interaction of the *Xenopus* Y-box protein FRGY2 with mRNA. *Journal of Biological Chemistry* 271(37): 22706-22712.
 51. Matsumoto, K., Tanaka, K.J., Aoki, K., Sameshima, M., and Tsujimoto, M. (2003) Visualization of the reconstructed FRGY2-mRNA complexes by electron microscopy. *Biochemical and Biophysical Research Communications* 306: 53-58.
 52. Matsumoto, K., Tanaka, K.J., and Tsujimoto, M. (2005) An acidic protein, YBAP1, mediates the release of YB-1 from mRNA and relieves the translational repression activity of YB-1. *Molecular and Cellular Biology* 25(5): 1779-1792.
 53. Matsumoto, K. and Wolffe, A.P. (1998) Gene regulation by Y-box proteins: coupling control of transcription and translation. *Cell Biology* 8: 318-323.
 54. Melchior, F. (2000) SUMO-nonclassical ubiquitin. *Annual Review of Cell Development of Cell and Developmental Biology* 16: 591-626.
 55. Miura, K., Jin, J.B., and Hasegawa, P.M. (2007) Sumoylation, a post-translational regulatory process in plants. *Current Opinion in Plant Biology* 10: 495-502.
 56. Miura, K., Jin, J.B., Lee, J., Yoo, C.Y., Stirn, V., Miura, T., Ashworth, E.N., Bressan, R.A., Yun, D.J., and Hasegawa, P.M. (2007) SIZ1-mediated sumoylation of ICE1 controls *CBF3/DREB1A* expression and freezing tolerance in *Arabidopsis*. *The Plant Cell* 19(4): 1403-1414.

57. Moss, E.G., Lee, R.C., and Ambros, V. (1997) The cold shock domain protein LIN-28 controls developmental timing in *C. elegans* and is regulated by the lin-4 RNA. *Cell* 88: 637-646.
58. Moss, E.G. and Tang, L. (2003) Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. *Cell* 88: 637-646.
59. Murashige, T. and Skoog, F. (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiologia Plantarum* 15: 473-497.
60. Murray, M.G. and Thompson, W.F. (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* 8(19): 4321-4325.
61. Mussnug, J.H., Wobbe, L., Elles, I., Claus, C., Hamilton, M., Fink, A., Kahmann, U., Kapazoglou, A., Mullineaux, C.W., Hippler, M., Nickelsen, J., Nixon, P.J., and Kruse, O. (2005) NAB1 is an RNA binding protein involved in the light-regulated differential expression of the light-harvesting antenna of *Chlamydomonas reinhardtii*. *The Plant Cell* 17: 3409-3421.
62. Nakashima, K., Kanamaru, K., Mizuno, T., Horikoshi, K. (1996) A novel member of the cspA family of genes that is induced by cold shock in *Escherichia coli*. *Journal of Bacteriology* 178: 2994-2997.
63. Nakaminami, K., Karlson, D.T., and Imai R. (2006) Functional conservation of cold shock domains in bacteria and higher plants. *Proceedings of the National Academy of Science USA* 103: 10122-10127.
64. Nakaminami, K., Sasaki, K., Kajita, S., Takeda, H., Karlson, D., Ohgi, K., and Imai, R. (2005) Heat stable ssDNA/RNA-binding activity of a wheat cold shock domain protein. *FEBS Letters* 579: 4887-4891.
65. Narusaka, Y., Nakashima, K., Shinwari, Z.K., Sakuma, Y., Furihata, T., Abe, H., Narusaka, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003) Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of *Arabidopsis* rd29A gene in response to dehydration and high-salinity stresses. *The Plant Journal* 34: 137-148.
66. Newkirk, K., Feng, W., Jiang, W., Tejero, R., Emerson, S.D., Inouye, M., et al (1994) Solution NMR structure of the major cold-shock protein (CspA) from *Escherichia coli*: identification of a binding epitope for DNA. *Proceedings of the National Academy of Science USA* 91: 5114-5118.
67. Nishiyama, T., Hiwatashi, Y., Sakakibara, K., Kato, M., and Hasebe, M. (2000) Tagged mutagenesis and gene-trap in the moss, *Physcomitrella patens* by shuttle mutagenesis. *DNA Research* 7: 9-17.
68. Phadtare, S., and Inouye, M. (1999) Sequence-selective interactions with RNA by CspB, CspC and CspE, members of the CspA family of *Escherichia coli*. *Molecular Microbiology* 33(5): 1004-1014.
69. Ranjan, M., Tafuri, S.R., and Wolffe, A.P. (1993) Masking mRNA from translation in somatic cells. *Genes and Development* 7: 1725-1736.
70. Richter, J.D., and Smith, L.D. (1984) Reversible inhibition of translation by *Xenopus* oocyte-specific proteins. *Nature* 309(5966?): 378-380.
71. Rock, C.D. and Zeevaart, J.A.D. (1991) The aba mutant of *Arabidopsis thaliana* is impaired in epoxy-carotenoid biosynthesis. *Proceedings of the National Academy of Science USA* 88: 7496-7499.
72. Sasaki, K., Kim, M.H., and Imai, R. (2007) *Arabidopsis* COLD SHOCK DOMAIN PROTEIN2 is a RNA chaperone that is regulated by cold and developmental signals. *Biochemical and Biophysical Research Communications* 364: 633-638.
73. Sato, N. (1995) A family of cold-regulated RNA-binding protein genes in the cyanobacterium *Anabaena variabilis* M3. *Nucleic Acids Research* 23(12): 2161-2167.
74. Saunders, N.F.W., Thomas, T., Curmi, P.M.G., Mattick, J.S., Kuczek, E., Slade, R., Davis, J., Franzmann, P.D., Boone, D., Rusterholtz, K., Feldman, R., Gates, C., Bench, S., Sowers, K., Kadner, K., Aerts, A., Dehal, P., Detter, C., Glavina, T., Lucas, S., Richardson, P., Larimer, F., Hauser, L., Land, M., and Cavicchioli, R. (2003) Mechanisms of thermal adaptation revealed from the genomes of the Antarctic Archaea *Methanogenium frigidum*

- and *Methanococcoides burtonii*. *Genome Research* 13: 1580-1588.
75. Schaefer, D.G., and Zryd, J.P. (1997) Efficient gene targeting in the moss *Physcomitrella patens*. *The Plant Journal* 11(6): 1195-1206.
 76. Schindelin, H., Marahiel, M.A., and Heinemann, U. (1993) Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein. *Nature* 364: 164-168.
 77. Schroder, K., Graumann, P., Schnuchel, A., Holak, T.A., Marahiel, M.A. (1995) Mutational analysis of the putative nucleic acid-binding surface of the cold-shock domain, CspB, revealed an essential role of aromatic and basic residues in binding of single-stranded DNA containing the Y-box motif. *Molecular Microbiology* 16: 699-708.
 78. Sheth, U. and Parker, R. (2003) Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* 300(5620): 805-808.
 79. Shindo, S., Sakakikara, K., Sano, R., Ueda, K., Hasebe, M. (2001) Characterization of a FLORICAULA/LEAFY homologue of *Gnetum parvifolium* and its implications for the evolution of reproductive organs in seed plants. *International Journal of plant sciences* 162(6): 1199-1209.
 80. Shinozaki, K. and Yamaguchi-Shinozaki, K. (1996) Molecular responses to drought and cold stress. *Current Opinion in Biotechnology* 7: 161-167.
 81. Sommerville, J. (1999) Activities of cold-shock domain proteins in translation control. *Bioessays* 21: 319-325.
 82. Steinert, P.M., Mack, J.W., Korge, B.P., Gan, S.Q., Haynes, S.R., and Steven, A.C. (1991) Glycine loops in proteins: their occurrence in certain intermediate occurrence in certain intermediate filament chains, loricrins and single-stranded RNA binding proteins. *International Journal of Biological Macromolecules* 13(3): 130-139.
 83. Stickeler, E., Fraser, S.D., Honig, A., Chen, A.L., Berget, S.M., and Cooper, T.A. (2001) The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4. *EMBO Journal* 20(14): 3821-3830.
 84. Stockinger, E.J., Gilmour, S.J., and Thomashow, M.F. (1997) *Arabidopsis thaliana* CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. *Proceedings of the National Academy of Science USA* 94: 1035-1040.
 85. Sutherland, B.W., Kucab, J., Wu, J., Lee, C., Cheang, M.C.U., Yorida, E., Turbin, D., Dedhar, S., Nelson, C., Pollak, M., Grimes, H.L., Miller, K., Badve, S., Huntsman, D., Blake-Gilks, C., Chen, M., Pallen, C.J., and Dunn, S. E. (2005) Akt phosphorylates the Y-box binding protein 1 at Ser102 located in the cold shock domain and affects the anchorage-independent growth of breast cancer cells. *Oncogene* 24(26): 4281-4292.
 86. Thomashow, M.F. (1999) Plant Cold Acclimation: freezing tolerance genes and regulatory mechanisms. *Annual Review of Plant Physiology and Plant Molecular Biology* 50: 571-599.
 87. Triqueneaux, G., Velten, M., Franzon, P., Dautry, F., and Jacquemin-Sablon, H. (1999) RNA binding specificity of Unr, a protein with five cold shock domains. *Nucleic Acids Research* 27(8): 1926-1934.
 88. Wang, N., Yamanaka, K., Inouye, M. (1999) CspI, the ninth member of the CspA family of *Escherichia coli*, is induced upon cold shock. *Journal of Bacteriology* 181: 1603-1609.
 89. Xia, B., Ke, H., and Inouye, M. (2001) Acquirement of cold sensitivity by quadruple deletion of the cspA family and its suppression by PNPase S1 domain in *Escherichia coli*. *Molecular Microbiology* 40(1): 179-188.
 90. Yamaguchi-Shinozaki, K. and Shinozaki, K. (1994) A novel cis-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress. *The Plant Cell* 6: 251-264.
 91. Yamanaka, K., Fang, L., Inouye, M. (1998) The CspA family in *Escherichia coli*: multiple gene duplication for stress adaptation. *Molecular Microbiology* 27: 247-255.

CURRICULUM VITAE

Kari Beth Thompson

West Virginia University, Division of Plant and Soil Sciences

University Address:

1090 Agriculture Science Building
P.O. Box 6108
Morgantown, WV 26506
Telephone (304) 293-6023 ext. 4337 (Laboratory)
E-mail: kthomp11@mix.wvu.edu

Stable Address:

26 Aylesworth Way
Grove City, PA 16127
Telephone (724) 458-8685 (Home)
E-mail: this.is.kari@gmail.com

PERSONAL

Age: 25
Birthdate: August 18th 1982

EDUCATION

West Virginia University, Morgantown, West Virginia
Attended from August 2005 to May 2008. M.S. Genetics and Developmental Biology candidate; Division of Plant and Soil Sciences. QPA: 3.82/4.00.

Slippery Rock University, Slippery Rock, Pennsylvania
Attended from August 2001 until May 2005. Graduated Magna Cum Laude with a B.S. Biology, and a minor in Chemistry.
QPA: 3.759/4.00.

Grove City Senior High School, Grove City, Pennsylvania
Attended from September 1997 to June 2001. Graduated as a Distinguished Scholar in the top 10% of the senior class, my rank was 7 out of 168 students.
GPA: 95.98/100.

FELLOWSHIPS

National Science Foundation 2006 East Asia and Summer Pacific Institute in collaboration with the Japan Society for the Promotion of Science Summer Program.
Completed ten weeks of intensive training and research at the National Institute for Basic Biology under the guidance of Dr. Mitsuyasu Hasebe. June 21st to August 21st, 2006.

APPLICABLE EXPERIENCE

Japan Science and Technology Agency (JST), Okazaki, Aichi, Japan
Working as a technician in the Hasebe Reprogramming Evolution Project of the Exporatory Research for Advanced Technology (ERATO). Beginning December 2007 to the present.

Plant Biology & Botany 2007 Joint Congress in Chicago

My poster entitled: Composition and Evolution of the Cold Shock Domain Proteins in Plants, was presented by Dr. Dale Karlson on July 8, 2007. (I was researching in Japan at the National Institute for Basic Biology at the time.)

West Virginia University's Davis College of Agriculture, Forestry & Consumer Sciences 11th Annual Student Research Conference & Creative Design Showcase.

Presented a poster on my research entitled: Composition and Evolution of the Cold Shock Domain Proteins in Plants, April 19, 2007.

West Virginia University's Davis College of Agriculture, Forestry & Consumer Sciences 10th Annual Student Research Conference & Creative Design Showcase.

Presented a poster on my research entitled: Evolution and Function of the Cold Shock Domain Proteins in Plants, April 20, 2006.

Pennsylvania State University: Huck Institute of the Life Sciences Techniques in Molecular Biology Workshop.

Completed an intensive two-week molecular biology workshop, May 16-27, 2005.

National Conferences on Undergraduate Research (NCUR).

Gave oral Presentation entitled: DNA Isolation and PCR for Use in Predicting Ectomycorrhizae Formation, April 22, 2005.

Pennsylvania Junior Academy of Science (PJAS).

Acted as a judge on February 26, 2004 and February 25, 2005.

Commonwealth of Pennsylvania University Biologists (CPUB); 36th Annual Meeting

Presented a poster on my research entitled: Investigation of the Use of Molecular Markers In the Plant and Fungal Partner to Predict Ectomycorrhizal Establishment, April 3, 2005.

13th Annual Sigma Xi Undergraduate Student Research and Creative Accomplishment Conference at Penn State Behrend.

Gave an oral presentation entitled: DNA Isolation for Use in Predicting Ectomycorrhizae Formation April 3, 2004.

Slippery Rock University Symposium for Research and Scholarship;

Presented a Poster on my research entitled Investigation of the Use of Molecular Markers in the Plant and Fungal Partner to Predict Ectomycorrhizal Establishment, March 26, 2004.

AWARDS

Received the Outstanding Student Award for Slippery Rock University at the *Commonwealth of Pennsylvania University Biologists (CPUB); 36th Annual Meeting*, 2005.

ACTIVITIES

At Slippery Rock University I joined organizations including the Lambda Lambda chapter of the Beta Beta Beta Biology Honor Society, the Phi Kappa Phi Honor Society, and the Golden Key International Honor Society.

**PERSONAL
INTERESTS**

Throwing pottery, reading, badminton, cross-stitch, canoeing, hiking, camping, and baking.

REFERENCES

Available upon request.