Study of C-repeat Binding Factors (CBF) structural elements conferring stability under adverse conditions

A Thesis Submitted to the

College of Graduate and Postdoctoral Studies

in Partial Fulfilment of the Requirements for

the Degree of Masters of Science in the

Department of Plant Sciences

University of Saskatchewan,

Saskatoon, Saskatchewan, Canada

By

Brent Puchalski

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this university may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in thesis in whole or part should be addressed to:

Head of the Department of Plant Sciences

51 Campus Drive

University of Saskatchewan

Saskatoon, Saskatchewan, Canada

S7N 5A8

ABSTRACT

Winter wheat is a crop with a lot of potential for the Canadian Prairies. It has many agronomic, environmental and economic advantages over spring wheat, such as a 20-30 % higher yield potential and a lower rate of *Fusarium* infections. However, wheat production on the Prairies is dominated by spring varieties because the current winter wheat cultivars do not accumulate sufficient frost resistance during cold acclimation in the autumn to survive the winters on a consistent basis. Efforts to improve frost tolerance genes in winter wheat via conventional breeding practices has been attempted but proven to be difficult as the cold tolerance in the available varieties cannot be exceeded.

A possible approach to increase frost hardiness in plants is through the optimization of CRT (C-repeat)-Binding Factor (CBF) expression, activities and stability under cold conditions. These master transcription factors upregulate a wide range of cold stress responses and represent about 40% of the difference in freezing tolerance observed between Canadian winter wheat cultivars with low and very high winter-hardiness. Cold-hardy winter wheat produces many CBF variants, but little information is available about the role of the individual proteins. From initial studies of the CBF family encoded by cold-hardy cultivar Norstar, several members display high resilience to extreme temperatures and denaturing conditions, which may be important for winter survival. The high CBF stability is also associated with anomalous slow migration on SDS-PAGE gels.

To further investigate the abnormal gel behavior of CBF12, various modifications to the protein were made and analyzed by SDS-PAGE. It was concluded that the conserved CMIII3, CMIII-1 motifs flanking the DNA-binding AP2 domain or the AP2 domain itself do not slow down CBF12 migration during SDS-PAGE. Rather, removal of the secondary structures within the AP2 region increased the degree of anomalous migration. Instead, it was found that changes in the ratio and distribution of charged residues and the degree of disorder predicted for CBF12 correlated with increased anomalous migration during SDS-PAGE. Similar features are also present for other stress response proteins and transcription factors. Whether the specific amino acid residues contributing to gel shifting play a role in enhancing cold tolerance *in planta* remains to be verified.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

CHAPTER 2: LITERATURE REVIEW

CHAPTER 3: MATERIALS AND METHODS

CHAPTER 4: RESULTS

LIST OF FIGURES

LIST OF TABLES

APPENDIX

LIST OF ABBREVIATIONS

ABA	:	Abscisic acid
AP2	:	Apetala2
ABF/AREB	:	Abscisic acid responsive element binding
BAC	:	Bacterial artificial chromosome
bHLH	:	Basic helix-loop-helix
CBF	:	C-repeat binding factor
CES	:	Combinational enhancer solution
COR	:	Cold regulated
CRT	:	C-repeat
DRE	:	Dehydration responsive element
DTT	:	Dithiothreitol
E. coli	:	*Escherichia coli*
EDTA	:	Ethylene diamine tetra acetic acid
EGTA	:	Ethylene glycol tetra acetic acid
EK/LIC	:	Enterokinase/ligation independent cloning
ERF	:	Ethylene-responsive element
ExPASy	:	Expert Protein Analysis System
Fr	:	Frost resistance
GRAVY	:	Grand average of hydrophobicity
His-tag	:	Poly-histidine tag
HOS1	:	High expression of osmotically responsive protein 1
ICE	:	Inducer of CBF expression
IPTG	:	Isopropyl β-D-1-thiogalactopyranoside
KDa	:	Kilo Dalton
LAC	:	Lactose
LB	:	Luria-Bertani broth
LT	:	Low temperature
LT_{50}	:	Median lethal temperature
PAGE	:	Polyacrylamide gel electrophoresis

PCR	:	Polymerase chain reaction
pI	:	pH(I); isoelectric point
PIF7	:	Phytochrome Interacting Factor 7
PrDOS	:	Protein Dis Order prediction System
QTL	:	Quantitative trait locus
RING	:	Really Interesting New Gene
S-Tag	:	Oligopeptide sequence of pancreatic RNaseA
SB	:	Super broth
SDS	:	Sodium dodecyl sulfate
SIB	:	Swiss Institute of Bioinformatics
SUMO	:	Small Ubiquitin-related modifier
TOC1	:	Timing of CAB expression1
VRN	:	Vernalisation

CHAPTER 1: INTRODUCTION

1.1 Wheat and the Canadian Prairies

Most of the calories in the human diet comes from cereal grains, which encompass a wide range of species adapted to specific climates around the globe. However, only three species, maize, rice and wheat, dominate the cultivated cereals. Maize accounts for one-third (33%) while rice and wheat contribute 28% and 26%, respectively, to the global cereal production estimated to 2010 million metric tons in 2015-16 (IGC 2017). Maize and rice are mostly cultivated in the warmer sub-tropical regions, whereas wheat is adapted to a wider range of climates ranging from the tropics to the edge of the Arctic circle (Worland and Snape, 2001).

Bread wheat (*Triticum aestivum* L.), which has evolved as an allo-hexaploid (AABBDD; 2n=6x=42) species can be distinguished into three different growth habits: spring, intermediate (facultative) and winter type. This differentiation has been a major factor in the wide adaptation of wheat cultivars to diverse climatic conditions. Spring wheat cultivars grown in Canada are seeded in the spring and harvested by the end of the summer. In contrast, winter wheat cultivars are seeded in the autumn and over-winter as seedlings during the cold months. Heading occurs in the late spring and the crop is harvested in the late summer ahead of the spring wheat. In general, winter wheat yields are about 20-30% higher than spring wheat, and thus are more profitable for grain producers.

There are several environmental benefits with winter wheat production as the young plants provide ground cover during the late fall, winter and early spring, which helps to reduce soil erosion. In the spring, winter wheat makes effective use of soil moisture from melting snow, which allows a rapid regrowth and out-competition of weeds. An early flowering time is also beneficial for winter wheat as it increases the chances of escaping Fusarium head blight infections occurring later in the summer. However, the ability to withstand severe low temperature (LT) stress during the harsh prairie winters is a major factor limiting winter wheat production in northern temperate climates, such as the Canadian prairies. The success of breeding wheat for higher winter hardiness has culminated, as no improvement for frost resistance has been obtained in wheat since the release of the Norstar cultivar in 1977.

1.2 Low temperature tolerance

LT tolerance is a major determinant for winter survival in overwintering plants (Fowler et al., 1999). The ability to withstand cold or freezing temperatures is acquired through a cold acclimation process in the fall, during which biochemical, physiological and structural changes result in enhanced protection from low temperature (Thomashow, 1999). In winter cereals, cold acclimation accompanies vernalization, which when saturated corresponds to maximum LT tolerance. The length of the delay to floral transition also influences the LT-tolerance level as an up-regulation of LT-tolerance genes for a longer period result in increased tolerance to LT-stress (Fowler et al., 1996). Short day conditions, which deter flowering during winter months, also increase LT-tolerance (Mahfoozi et al., 2000; Limin and Fowler, 2006). Upon vernalization saturation, some cold-regulated genes are gradually down-regulated resulting in LT-tolerance decline as the winter progresses (Fowler et al., 1996).

1.2.1 Role of vernalization and CBF for cold tolerance

Genetic mapping has shown that several loci on the wheat genome influence LT tolerance and winter survival (Sutka, 1994). However, the most significant quantitative trait loci (QTL) are located on group five chromosomes, where the recessive *Vrn-1* alleles (*vrn-A1*, *vrn-B1*, *vrn-D1*) control the winter growth habit, and thus are major QTL for LT tolerance. Additional frost resistance (*Fr*) loci are also found on group 5 chromosomes. *Fr-B1* and *Fr-D1* loci are located on chromosome arms 5BL (Toth et al 2003) and 5DL (Snape et al 2001), respectively. On chromosome 5A, *Fr-A1* is located close to *vrn-A1* (Galiba et al 1995) and it is possible that *Fr-A1* and *vrnA1* correspond to the same *Fr* locus (Francia et al 2004; Dhillon et al 2010). A second *Fr* locus on chromosome 5A, denoted *Fr-A2* in wheat, is located about 30 cM upstream of *Fr-A1*/*vrnA1* region and constitutes a major QTL for LT tolerance in diploid and hexaploid wheat (Vagufalvi et al 2003; Båga et al 2007), barley (Francia et al 2004) and meadow fescue (Ergon et al 2006).

The orthologous *Fr-A2* loci in cereals carry a cluster of C-repeat (CRT) binding factors (CBFs) genes (Båga et al 2007; Miller et al 2006; Skinner et al 2005), which are transcription factors first described as important regulators of LT tolerance in *Arabidopsis* (Thomashow, 1999). CBFs are characterized by an AP2 DNA binding domain flanked by two different conserved signature motifs (Jaglo et al 2001). A nuclear localization signal is generally found upstream of the AP2 domain

and the C-terminal region may carry transcriptional activation domain(s). In summary, (i) the *FrA2* locus on 5A explains 40% of the difference in cold tolerance between near isogenic lines with low and high frost tolerance; and (ii) CBF as transcription factors play a major role in gene expression during cold acclimation and LT tolerance (Båga et al., 2007). These observations make CBF a key target to enhance cold tolerance in cereals.

1.2.2 Challenges in studying CBF

The wheat *Fr-A2* has more than twenty CBF genes that can be divided into four phylogenetic groups, each containing two or more subgroups (Badawi et al 2008). A large number of these CBF genes are present at homeologous *Fr-A2* in barley and rye (Campoli et al 2009; Choi et al 2002; Schlegel 2013). In cold-sensitive cereals like maize and rice, which lack winter habit, the number of CBF genes are low (Tondelli et al. 2011). Studies of CBF genes from cold-hardy cereals are mainly focused on transcriptional profiling during cold acclimation (Gulick et al., 2005; Laudencia-Chingcuanco et al 2011) and very few reports are available on CBF diversity in terms of function, stability, abundance or modification (Novillo et al. 2007; Zhao et al. 2016). To completely elucidate the role of CBF in low temperature tolerance it is imperative to study the wheat CBF functional properties such as binding to the promoter region of the genes coding the target cold regulated proteins (COR). Based on previous DNA binding studies, most CBF recognize the CRT/DRE (CCGAC) motif present in promoters of many COR genes (Thomashow 1999). However, a slightly different motif (GTCGAC) is recognized by barley CBF2 produced under non-stress conditions (Xue 2003). Some CBF interact with the CRT/DRE motif in a temperature dependent manner. For example, the barley HvCBF1 and HvCBF3 bind the CRT/DRE element at both 20°C and 2°C, whereas members of CBF subgroup four only bind their target at low temperatures (Skinner et al 2005). Such differential binding has led to the speculation that CBF functional differences may underlie much of the variation in LT tolerance seen in plants (Xue 2003; Skinner et al 2005).

CBFs are present in minute quantities in plant tissues and must therefore be obtained in a recombinant form for functional studies. For this purpose, several of the winter wheat cv. Norstar CBF genes were introduced into protein expression vectors for production of various TrxHisStagged CBF in *E. coli* (Jain 2013; Gangopadhyay, 2015). However, purification of over-produced recombinant CBFs from *E. coli* lysates is challenging as the proteins have a high

tendency to aggregate, leaving only a small amount in soluble form. Due to CBF accumulation in protein bodies, the yield of soluble CBF from a bacterial lysate is generally low. Nevertheless, analysis of a few successfully purified TrxHisS-CBF show the protein can maintain functionality upon exposure to either extreme temperatures (boiling or -80°C) or 6 M urea, treatments that would cause inactivation of most proteins (Jain, 2013). The functional stability of the TrxHisS-CBFs also appears to correlate with the degree delayed protein migration on SDS-PAGE, an odd behavior displayed by most wheat CBFs (Jain, 2013). The reason behind anomalous gel migration observed for certain proteins is not understood. A range of factors are reported to contribute to the phenomenon of slow gel migration, such as (i) presence of a higher ratio of acidic residues, (ii) dense hydrophobic groups, (iii) post-translational modifications, (iv) surface distribution of charged residues, (v) elongated structures and (vi) intrinsically disordered regions (Iwashita et al., 2015; Lee et al., 2014; Rath et al., 2013, 2009; Takano et al., 1988). Proteins with these features may also be resilient to denaturation by temperature and be reactive under adverse conditions (De Santi et al., 2016; Gatti-Lafranconi et al., 2010; Jiang et al., 2016; Wang et al., 2016) and are typically present within extremophiles (De Santi et al., 2016; Gatti-Lafranconi et al., 2010; Jiang et al., 2016; Wang et al., 2016).

In an effort to understand the slow gel migration observed for the wheat CBFs, various causes were tested in a previous study (Jain 2013). Analysis of CBF12 threated with Schiff's base and calf intestinal phosphatase ruled out glycosylation or phosphorylation as causes for the slow gel migration. Incubating CBF12 with metal-chelating agents such as EDTA and imidazole also had no effect on alleviating the slow gel migration. The possibility that denaturation of analyzed TrxHisS-CBF12 samples was incomplete prior to SDS-PAGE was also tested by treating the protein with different concentrations of SDS, urea and heating the protein to 98°C for different time periods (Gangopadhyay, 2015; Jain, 2013). None of these methods or a combination of them reduced the slow migration pattern of CBF12. CBF12 variants with different truncations of the Cterminal region were also analyzed to possibly identify protein sequences contributing to the slow gel migration. Some effect on protein migration during SDS-PAGE gel was demonstrated for the C-terminal region of CBF12, but the effect was small (Jain, 2013). Thus, the N-terminal region containing the AP2 DNA binding domain flanked by two unique motifs, the CMIII-3 and CMIII-1, was thought to have a major role in providing protein stability and causing the slow gel migration of CBF12 during SDS-PAGE analysis (Jain 2013).

1.3 Hypothesis

The structure of the resilient AP2 DNA binding domain of CBF12 causes aberrant migration during SDS-PAGE analyses.

1.4 Objectives

To genetically modify different segments of the CBF12 coding region and study the effect on gel mobility for encoded CBF12 variants by SDS-PAGE.

CHAPTER 2: LITERATURE REVIEW

2.1 Origin of wheat

Wheat is an important crop for food and feed production and contributes more than 20% of the calories present in human diets (FAO Statistical Yearbook 2012). While maize and rice have higher global production volumes than wheat, the wheat share has lately increased due to the versatility and diversity of its products (Shewry, 2009).

Wheat (*Triticum aestivum* L.) first emerged during the early Eocene epoch, then evolved and disseminated throughout the world during the Eocene-Oligocene boundary (Fawcett et al., 2009). This period was noted for being climatically diverse, with the early Eocene being warmer than the succeeding period (Sandve and Fjellheim, 2010). Cultivation of wheat was initially centered to the Fertile Crescent in the Middle East, which consists of the modern boundaries of Iraq, Iran, Turkey and Syria (Harlan and Zohary, 1966; Heun et al., 1997). Similar to other domesticated cereals, the loss of shattering was important for wheat to develop as an agricultural crop (Dubcovsky and Dvorak, 2007; Shewry, 2009). This trait was acquired by a mutation in the *brittle rachis* (*Br*) locus and vital for early farmers to harvest their crops with minimal yield loss and effort (Nalam et al., 2006). In addition, free threshing heads were also an important trait for wheat domestication (Shewry, 2009). Further traits, such as shorter stems, erect growth, non-dormant seeds, high yield, large seed size, uniform germination and disease resistance, were subsequent traits for wheat improvement (Feldman, 2001, Preece et al., 2016). In the early agriculture, einkorn wheat (*Triticum monococum*) and emmer wheat (*Triticum turgidum* ssp. *dicoccoides*) were the preferred species (Dubcovsky and Dvorak, 2007; Heun et al., 1997). Emmer wheat (AABB genomes) was derived from an earlier hybridization of *Triticum uratu* Them. Ex Gandil and *Aegilops speltiodes* (Tausch) Gren (Dubcovsky and Dvorak, 2007; Golovnina et al., 2007) (Fig. 2.1). Subsequently, emmer wheat hybridized with the DD genome donor, *Aegilops tauschii,* to create the modern day hexaploid (AABBDD) bread wheat (Dubcovsky and Dvorak, 2007; Talbert et al., 1998). The transition from einkorn and emmer wheat to hexaploid wheat species was aided by the polyploid species being more adaptable and more tolerant of a wider range of environments compared to the diploid species (Fawcett et al., 2009). Thus, bread wheat cultivation spread to areas from the subtropics to the temperate climate zones (Chopra and Paroda, 1986). Nowadays, wheat has the

widest range of production areas and is the most adaptable among the cereal crops (Gill et al., 2004; Graham and Patterson, 1982).

While wheat can survive in widely diverse climates, being economically productive is challenging within its northern ranges. On the Canadian Prairies, the growing season is only 150-160 days and winter wheat can occasionally experience 100% winterkill (http://www.agr.gc.ca/eng/science-and-innovation/agricultural-practices/agriculture-andclimate/future-outlook/climate-change-scenarios/length-of-growing-season-in-the-prairieregion/?id=1363104179472, Fowler, 2002). To improve the viability of winter wheat production within Canada, it is necessary to obtain a better understanding of the molecular and genetic mechanisms underlying field survival during the harsh prairie winters.

2.2 Vernalization requirement and its contribution to cold tolerance

Genetic mapping studies have determined that *Vrn-1* loci, located on chromosome 5A, 5B and 5D, control the vernalization requirement and is one of the principal genetic determinants for cold tolerance in winter wheat (Fowler et al., 1996; Galiba et al., 1995; Law et al., 1976; Snape et al., 1985). In short, *Vrn-1* distinguishes the growth habit between spring and winter wheat. Both spring and winter wheat undergo a transition from vegetative to reproductive growth to complete their life cycle. Spring wheat accomplishes this via monitoring changes in photoperiod and ambient temperature (Winfield et al., 2009). Winter wheat, however, must experience cold temperatures over a variable length of time depending upon the cultivar to effectively transition from vegetative to reproductive state. Winter cereals, compared to spring varieties, are delayed in apical development and will not form a double ridge structure till the vernalization requirement has been met (Limin and Fowler, 2006). Reproductive tissues are known to be more sensitive to low temperature stress than vegetative tissues (Galiba et al., 2009). Thus, the vernalization requirement prevents sensitive tissues from being produced during the winter (Fowler et al., 1996). Vernalization is also associated with partially regulating other low temperature adaptations collectively referred to as winter hardening (Fowler et al., 1996). This results in increased cold tolerance that is maintained during the vernalization period till it reaches the vernalization saturation point, after which LT tolerance starts to decrease (Fowler et al., 1996). A longer

A common ancestor (2n=14)

Triticum
AA

Aegilops

Ae. speltoides
SS

T.boeoticum
$A^b A^b$

T. urartu
$A^u A^u$

Ae. speltoides ▪
GG

Ae. speltoides ★
BB

Ae. squarrosa ●
DD

T. araraticum (2n=28) ▪
$GGA^b A^b$

T. monococcum
$A^b A^b$

T. dicoccoides (2n=28) ★
$BBA^u A^u$

T. zhukovskyi (2n=42) ▪
$GGA^b A^b A^b A^b$

T. aestivum (2n=42) ★●
$BBA^u A^u DD$
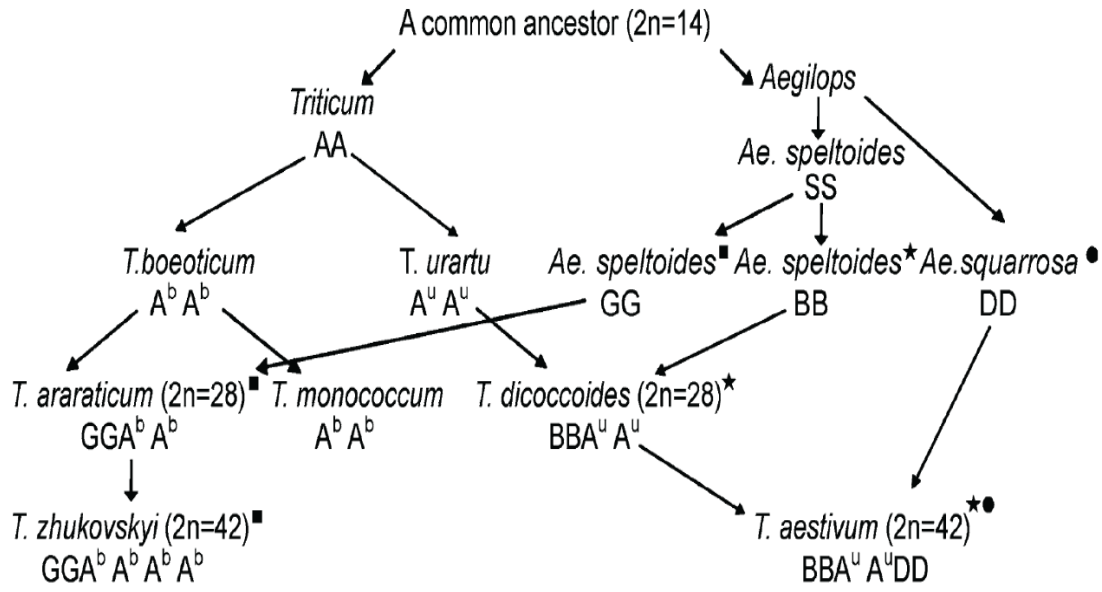
Fig. 2.1 Diagram illustrating the evolution of wheat and related species. Species names and their genome compositions are indicated. Path of inheritance for the D, B and G genomes is indicated by the circle, star and rectangle symbols. Figure is adapted and modified from Golovnina et al. 2007.

vernalization requirement therefore results in winter hardiness being maintained for a longer period during winter (Fowler, 2008). Throughout the vernalization period, leaf primordia are continually produced, which results in a higher final leaf number on the mature plant (Danyluk et al., 2003).

The spring/winter growth habit is determined by allele combinations at *Vrn-1A*, *Vrn-1B* and *Vrn-1D* (Fowler et al., 1996; Galiba et al., 1995; Law et al., 1976; Snape et al., 1985). Each gene influences the strength of the vernalization requirement to varying degrees. A dominant *Vrn-1A* allele abolishes the vernalization requirement completely. Dominant *Vrn-1B* and *Vrn-1D* alleles result in weak vernalization requirement of 15 to 30 days, while if all three alleles are recessive, the complete winter habit requires 45 to 60 days of cold temperatures to complete the vernalization requirement and transition to reproductive phase in the spring. Spring cereals, such as the spring wheat Glenlea and the spring rye Gazelle, are able to attain their maximal cold tolerance by 21 days at 4°C, with LT_{50} values of -5°C to -7°C respectively (Fowler et al., 1996). The winter wheat Norstar and the winter rye Puma require approximately 49 days to attain their full potential of cold tolerance, reaching LT_{50} values of -20°C and -25°C respectively (Fowler et al., 1996). In addition to group 5 chromosomes, QTLs at genomic regions on 4A, 6B and 6D are also implicated in altering length of vegetative growth in a Norstar x winter Manitou population (Båga et al., 2009). The *Vrn-1* loci are correlated with the activation of genes associated with cold tolerance, such as the *Wcs120* family (Fowler et al., 1996). Both spring and winter cereals exhibit a burst of transcriptional cold tolerance response elements during cold stress. These transcript numbers rapidly decline in spring cultivars, while winter cultivars maintain high transcription rates for the entire vernalization period (Fowler et al., 1996). Even when the vernalization requirement has been reached for the winter cereals, the cold responsive transcripts do not fall below the maximum achievable transcripts in the spring cultivars (Fowler et al., 1996).

2.3 Physical mapping of low temperature QTLs in wheat

A major QTL (*Fr*-A2) for LT tolerance in wheat is located upstream of *Vrn-1* on the long arm of the 5A chromosome (Båga et al., 2007; Galiba et al., 1995). An orthologous locus is also present in barley (Skinner et al., 2005) and rye (Jaglo et al. 2001; Campoli et al. 2009). *Fr-A2* is 30 cM proximal to *vrn-1* on chromosome 5A, and explains about 40% of the variation in LT tolerance observed in a Norstar x winter Manitou population. This region, across multiple maps and different

cereals, is correlated with a high density of CBF genes in cold-hardy species (Båga et al., 2007; Miller et al., 2006).

2.4 CBF and its role in cold tolerance

The CBF genes present at *Fr-A2* encode transcription factors of the APETALA2/ ethyleneresponsive element binding factor (AP2/ERF) family and control the expression of COR genes harboring the CRT/DRE cis regulatory element (Fig. 2.2) (Baker et al., 1994; Stockinger et al., 1997; Yamaguchi-Shinozaki and Shinozaki, 1994). About 84% of Arabidopsis genes which are up-regulated 15-fold during cold acclimation belong to the CBF regulon and 79% of the active metabolome associated with cold tolerance in *Arabidopsis* is related to CBF activity (Cook et al., 2004; Vogel et al., 2005).

2.5 Structure of CBF and its mechanism of action

CBFs, like other members in the AP2/ERF superfamily, are characterized by the presence of three β-sheets and a proceeding α-helix domain in the active site of 60 amino acids, with an arginine and a tyrosine in key complementary positions for DNA binding (Allen et al., 1998). These structural components interact with repeated sequences GCC/CCG (CRT) (Allen et al., 1998). Arginine and tryptophan residues contained within the β-sheet allow the protein to bind with the major groove of the DNA (Allen et al., 1998). It was initially believed that AP2/ERF proteins were found only in plants, but they show similarities to endonucleases present in cold tolerant bacterial species (Hake et al., 2004; Rashid et al., 2012). The lack of introns for most genes coding for AP2/ERF proteins suggests CBF genes likely have a bacterial origin (Hake et al., 2004; Rashid et al., 2012). However, certain CBF genes have introns that are highly conserved across plant families (Nakano et al., 2006). The target for CBF binding is the CRT element usually contained in the upstream regulatory region of COR genes, which respond to cold exposure (Stockinger et al., 1997; Vogel et al., 2005). It is generally believed that the COR genes have a major influence on cold tolerance in plants (Cook et al., 2004), which makes CBF genes an obvious target for improvement of cold tolerance in plants. This possibility is encouraged by increased cold tolerance levels in transgenic Arabidopsis plants over-expressing CBF genes without cold acclimation (Liu et al., 1998). However, the strategy is not straight-forward as negative effects such as stunted growth may occur if CBF over-expression in plants is too high (Liu et al., 1998).

2.6 Activation of CBFs by ICE1

CBF genes are under the control of several upstream factors that regulate CBF expression levels. One of these regulators is the Inducer of CBF expression 1 (*ICE 1*), which belongs to the basic helix-loop-helix transcription (bHLH) family. Similar to other members of this family, ICE1 interacts with MYC recognition elements present in promoters of target genes such as CBF3 in Arabidopsis (Chinnusamy et al., 2003). Aside from *ICE1*, there is also a *ICE2*, which also influences CBF expression in Arabidopsis and shares many of the same functional groups of *ICE1* and is constitutively expressed (Fursova et al., 2009). Other CBF genes may also be regulated by ICE1-like factors (Miura and Furumoto, 2013). *ICE1* is localized within the nucleus and is constitutively expressed under normal growth conditions (Chinnusamy et al., 2003). Overexpression of *ICE1* results in a greater degree of cold tolerance in transformed Arabidopsis plants (Fursova et al., 2009). Regulation of *ICE1* is handled by *HOS9* and *SIZ1*, which regulate ubiquitination and sumoylation of target CBFs, respectively (Jung et al., 2014; Miura et al., 2007). Via *HOS9*, which encodes a RING finger E3 ligase, *ICE1* is degraded by the 26S proteasome abolishing any up-regulation of ICE1 downstream pathways (Ishitani et al., 1998; Jung et al., 2014; H. Lee et al., 2001). Sumoylation of the *ICE1* protein by *SIZ1*, which encodes a SUMO E3 ligase, stabilizes ICE1 by preventing ubiquitylation of ICE1 (Miura et al., 2007). Due to this regulation by ICE1 and SIZ1, plants with a *Siz1* mutation are hypersensitive to cold temperature stress (Miura et al., 2007). Both *HOS9* and *SIZ1* work in concert with one another to fine tune the expression of *ICE1*, striking a balance between enabling expression and preventing deleterious effects (Thomashow, 2010).

This balance characterizes the transient cold-shock expression of CBFs, whereby transcription increases rapidly till peaking within two 2 to three h, followed by a decrease in levels down to a few times greater than the expression level in plants at normal growth temperatures (Gilmour et al., 1998; Thomashow, 2010). In wheat, there are also two different *ICE1*-like genes, *TaICE41* and *TaICE87*, which are 50% and 47% homologous to the Arabidopsis *ICE1* and *ICE2*, respectively and have similar function and regulation (Badawi et al., 2008).
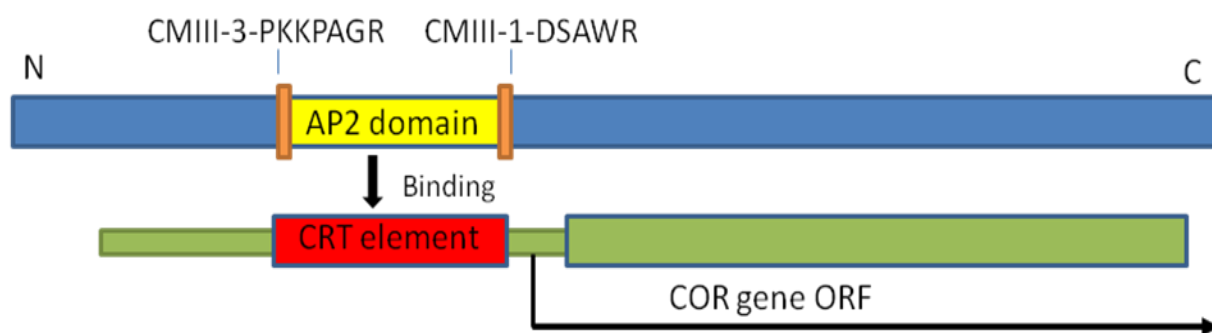
Fig. 2.2 Diagram of CBF12 and its regulation of COR genes. Schematic illustration of CBF DNAbinding region composed of the AP2/ERF domain flanked by conserved signature motifs, CMIII3 and CMIII-1 (Skinner et al., 2005) is shown at the top. The AP2-recognition site, the CRT element, located in the regulatory region of target COR genes is shown below.

2.7 Environmental influences on CBF expression

While *ICE*-like genes are able to induce CBF expression, MYB15 is noted for being an active suppressor (M. Agarwal et al., 2006). MYB15 can bind to both ICE1 and the promoters of the CBF genes and thereby prevent CBF expression (M. Agarwal et al., 2006). The MYB proteins are part of the abscisic acid (ABA)-dependent pathways for stress responses (P. K. Agarwal et al., 2006). ABA-dependent elements paradoxically however, can also increase cold tolerance. ABAinsensitive and -deficient Arabidopsis mutants, such as *abi-1* and *aba-1*, have reduced cold tolerance (Gilmour and Thomashow, 1991; Heino et al., 1990). Exogenous application of ABA is also noted to increase cold tolerance (Baker et al., 1994). The effect of ABA provides a more modest increase as compared to constitutively expressed CBFs (Baker et al., 1994). ABA responsive genes are regulated by basic leucine zipper transcription factors called ABFs or AREBs, which bind to the promoter sequence of C/TACGTGGC in their target genes, including the CBFs (Xiong et al., 2002).

Light conditions have a large role in the regulation of CBF expression. These influences act via systems sensing the chloroplast redox status, light quality and intensity and the circadian clock (Fig 2.3) (Kurepin et al., 2013). Some of these light-induced signals lead to spikes in the concentration of cytosolic Ca^{2+} ions that activate kinase signaling cascades (Knight and Knight, 2000). The effect of calcium is further demonstrated by antiporters which regulate the out-flux of calcium, such as *Calcium Exchanger 1*, whereby if it is knocked out results in an increase in cold tolerance (Catalá et al., 2003). Furthermore, phytochrome interacting factor (PIF) 7 negatively regulates CBF3 expression in Arabidopsis (Kidokoro et al., 2009). PIF7 is known to interact with Timing of CAB Expression1 (TOC1), connecting CBF regulation with the circadian cycle, whereby co-expression of PIF7 and TOC1 reduces CBF3 expression significantly more than PIF7 expressed alone (Kidokoro et al., 2009). The expression patterns for CBF1, CBF2 and CBF 3 over 36 h matches the patterns of Circadian Clock-Associated 1(CCA1) and Late Elongated Hypocotypl (LHY), which are components of the central oscillator in the circadian clock, (Fowler et al., 2005; Keily et al., 2013; Kidokoro et al., 2009). The CBFs therefore, integrate information from many different biochemical sources to help define cold tolerance and are vital regulators in overwintering plants (Fig 2.3). CBFs do more than just interact with cold tolerance genes, they also can bind to other transcription factors responsive to cold and drought stress, such as the dehydrins (Wise and
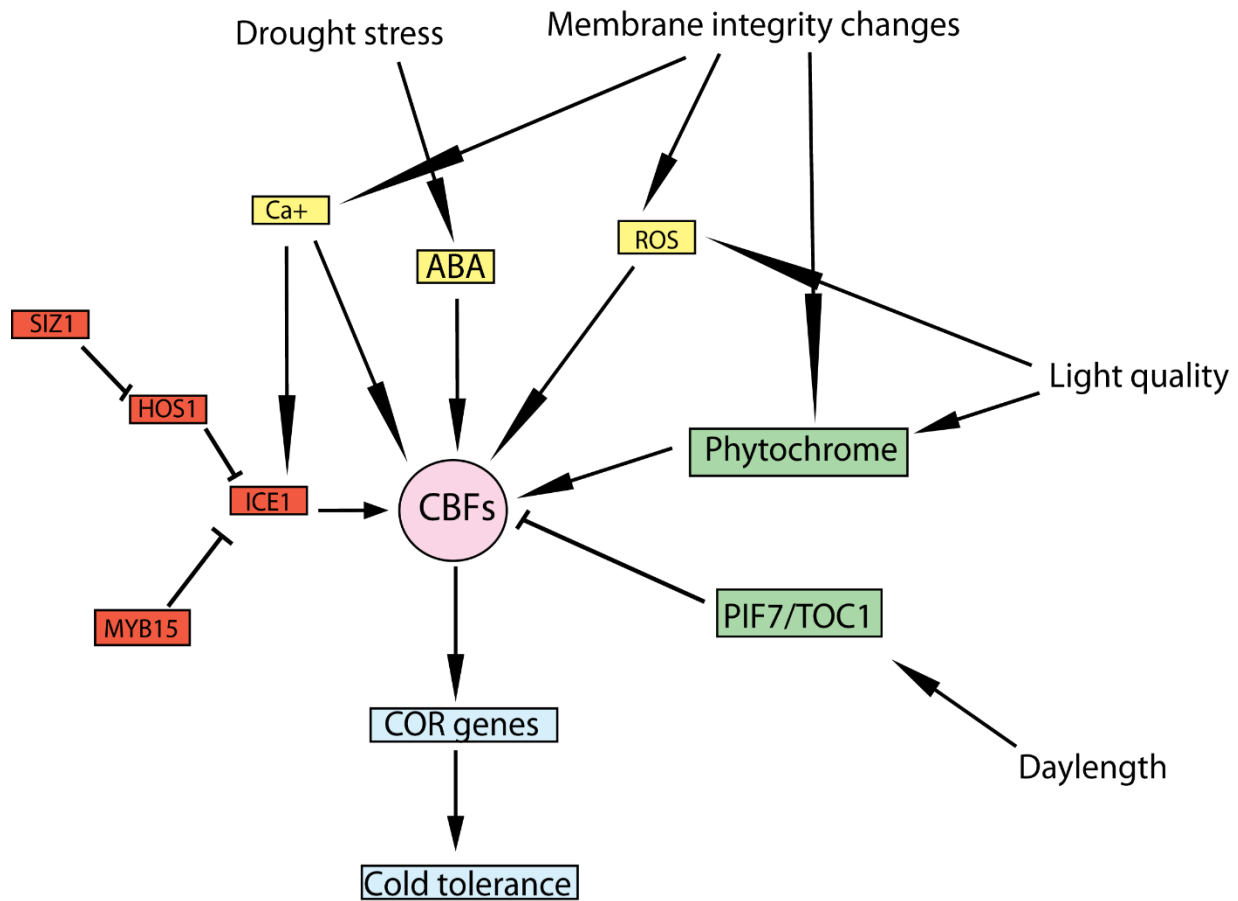
Fig. 2.3 Schematic illustration of some pathways and environmental factors affecting CBF expression. Environmental factors are designated as unbounded text. Factors responsive to light and photosynthesis are bounded by green boxes. Aspects responsive to drought and freezing stress are represented by yellow boxes. Red boxes correspond to regulatory factors which directly regulate CBF expression. Blue boxes represent CBF targets.

Tunnacliffe, 2004). In this way, multiple pathways can come together for a dynamic response involving cold stress.

2.8 Adaptation to freezing temperatures and CBF variation

The genetic basis of adaptability of organisms to varying environments is not well understood. *Arabidopsis thaliana*, a model plant with considerable genetic resources is naturally adapted to a broad range of latitudes and climatic conditions (Koornneef et al 2004). Arabidopsis accessions collected from diverse regions of native distribution show clinal patterns of variation associated with latitude and temperature. A highly significant positive relation is noted between freezing tolerance and latitude (Zhen and Ungerer 2008). In a study of a RIL population derived from a cross between *A. thaliana* genotypes adapted to widely different European climates, genetic mapping identified 15 fitness QTL and only five of these revealed tradeoffs for local adaptation (Ågren et al 2013). In a study of southern European Arabidopsis accessions, CBF genes showed 1.5 to 4.6 times greater occurrence of nucleotide polymorphism that were regulatory or nonsynonymous, thus significantly reducing expression of key COR genes, such as COR15A, COR6.6 and COR78 (Zhen and Ungerer, 2008). The observed polymorphisms were symptomatic of a lack of selection pressure, not selection against CBFs (Zhen et al., 2011). A high expression of CBFs in a warm environment has negative effects on the plants such as promotion of stunted growth, and CBFs may constitute a drain on available resources under such circumstances (Liu et al., 1998).

Within grasses, to which wheat belongs, a considerable variation to survive under extreme conditions is possible due to the genome duplication that occurred during the Cretaceous –Tertiary boundary mass extinction (Fawcett et al 2009). Under extreme environmental conditions, increased gene content can be associated with higher rate of evolution and adaptation due to functional redundancy that offers an evolutionary advantage for plants to survive (Crow and Wagner 2006). The ability of winter cereals to survive under extreme cold temperatures can be attributed to a higher number of CBF genes (Knox et al., 2010; Sieber et al., 2016; Zhu et al., 2014). In Norstar, 60 CBF genes are known to be present (Mohseni et al., 2012), with 27 of these genes being paralogous and two genes being pseudogenes. Some of these genes have features that appear to be detrimental to binding, such as substitutions or deletions within the AP2 domain or CMIII-3 motif (Mohseni et al., 2012). As such, to improve cold tolerance, an attractive approach would be to

determine which CBF genes optimally enhance cold tolerance and then increase their copy number within the genome.

2.9 Physiological basis of psychrophilic proteins

During the process of cold acclimation, CBFs induce numerous physiological processes which provides the plant with protection against freezing temperatures. Such changes can include the substitution of saturated phospholipids by unsaturated ones within the plasma membranes of cells and the accumulation of osmolytes (Survila et al., 2010; Thomashow, 1999; Uemura and Steponkus, 1999). Proteins generally however, are less reactive at low temperatures and cannot be as readily altered. Processes which work effectively at room temperature are difficult to perform below freezing. At sufficiently cold temperatures, protein can also become denatured, completely preventing its function (Crosatti et al., 2013). Proteins are the effectors of mRNA transcripts, so it is important to consider their activity and accumulation. To counteract low or high temperature stress, specifically adapted proteins are required to retain function. One strategy is simplifying the degree of binding between the substrate and the active site and thereby decreasing the activation energy required (Feller and Gerday, 1997). Generally, for psychrophilic and psychrotrophic enzymes, the active site will be highly conserved compared with their mesophilic counterparts. Instead, the rest of the protein structure is made more open and flexible (D'Amico et al., 2001). Greater flexibility allows a protein to couple and decouple with its substrate fast enough to mitigate low temperature inhibition. Improved flexibility can be achieved by lowering the number of ionic pairs, aromatic and hydrophobic interactions and decreasing the degree of hydrogen bonding occurring within the structure. As such, fewer arginine and proline and more glycine residues will be present (D'Amico et al., 2002).

Becoming more flexible however also results in the protein being more unstable. The halflife of many cold tolerant proteins is under 12 min at 50^0C and some can denature at only 35^0C (Siddiqui and Cavicchioli, 2006). However, stress related proteins, like the CBFs and dehydrins, are thermostable (Alonso et al., 2002; Das and Pappu, 2013; Graether and Boddington, 2014; Kawamoto et al., 2007; Tantos et al., 2009, 2013; Uversky, 2002).

Proteins isolated from psychrophiles are noted for being unstable *in vitro*, and it has been demonstrated that there are some highly resilient enzymes which are most active at low

temperatures (Gatti-Lafranconi et al., 2010). Additionally, proteins such as an aspartase isolated from the psychrophile *Cytophaga* sp. KUC-1 show greater resistance to denaturation by high temperatures and pH compared to mesophile equivalents (Kazuoka et al., 2003). Only a few changes to the amino acid composition confer this stability (Schmid, 2011).

Changes which increase the degree of hydrophobicity of the protein interior, add and stabilize secondary structures, improve compactness of the protein and increase the number of ionic and electrostatic interactions increase the stability of a protein (Feller and Gerday, 1997; Hazel and Prosser, 1974; Kazuoka et al., 2003; Oikawa et al., 2003; Somero, 1995; Struvay et al., 2013). Improving stability for mesophiles and thermophiles is achieved by enhancing the hydrophobicity of the protein and making its structure more compact (Hazel and Prosser, 1974; Somero, 1995). Enhanced hydrophobicity is also associated with a proteins possessing more secondary structures (Kazuoka et al., 2003; Oikawa et al., 2003). Against all conditions, secondary structures such as C-terminal alpha coils, hairpin loops and compacted, overlapping beta sheets can drastically improve a proteins resistance to denaturation by heat or solvents (Alonso et al., 2002; Arnórsdóttir et al., 2002; Pudney et al., 2003; Rath et al., 2009; Struvay et al., 2013). Hydrophobic interactions however become weaker than hydrophilic interactions at low temperatures, so possible gains may be very small (Siddiqui and Cavicchioli, 2006). Changes which result in a stronger yet dispersed charge have been demonstrated to be more effective for improving cold temperature protein stability, although all changes to stability are cumulative (Das and Pappu, 2013; Feller and Gerday, 1997). Point mutations which replace an uncharged amino acid with a negatively charged one, such as aspartic acid, can result in improved stability provided the hydrophobic cores are preserved (Alves et al., 2004; Takano et al., 1988). An exception to this is intrinsically disordered proteins lacking in secondary or tertiary structure at physiological conditions (Kalthoff, 2003). Despite this, they can be resilient against high and low temperature denaturation (C. Lee et al., 2014; W. K. Lee et al., 2014; Tantos et al., 2013, 2009). This is due to a combination of high net charge and low hydrophobicity such that a hydrophobic core does not exist (Kalthoff, 2003; Tantos et al., 2009). How they are able to retain stability despite their lack of structure is due to stronger interactions between the protein, its substrates and the solvent, and lacking a hydrophobic core which can be irreversibly denatured (Kalthoff, 2003; Tantos et al., 2009, 2013). For plants, these characteristics typically define cold shock and dehydration responsive elements, such as dehydrins (Graether and Boddington, 2014; Kazuoka and Oeda, 1992). These characteristics are also shared by numerous

transcription factors, allowing for high enzymatic affinity despite the conditions (Alonso et al., 2002; Das and Pappu, 2013; Kawamoto et al., 2007; Tantos et al., 2009, 2013; Uversky, 2002).

2.10 Anomalous migration of proteins during SDS-PAGE

A primary amino acid sequence which allows an individual protein to be resilient to a range of conditions, can also result in anomalous migration during SDS-PAGE (Alonso et al., 2002; Alves et al., 2004; Iwashita et al., 2015; Manning and Colón, 2004). Sodium dodecyl sulfate (SDS) is a unique denaturing detergent in that it binds to amino acids on average at a ratio of 1.4 g : 1 g of protein irrespective of protein composition (Imamura, 1988). This binding occurs in two stages, differentiated by being either below or near the critical micelle concentration, which for SDS is typically 7 mM (20% w/v) in water (Manning and Colón, 2004). At the first stage, there are ionic interactions between basic amino acids and the charged heads of dodecyl sulfate ions. At the critical micelle concentration, surfactants and the protein undergo cooperative binding, whereby multiple molecules of SDS bind to each amino acid and as more SDS becomes bound to the protein, it becomes easier for SDS molecules to bind to the protein. Micelles of SDS form along the denatured protein like "beads on a string" at regular intervals, with hydrophobic components buried within the micelle and hydrophilic parts wrapped around the surface. Attached to these micelles, the native protein conformation changes to mixture of random coils and induced α-helixes. If the protein is completely denatured, SDS will be able to normalize the charges and hydrophobicity of the amino acids and have the protein migrate through the gel purely based on size. While SDS in general can mitigate properties of individual amino acids, there are structural factors which inhibit either the binding of SDS or preventing SDS from accessing regions of the protein structure. These proteins will migrate either faster or slower through the polyacrylamide gel than their predicted molecular weight, referred to as anomalous migration or gel shifting (Rath et al., 2009). Proteins with a greater ratio of charged to hydrophobic amino acids, especially negatively charged groups, will bind to surfactants differently (Alves et al., 2004; Armstrong and Roman, 1993; Garcia-Ortega et al., 2005; Iwashita et al., 2015; Takano et al., 1988). It is postulated that this is due to negative charges repelling the dodecyl sulfate ions. Entire structural motifs can also result in gel shifts, such as having elongated or transmembrane domains and overlapping βsheets (Manning and Colón, 2004; Rath and Deber, 2013). Combining aspects of structure and amino acid composition, the presence of high concentrations of proline at the N or C-terminus can result in greater stability as

well as anomalous migration, due to stabilization or disruption of αhelical structures within the protein against SDS binding (C. Lee et al., 2014; Takano et al., 1988). Post-translational modification may also result in anomalous migration, due to the incorporation of phosphate groups, metal ions, carbohydrate or lipid groups and others to the protein (Adamson et al., 1992; Bretscher, 1971; Georgieva and Sendra, 1999; Hayward et al., 2002; Hosoi et al., 1995; Shi et al., 2012). These modifications can either increase or decrease the rate by which a protein migrates by the availability of the protein to SDS (Hayward et al., 2002). The size of the protein also change due to the addition of the modification (Bretscher, 1971). All these described structures and modifications can work in tandem or in varying degrees, to alter the migration pattern of proteins during SDS PAGE separation (Shi et al., 2012; Takano et al., 1988). Anomalous migration can also be associated with the propensity of proteins to aggregate during purification and within tissues as amyloid fibrils, or if the protein is intrinsically disordered (Amani and Naeem, 2013; Kalthoff, 2003; Karan et al., 2013; W. K. Lee et al., 2014; Manning and Colón, 2004; Ohshida et al., 2016; Uversky and Dunker, 2010).

2.11 Production of recombinant proteins in *E. coli*

Transcription factors, like the CBFs, are only required in minute quantities of proteins within the organism to function, therefore it is a challenge to purify adequate quantities to study their function (Harrison et al., 1991). To overcome this challenge, *in vitro* production of recombinant CBF is preferred. Since CBF genes do not have introns, they should express and process properly in a prokaryotic system (Hake et al., 2004; Rashid et al., 2012). For applications involving recombinant proteins, an *E coli* expression system based upon the *lac* operon and induced via IPTG is used. IPTG allows for consistent and constant over-expression of the desired protein, since it cannot be metabolized by bacteria (Hockney, 1994). For each protein however, some optimization may be required.

To facilitate purification of *in vitro* produced recombinant protein, a tag is attached to the desired protein. Previous studies of wheat CBFs used thioredoxin (Trx), histidine and S15-peptide tags to enable purification (Jain, 2013; Gangopadhyay, 2015). The Trx tag improved protein solubility, whereas the S-tag enabled purification using immunoaffinity columns and the His tag allowed purification using Ni^+ affinity chromatography. To analyze the SDS-PAGE migration of purified CBFs it was suggested that tags like the Trx tag might influence the migration of CBF on

the SDS-polyacrylamide gels (Costa et al., 2014). However, His tag does not affect the migration of recombinant proteins on SDS-PAGE gels (Alves et al., 2004; Gupta et al., 2013; Novak et al., 2013).

2.12 Tricine and Glycine SDS-PAGE

When visualizing proteins during SDS-PAGE, care must be taken in using the most accurate technique. Generally, SDS-PAGE refers to the glycine-SDS-PAGE whereby the ion front of the reducing polyacrylamide gel is determined by chlorine ions and glycine through a discontinuous buffer system (Laemmli, 1970). This system is able to separate and visualize proteins, ranging from 10 kDa to 500 kDa (Laemmli, 1970). Its versatility is due to the pore size of the polyacrylamide gels depending upon the concentration and cross-linking ratio of acrylamide:bisacrylamide. For proteins which are smaller than 20 kDa however, glycine-SDS-PAGE gels lack the resolution required to separate them (Schägger and von Jagow, 1987). Smaller proteins either run off the gel with the glycine ion front or, if higher concentrations of acrylamide are used in the resolving gel, can result in protein smearing or streaking (Schägger and von Jagow, 1987). The development of Tricine SDS-PAGE provided a method to efficiently separate proteins ranging from 30 kDa down to 1 kDa (Schägger, 2006; Schägger and von Jagow, 1987). This is due to the differences between glycine and tricine fronts within the stacking gels during SDS-PAGE. Glycine migrates slowly through the acidic stacking gel, facilitating stacking of larger proteins (Schägger and von Jagow, 1987). Proteins below 20 kDa in size however do not properly disassociate from the bulk of the migrating SDS and thus do not stack correctly (Schägger and von Jagow, 1987). The stacking limit for glycine-SDS-PAGE therefore is generally too high, though there are methods to alter this, such as using multiple gradient stacking gels to continuously destack the proteins according to size (Schägger, 2006). Tricine migrates much faster compared to glycine at typical pH values for the stacking and resolving gels, even though it has a greater molecular mass, due to more tricine being present in the anionic form (Schägger and von Jagow, 1987). This creates a lower stacking limit range, without requiring changes in the pH values of the polyacrylamide gels, additional stacking gels or increased concentrations of acrylamide (Schägger, 2006; Schägger and von Jagow, 1987). The limitation of this procedure is the loss of resolution at greater molecular masses.

2.13 *In Silico* analysis of protein structural properties

Through the continued efforts of molecular biologists and biochemists, a large wealth of data exists on the structure and capabilities of a wide range of genes and proteins. To facilitate and streamline further research, it is not always necessary to experimentally determine every feature a protein possesses. Instead, provided that the properties in question are wholly determined by the primary sequence, aspects such as size, net charge, PI, hydrophobicity and presence or absence of secondary structures can be computationally determined. The accuracy of these predictions however do not replace experimentally determined results for full proteins (Audain et al., 2015). To predict pI, it was found that prediction algorithms such as Bjellqvist, Branca, Cofactor, and support vector machine approaches correlated poorly with experimental data on protein pI, with r^2 values ranging from 0.6 to 0.15 depending on the algorithm (Audain et al., 2015). Against peptide sequences however, the algorithms are much more successful, at r^2 values of 0.96 between the predicted and experimental results (Audain et al., 2015). Thus computation methods for pI cannot be completely relied, however, they can be used to identify outliers from experimental data and modified sequences of proteins (Audain et al., 2015). Other features, such as protein aggregation and protein disorder, are difficult to determine experimentally, often requiring methodologies such as NMR spectroscopy (Dyson and Wright, 2005). It is important to know if these features are present as soon as possible, to minimize frustration in subsequent experiments (Linding et al., 2004). Protein disorder generally is suspected not on a specific parameter but rather the absence of normal protein behavior, such as anomalous migration during SDS-PAGE and aggregation during purification (C. Lee et al., 2014; W. K. Lee et al., 2014; Tantos et al., 2013, 2009). Disordered regions of a protein are characterized by higher frequencies of hydrophobic and charged residues or low sequence complexity due to repeated motifs (Dyson and Wright, 2005). Support vector machine algorithms featuring a sliding feature space to define a position-specific score matrix for the amino acid sequence is combined with a template search of known intrinsically disordered protein families to predict disordered regions (Ishida and Kinoshita, 2007). This Protein DisOrder prediction System, or PrDOS, predicted disordered regions at 90% accuracy, especially if they are short in length (Ishida and Kinoshita, 2007). A protein's propensity to aggregate can be defined by simple correlations of physio-chemical properties such as hydrophobicity, presence of multiple β-sheets and charge (Rousseau et al., 2006). TANGO is a statistical mechanics algorithm to predict protein aggregation based upon the aggregation propensity score of amino acid sequence and

sampled fragments of helixes or hairpin loops, at 90% accuracy (Fernandez-Escamilla et al., 2004; Rousseau et al., 2006). TANGO however can accurately compare mutants or modified proteins of a common sequence; it cannot be used to compare distinctly different proteins (Fernandez-Escamilla et al., 2004). Hydrophobicity and hydrophilicity can also be calculated by Grand average of hydrophobicity (GRAVY), which uses a moving-segment approach of a predetermined length to average the hydrophobicity of the entire amino acid sequence (Kyte and Doolittle, 1982). It is a simple calculation compared to TANGO, but cannot comment on protein secondary structure. Data on changes in the hydrophilicity can assist in the determination of protein disorder, since intrinsically disordered proteins are highly hydrophilic (Tantos et al., 2009). Computational approaches and bioinformatic tools may not have the same fidelity as experimental approaches, yet are still capable to make predictions in situations where experimental approaches are difficult and tedious.

CHAPTER 3: MATERIALS AND METHODS

3.1 *E. coli* strains

All *E. coli* strains used in this study were obtained as chemically competent cells from commercial sources (Table 3.1). The K12 NovaBlue cells were used for propagation of vectors constructed in the study. The K12 DH5α™-T1^R strain (ThermoFisher Scientific, Waltham, MA, USA) expresses an active McrBC endonuclease and was the initial host for plasmids modified by GeneArt site-directed mutagenesis. The type B BL21(DE3) and BLR(DE3)/pLysS strains are deficient in Lon and OmpT proteases and were therefore selected as host strains for protein expression vectors. The λDE3 prophage carried by the type B strains expresses the T7 RNA polymerase required for recombinant protein production and is regulated by an isopropyl β-D-1thiogalactopyranoside (IPTG)-inducible *lacUV5* promoter. In the absence of IPTG, "leaky" expression from the T7 RNA polymerase gene was suppressed by the presence of the pLysS plasmid encoding low levels of T7 lysozyme, a natural inhibitor of T7 RNA polymerase.

Each *E.coli* strain was typically transformed with ~1 ng plasmid following protocols supplied with the competent cells. NovaBlue and DH5α-T1 transformants were selected on solid Luria Broth (LB) media (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, 15 g/L agar, pH 7.0-7.2) containing 100 µg/mL carbenicillin. BL21(DE3) cells transformed with protein expression vectors were selected on a richer solid media, Super Broth (SB), consisting of 25 g/L tryptone, 15 g/L yeast extract, 5 g/L NaCl, 1 % (w/v) glucose, 15 g agar/L, pH of 7.1-7.2 and supplemented with 100 µg/mL carbenicillin. BLR(DE3)/pLysS transformants were selected on solid SB media containing 12.5 µg/mL tetracycline, 34 µg/mL chloramphenicol and 100 µg/mL carbenicillin.

3.2 Purification of genomic DNA and plasmids

Isolation of genomic DNA from young leaves of winter wheat cv. Norstar was done based on a CTAB method (Doyle and Doyle 1987) with modifications. The presence of high molecular weight DNA (>20 kb) and absence of detectable RNA in the genomic DNA preparations was verified by 0.8% agarose gel electrophoresis. For preparation of plasmid DNA, NovaBlue cells transformed with plasmid of interest were grown with shaking at 37°C for 10-15 hours in 2 ml LB media supplied with appropriate antibiotics. The cells were pelleted by centrifugation (12,000 x g,

Table 3.1 *Escherichia coli* strains used in the study

Strain	Type	Genotype	Antibiotic resistance	Usage
NovaBlue*	K12	*end*A1 *hsd*R17(r_{K12}^- m_{K12}^+) *sup*E44 *thi*-1 *rec*A1 *gyr*A96 *rel*A1 *lac* F'[*pro*A$^+$B$^+$ *lac*I^q ZΔM15::Tn10]	Tet^R	Plasmid propagation
DH5α-T1**	K12	F$^-$ ϕ80*lacZ*ΔM15 Δ(*lac*ZYA*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_{K12}^-, m_{K12}^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 *ton*A	None	Site-directed mutagenesis
BL21(DE3)**	B	F$^-$ *ompT gal dcm lon hsdS$_B$($r_B^-m_B^-$)* λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB$^+$*]$_{K-12}$(λ^S)	None	Expression of recombinant proteins
BLR(DE3)/ pLysS**	B	F- ompT gal dcm hsdSB(rB- mB-) Δ(srlrecA)306::Tn10 λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS) pLysS[T7p20 orip15A]	Tet^R Cm^R	Expression of recombinant proteins

Provided by *Millipore Sigma (St. Louis, MO, USA), **(ThermoFisher Scientific, Waltham, MA, USA).

5 min) and plasmids in cell pellet were extracted using the AccuPrep Plasmid Mini Extraction Kit (Bioneer, Deajeon, Republic of Korea) according to manufacturer's instructions. Prepared plasmids were dissolved in TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA), except for plasmids to be analyzed by DNA sequencing, for which 10 mM Tris-HCl pH 8.0 was used. The quality and concentration of genomic and plasmid DNA preparations was determined by UV spectrophotometry using a DU800 spectrophotometer (Beckman Coulter, Fullerton, CA, USA). DNA concentrations were calculated from the formula: $(A_{260}-A_{320})$ x dilution factor x 50 µg/mL. DNA samples with A_{260}/A_{280} ratios ~1.80 and A_{260}/A_{230} ratio >1.8 were considered to be of good quality.

3.3 Agarose gel electrophoresis and DNA fragment isolation

All DNA samples analyzed by agarose gel electrophoresis were dissolved in 1x sample buffer (0.025 g/mL Ficoll 400, 0.12% Orange G) prior to loading. To allow for size determination of separated DNA fragments, samples containing the GeneRuler 1 kb DNA Ladder (0.25-10.0 kb) or the MassRuler DNA Ladder Mix (0.08 -10.0 kb), both supplied by ThermoFisher Scientific (Waltham, MA, USA), were used as molecular weight markers. The gel electrophoresis was performed using 50-120 ml 1% (w/v) agarose gels containing 0.5 µg/mL ethidium bromide and 1x TAE running buffer (40 mM Tris-HCL, 20 mM acetic acid, 1 mM EDTA, pH 8.3) and applying a constant voltage (80-120 V) for 30-90 min. Thereafter, a ChemiDoc Touch Imaging System (BioRad, Hercules, CA, USA) was used to expose the gel to UV light for band visualization and determination of DNA fragment sizes.

For isolation of DNA fragments from the agarose gel, the EtBr-stained DNA bands were visualized by low intensity UV light and excised in the protection of a UV-safe shield. The AccuPrep Gel Purification Kit (Bioneer Corp., Deajeon, Republic of Korea) was used for extracting the DNA from the agarose slices following supplier's instructions. The DNA concentration in the preparations was determined by UV spectrophotometry. For dilute DNA samples, an aliquot of the purified DNA fragments was electrophoresed on a 1% agarose gel together with a sample containing standard fragments of known masses (MassRuler DNA Ladder Mix). Based on the intensity of standard and sample bands, the DNA concentration in gel-purified sample was estimated.

3.4 Enterokinase/ligation independent cloning (EK/LIC)

Various fragments of the CBF12 coding sequence were generated by PCR amplification for cloning into the protein expression pET-32Ek/LIC vector (Fig. 3.1A) supplied by Millipore Sigma (St. Louis, MO, USA). The template used for all amplifications was the TaCBF12 gene carried by the BAC clone 3149-L3 (Genbank accession JF729236) isolated from a genomic library of winter wheat (*Triticum aestivum*) cv. Norstar (Ratnayaka et al., 2005). The sequence-specific primers required for the amplification were designed using the Primer3 software (Untergasser et al. 2012) and consisted of a 12-16 nucleotide long adapter sequence followed by 13-18 nucleotides of the CBF12-specific sequence. Inclusion of adaptor sequence on amplified product was done to enable subsequent enterokinase/ligation independent cloning (Ek/LIC) into linearized pET-32Ek/LIC as outlined in Fig. 3.1B. The forward and reverse primers used to generate each insert are listed in Appendix 1 and were provided by Eurofins MWG Operon (Eurofins Genomic, Louisville, KY, USA).

The PCR amplification for production of pET32-Ek/LIC inserts was done in 25 µL reaction volumes consisting of 1x *Pfu* PCR buffer (20 mM Tris-HCL, pH 8.8, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 0.1% Triton X–100, 0.1 mg/mL BSA), 1.2 mM of $MgSO_4$, 0.2 mM of each dNTP, 1 x CES (0.54 M betaine, 1.34 M DTT, 1.34% (v/v) DMSO and 11 µg/mL BSA), 5 µM each of forward and reverse PCR primers, 1 ng of 3146L3 BAC DNA (or 0.1 µg of genomic wheat DNA) and 1 unit of the *Pfu* proofreading DNA polymerase (ThermoFisher Scientific, Waltham, MA, USA). The 1 x CES buffer promotes melting of GC-rich templates and was required for all reactions containing the CBF12 coding fragments (74.9% GC overall) as template. The amplifications were performed using an Eppendorf AG thermocycler (Eppendorf, Hamburg, Germany) with ramping adjusted to 2.5^{o}C s^{-1} for heating and 1.5^{o}C s^{-1} for cooling. An initial denaturation of template was done at 95°C for 4 min followed by 32 cycles of denaturation at 94°C for 1 min, annealing at 6063°C for 30 sec, extension at 72°C for 1 min and a final extension at 72°C for 10 min to conclude the reaction. The specific annealing temperature for each primer pair and size of PCR product is given in Appendix 1. The generated PCR products were treated with T4 DNA polymerase to generate adapter overhangs that were compatible with the linearized pET32-Ek/LIC vector (Fig. 3.1). In this reaction, 0.2 pmol PCR product was incubated at 22°C for 30 min in 1x T4 polymerase buffer (330 mM Tris-acetate, 660 mM K acetate, 100 mM Mg acetate pH 7.8), 2.5 mM dATP, 1
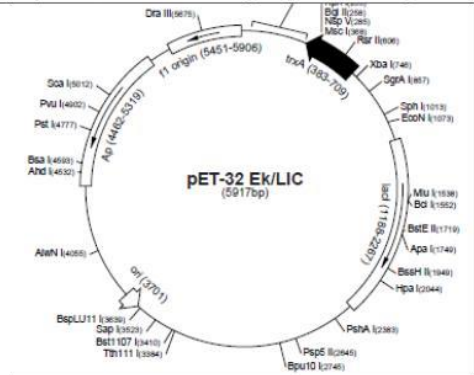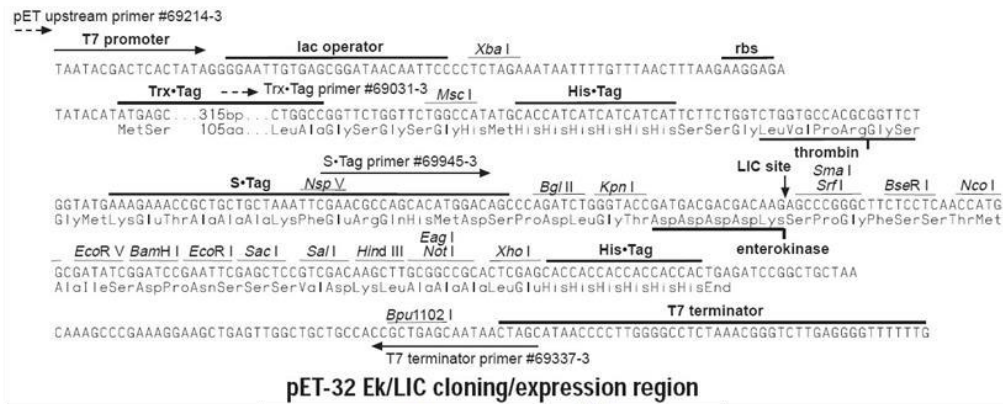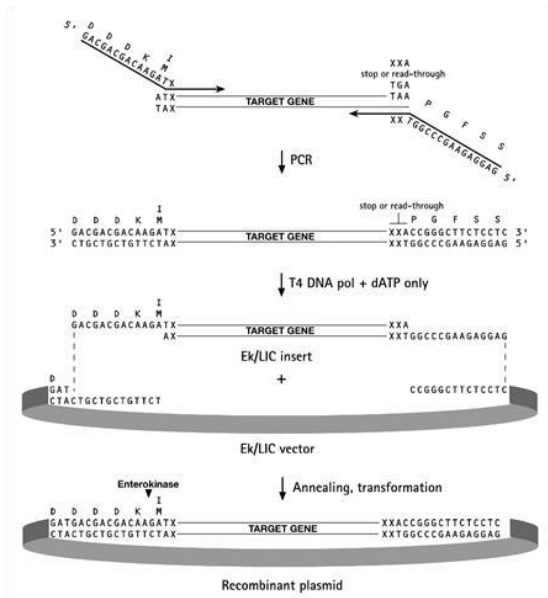
Fig. 3.1 Production of protein expression vectors using Ek/LIC strategy. (**A**) Cloning vector pET32Ek/LIC with DNA sequence around LIC insertion site. (**B**) Strategy for Ek/LIC cloning. Figures were adapted from pET-32Ek/LIC information pamphlet and Ek/LIC Cloning Kit manual, both supplied by Millipore Sigma (St. Louis, MO, USA).

unit T4 polymerase and 5 mM DTT in a total volume of 20 µL. The reaction was then heat inactivated at 75°C for 20 min. Annealing of insert and vector was done at 22 °C for 5 min upon addition of 2 µL aliquot T4 polymerase-treated PCR product (0.02 pmol) to 1 µL 100 ng/µL (0.02 pmol) linearized pET-32 EK/LIC vector and 1 µL 25 mM EDTA. The annealed circular product was transformed into competent Novablue cells. Successful integration of desired insert into vector was verified by DNA sequencing.

3.5 Conversion of pTrxHisS-L3-CBF12 type constructs to pHis-L3-CBF12 type constructs

Expression vectors encoding TrxHisS-CBF12 and variants thereof were converted to vectors encoding the corresponding CBF12 polypeptides, but carrying His tags as outlined in Fig. 3.2. The construction was initiated by restricting 2.0 µg of each pTrxHisS-CBF12 type vector with BglII and BamHI (ThermoFisher Scientific, Waltham, Mass., USA) to release a CBF12 DNA fragment flanked by 27 bp upstream CBF12 start codon and 33 bp downstream stop codon. Separation of the BglII-BamHI fragment from the vector part was done by 1% agarose gel electrophoresis followed by DNA fragment isolation from the gel and DNA quantification. An aliquot (~0.005 pmol) of each fragment was ligated at 16 °C for 3 hours with 2.5 nmol BamHI-digested and dephosporylated pET-15b vector in 20 µL volume containing 1x T4 DNA ligase buffer and 1.0 units T4 DNA ligase (New England BioLabs, Ipswich MA, USA). This ligation generated an inframe fusion between pET15b-encoded His-tag and introduced CBF12 sequence.

Dephosphorylation of pET-15b vector used in the ligation was done by incubating 1 µg BamHIdigested pET-15b in a 50 µL reaction volume containing alkaline phosphatase buffer (50 mM BisTris-Propane HCl, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$ pH 6.0) and 2 units Antartic Phosphatase (New England BioLabs, Ipswich MA, USA) for 15 min at 37°C. The phosphatase was subsequently inactivated by heating at 65°C for 5 min.

3.6 Site-directed mutagenesis

Site-directed mutagenesis of pHis-CBF12 vector was done to insert stop codons or restriction enzyme sites (*NheI* or *SmaI*) at various locations along the CBF12 coding sequence. To generate the mutations, two different methods were used as described below and summarized in Appendix 1. Both techniques required mutagenesis (Mut) primer pairs (Appendix 1) which were designed using Primer3 software (Untergasser et al 2012) or the GeneArt® Primer and Construct tool

(ThermoFisher Scientific, Waltham, MA, USA). All primers were provided by Eurofins MWG Operon (Eurofins Genomic, Louisville, KY, USA).

3.6.1 Method 1 (Q5 mutagenesis)

The primer pair used in the reactions contained more than eight non-overlapping bases at the 3' ends, and terminated by at least one G or C residue. A complementary mutated site was placed on each primer and located more than four bases from the 5' end and at least eight bases from the 3' end. The Q5 mutagenesis reactions were performed in 50 µL volumes containing 1x Q5 reaction buffer (New England BioLabs, Ipswich MA, USA), 0.2 mM of each dNTP, 0.5 µM of each mutagenesis primer, 1x Q5 high GC enhancer (New England BioLabs, Ipswich MA, USA), 1 ng pHis-CBF12 DNA, and 1 unit of Q5 high-fidelity DNA polymerase (New England BioLabs, Ipswich MA, USA). The PCR reactions were done as described (Appendix 1). Samples (15 uL) of the mutagenesis reactions were analyzed by 1% agarose gel electrophoresis to confirm successful amplification before a 1.0-uL aliquot was used for transformation of NovaBlue cells. The presence of desired mutation in selected clones was confirmed by DNA sequence analysis of purified plasmids from selected clones.
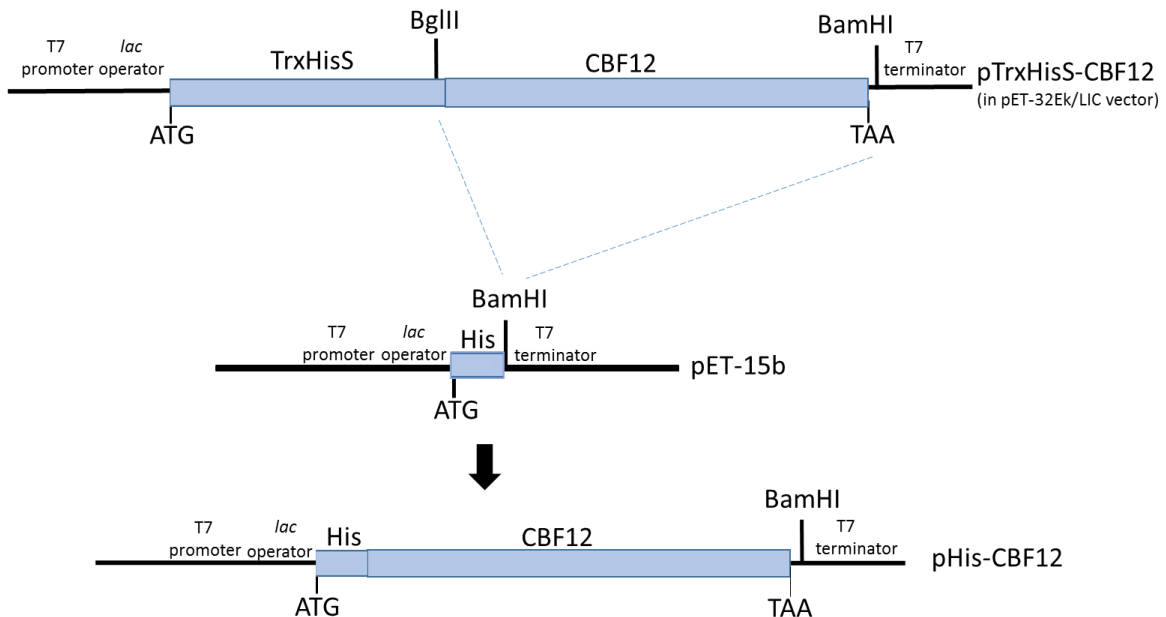
Fig. 3.2 Strategy for conversion of pTrxHisS-CBF12 constructs to pHis-CBF12 versions.

3.6.2 Method 2: GeneArt site-directed mutagenesis

Both forward and reverse primers synthesized for each mutagenesis reaction were 100% complementary to each other, 35-36 nucleotide long, and contained a centrally located mutated region (Appendix 1). The mutations involved four or five nucleotides, where novel *NheI* (CGATCG) recognition sites were introduced into CBF12 coding sequence. The mutagenesis reactions were done using the GeneArt site-directed mutagenesis plus kit (ThermoFisher Scientific, Waltham, MA, USA) and performed as schematically illustrated in Fig. 3.3. The methylation and PCR amplification steps occurred in the same reaction mixture of 20 µL containing 1x *Pfu* buffer (20 mM Tris-HCL, pH 8.8, 10 mM $(NH_4)SO_4$, 10 mM KCl, 0.1% (w/v) Triton X–100, 0.1 mg/mL BSA), 1x PCR Enhancer, 1x CES (0.54 M betaine, 1.34 M DTT, 1.34 % DMSO and 11 µg/mL BSA) 1.6 units of DNA methylase, 1x S-adenosine methionine, 2 mM dNTPs containing 7-deazadGTP, 1.5 mM $MgSO_4$, 5 pmol each of forward and reverse primers, 50 ng of 3146L3 BAC DNA. For the initial methylation reaction, the reactions mixture was incubated at 37^oC for 15 min and terminated by DNA denaturation at 95^oC for 2 min. Thereafter, 32 PCR cycles of denaturation at 94^oC for 20 sec, annealing at 69.9^oC for 30 sec, and extension at 72^oC for 10 min were done and concluded by a final extension at 72^oC for 10 min. During the PCR cycling, the ramp settings were 2.5^oC s^{-1} for heating and 1.5^oC s^{-1} for cooling. Following amplification, a 10µL sample of the reaction was analyzed by agarose gel electrophoresis to confirm presence of reaction product. A 4-µL aliquot of each successful PCR reactions was subjected to an *in vitro* recombination reaction performed in 20 µL reaction volume containing buffer and 1x enzyme mix supplied with the GeneArt site-directed mutagenesis kit. The recombination reaction was performed at room temperature for 10 min and terminated by addition of EDTA to 0.025 mM concentration. DH5α™-T1^R cells transformed with mutagenized plasmids were grown in LB media containing 100 µg/mL of carbenicillin.
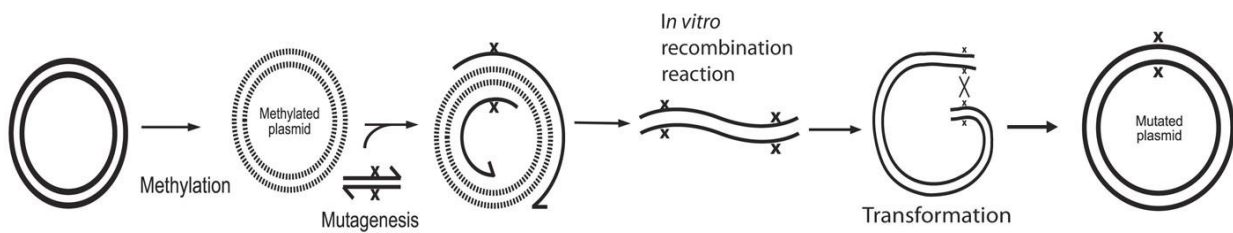
Fig. 3.3 Outline of strategy for GENEART® Site-Directed Mutagenesis System. Figure modified from user manual.

3.7 Sub-cloning of mutagenized pHis-CBF12 fragments

Two different pHis-CBF12 plasmids with NheI sites inserted by mutagenesis before CMIII-3 and before CMIII-1 (Appendix 1), respectively, were used to construct pHis-CBF12-DelAP2 (Figs. 3.1 and 3.2). Both mutant plasmids were digested NheI and KpnI, of which KpnI is a common and single site on both plasmids. The digest were done in Cutsmart buffer (New England Biolabs, Ipswich Mass., USA) and enzyme activities were removed upon digestion using a QIAquick Nucleotide Removal Kit (QIAGEN, Hilden, Germany). Ligation of DNA fragments was done usingT4 DNA ligase (New England Biolabs, Ipswich Mass., USA) and ligated vectors were transformed into DH5α cells. Screening of transformants by restriction digests led to identification of clones with CMIII-3 to AP2 DNA region deleted. DNA sequence analysis of insert confirmed the identity of pHis-CBF12-DelAP2.

The deletion of the CMIII-3 sequence was obtained by introducing SmaI site after CMIII-3 in pHis-CBF12-MutCMIII-3 followed by restricting resulting plasmid with NheI and SmaI, fillingin ends using blunting enzyme mix of CloneJet PCR Cloning Kit (ThermoFisher Scientific, Waltham, MA, USA) and ligating the DNA fragments. Screening of clones obtained from transformation of ligation into DH5α cells was done by restriction enzyme digest and led to the identification of desired clone (pHis-CBF12-DelCMIII-3), which was further verified by DNA sequence analysis.

The pHis-CBF12-DelCMIII-1 construct was obtained by introducing a NheI/BmtI site after the CMIII-1 sequence followed by introducing a SmaI site before CMIII on the resulting construct. The plasmid with the two new restriction sites was digested with SmaI and BmtI in Cutsmart buffer followed by blunting the ends using blunting enzyme mix. Upon ligation of the digestion products and transformation into DH5α cells, the desired clones with CMIII-1 sequence removed were identified by restriction digests. Insert carried by the final construct pHisCBF12-DelCMIII-1 was validated by DNA sequencing.

3.8 DNA sequencing

The DNA sequence of PCR products cloned in pET-32Ek/LIC and pET-15b vectors constructed in the study were DNA sequenced by Eurofins Operon (Eurofins Genomic, Louisville, KY, USA). Sequencing reactions were done with the "power read" option to ensure good quality sequence data from the GC-rich templates. All sequencing was done with the forward T7 and the reverse T7 terminator sequencing primers (Appendix 1), which match sequences located upstream of T7

promoter and at T7 terminator site, respectively, both present on pET32-Ek/LIC and pET15b vectors (Figs. 3.1A and 3.2). The DNA sequencing data was analyzed using the SeqMan module of the DNASTAR Lasergene8 software (DNASTAR Inc., Madison, WI, USA).

3.9 Production of His-CBF12 polypeptide variants in *E. coli* cells

The assembled protein expression vectors were transformed into BL21(DE3) or BLR(DE3)/pLysS cells were inoculated onto solid SB media containing 100 µg/mL carbenicillin (BL21(DE3) transformants) or 100 µg/mL carbenicillin, 12.5 µg/mL tetracycline and 34 µg/mL chloramphenicol (BLR(DE3)/pLysS transformants) followed by incubation at 28^oC overnight. An aliquot of the cultivated cells was inoculated into 25 mL liquid SB media with antibiotics to a starting OD_{600} of about 0.05. The cultures were grown with shaking at 26^oC until OD_{600} reached 0.5, at which time IPTG was added to a final concentration of 0.5 mM to induce a two-hour overexpression of protein encoded from the expression vector. Samples were collected from the cultures immediately before (1 ml) and at the end of protein induction (0.25 ml) and centrifuged at 12,000 x g for 5 min to pellet the cells. The cell content in the samples were equalized by suspending each pellet in 50 µL SDS sample buffer (50 mM Tris-HCL, pH 6.8, 2% SDS, 10% glycerol, 5% β-mercaptoethanol) per 0.5 OD_{600} value in culture. The cell extracts were stored at 20^oC until being analyzed by SDS-PAGE.

3.10 SDS-PAGE analysis

The analysis of cell extracts was done using a GibcoBRL vertical gel electrophoresis system (ThermoFisher Scientific, Waltham, MA, USA) with 12 x 19 cm gels and 15 wells/gel. For all recombinant proteins larger than 25 kDa, the gels were prepared and run according to Laemmli protocol (Laemmli, 1970) using a standard glycine buffer (25 mM Tris-HCl, 0.25M glycine, 0.1% SDS pH 8.3) and 30:0.8 polyacrylamide:bisacrylamide crosslinking in both separating (12-15%) and stacking (5%) gels. Prior to loading, the cell extract samples in 1x SDS gel loading buffer were boiled at 95^oC for 5 min and centrifuged at 11,000 x g for 2 min before 20 µL of supernatants were applied to gel. Samples of denatured standard protein samples were also applied on each gel. The electrophoresis was typically run for a total of 180 mAh.

Analysis of proteins smaller than 25 kDa were done using either 20% (30:0.8) glycine-SDSPAGE gels as described above or 16% (30:0.8) tricine-SDS-PAGE (Schägger, 2006) gels to allow for better focusing of separated polypeptides. For the tricine-SDS-PAGE, 10 µL cells extracts were loaded and the gel was initially run at 30 V for 1 h and then at 120 V for 13 h. The buffer system employed consisted of an anode (0.1 M Tris-base, 22.5 mM HCl, pH 8.9) and a cathode (0.1 M Tris, 0.1 M tricine, 0.1% SDS, pH 8.25) buffer.

All gels were upon electrophoresis incubated in fixation solution (45% methanol, 9% glacial acetic acid) for 15 min at room temperature, and thereafter, stained for 30-45 min in fixation solution containing 0.06% (w/v) Coomassie Brilliant Blue R-250 (Sigma-Aldrich, St. Louis, MO, USA) and de-stained in 5% (v/v) acetic acid till bands became clearly visible. Imaging of gel was done using a Chemdoc Touch Imaging System (Bio-Rad, Hercules, CA, USA). The Image Lab software supplied with the instrument was used for determination of apparent molecular weights for separated polypeptides. The migration of the samples was also physically measured via a ruler and lightbox, and compared to the migration of Bio-Rad molecular weight standards (Bio-Rad, Hercules, CA, USA) on the gel.

3.11 Analysis of polypeptides via bioinformatic tools

The protein sequences of His-CBF12 variants produced in the study were analyzed via bioinformatics tools available from the following webservers: Expert Protein Analysis System (ExPASy) Bioformatics Resource Portal (Artimo et al., 2012) at the Swiss Institute of Bioinformatics, Alphalyse (https://www.alphalyse.com/), TANGO (http://tango.crg.es/) and the Protein DisOrder prediction System (PrDOS; http://prdos.hgc.jp/cgi-bin/top.cgi). The ExPASy ProtParam tool, which uses pK values of amino acids determined by analysis of polypeptide migration in an immobilized pH gradient gel environment (Bjellqvist et al. 1994) was used for calculation of the theoretical pI for each protein. Molecular weights was determined by adding the average isotopic mass for the amino acids using gpmaw-lite tool at Alphalyse website. Protein disorder was determined using the PrDOS program, which is based on the local amino acid sequence information and information from template proteins (Ishida and Kinoshita, 2007). The protein propensity to aggregate was determined by a statistical algorithm, TANGO (FernandezEscamilla et al., 2004; Rousseau et al., 2006; Linding et al. 2004) using default settings. The

TANGO analysis is based on physico-chemical principles of secondary structure formation assuming the core regions of an aggregate are fully buried. The Grand Average of Hydropathy (GRAVY) score for each protein was calculated using the ExPASy ProtParam tool and represents the sum of hydropathy values (Kyte and Doolittle 1982) of all the amino acids, divided by the number of residues in the protein sequence. The average flexibility of the proteins was calculated using the ExPASy ProtScale tool (Bhaskaran and Ponnuswamy, 1988) using a window size of five and a relative weight of window edge of 100%. The flexibility scores along the protein sequence were displayed as a two-dimensional plot.

CHAPTER 4: RESULTS

4.1 Production of expression vectors

The CBF12 gene studied was derived from *TaCBF12* carried by BAC clone 3149-L3-CBF12 (Genbank accession JF729236) isolated from a genomic library of winter wheat (*Triticum aestivum*) cv. Norstar (Ratnayaka et al., 2005). Similar to other CBF genes, *TaCBF12* lacks introns and has a prokaryotic origin (Hake et al., 2004; Rashid et al., 2012); thus, no issues were foreseen for expression of *TaCBF12* in a prokaryotic host. A protein expression vector, pTrxHisS-CBF12, encoding a full-length TrxHisS-tagged CBF12 fusion protein was prepared in an earlier study (Jain 2013). A high affinity for the CRT target (CCGAC) *in vitro* binding assays was demonstrated by purified TrxHisS-CBF12 showing that CBF12 does not require any specific eukaryotic modifications for activity (Jain 2013). The presence of the TrxHisS tag does not seem to obstruct CRT recognition as essentially the same amount of CBF12-CRT interaction is observed in the presence as in the absence of the tag. TrxHisS-CBF12 appears to be a very stable protein as it is fully functional upon exposure to boiling temperature, -80°C freezing and 6 M urea treatment (Jain 2013), which are conditions that irreversibly denature most proteins. However, an odd behavior is displayed by TrxHisS-CBF12 during SDS-PAGE; its migration is much slower than expected from its actual molecular mass. The delayed migration suggests the CBF12 polypeptide obtains a non-global structure and/or binds SDS unevenly upon heating in 2.0 % SDS buffer prior to gel electrophoresis. Thus, it was of interest in this study to determine if the anomalous gel migration could be associated with a specific CBF12 domain. If so, could this region be responsible for the high stability displayed by CBF12? To study the influence of the different CBF12 domains on protein migration during SDS-PAGE, a series of expression vectors encoding variant N-terminal, AP2 domain, AP2 signature motifs (CMIII-3 and CMIII-1) and C-terminal regions were analyzed.

Inclusion of a peptide tag (e.g. Trx, His or S tag) is of advantage if a recombinant protein produced in *E. coli* is going to be purified from a bacterial extract by affinity chromatography or be analyzed using antibodies directed towards the tag. Generally, the tag does not interfere with the rest of a fusion protein, and is therefore of little concern for functional studies. However, with regard to the produced TrxHisS-CBF12, there was a possibility that the long TrxHisS peptide (158 amino acids) could stabilize the CBF12 structure through its thioredoxin activity and/or cysteine bonds provided by the Trx peptide. To exclude such possibilities, the available pTrxHisS-CBF12

and its variant constructs were modified in this study to encode CBF12 fusion proteins with a rather short N-terminal extension (32 amino acids) including six His residues that make up the His tag.

4.2 Constructs encoding C-terminal deletions of CBF12

To observe the effect of C-terminal deletions on CBF12 migration in SDS-PAGE gels, a set of 11 plasmids expressing truncated version of CBF12 polypeptide were produced. Six of these constructs were obtained from an existing series of constructs producing full-length TrxHisSCBF12 polypeptides and variants with C-terminal truncations (Jain, 2013). These constructs and pTrxHisS-CBF12-C11 produced in this study (Fig. 4.1), were all converted to pHis-L3-CBF12 variants by one cloning step as outlined in Fig. 3.2. The generated expression constructs were denoted pHis-CBF12-C2, -C4, -C6, -C8, -C9, and –C10 (Fig. 4.1). A control plasmid, pHis-adapter (Fig. 4.1) was also made from an available pTrxHisS-adapter construct. Four additional Ctruncated variants of pHis-CBF12 were produced by introduction of stop codons into the CBF12 coding sequence carried by pHis-CBF12. These mutated pHis-CBF12 vectors were denoted pHisCBF12-C1, -C3, -C5, and –C7 (Fig. 4.1). All expression cassettes encoding C-terminal truncations of His-CBF12 were verified by DNA sequencing and deduced protein sequences of encoded polypeptides are presented in Fig. 4.2.

4.3 Constructs encoding N-terminal deletions of CBF12

It was hypothesized from previous studies that the N-terminus of CBF12 might play a more significant role in providing stability than other regions on CBF12 (Jain 2013). To test this theory, a set of four constructs encoding TrxHisS-tagged CBF12 polypeptides with 13, 22, 30 and 47 amino acids, respectively, removed from the N-terminal end of CBF12 were produced. The inserts for these vectors were generated by PCR amplifications using forward CBF12-specific primers (NF1-NF4) and a common reverse TaCBF12 end terminus primer (Appendix 1). Upon cloning the PCR amplified fragments (Fig. 4.3, lanes 2-5) into pET32-Ek/LIC vector by the Ek/LIC cloning system (Fig. 3.1), the resulting vectors were converted to pHis-CBF12 type constructs as outlined in Fig. 3.2. The resulting vectors were denoted pHis-CBF12-N1 to –N4 (Fig. 4.1) and their encoded variant CBF12 protein sequences are shown in Fig. 4.2.

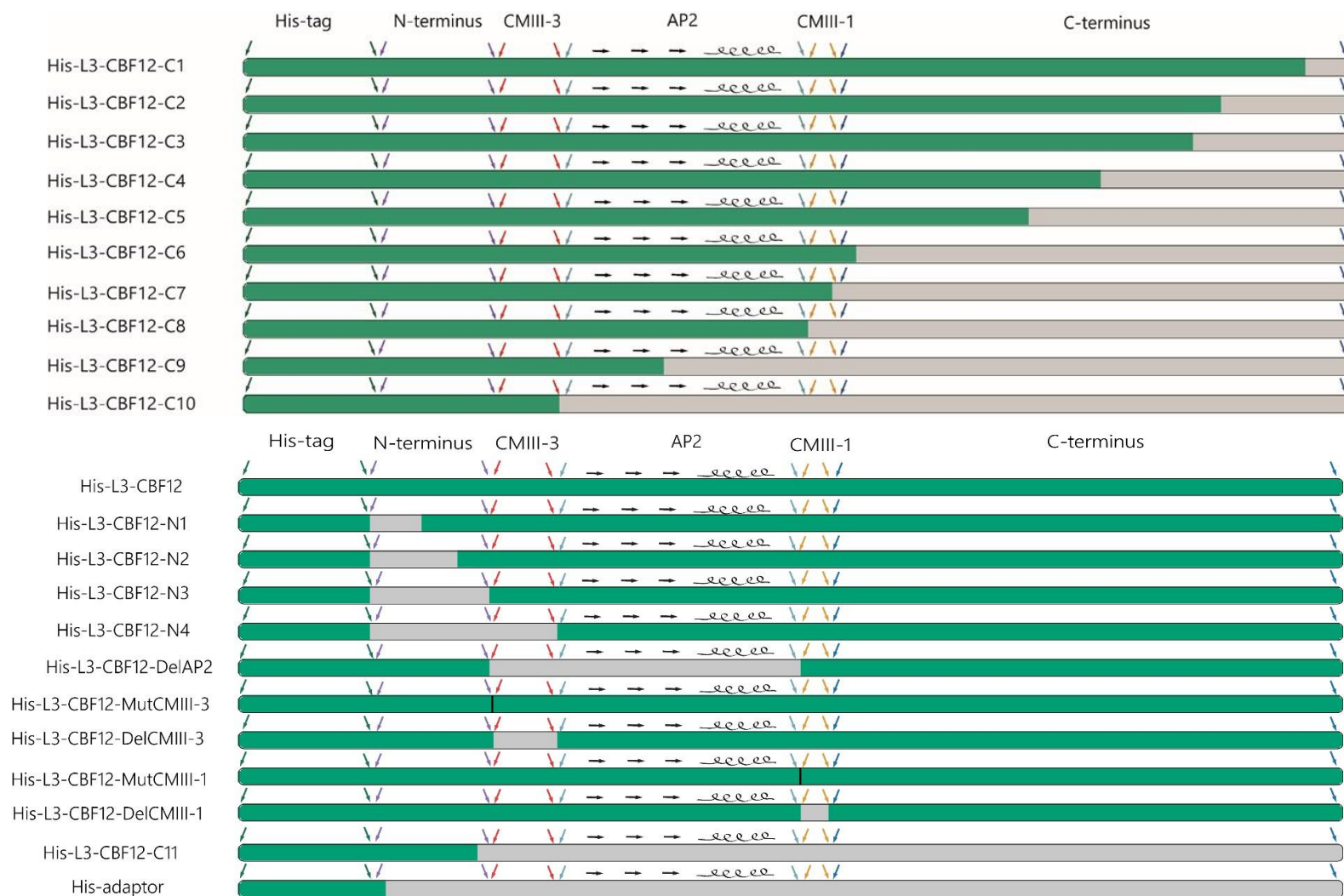**Fig. 4.1** Graphical representation of full-length and mutant His-CBF12 variants produced in *E. coli*. The location of domains carried by the full-length His-CBF12 are indicted at the top. Secondary structures within the DNA-binding AP2 domain are shown by black arrows (beta sheets) and a black coil (alpha helix). Sequences present on protein are represented by green areas, whereas grey areas indicted region missing.

His-L3-CBF12	MGSSHHHHHHSSGLVPRGSHMLEDLGTDDDDKMDTGPERNWNSPASPPSSLEQGMPTSPASPTPKRPAGRTKFKETRHPV	80
His-L3-CBF12-N1	MGSSHHHHHHSSGLVPRGSHMLEDLGTDDDDKI-------------PPSSLEQGMPTSPASPTPKRPAGRTKFKETRHPV	67
His-L3-CBF12-N2	MGSSHHHHHHSSGLVPRGSHMLEDLGTDDDDKM----------------------PTSPASPTPKRPAGRTKFKETRHPV	58
His-L3-CBF12-N3	MGSSHHHHHHSSGLVPRGSHMLEDLGTDDDDKM-------------------------------PKRPAGRTKFKETRHPV	50
His-L3-CBF12-N4	MGSSHHHHHHSSGLVPRGSHMLEDLGTDDDDKM---	33
His-L3-CBF12-DelAP2	MGSSHHHHHHSSGLVPRGSHMLEDLGTDDDDKMDTGPERNWNSPASPPSSLEQGMPTSPASAS-----------------	63
His-L3-CBF12-MutCMIII-3	MGSSHHHHHHSSGLVPRGSHMLEDLGTDDDDKMDTGPERNWNSPASPPSSLEQGMPTSPASASPKRPAGRTKFKETRHPV	80
His-L3-CBF12-DelCMIII-3	MGSSHHHHHHSSGLVPRGSHMLEDLGTDDDDKMDTGPERNWNSPASPPSSLEQGMPTSPASARG----------------	64
His-L3-CBF12-MutCMIII-1	MGSSHHHHHHSSGLVPRGSHMLEDLGTDDDDKMDTGPERNWNSPASPPSSLEQGMPTSPASPTPKRPAGRTKFKETRHPV	80
His-L3-CBF12-DelCMIII-1	MGSSHHHHHHSSGLVPRGSHMLEDLGTDDDDKMDTGPERNWNSPASPPSSLEQGMPTSPASPTPKRPAGRTKFKETRHPV	80
His-L3-CBF12-C11	MGSSHHHHHHSSGLVPRGSHMLEDLGTDDDDKMDTGPERNWNSPASPPSSLEQGMPTSPA	60
His-adaptor	MGSSHHHHHHSSGLVPRGSHMLEDLGTDDDDKMDMGS	37

His-L3-CBF12	FHGVRRRGSNGRWVCEVRVPGKRGERLWLGTHVTAEAAARAHDAAMLALYGRTPAARLNYPDSAWLLAVPSSLSDLADVR	160
His-L3-CBF12-N1	FHGVRRRGSNGRWVCEVRVPGKRGERLWLGTHVTAEAAARAHDAAMLALYGRTPAARLNYPDSAWLLAVPSSLSDLADVR	147
His-L3-CBF12-N2	FHGVRRRGSNGRWVCEVRVPGKRGERLWLGTHVTAEAAARAHDAAMLALYGRTPAARLNYPDSAWLLAVPSSLSDLADVR	138
His-L3-CBF12-N3	FHGVRRRGSNGRWVCEVRVPGKRGERLWLGTHVTAEAAARAHDAAMLALYGRTPAARLNYPDSAWLLAVPSSLSDLADVR	130
His-L3-CBF12-N4	FHGVRRRGSNGRWVCEVRVPGKRGERLWLGTHVTAEAAARAHDAAMLALYGRTPAARLNYPDSAWLLAVPSSLSDLADVR	113
His-L3-CBF12-DelAP2	---DSAWLLAVPSSLSDLADVR	82
His-L3-CBF12-MutCMIII-3	FHGVRRRGSNGRWVCEVRVPGKRGERLWLGTHVTAEAAARAHDAAMLALYGRTPAARLNYPDSAWLLAVPSSLSDLADVR	144
His-L3-CBF12-DelCMIII-3	FHGVRRRGSNGRWVCEVRVPGKRGERLWLGTHVTAEAAARAHDAAMLALYGRTPAARLNYPDSAWLLAVPSSLSDLADVR	160
His-L3-CBF12-MutCMIII-1	FHGVRRRGSNGRWVCEVRVPGKRGERLWLGTHVTAEAAARAHDAAMLALYGRTPAARLNYPGSAWLLAVPSSLSDLADVR	160
His-L3-CBF12-DelCMIII-1	FHGVRRRGSNGRWVCEVRVPGKRGERLWLGTHVTAEAAARAHDAAMLALYGRTPAARLNYP-------LPSSLSDLADVR	153

His-L3-CBF12	RAAIGAVVDFLRRQEAGASAGAVAEEAHVDGIASAASAPDNASSSAAAAHSQPPCANAGYEVPDALCHDMFELHTSGEMD	240
His-L3-CBF12-N1	RAAIGAVVDFLRRQEAGASAGAVAEEAHVDGIASAASAPDNASSSAAAAHSQPPCANAGYEVPDALCHDMFELHTSGEMD	227
His-L3-CBF12-N2	RAAIGAVVDFLRRQEAGASAGAVAEEAHVDGIASAASAPDNASSSAAAAHSQPPCANAGYEVPDALCHDMFELHTSGEMD	218
His-L3-CBF12-N3	RAAIGAVVDFLRRQEAGASAGAVAEEAHVDGIASAASAPDNASSSAAAAHSQPPCANAGYEVPDALCHDMFELHTSGEMD	210
His-L3-CBF12-N4	RAAIGAVVDFLRRQEAGASAGAVAEEAHVDGIASAASAPDNASSSAAAAHSQPPCANAGYEVPDALCHDMFELHTSGEMD	193
His-L3-CBF12-DelAP2	RAAIGAVVDFLRRQEAGASAGAVAEEAHVDGIASAASAPDNASSSAAAAHSQPPCANAGYEVPDALCHDMFELHTSGEMD	162
His-L3-CBF12-MutCMIII-3	RAAIGAVVDFLRRQEAGASAGAVAEEAHVDGIASAASAPDNASSSAAAAHSQPPCANAGYEVPDALCHDMFELHTSGEMD	240
His-L3-CBF12-DelCMIII-3	RAAIGAVVDFLRRQEAGASAGAVAEEAHVDGIASAASAPDNASSSAAAAHSQPPCANAGYEVPDALCHDMFELHTSGEMD	224
His-L3-CBF12-MutCMIII-1	RAAIGAVVDFLRRQEAGASAGAVAEEAHVDGIASAASAPDNASSSAAAAHSQPPCANAGYEVPDALCHDMFELHTSGEMD	240
His-L3-CBF12-DelCMIII-1	RAAIGAVVDFLRRQEAGASAGAVAEEAHVDGIASAASAPDNASSSAAAAHSQPPCANAGYEVPDALCHDMFELHTSGEMD	233

His-L3-CBF12	AGTYYADLAQGLLLEPPPPPSSGASSERGDDAALWNH	277
His-L3-CBF12-N1	AGTYYADLAQGLLLEPPPPPSSGASSERGDDAALWNH	264
His-L3-CBF12-N2	AGTYYADLAQGLLLEPPPPPSSGASSERGDDAALWNH	255
His-L3-CBF12-N3	AGTYYADLAQGLLLEPPPPPSSGASSERGDDAALWNH	247
His-L3-CBF12-N4	AGTYYADLAQGLLLEPPPPPSSGASSERGDDAALWNH	230
His-L3-CBF12-DelAP2	AGTYYADLAQGLLLEPPPPPSSGASSERGDDAALWNH	199
His-L3-CBF12-MutCMIII-3	AGTYYADLAQGLLLEPPPPPSSGASSERGDDAALWNH	277
His-L3-CBF12-DelCMIII-3	AGTYYADLAQGLLLEPPPPPSSGASSERGDDAALWNH	261
His-L3-CBF12-MutCMIII-1	AGTYYADLAQGLLLEPPPPPSSGASSERGDDAALWNH	277
His-L3-CBF12-DelCMIII-1	AGTYYADLAQGLLLEPPPPPSSGASSERGDDAALWNH	270

Fig. 4.2 Detailed sequence information for full-length and mutated His-CBF12 variants produced in E. coli. Residues in color indicate sequence deviations from His-CBF12 sequence shown at the top. Dotted lines represent missing amino acid residues.
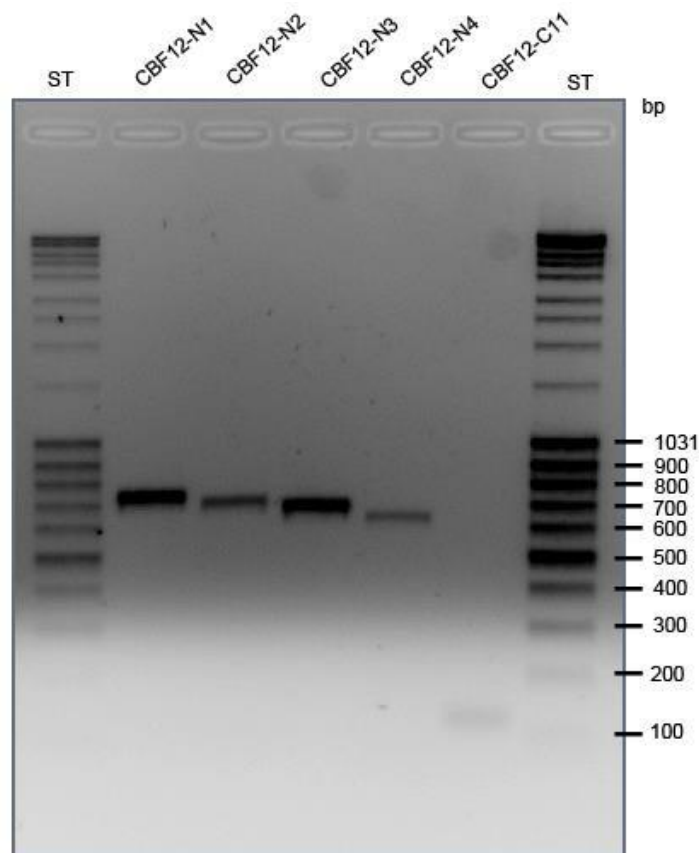
Fig. 4.3 Analysis of PCR amplified fragments by 1% agarose gel electrophoresis. Samples (10 µL) of PCR products generated to encode deletions within CBF12 coding sequence were analyzed. Products encoding N-terminal deletions to CBF12 were analyzed in lanes 2-5, and product producing a C-terminal deletion to CBF12 was analyzed in lane 6. The MassRuler DNA ladder mix (ThermoFisher Scientific (Waltham, MA, USA) was loaded in lanes 1 and 7.

4.4 Constructs encoding mutations to signature motifs and AP2 domain of CBF12

Aside from implicating the N-terminal or the C-terminal regions of CBF12, constructs testing the influence of the DNA-binding AP2 domain and signature motifs flanking AP2 (CMIII-3 and CMIII-1; Fig. 4.1) on His-CBF12 migration during SDS-PAGE were made. As modifications of these domains could not be done by direct cloning, novel restriction recognition sites had to be introduced at sites bordering the targeted sequence motifs. Through the use of two different mutagenesis strategies and primers including a restriction site (Table 3.2), pHis-CBF12 variants were produced with novel *NheI/Bmt1* or *SmaI* sites. Fig. 4.4 demonstrates successful PCR mutagenesis using GeneArt site-directed mutagenesis to generated five different linearized versions of the pHis-CBF12 vector with different sites mutated. Upon circulation of mutated vectors *in vivo*, constructs with *NheI* sites introduced prior to CMIII-3 (pHis-CBF12-MutCMIII-3 with encoded P_{62}-T_{63} before CMIII-3 motif is altered to A_{62}-S_{63}; Figs. 4.1 and 4.2), after CMIII-1 (pHis-CBF12-MutCMIII-1 with encoded L_{146}-A_{147} after CMIII-1 motif is altered to R_{146}-$Stop_{147}$; Figs. 4.1 and 4.2), respectively, were produced. Similarly, *SmaI* sites were introduced after CMIII-3 (pHis-L3-MutCMIII-3 with encoded V_{80} altered to a G_{80}) and before CMIII-1 (pHis-CBF12MutCMIII-1 with DSAWL of motif altered to GSAWL; Figs. 4.1 and 4.2), respectively. Through a subsequent sub-cloning procedure utilizing the generated *NheI* and *SmaI* sites, constructs encoding His-CBF12 lacking CMIII-3 (pHis-CBF12-DelCMIII-3), CMIII-1 (pHis-CBF12DelCMIII-1) or both CMIII-3 and AP2 domain (pHis-L3-CBF12-DelAP2) were generated (Figs. 4.1 and 4.2).

4.5 Expression of recombinant CBF proteins in *E. coli*

A graphical representation and detailed protein sequence information for all His-tagged CBF12 variants produced is given in Figs. 4.1 and 4.2. The growth and induction of protein expression was done following previously established protocol with minor adjustments (Jain 2013). As demonstrated earlier, growth in SB media at a relatively low growth temperature (26°C), use of a moderate amount of IPTG (0.5 mM) for protein induction for two hours was appropriate for detecting induced proteins in cell extracts by SDS-PAGE. The growth curves for the different cultures analyzed were very similar both before and after induction with IPTG (Fig. 4.5), suggesting overproduction of the recombinant proteins did not induce toxic effects on the cells.

41
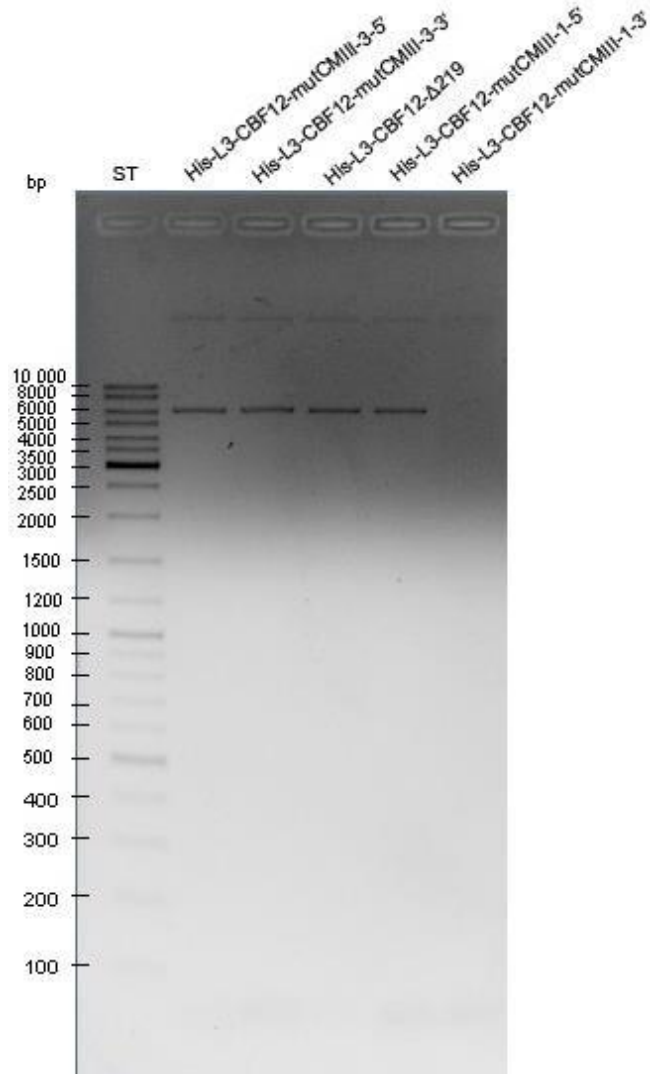
Fig. 4.4 Analysis of PCR amplicons generated by GeneArt site-directed mutagenesis. 1% agarose gel showing migration of expected 6,446 bp PCR product with site affected indicated was analyzed in lanes 2-6. Samples loaded on gel contained 10 μL PCR product. His-CBF12-CMIII-1R product was not visible on the gel due to a low, but usable, amount DNA in the sample. The Generuler 1kb DNA ladder (ST) was used as standard.
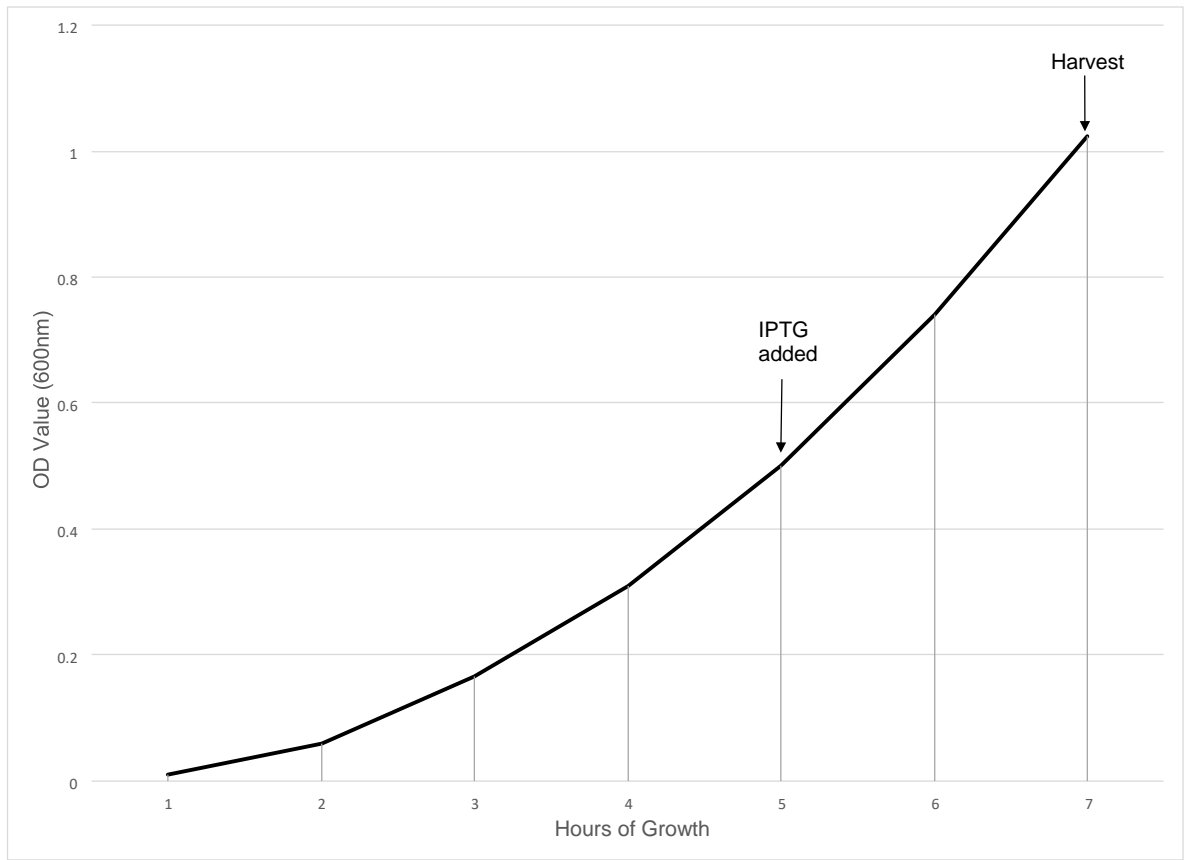
Fig. 4.5 Growth curve for BL21(DE3)pLysS cells over-expressing His-CBF12. Time points for IPTG induction and cell harvest are indicated.

Successful induction of recombinant protein production was obtained for all CBF12 fusion, constructs, but variations in the amount of fusion protein produced was observed. Cultures producing His-CBF12-C6, C7 and C8 were less reproducible than the other cultures due to weak protein bands on gels. Typically, gels containing these fusion proteins needed to be de-stained for a long time for polypeptide bands to be visible. The low protein accumulation during induction may indicate that these proteins were less stable in *E. coli* than the other recombinant proteins produced. Due to the difficulties experienced with the three constructs, fewer gels were analyzed for His-CBF12-C8 (four gels), -C10 (three gels), and –C11 (three gels) as compared to the other fusion protein (five gels). The full-length His-L3-CBF12 used as control in most gels was replicated eleven times.

4.6 Effect of CBF12 mutations on protein migration during SDS-PAGE

The anomalous migration of TrxHisS-CBF is observed for both purified protein and when present in total *E. coli* cell lysates (Gangopadhyay, 2015, Jain et al. 2013), which shows *E. coli* proteins do not affect TrxHisS-CBF12 migration during SDS-PAGE. This conclusion has also been made for His-CBF12 fusions. Thus, gel migration for the various His-CBF12 proteins was done by analysis of total *E. coli* cell lysates in this study. This choice was also made considering the difficulties in obtaining sufficient quantities of soluble TrxHisS-CBF12 required for protein purification by affinity chromatography (Gangopadhyay et al. 2015, Jain et al. 2013) and HisCBF12.

The SDS-PAGE gel analysis of the full-length protein, His-CBF12, in induced *E. coli* cell extract showed, like the previously reported TrxHisS-CBF12, slow gel migration (Fig. 4.6A). With a theoretical molecular weight of 29.3 kDa, His-CBF12 migrated consistently with an apparent molecular weight of about 41.3 kDa, thus 41.1% slower than expected. His-CBF12 proteins with truncations at the N-terminal end (His-CBF12-N1 to -N4; Fig. 4.1) migrated faster than the fulllength His-CBF12 due to smaller masses (Fig. 4.6), but the sequential decrease in apparent molecular weight for all four truncated proteins correlated with the amount of amino acids removed from each polypeptide. Thus, both His-CBF12-N1 and N2 were virtually identical to His-CBF12 in their degree of anomalous migration (Fig. 4.7). Migration of His-L3-CBF12-N3 and His-L3CBF12-N4 were substantially more variable, but not significantly different from the full-length fusion protein. Changes in the distribution of the variation were also inconsistent, with His-L3-
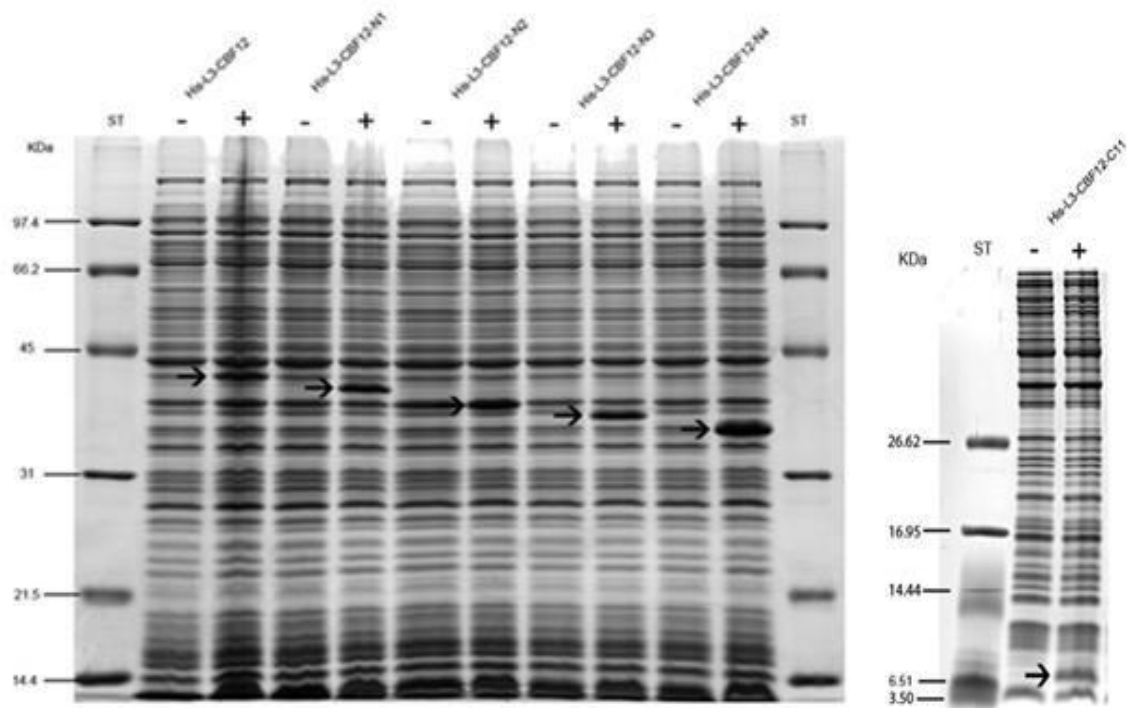
Fig. 4.6 Effect of CBF12 N-terminal truncations on protein mobility in SDS-PAGE gel. Coomassie Blue-stained protein gel with total *E. coli* cell extract samples from un-induced (-) and two-hour induced cultures (+) cultures analyzed. (A-Left Fig.) 12% SDS-PAGE gel with samples from cultures producing His-CBF12 and four of its N-terminal truncations of CBF12 (N1 to –N4). (B-Right Fig.)16.0% Tricine SDS PAGE gel with samples from culture producing His-CBF12-C11. All samples analyzed originated from equal amounts of cells. Arrows indicate migration of induced fusion proteins. The protein standards (ST) used were (A) Broad Range SDS-PAGE Molecular Weight Standards (Bio-Rad, Hercules, CA, USA) and (B) Bio-Rad Polypeptide SDS-PAGE Molecular Weight Standards.
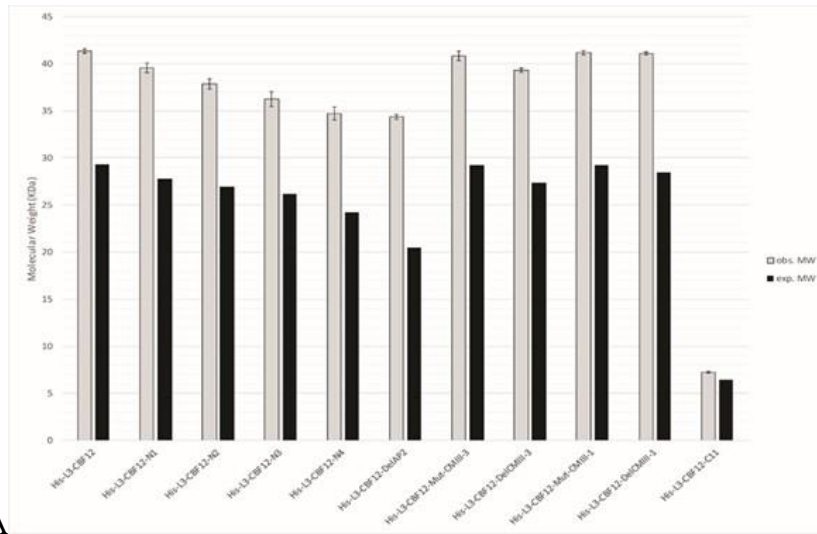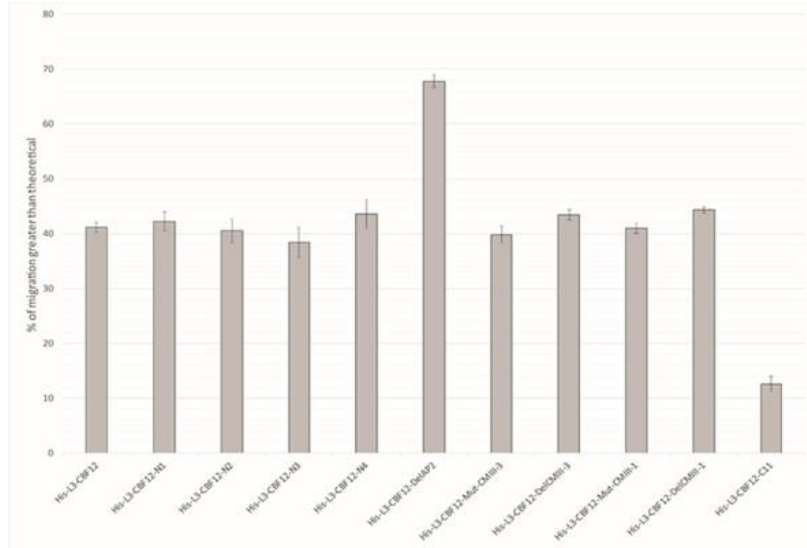
Fig. 4.7 Effects of N-terminal, CMIII and AP2 mutations on CBF12 gel migration. A. Histogram showing differences between molecular weights determined from SDS-PAGE gels and theoretical molecular weights. Error bars represent the 95% confidence interval for each construct, un-pooled as a population. (B) Histogram showing percent difference between apparent and theoretical molecular weights.

CBF12-N3 decreasing in mean degree of anomalous migration and His-L3-CBF12-N4 instead increasing. When the anomalous migration is corrected for, the measured molecular weight differed from the expected value by, at most, only two percent, a difference that lies within the error of protein mass determination by SDS-PAGE. In summary, the removal of the N-terminus sequences from the first amino acid of CBF12 to the AP2 domain could not be correlated with delayed gel migration for His-CBF12.

To further test the effect of the N-terminal sequences on gel migration, a His-CBF12 variant consisting only of the N-terminal sequences of CBF12 (pHis-CBF12-C11; Fig. 4.1) was analyzed. This construct proved challenging to visualize on SDS-PAGE due to its low molecular weight (6.45 kDa predicted from amino acid sequence). His-CBF12-C11 did not separate properly on glycine SDS-PAGE gels (Laemmli 1970), even at higher acrylamide percentages, and was only visible as a smear on the gel (data not shown). Thus, for this protein, a Tricine SDS-PAGE system (Schägger, 2006) was used for analysis. This gel system has the greatest resolution for proteins less than 30 kDa of the available gel-based electrophoretic systems. Due to the lower stacking limit provided by the Tricine SDS-PAGE system as compared to glycine SDS-PAGE, the separation of His-CBF12-C11 from other protein bands on the gel was possible (Fig. 4.6B).

On a 16% Tricine gel, His-CBF12-C11 was found to run only 12.5% slower than expected and showed the best agreement between apparent and predicted molecular weight for all recombinant proteins analyzed in the study (Figs. 4.6, 4.7). Although the confidence intervals and standard deviation for His-CBF12-C11 did not overlap with the expected molecular weight, yet 1-4 kDa could be considered within the error of size determination for proteins separated by SDS-PAGE (Jain, 2013). Exact determination was also hampered by His-CBF12-C11 being visualized near the bottom of the 16% tricine SDS-PAGE gels, leading to a greater possible error of determination. Nevertheless, compared to the other constructs producing much larger delays in migration, HisCBF12-C11 gel behavior strongly indicated that the N-terminus contributed little to the gel shifting phenomenon.

The CMIII-3 (PKRPAGRTKFKETRHP) and CMIII-1(DSAWL) flanking motifs distinguish CBFs from other AP2 domain proteins (El Kayal et al., 2006). These motifs were in conjunction with the N-terminal region indicated to reduce CBF12 gel migration in previous studies (Gangopadhyay, 2015, Jain et al. 2013). However, a comparison of His-CBF12- N3 and His-CBF12-N4 gel migration indicted no role for CMIII-3 in delaying protein migration (Fig. 4.6).

This was further emphasized by analysis of His-CBF12-MutCMIII-3, for which two amino acids prior to CMIII-3 motif were altered; nevertheless, the mutated protein migrated approximate to full-length His-CBF12 (Fig. 4.8). Nor could an alteration to the first residue of CMIII-1 (HisCBF12-MutCMIII-1) alter the protein gel migration pattern (Fig. 4.8). Further on, analysis of proteins variants lacking CMIII-3 (His-CBF12-DelCMIII-3) or CMIII-1 (His-CBF12-DelCMIII3) sequences, respectively, indicated no role for any of the two CMIII motifs in delaying gel migration (Fig. 4.7, 4.8) Rather, a slight increase in migrational delay was noted for both HisCBF12-DelCMIII-3' (+2.3%) and His-CBF12-DelCMIII-1 (+3.3%). Unlike His-CBF12-N4, these small increases in gel migration delay seen for both CMIII deletion constructs were found to be significant when compared to His-CBF12. Contrary to what was predicted, the removal of the CMIII-3 and CMIII-1 motifs exuberated rather than diminished the anomalous gel migration, although they reflected little the observed gel shifting associated with His-CBF12 as a whole.

In contrast to the subtle effects noted for His-CBF12 variants lacking CMIII motifs, removal of both the CMIII-3 motif and AP2 domain (His-CBF12-DelAP2) displayed a more drastic effect (Fig. 4.8). Protein resulting from this deletion migrated 67.8% slower than estimated, 20% slower than and eighteen standard deviations above the mean of the His-L3-CBF12 gels (Figs. 4.7). The dramatic increase in the degree of anomalous migration associated with the removal of the AP2 region suggested that the DNA-binding domain exerts a stabilizing effect on the CBF12 structure. Based on these results, it was concluded that the N-terminal and the DNA binding domain do not contribute on the anomalous migration observed for CBF12. At worst, their removal serve only to exuberate the effect.

All C-terminal truncations of His-CBF12 yielded proteins that were more variable in their abundance in cell extracts (Fig. 4.9 to 4.10) compared to proteins with deletions affecting the Nterminal region, CMIII motifs or the AP2 domain. For C-terminal truncations extending beyond His-L3-CBF12-C6, the visualization of overexpressed polypeptides on a 12% SDS-PAGE gel proved challenging. These small proteins migrated near the edge of the gel and had a tendency to smear, which caused higher variation in their migration between gels. On a standard 12% SDSPAGE gel, molecular weight determinations for His-L3-CBF12-C6 had a standard error of 3.1, but was decrease to 1 by using the 20% SDS-PAGE gels providing higher resolution of separated proteins (Figs. 4.9, 4.10). An exception to this was His-CBF12-C7, which was still successfully visualized on a 12% SDS-PAGE gel, at a standard error of 1.9 (Fig. 4.9).
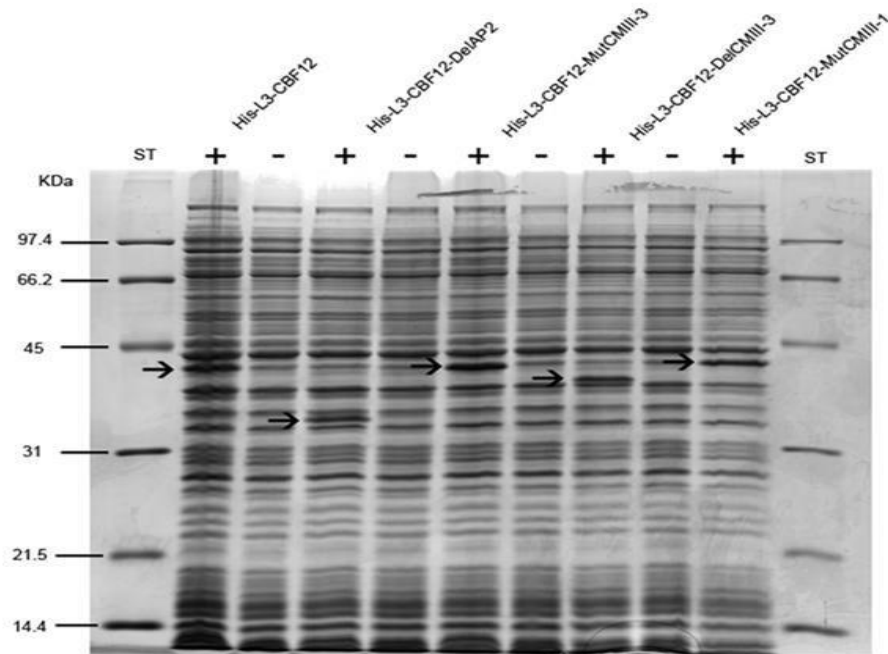
Fig. 4.8 Effects of mutations affecting CMIII and AP2 domains on His-CBF12 gel migration. Coomassie Blue-stained 12% SDS-PAGE gel of samples containing *E. coli* cell extracts from uninduced (-) and two-hour (+) induced cultures as indicated above. All samples originated from equal amounts of cells. Arrows indicate migration of induced fusion proteins. Migration of the Broad Range SDS-PAGE Molecular Weight Standards (Bio-Rad, Hercules, CA, USA) is shown to the left.
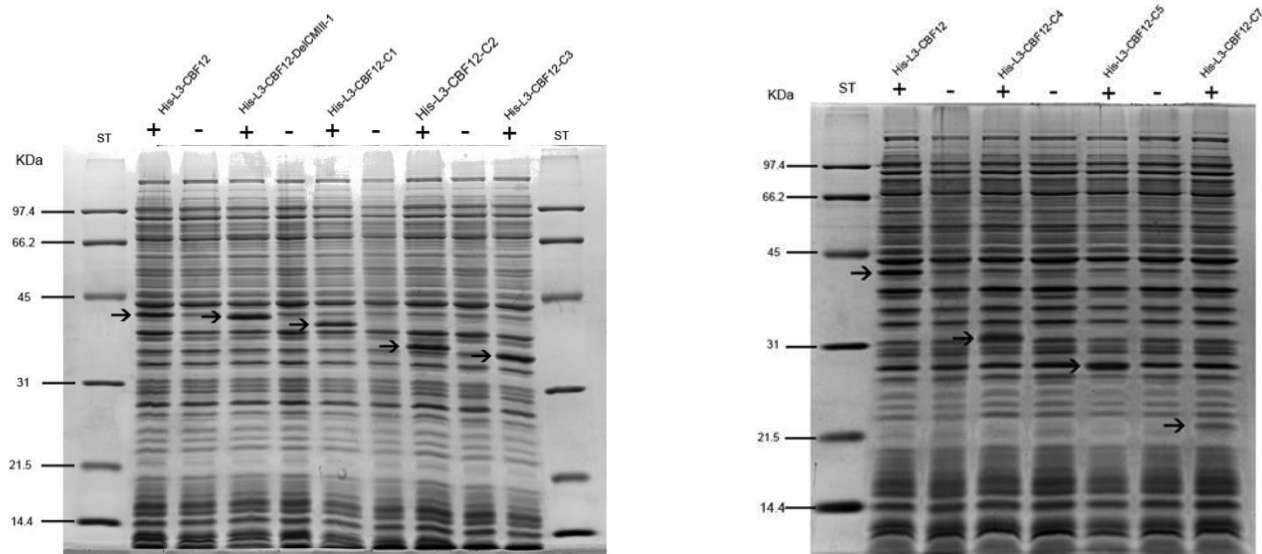
Fig. 4.9. Gel migration displayed by His-CBF12 with CMIII-1 deleted or lacking different segments of the C-terminal region. Coomassie Blue-stained 12% SDS-PAGE gel of samples containing *E. coli* cell extracts from un-induced (-) and two-hour (+) induced cultures as indicated above. All samples originated from equal amounts of cells. Arrows indicate migration of induced fusion proteins. Migration of the Broad Range SDS-PAGE Molecular Weight Standards (Bio-Rad, Hercules, CA, USA) is shown to the left.
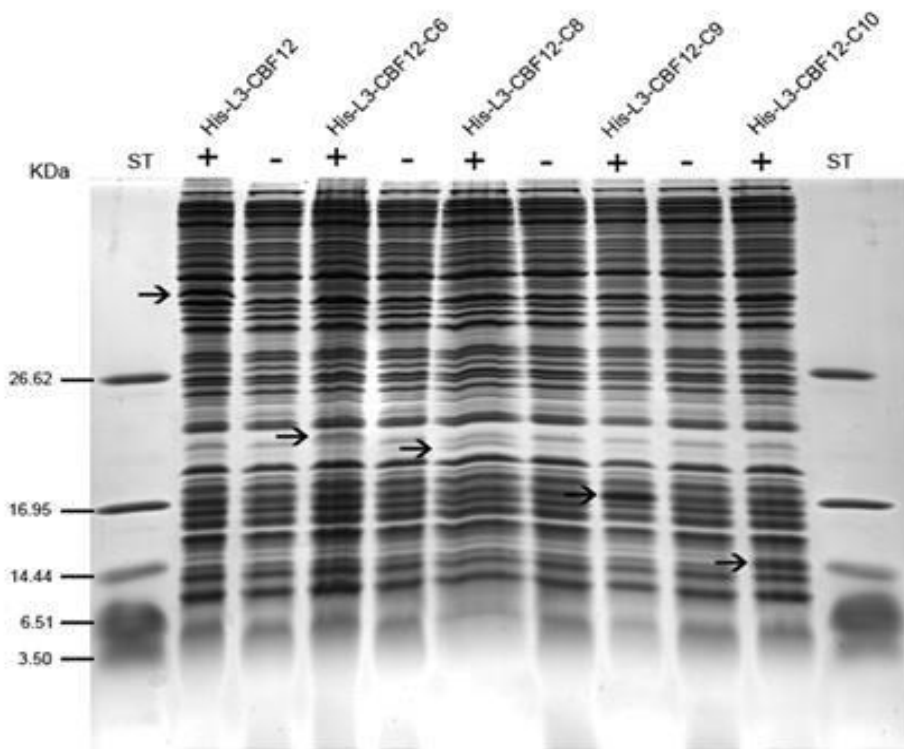
Fig. 4.10 Glycine SDS-PAGE (20%) analysis of His-CBF12 with C-terminal truncations. Coomassie Blue-stained 20% SDS-PAGE gel of His-CBF12-C6, -C8, -C9 and -C10 from bacterial lysates. Samples indicated were from un-induced (–) and two-hour (+) induced cells. An equal amount of cells was analyzed in each sample. Arrows indicate migration of induced protein. St: Biorad Polypeptide SDS-PAGE Molecular Weight Standards.

The last 40 residues on CBF12 did not seem to cause any significant delay of gel migration as His-CBF12-C1 through to C3 showed delay of migration (around +40%) as the full-length HisCBF12 (Figs. 4.9, 4.10, 4.11). Additional truncations of the C-terminal region (His-CBF12-C4 through to His-CBF12-C6) gradually reduced to 36.8% (-C4), 34.5% (-C5) and 29.9% (-C6) (Figs. 4.9, 4.10). The data suggests this 61-residue region leading up to the AP2 domain has some influence on slowing down CBF12 movement through the gel.

While the degree of anomalous migration was lower for His-CBF12-C5 and –C6, an additional 43-residue reduction of C-terminal sequence (His-CBF12-C7) increased the anomalous migration back to that of the full-length His-CBF12 (~40%). The removal of the following CMIII-1 motif (His-CBF12-C8) resulted in an observed migration delay similar to His-CBF12-C6 (Fig. 4.10, 4.11). His-CBF12-C9, where half of the AP2 domain was removed, exhibited a slightly higher anomalous migration (46.2%) to the full-length recombinant protein (Fig. 4.10, 4.11). Removing the entire AP2 domain, leaving just the N-terminus and the CMIII-3 motif intact for His-CBF1210, dramatically increased the delay to 68.6% slower, matching that of His-CBF12- DelAP2. This further demonstrated that the structure of the AP2 domain counteracts anomalous migration. In regard to the C-terminus itself, it was found that it does contribute to gel shifting by delaying protein migration during SDS-PAGE, however it is not a major contributor.

4.7 Comparison to previous study of CBF12 gel migration

His-CBF12-C1, -C2, -C4, -C5 and -C8 proteins analyzed in this study were previously analyzed as TrxHisS-tagged proteins (Jain 2013). The TrxHisS tag was used to alleviate aggregation and formation of inclusion bodies in *E. coli*, which is a concern if the recombinant protein is going to be purified from the bacterial lysate. In this context it is interesting to note that aggregation-prone proteins like CBF12 in general exhibit various degree of anomalous migration during SDS-PAGE (Amani and Naeem, 2014, 2013; Rath et al., 2013; Rath and Deber, 2013). A comparison of the migration pattern for His-tagged and TrxHisS-tagged CBF12 with the same Cterminal truncations showed similar reductions in gel migration as more sequences from the Cterminal end of CBF12 were removed. For both types of CBF12 fusions, the bulk of the gel shifting, however would not be alleviated by C-terminal truncations. A more dramatic effect on gel migration was demonstrated in truncating through the CMIII-1 motif and the AP2 domain was demonstrated in this study, but these domains could not be associated with slow gel migration.
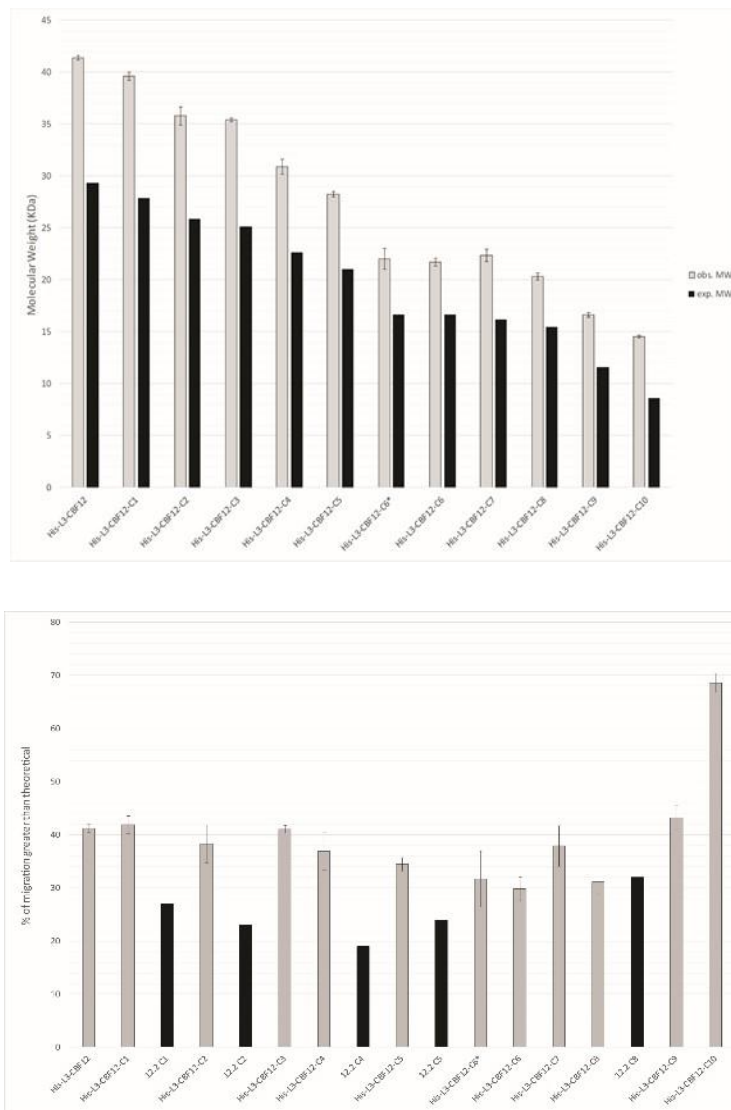
Fig. 4.11 Effect of His-CBF12 C-terminal deletions of on gel migration. A. Histogram showing differences between molecular weights determined by SDS-PAGE and theoretical molecular weights. Error bars represent the 95% confidence interval for each construct, unpooled as a population. (B) Histogram showing percent difference between apparent and theoretical molecular weights. *migration of corresponding TrxHisS-tagged CBF12 variant data (Jain 2013).

Attempts to determine the gel migration for the tag alone were made but the small polypeptide size (calculated molecular mass of 4.05 kDa) may have eluded detection by Tricine-SDS-PAGE, which has a strong resolution limit of 5 kDa (Schägger and von Jagow, 1987). As the TrxHisS-tag in isolation migrates closely to expected molecular weight (Jain 2013), the His-tag present of proteins analyzed in this study was not expected to cause the slow gel migration.

4.8 *In silico* analysis of the constructs

In analyzing the proteins expressed by the various constructs with SDS-PAGE, it was found that no construct eliminated the anomalous migration. The element which indicated a definite role was the C-terminus, whereby its reduction resulted in gel shifting decreasing by 11%. Other elements, such as the AP2 domain, proved to be detrimental when removed and increased the degree of anomalous migration. Thus, other properties of the protein which were not affected by the constructs generated so far must be considered or what facets were significantly altered by a specific construct. Gel shifting can be caused by numerous factors, such as possessing a higher ratio of acidic or charged residues, the presence of dense hydrophobic groups, post-translational modification, surface distribution of charged residues, having elongated structures, being intrinsically disordered and having a lower isoelectric point (Iwashita et al., 2015; C. Lee et al., 2014; Rath et al., 2013, 2009; Takano et al., 1988). To explore these facets, the sequences of the generated constructs were analyzed using freely available bioinformatic tools and protein calculators, such as the ProtParam and ProtScale tools on the ExPASy portal for molecular weight, theoretical pI, theoretical protein disorder (PrDOS), and grand average of hydropathy (GRAVY) score and predicted aggregation potential TANGO for. Further analysis into aspects involving secondary structure were not attempted. It was presumed that dense hydrophobic regions, like the AP2 domain and its flanking motifs, would be the principle determinant for gel shifting. This is due to residues being made inaccessible for binding with SDS due to hydrophobic packing, both by the action of hydrophobic secondary structures and general protein aggregation. It was found that these factors did not correlate with what was observed for migrational behavior. Based upon TANGO, proteins which were noted to have altered migrational behavior compared to the wild type had lower predicted aggregation scores, such as His-CBF12-DelAP2 (Fig. 4.12). This was also true with His-CBF12-C10, which had an aggregation score of 0, yet it is one of the slowest migrating constructs. Other constructs with the same aggregation score, His-L3-CBF12-C9 and
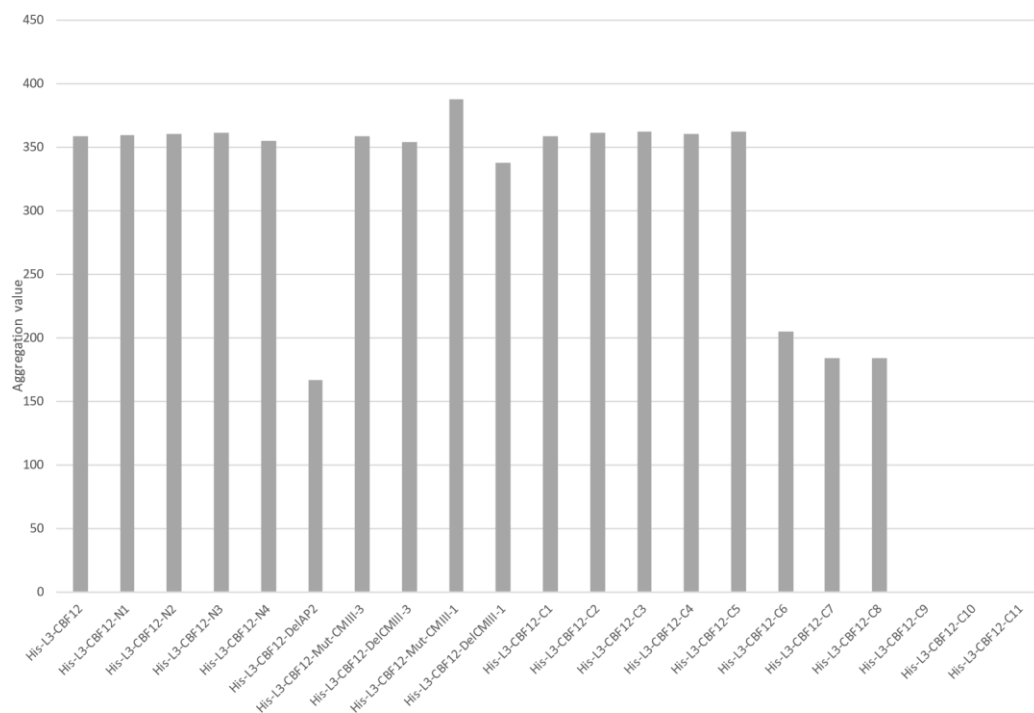
Fig. 4.12 Aggregation scores for the His-CBF12 variants. The aggregation propensities of the hydrophobic secondary structures present in the protein were calculated according to TANGO (http://tango.crg.es), assuming that all core regions are completely buried. Aggregation value is the sum of the percentages of the statistical weight for each residue to be aggregation-prone. A higher value for a polypeptide indicates a greater likelihood to form inclusion bodies.

C11, both migrated faster than His-L3-CBF12-C10. Likewise, hydrophobicity as predicted by GRAVY score did not correlate with the observed migration. An increase in hydrophilicity to 1.37 from -0.54 during the removal of the C-terminus was the only change elicited (Fig. 4.13). Therefore, hydrophobicity and associated secondary structure likely do not appear to play a role in gel shifting for CBF12 or actively work against it.

During SDS-PAGE, normalizing the charges of the protein is crucial for accurately determining the molecular weight of the protein. Acidic proteins migrate slower during SDSPAGE than expected, due to interfering with the binding of SDS to the protein (Alonso et al., 2002; Alves et al., 2004; Armstrong and Roman, 1993; Garcia-Ortega et al., 2005). The pI of CBF12 is 5.94, not notably acidic enough to warrant it as a cause. Removal of the N-terminus and the CMIII3 and CMIII-1 motifs did not alter the calculated pI value significantly (Fig. 4.14). The pI steadily became more basic as the C-terminus was removed. This difference was substantial between the constructs His-CBF12-C4 to His-CBF12-C9, ranging from 8.07 to 10.1 (Fig. 4.14). Constructs C10 and C11 however reverted to within the range of His-CBF12, at pI values of 6.69 and 5.42 respectively (Fig. 4.14). There is little correlation to be found with this trend compared to the observed migrations of the constructs. While the increase in pI matches the reduction in migrational delay from His-CBF12-C4 to His-CBF12-C8, His-CBF12-C9 rebounds to 43.2% despite a pI of 9.84. The exception to this was His-CBF12-DelAP2. This construct was more acidic compared to the wild type, at a pI of 4.45 (Fig 4.15). This decrease in pI is correlated with a significant increase in the observed migrational delay, at 67.8% slower. While it is the only construct to become more acidic, the change is not enough for His-CBF12-DelAP2 to be considered equal to an acidic protein. Distribution of charge is more telling of the impact of this change. Half of the basic residues for CBF12 are situated within the AP2 domain and CMIII-3 motif (Fig. 4.15). Loss of these residues and the influence of acidic residues becoming more dominant may explain the observed gel shift.

In the previous study done by Jain (2013), it was found that the C-terminus of the L3-CBF12 protein was characterized as elongated via circular dichromism. This elongation was postulated as playing a role in the observed anomalous migration, though it does not explain the full phenomenon. In light of this, analysis was done to determine the average flexibility of each of the constructs, using bioinformatics tools. Flexibility itself does not correlate directly with anomalous migration but extended and resilient elongated regions certainly can contribute.
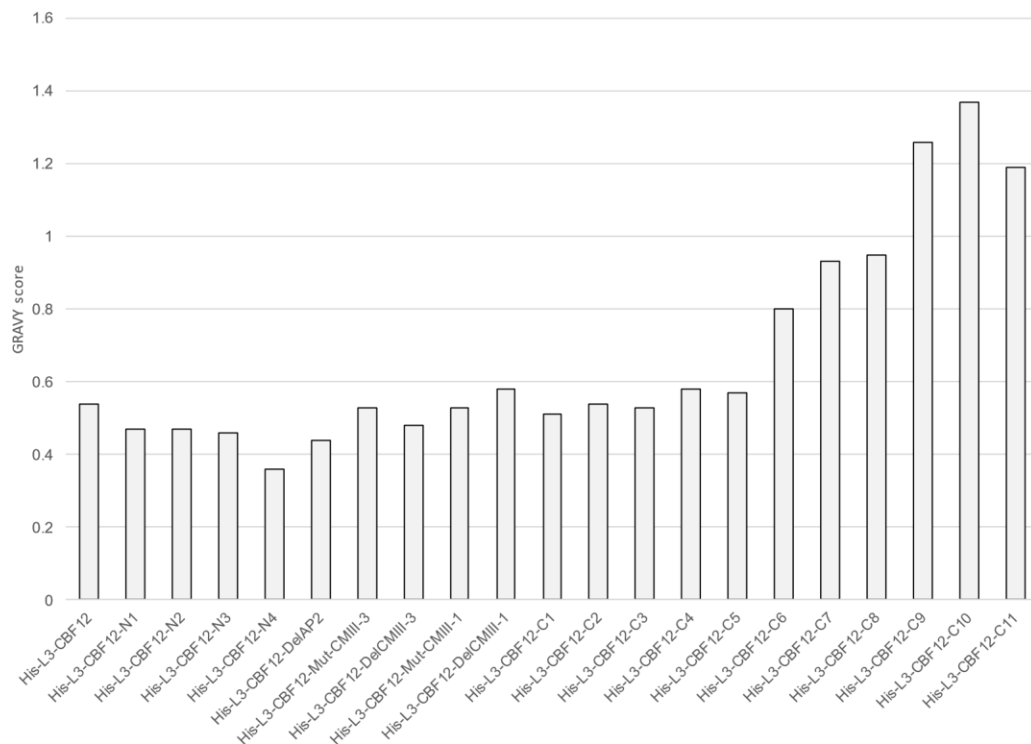
Fig. 4.13 GRAVY (Grand average of hydropathy) scores for the His-CBF12 variants. All GRAVY scores indicated here are negative in value and a higher score corresponds to an increase in hydrophilic properties.
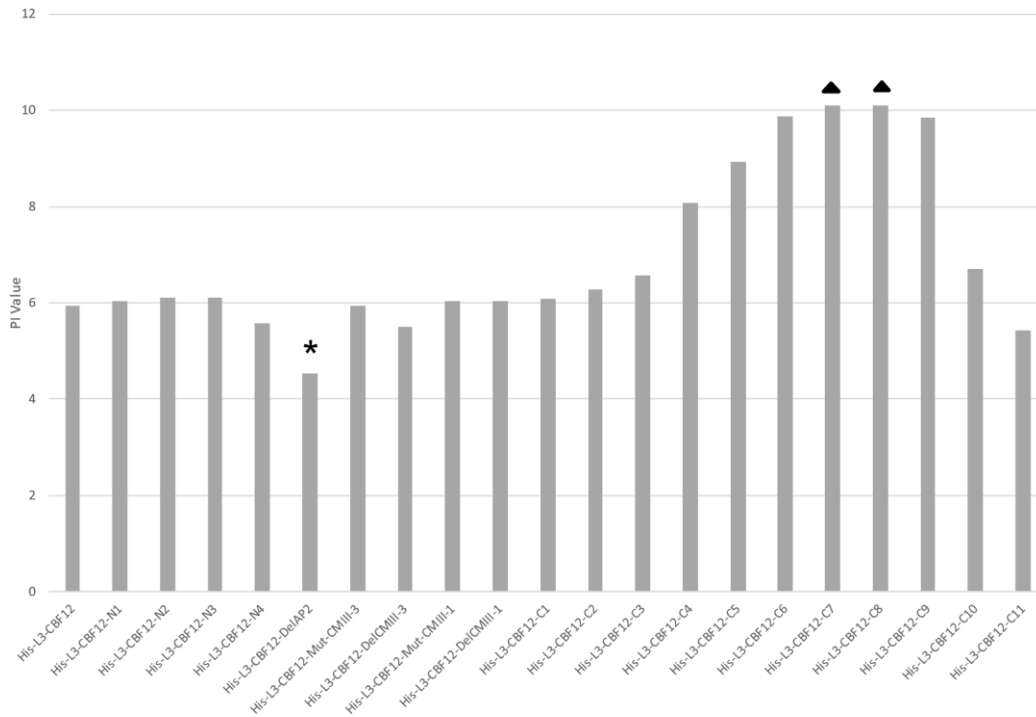
Fig. 4.14 pI values determined for the His-CBF12 variants. Determination of pI were done according to Bjellqvist et al. 1994, using the pI/MW tool on the Expasy web portal. The asterisk indicates the lowest pI value and triangles indicate the highest pI values.

His-L3-CBF12

1	MGSSHHHHHH	SSGLVPRGSH	MLEDLGTDDD	DKMDTGPERN	WNSPASPPSS	50
51	LEQGMPTSPA	SPTPKRPAGR	TKFKETRHPV	FHGVRRRGSN	GRWVCEVRVP	100
101	GKRGERLWLG	THVTAEAAAR	AHDAAMLALY	GRTPAARLNY	PDSAWLLAVP	150
151	SSLSDLADVR	RAAIGAVVDF	LRRQEAGASA	GAVAEEAHVD	GIASAASAPD	200
201	NASSSAAAAH	SQPPCANAGY	EVPDALCHDM	FELHTSGEMD	AGTYYADLAQ	250
251	GLLLEPPPPP	SSGASSERGD	DAALWNH			277

Pink = Acidic amino acids   Blue = Basic amino acids

Fig. 4.15 Distribution of charged residues within full-length His-CBF12. The underlined section corresponds to the AP2 domain and the CMIII-3 motif, which contains half of the basic residues of the protein. A substantial amount of the acidic residues are contained within the carboxyl terminus.

It was found for all the N-terminus truncations that they all cut through similar regions of averaged flexibility (Fig. 4.16). Average flexibility predominately stayed between a score of 0.4 and 0.5. The most flexible region consisted of three proline residues at the beginning of HisCBF12-N1. Low flexibility residues, such as tryptophan and phenylalanine, were cancelled out by being surrounded by extended regions of high flexibility residues. There were also few low flexibility amino acids present within the N-terminus sequences. From the start of the CBF12 gene through to His-CBF12-N4, a combined region of 61 residues, there were only eight low flexibility amino acids. In contrast, the AP2 domain has 19 low flexibility amino acids within its 63-residue sequence. Thus, the N-terminus is characterized by relatively uniform high flexibility. There is little evidence of any secondary structure, according to flexibility, that would promote anomalous migration during SDS-PAGE.

The C-terminus truncations present a more nuanced story regarding flexibility. The initial truncations are similar to what was measured for the N-terminus constructs (Fig. 4.16). From the end of the protein to His-L3-CBF12-C3, there are only eight low flexibility residues over a 41residue section. There are also some sections demonstrating average flexibility values over 0.5, such as a section of proline residues at positions 256 to 260. These regions however are equally as sparse, with only three such regions over the span. Like with the N-terminus truncations, the removal of these segments of residues had no observed effect on anomalous migration.

In going from His-CBF12-C3 to His-CBF12-C4, an additional eight low flexibility residues are removed. The sequence removed is characterized by a generally lower average flexibility compared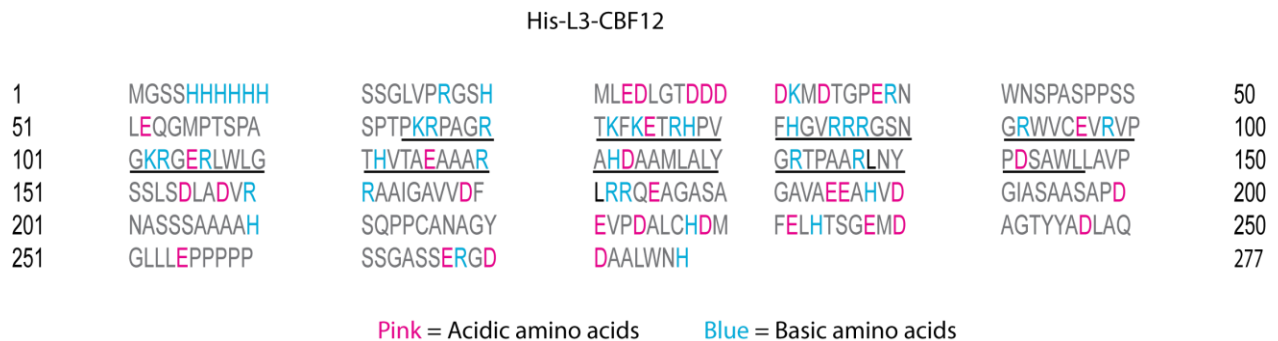 to the previously discussed constructs. This increased removal of low flexibility amino acids is correlated with a slight, yet significant, reduction in the observed anomalous migration (Fig. 4.11). This trend is continued for the subsequent truncations of His-CBF12-C5 and -C6. By -C6, an additional 13 low flexibility amino acids have been removed. Many of these amino acids also form small clusters, rather than being isolated in high flexibility regions. There were an equal number of high flexibility regions as was found for His-CBF12-C1 through C3 (Fig. 4.12). The centers for high and low flexibility were also evenly distributed throughout the sequence. This is suggestive of the elongated regions found previously by Jain (2013), in that strands of flexible amino acids are anchored into an extended configuration by clusters of inflexible residues. These regions thus may be inhibitory to migration during SDS-PAGE.
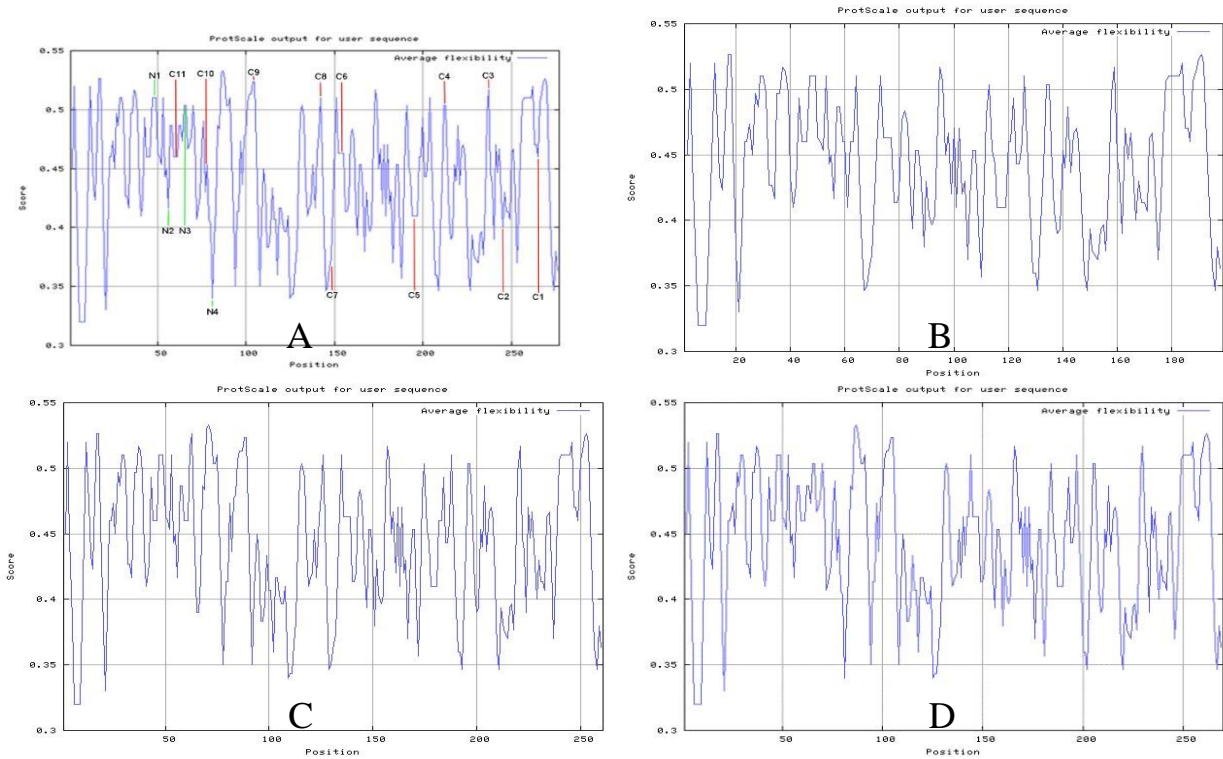
Fig. 4.16. Calculated average flexibility for the truncations and deletion mutant constructs of HisL3-CBF12. Average calculated flexibility derived from Bhaskaran and Ponnuswamy (1988). Average flexibility of the full-length L3-CBF12 construct (A), His-L3-CBF12-DelAP2 (B), HisL3-CBF12-DelCMIII-3 (C) and His-L3-CBF12-DelCMIII-1 (D) are shown. The end-points for Nterminal (N1-N4) and C-terminal truncations (C1-C11) are delineated by vertical green or red lines are indicated for the full-length protein.

The removal of the entire C-terminus and truncating into the CMIII-1 and subsequent AP2 domain were not indicative of any major changes to the flexibility of the protein (Fig. 4.16). Going from His-CBF12-C7 to His-CBF12-C6 is only a difference of seven amino acids. Nevertheless, the apparent migration difference increases (Fig. 4.11). The AP2 domain represents the largest region of low flexibility within the protein. Their removal, as with His-CBF12-C8 through HisCBF12-C10, removes the bulk of the aromatic and hydrophobic residues of the protein. This is correlated with a dramatic increase in anomalous migration behaviour. Hence for C-terminus truncations, features which can potentially enhance protein elongation are correlated with increased anomalous migration.

Considering the deletion constructs, there was no notable difference from the full-length construct aside from the removal of the indicated motifs from the chart (Fig. 4.16). It was hoped that the removal of a motif would result in a frame shift within the analysis corresponding with the anomalous migration data. Alas, the removal of the CMII-3, CMIII-1 or AP2 domain did not substantially alter average flexibility as determined.

While the change in migration for His-CBF12-DelAP2 could be attributed with changes in the charge of the polypeptide, this possibility fails to explain the increase in anomalous behavior observed for the constructs His-CBF12-C9 and C10. Instead, the degree of disorder is the property that changed with these constructs. Per the algorithms of PrDOS, the full-length native His-CBF12, while containing distinct disordered regions, is not an intrinsically disordered protein. This is in the consideration that a disordered protein contains more disordered residues than otherwise, of which His-CBF12 contains only 39% disordered residues for its total sequence (Fig. 4.17). HisCBF12-C9 and C10 had 52% and 93% of its total sequence as disordered residues, such that they may be considered intrinsically disordered regarding SDS-PAGE. All other constructs were comparable in disorder with the full length native protein. It must be noted that His-CBF12-C11 was also calculated to have 90% of its total sequence in disordered residues yet did not display the elevated migrational delay which defined the previous two constructs.
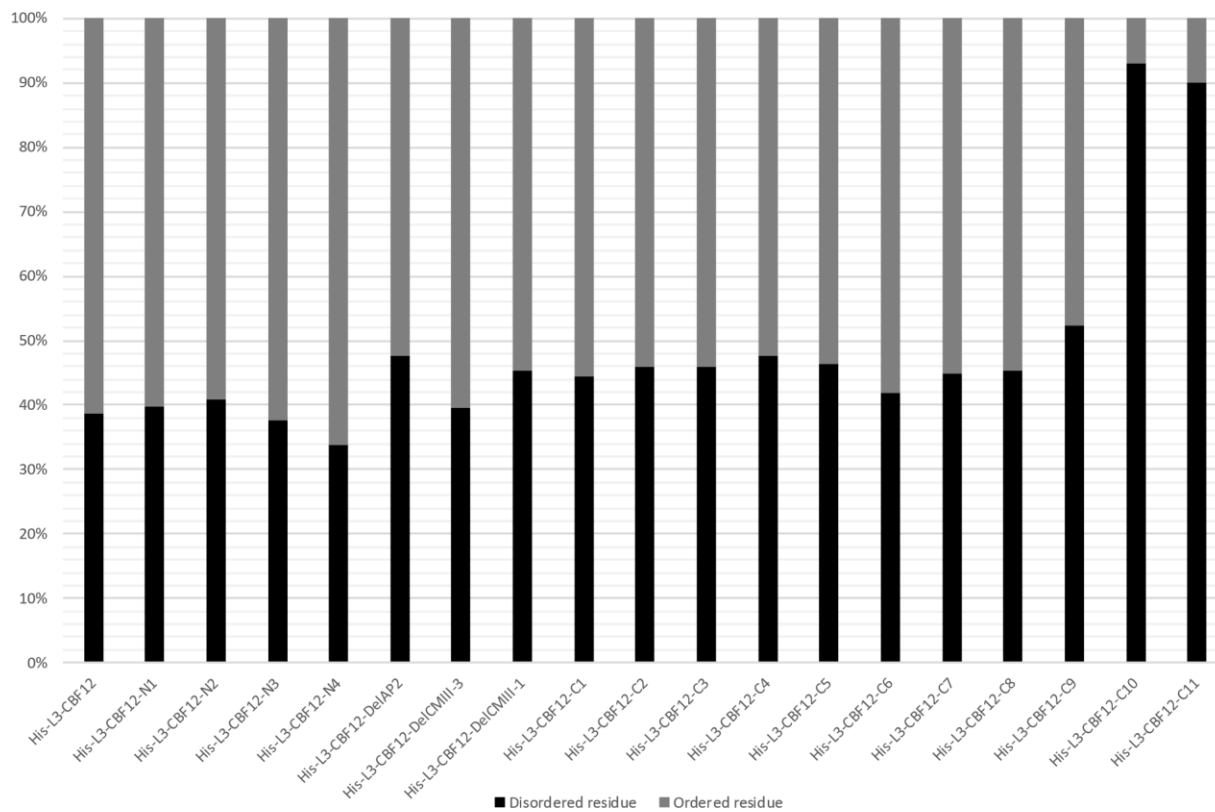
Fig. 4.17 Percentage of disordered to ordered residues for the His-CBF12 variant proteins. The values were calculated by PrDOS (Ishida, and Kinoshita, 2007). Both -C10 and -C11 variants were predicted to be characterized as intrinsically disordered proteins. All other proteins were classified as ordered.

CHAPTER 5: DISCUSSION

5.1 Potential role of pI in modulating gel migration

When the truncated and mutated constructs were initially designed, it was hypothesized that the structural motifs CMIII-3 or CMIII-1 or the N-terminus would be at least partially responsible for the observed delayed gel shift. According to previous work done by Jain (2013), the delayed migration of the CBF proteins within wheat was due to the proteins either being elongated and/or the effect of dense, hydrophobic secondary structural elements like the AP2 domain. To verify the potential cause, constructs were produced whereby the amino-terminus, the AP2 domain, the CMIII-3 and CMIII-1 motifs, and the carboxyl terminus of the CBF12 protein were removed sequentially. These constructs were assembled using the pET-15b vector and successfully transformed into either DH5α or Novablue™ *E. coli* cells. All generated constructs were successfully induced and expressed in BLR(DE3)/pLysS or BL21(DE3) cells, as confirmed by SDS-PAGE analysis of the cell extracts. Isolation and purification of CBFs such as CBF12 have proven to be challenging due to its propensity to aggregate or degrade (Gangopadhyay, 2015). Purification however has no effect on migrational delay (Gangopadhyay, 2015, Jain et al. 2013), therefore gel migration studies here were done using total cell lysates.

Many of the alterations made to His-CBF12 did not result in changes to the degree of delay during SDS-PAGE. All truncations involving the N-terminus had virtually no effect. Likewise, the deletions and mutations involving CMIII-3 (PKRPAGRTKF) and CMIII-1 (DSAWL) had no significant influence. The other mutational and deletion constructs would maintain the approximate 40% difference between the observed and estimated molecular weight. His-CBF12DelAP2 and His-CBF12-C10 were noted exceptions among the constructs, in both predicted properties and the migration during electrophoresis observed.

It was predicted that secondary structure rich elements like the AP2 domain would be contributing to any observed gel shifting, by interfering with the degree of binding of SDS molecules to the protein backbone (Rath et al., 2009; Rath and Deber, 2013; Shi et al., 2012). This is especially relevant considering the AP2 domain, which has three beta-sheets and an alpha helix in close proximity (Allen et al., 1998). Instead, removal of the AP2 domain (His-CBF12-delAP2) resulted in the opposite effect with migration being 68% slower than the estimated molecular weight. This construct also had the CMIII-3 motif removed as well, which may have played a role.

Considering however, that His-CBF12-DelCMIII-3 had little effect on correcting migration, it may be appropriate to assume that the influence of the CMIII-3 (PKRPAGRTFK) motif on gel shifting is marginal. Therefore, removing the AP2 domain is the cause for this increase. If the presence of dense regions of secondary structure was not a contributing factor, other aspects are driving the migration delay. Post-translational modification, such as phosphorylation or glycosylation, can lead to gel shifting (Adamson et al., 1992; Bretscher, 1971; Georgieva and Sendra, 1999; Hayward et al., 2002; Hosoi et al., 1995; Shi et al., 2012). Previous studies however, have eliminated this possibility for CBF12 (Jain et al., 2013). Of all the generated constructs, His-CBF12-DelAP2 is the only one which trends more acidic for its pI value compared to the full-length protein. More than half of all the basic residues of the protein are contained within the CMIII-3 motif and the AP2 DNA binding domain. Removing these two elements results in a more acidic pI and a greater distribution of negative charge throughout the polypeptide. More acidic proteins are noted to migrate slower during electrophoresis compared to neutral or basic proteins (Alves et al., 2004; Armstrong and Roman, 1993; Garcia-Ortega et al., 2005; Kang et al., 2008). The same could be said for proteins where negative charges are evenly distributed (Das and Pappu, 2013). This is hypothesized to be due to the negative charges repealing the dodecyl sulfate ions but the full nature of this phenomenon has yet to be elucidated. The other constructs either remained close to the original pI value or became more basic in pI as sections were removed. The increase in the value of pI for the C-terminus constructs was correlated with a decrease in the degree of gel shifting but the effect was minimal.

Other potential factors that could contribute to gel shifting, such as disorder, hydrophobicity and potential for aggregation (Amani and Naeem, 2013; Kalthoff, 2003; Karan et al., 2013; Kazuoka et al., 2003; C. Lee et al., 2014; W. K. Lee et al., 2014; Manning and Colón, 2004; Ohshida et al., 2016; Oikawa et al., 2003; Rath et al., 2009; Tantos et al., 2009, 2013; Uversky and Dunker, 2010), were not significant for the N-terminus truncations and the mutation and deletion constructs (Figs. 4.12, 4.13, 4.17). While changes in the pI and the charge distribution of the protein may explain the increase in the migrational delay of His-CBF12-DelAP2, it fails to further any understanding about the anomalous behavior of the native protein. Further experiments would be required to evaluate the role in which charged residues contribute to the anomalous migration behavior of CBF12. This would entail determining which acidic residues are most exposed, and thus be more disruptive to SDS binding, and substituting these with neutral amino acids, as was

done with the 67 amino acid cold-shock protein *Bs*-CspB from *Bacillus subtilis* (Martin et al., 2002). Another approach would be neutralizing negative charge of the protein altogether (Alves et al., 2004; Rae and Elliott, 1986). In regards to charge interactions within a protein, a single amino acid substitution can have a profound effect on its migration during SDS-PAGE (Martin et al., 2002, 2001). Aside from anomalous migration, charge distribution within CBF12 could also explain other known properties. Highly acidic proteins are known to be resistant to low or high temperature denaturation and are frequently associated with extremophiles (Alonso et al., 2002; Alves et al., 2004; Kim et al., 2016; Rao et al., 2009). Some also, during purification, are prone to aggregate (Alonso et al., 2002). Considering the challenging environments that these other organisms face, the ability for a protein to operate efficiently at cold temperatures may rely in part to its charge distribution or acidity, in whole or in part. To approach this question though would require a more in-depth understanding of the secondary and tertiary structures of the CBF12 protein. Considering the other constructs, there are also other factors to be considered which contribute to the anomalous behavior during SDS-PAGE, as demonstrated by the C-terminus.

5.2 C-terminus truncations and protein elongation

Unlike the previously described constructs, sequential truncations from the C-terminus resulted in a gradual reduction in the gel shifting of CBF12, reducing delay by 11% with HisCBF12-C6. Thus, the C-terminus partially explained the observed anomalous migration of CBF12. How it contributes was not illuminated by the factors considered in this study. It was previously found that this region of the protein was highly elongated, as determined by dynamic light scattering analysis (Jain, 2013). It must be concluded that this reduction in delay corresponds to the elimination of this elongated region. Proteins which possess elongated regions are noted for their delayed migration, due to their structure ultimately possessing a higher Stokes radius than what a globular protein of a similar size would entail (Alonso et al., 2002). An elongated Cterminus can describe a fraction of the anomalous migration, although a substantial portion of the effect remains unsubstantiated (Furthmay and Timpl, 1971).

5.3 Exposure of the DNA binding domains

Each truncation from the C-terminus resulted in a gradual decrease in the observed gel shift during SDS-PAGE thus far. Removal of the entire C-terminus, as with His-CBF12-C7, however,

resulted in the restoration of the gel shift to 37.9%, from 29.9% in His-CBF12-C6. This recovery was reset to 31.1% by the next truncation, His-CBF12-C8, which removed the CMIII-1 motif along with the rest of the C terminus. While the removal of the CMIII-3 motif did not affect the migration difference, it being at the end of the polypeptide seemed to have a substantial effect. None of the other aspects considered, such as pI or hydrophobicity, have any significant deviation for this construct. CMIII-1 being at the end of the polypeptide may be disruptive enough to prevent proper migration. A similar phenomenon may also explain the anomalous migration behavior observed for His-CBF12-C9 and His-CBF12-10, where the core of the AP2 domain and the end of the CMIII-3 motif is exposed, with the difference in degree of migration being explain by how disruptive the exposed hydrophobic motif is to the proper binding of SDS to the protein. Typically, these structures are expected to be buried in the hydrophobic core. Nothing further indicated by information gleaned from the primary structure though is indicative of what the cause of this may be.

5.4 Role of disorder in modulating gel migration

Another possibility which could explain the anomalous migration behavior is that the smaller constructs, C9 through C11, are notably more disordered than the others. A protein being more disordered is a probable cause for anomalous migration since intrinsically disordered proteins will migrate 1.2 to 1.8 times slower than predicted via sequence data during SDS-PAGE separation (Uversky and Dunker, 2010). This is due to their diffuse structure and the larger hydrodynamic dimensions which typifies this class of protein (Graether and Boddington, 2014; Receveur-Bréhot et al., 2006). The other constructs ranged from 34% to 48% predicted disordered residues versus ordered residues. By C9, the disordered residues reached 55%, and C10 and C11 attained 92% and 90% respectively. This rise in disorder correlates with the drastic increase in migration delay for the C9 and C10 migrations. In consideration for C10, the degree of gel retardation observed matches that of His-CBF12-DelAP2. Thus, some element must be additive to what effect defines CBF12 anomalous migration behavior, of which protein disorder may be a candidate. Anomalous migration during SDS-PAGE separation alone however, is a poor indicator for protein disorder (Uversky and Dunker, 2010). The degree of migrational delay observed during separation in either SDS-PAGE or size exclusion chromatography can vary depending on if the protein is in a native molten globule, a native premolten globule or a native coil (Uversky and Dunker, 2010). The native

CBF12 protein must also be acknowledged that it is not an intrinsically disordered protein. Based on dynamic light scattering and circular dichromism however, the native CBF12 protein was confirmed to be highly elongated (Jain et al., 2013). Combined with the computational analysis, it is indicative of disordered regions within the protein, of which the C9 through C11 truncations may have isolated. Similar regions of disorder are found within numerous transcription factors and stress response proteins (Battaglia et al., 2008; Hincha and Thalhammer, 2012; Wise and Tunnacliffe, 2004). For drought and cold-shock proteins, it is postulated that the disordered regions allow them to use surfaces like membranes or carbohydrates rather than water to stabilize their structure (Battaglia et al., 2008). This allows them to remain operational during severe drought conditions (Battaglia et al., 2008). Numerous transcription factors also utilize disordered regions, primarily located in the N-terminus (Tantos et al., 2012; Uversky, 2002). Under stressful conditions, this feature can facilitate DNA binding (Tantos et al., 2012). This is due to disordered regions being less sensitive to perturbations in the reaction conditions yet retaining a high enzymatic specificity via their structural flexibility (Huang and Liu, 2010). Neither of these possibilities were confirmed at this time. It does suggest however, that the characteristics which cause gel shifting may be an adaptive feature to allow for function in stressed conditions.

His-CBF12-C11 also exhibited a high degree of disorder. Its degree of anomalous migration however was more relative to the predicted size, 7.2 kDa compared to 6.4 kDa respectively. It may be too small for a substantial effect on migration to materialize. Small disordered protein fragments still can exhibit anomalous SDS-PAGE migration (Iakoucheva et al., 2001). A third of the protein however consists of the histidine tag, which while predicted to be disordered, will not affect SDSPAGE migration and run as predicted (Alves et al., 2004; Gupta et al., 2013; Novak et al., 2013).

There is also a relationship between a disordered protein fragment's size and degree of gel shift observed, with larger fragments experiencing a greater effect during size separation (Iakoucheva et al., 2001; Uversky and Dunker, 2010). To these other fragments, His-CBF12-C11 has a comparable migrational delay (Iakoucheva et al., 2001). That there is such substantial difference in migrational delay between His-CBF12-C11 and C10 though may be concerning. The difference between these two constructs is the presence of the CMIII-3 motif. This motif however has been demonstrated previously to not have an influence on the anomalous migration of the protein. It was previously suggested that the presence of either the CMII-3 or CMIII-1 motifs at the Cterminus

end of the protein may disrupt proper binding of SDS. Per PrDos, the CMIII-3 motif consists of disordered residues. These additional residues may amplify the observed effect (Uversky and Dunker, 2010). The presence of additional proline residues, particularly at the end of the polypeptide as observed in His-CBF12-C8 and C10, were found to be disruptive of helical structures (C. Lee et al., 2014). SDS is known to preferentially bind to induced helical structures, thus these end terminus proline residues may be more disruptive relative to other potential disordered residues (Rath et al., 2009).

5.5 Difference between N-terminus and C-terminus constructs

Of the constructs, only His-CBF12-C11, which consists of the 32 amino acid His tag and the 28 amino acid N-terminus of the CBF12 protein, ran with the lowest anomalous migration during SDS-PAGE. In addition, the excision of sequentially larger sections of the amino terminus had no effect on the degree of migration. The observed decrease between each of these constructs was linear, decreasing strictly based on the reductions in total protein size.

5.6 Conclusions

Of the possibilities presented, it can be concluded that the amino terminal of the protein likely has no gel shifting effect. This is likely not true for the other structural components. The AP2 domain was predicted to be a contributor to the anomalous migration, due to its secondary structure elements. Its removal however resulted in the opposite effect in the context of CBF12. This was verified in both His-CBF12-DelAP2 and His-CBF12-C10, where the loss of the AP2 domain resulted in the largest increase in the observed migrational delay. Therefore, the AP2 domain counteracts whatever is responsible for anomalous migration of recombinant CBF12 proteins. Other domains which were presumed to have an effect when removed, such as the CMIII-3 and CMIII-1 motifs, did not change the degree of migration. Compared to the N-terminus truncations, the C-terminus truncations exhibited a minor effect, reducing anomalous migration by 11% when completely removed from the protein. While the C-terminus contributes to gel shifting, it does not define it. All attempts within this study to selectively isolate the causal sequences of the gel shift within the primary sequence of CBF12 have failed. All constructs still demonstrate a slower migration than estimated during SDS-PAGE separation. It must be concluded that no isolated sequence contributes to gel shifting but rather a property or sets of properties which can define all

other sequences, except for the DNA binding domain. Of the properties examined, both changes in pI and the ratio of disordered: ordered residues in the protein correlated with the increases in migrational delay observed for His-CBF12-DelAP2 and His-CBF12-C10. Both disorder and pI may potentially contribute but likely do not define the anomalous migration for the full-length protein. The most likely candidate for the full-length CBF12's anomalous migration is distribution of charge, considering the even distribution of acidic amino acids throughout the protein. To test this possibility further though would require more knowledge regarding what residues are exposed on the protein surface and the protein's structure in general.

5.7 Future Work

Future work on this phenomenon would involve the further characterization of the protein in regards to its secondary and tertiary structure. Purification and analysis of the prepared constructs via dynamic light scattering or circular dichromism spectra would be a natural next step. More immediate, replacing elements such as the C-terminus with components known to not exhibit anomalous migration in SDS-PAGE could better differentiate what roles the components present in CBF12 contribute to the gel behavior. The subsequent decrease in the total size of each truncated construct could be a confounding variable in this phenomenon. Splicing on a different C-terminus, for example, will be able to elucidate better the role without the loss in size.

CHAPTER 6: REFERENCES

Adamson, P., Marshall, C.J., Hall, A., Tilbrooks, P.A., 1992. Post-translational Modifications of p21 rho Proteins. Biochemistry 267, 20033–20038.

Agarwal, M., Hao, Y., Kapoor, A., Dong, C.H., Fujii, H., Zheng, X., Zhu, J.K., 2006. A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance. J. Biol. Chem. 281, 37636–37645.

Agarwal, P.K., Agarwal, P., Reddy, M.K., Sopory, S.K., 2006. Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. Plant Cell Rep. 25, 1263–1274.

Ågren, J., Oakley, C.G., McKay, J.K., Lovell, J.T., Schemske, D.W., 2013. Genetic mapping of adaptation reveals fitness tradeoffs in Arabidopsis thaliana. Proc. Natl. Acad. Sci. USA. 110, 21077–21082.

Allen, M.D., Yamasaki, K., Ohme-Takagi, M., Tateno, M., Suzuki, M., 1998. A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. EMBO J. 17, 5484–5496.

Alonso, L.G., García-Alai, M.M., Nadra, A.D., Lapeña, A.N., Almeida, F.L., Gualfetti, P., De Prat-Gay, G., 2002. High-risk (HPV16) human papillomavirus E7 oncoprotein is highly stable and extended, with conformational transitions that could explain its multiple cellular binding partners. Biochemistry 41, 10510–10518.

Alves, V.S., Pimenta, D.C., Sattlegger, E., Castilho, B.A., 2004. Biophysical characterization of Gir2, a highly acidic protein of Saccharomyces cerevisiae with anomalous electrophoretic behavior. Biochem. Biophys. Res. Commun. 314, 229–234.

Amani, S., Naeem, A., 2013. Understanding protein folding from globular to amyloid state: Aggregation: Darker side of protein. Process Biochem. 48, 1651–1664.

Amani, S., Naeem, A., 2014. Deciphering aggregates, prefibrillar oligomers and proto fibrils of cytochrome c. Amino Acids 46, 1839–1851

Armstrong, D.J., Roman, A., 1993. The anomalous electrophoretic behavior of the human papillomavirus type 16 E7 protein is due to the high content of acidic amino acid residues. Biochem. Biophys. Res. Commun. 192, 1380–1387.

Arnórsdóttir, J., Smáradóttir, R.B., Magnússon, Ó.T., Thorbjarnardóttir, S.H., Eggertsson, G., Kristjánsson, M.M., 2002. Characterization of a cloned subtilisin-like serine proteinase from a psychrotrophic Vibrio species. Eur. J. Biochem. 269, 5536–5546.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., Stockinger,H., 2012. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 40, 597–603.

Audain, E., Ramos, Y., Hermjakob, H., Flower, D.R., Perez-Riverol, Y., 2015. Accurate estimation of isoelectric point of protein and peptide based on amino acid sequences. Bioinformatics 32, 821–827.

Badawi, M., Reddy, Y.V., Agharbaoui, Z., Tominaga, Y., Danyluk, J., Sarhan, F., Houde, M., 2008. Structure and functional analysis of wheat ICE (inducer of CBF expression) genes. Plant Cell Physiol. 49, 1237–1249.

Båga, M., Chodaparambil, S. V, Limin, A.E., Pecar, M., Fowler, D.B., Chibbar, R.N., 2007. Identification of quantitative trait loci and associated candidate genes for low-temperature tolerance in cold-hardy winter wheat. Funct. Integr. Genomics 7, 53–68.

Båga, M., Fowler, D.B., Chibbar, R.N., 2009. Identification of genomic regions determining the phenological development leading to floral transition in wheat (*Triticum aestivum* L.). J. Exp. Bot. 60, 3575–3585.

Baker, S.S., Wilhelm, K.S., Thomashow, M.F., 1994. The 5'-region of Arabidopsis thaliana cor15a has cis-acting elements that confer cold-, drought- and ABA-regulated gene expression. Plant Mol. Biol. 24, 701–713.

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

Bhaskaran, R., Ponnuswamy, P., K. 1988. Positional flexibilities of amino acid residues in globular proteins. Int. J. Peptide Protein Res. 32, 241-255.

Bjellqvist, B., Basse, B., Olsen, E., Celis, J. E. 1994. Reference points for comparisons of twodimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. Electrophoresis. 15, 529-539.

Bretscher, M.S., 1971. Major Human Erythrocyte Glycoprotein spans the Cell Membrane. Nat. New Biol. 231, 229–232.

Campoli C, Matus-Cádiz MA, Pozniak CJ, Cattivelli L, Fowler DB (2009) Comparative expression of Cbf genes in the Triticeae under different acclimation induction temperatures. Mol Genet Genomics 282, 141–152.

Catalá, R., Santos, E., Alonso, J.M., Ecker, J.R., Martínez-Zapater, J.M., Salinas, J., 2003. Mutations in the Ca^{2+} /H^+ Transporter CAX1 Increase CBF/DREB1 Expression and the ColdAcclimation Response in Arabidopsis. Plant Cell 15, 2940–2951.

Chinnusamy, V., Ohta, M., Kanrar, S., Lee, B.-H., Hong, X., Agarwal, M., Zhu, J.-K., 2003. ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. Genes Dev. 17, 1043–1054.

Choi D. W., Rodriguez E. M., Close T. J. 2002. Barley *Cbf3* gene identification, expression pattern, and map location. Plant Physiol 129, 1781–1787.

Chopra, R., C. and Paroda, R., S. 1986. Approaches for incorporating drought and salinity resistance in crop plants. Oxford and IBM publishing company.

Cook, D., Fowler, S., Fiehn, O., Thomashow, M.F., 2004. A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of Arabidopsis. Proc. Natl. Acad. Sci. USA 101, 15243–15248.

Costa, S., Almeida, A., Castro, A., Domingues, L., 2014. Fusion tags for protein solubility, purification, and immunogenicity in Escherichia coli: The novel Fh8 system. Front. Microbiol. 5, 1–20.

Crosatti, C., Rizza, F., Badeck, F.W., Mazzucotelli, E., Cattivelli, L., 2013. Harden the chloroplast to protect the plant. Physiol. Plant. 147, 55–63.

Crow, K.D., Wagner, G.P., 2006. What is the role of genome duplication in the evolution of complexity and diversity? Mol. Biol. Evol. 23, 887–892.

D'Amico, S., Claverie, P., Collins, T., Georlette, D., Gratia, E., Hoyoux, A., Meuwis, M.-A., Feller, G., Gerday, C., 2002. Molecular basis of cold adaptation. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 357, 917–925.

D'Amico, S., Gerday, C., Feller, G., 2001. Structural Determinants of Cold Adaptation and Stability in a Large Protein. J. Biol. Chem. 276, 25791–25796.

Danyluk, J., Kane, N. a, Breton, G., Limin, A.E., Fowler, D.B., Sarhan, F., 2003. TaVRT-1, a putative transcription factor associated with vegetative to reproductive transition in cereals. Plant Physiol. 132, 1849–1860.

Das, R.K., Pappu, R. V, 2013. Conformations of intrinsically disordered proteins are infl uenced by linear sequence distributions of oppositely charged residues. Proc Natl Acad Sci USA 110, 13392–13397.

De Santi, C., Leiros, H.K.S., Di Scala, A., de Pascale, D., Altermark, B., Willassen, N.P., 2016. Biochemical characterization and structural analysis of a new cold-active and salt-tolerant esterase from the marine bacterium *Thalassospira* sp. Extremophiles 20, 323–336.

Dhillon, T., Pearce, S. P., Stockinger, E. J, Distelfeld, A., Li, C., Knox, A.K, Vashegyi, I., Vágújfalvi, A., Galiba, G., Dubcovsky, J. 2010. Regulation of freezing tolerance and flowering in temperate cereals: the VRN-1 connection. Plant Physiol 153, 1846–1858.

Doyle, J. J., Doyle J. L. 1987. Isolation of plant DNA from fresh tissue. Phytochemical Bulletin 19, 11-15.

Dubcovsky, J., Dvorak, J., 2007. Genome Plasticity a Key Factor. Science 316, 1862–1866.

Dyson, H.J., Wright, P.E., 2005. Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. 6, 197–208.

El Kayal, W., Navarro, M., Marque, G., Keller, G., Marque, C., Teulieres, C., 2006. Expression profile of CBF-like transcriptional factor genes from Eucalyptus in response to cold. J. Exp. Bot. 57, 2455–2469.

Ergon, A., Fang, C,,Jorgenson, O., Aamlid, T.S., Rognli, O.A. 2006. Quantitative trait loci controlling vernalization requirement, heading time and number of panicles in meadow fescue (*Festuca pratense* Huds). Theor. Appl. Genet. 112, 232-242.

Fawcett, J.A., Maere, S., Peer, Y. Van De, 2009. Plants with double genomes might have had a better chance to survive the Cretaceous – Tertiary extinction event. Proc. Natl. Acad. Sci USA 106, 5737–5742.

Feldman, M. 2001. Origin of cultivated wheat 0 A history of wheat breeding. In Bonjean, P.A. and Angus, J.W. (eds) The world wheat book. Lavosier Publishing, Paris, France. pp 3 - 56

Feller, G., Gerday, C., 1997. Psychrophilic enzymes: Molecular basis of cold adaptation. Cell. Mol. Life Sci. 53, 830–841.

Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J., Serrano, L., 2004. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat. Biotechnol. 22, 1302-1306.

Fowler, D. B., 2002. Winter wheat production manual. Accessed Feb 14th, 2018. www.usask.ca/agriculture/plantsci/winter_cereals/winter-wheat-production-manual/.

Fowler, D.B., 2008. Cold Acclimation Threshold Induction Temperatures in Cereals. Crop Sci. 48, 1147-1153.

Fowler, D.B., Chauvin, L.P., Limin, A.E., Sarhan, F., 1996. The regulatory role of vernalization in the expression of low-temperature-induced genes in wheat and rye. Theor. Appl. Genet. 93, 554-559.

Fowler, D.B., Limin, A.E..; Ritchie, J.T. 1999. Low-temperature tolerance in cereals: Model and genetic interpretation. Crop Science 39: 626-633.

Fowler, S.G., Cook, D., Thomashow, M.F., 2005. Low temperature induction of Arabidopsis CBF1, 2, and 3 is gated by the circadian clock. Plant Physiol. 137, 961–968.

Francia, E., Rizza, F., Cattivelli, L., Stanca, A.M., Galiba, G., Toth, B, Hayes, P.M., Skinner, J.S., Pecchioni, N. 2004. Two loci on chromosome 5H determine low-temperature tolerance in Nure (winter) x Tremois (spring) barley map. Theoretical and Applied Genetics 108, 670-680.

Furthmay, H., Timpl, R. 1971. Characterization of collagen peptides by sodium dodecylsulfatepolyacrylamide electrophoresis. Anal Biochem 41, 510-516.

Fursova, O. V, Pogorelko, G. V, Tarasov, V.A., 2009. Identification of ICE2, a gene involved in cold acclimation which determines freezing tolerance in Arabidopsis thaliana. Gene 429, 98–103.

Galiba, G., Quarrie, S.A., Sutka, J., Morgounov, A., Snape, J.W., 1995. RFLP mapping of the vernalization (Vrn1) and frost resistance (Fr1) genes on chromosome 5A of wheat. Theor. Appl. Genet. 90, 1174–1179.

Galiba, G., Vágújfalvi, A., Li, C., Soltész, A., Dubcovsky, J., 2009. Regulatory genes involved in the determination of frost tolerance in temperate cereals. Plant Sci. 176, 12–19.

Gangopadhyay, S., 2015. *In vitro* production of selected C-repeat binding factors (CBF) encoded by group five chromosomes in winter wheat *cv* Norstar. Master's thesis. University of Saskatchewan, Saskatoon, Canada.

Garcia-Ortega, L., De los Rios, V., Martinez-Ruiz, A., Onaderra, M., Lacadena, J., Martinez del Pozo, A., Gavilanes, J.G., 2005. Anomalous electrophoretic behavior of a very acidic protein: Ribonuclease U2. Electrophoresis 26, 3407–3413.

Gatti-Lafranconi, P., Natalello, A., Rehm, S., Doglia, S.M., Pleiss, J., Lotti, M., 2010. Evolution of Stability in a Cold-Active Enzyme Elicits Specificity Relaxation and Highlights SubstrateRelated Effects on Temperature Adaptation. J. Mol. Biol. 395, 155–166.

Georgieva, E., Sendra, R., 1999. Mobility of Acetylated Histones in Sodium Dodecyl. Anal. Biochem 402, 399–402.

Gill, B.S., Appels, R., Botha-Oberholster, A.M., Buell, C.R., Bennetzen, J.L., Chalhoub, B.,

Chumley, F., Dvořák, J., Iwanaga, M., Keller, B., Li, W., McCombie, W.R., Ogihara, Y., Quetier, F., Sasaki, T., 2004. A workshop report on wheat genome sequencing: International genome research on wheat consortium. Genetics 168, 1087–1096.

Gilmour, S.J., Thomashow, M.F., 1991. Cold acclimation and cold-regulated gene expression in ABA mutants of Arabidopsis thaliana. Plant Mol. Biol. 17, 1233–1240.

Gilmour, S.J., Zarka, D.G., Stockinger, E.J., Salazar, M.P., Houghton, J.M., 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. Plant J 16, 433–42.

Golovnina, K.A., Glushkov, S.A., Blinov, A.G., Mayorov, V.I., Adkison, L.R., 2007. Molecular phylogeny of the genus *Triticum* L. Plant Syst. Evol. 264, 195–216.

Graether, S.P., Boddington, K.F., 2014. Disorder and function: a review of the dehydrin protein family. Front. Plant Sci. 5, 576.

Graham, D., Patterson, B.D., 1982. Responses of Plants to Low, Non-freezing Temperatures: Proteins, Metabolism, and Acclimation. Ann. Rev. Plant Biol. 33, 347–372.

Gulick, P. J., Drouin, S., Yu, Z., Danyluk, J., Poisson, G., Monroy, A. F., Sarhan, F. 2005. Transcriptome comparison of winter and spring wheat responding to low temperature. Genome 48, 913–923.

Gupta, N., Shrestha, A., Panda, A.K., Gupta, S.K., 2013. Production of tag-free recombinant fusion protein encompassing promiscuous T cell epitope of tetanus toxoid and dog zona pellucida glycoprotein-3 for contraceptive vaccine development. Mol. Biotechnol 54, 853–862.

Hake, S., Magnani, E., Sjo, K., 2004. From Endonucleases to Transcription Factors: Evolution of the AP2 DNA Binding Domain in Plants. Plant Cell 16, 2265–2277.

Harlan, J.R., Zohary, D., 1966. Distribution of Wild Wheats and Barley. Science 153, 1074–1080.

Harrison, M.J., Lawton, M.A., Lamb, C.J., Dixon, R.A., 1991. Characterization of a nuclear protein that binds to three elements within the silencer region of a bean chalcone synthase gene promoter. Proc. Natl. Acad. Sci. USA 88, 2515–2519.

Hayward, L.J., Rodriguez, J.A., Kim, J.W., Tiwari, A., Goto, J.J., Cabelli, D.E., Valentine, J.S., Brown, R.H., 2002. Decreased metallation and activity in subsets of mutant superoxide dismutases associated with familial amyotrophic lateral sclerosis. J. Biol. Chem. 277, 15923–15931.

Hazel, J.R., Prosser, C.L., 1974. Molecular mechanisms of temperature compensation in poikilotherms. Physiol. Rev. 54, 620–677.

Heino, P., Sandman, G., Lång, V., Nordin, K., Palva, E.T., 1990. Abscisic acid deficiency prevents development of freezing tolerance in Arabidopsis thaliana (L.) Heynh. Theor. Appl. Genet. 79, 801–806.

Heun, M., Schafer-Pregl, R., Klawan, D., Castagna, R., Accerbi, M., Borghi, B., Salamini, F.,1997. Site of Einkorn Wheat Domestication Identified by DNA Fingerprinting. Science 278, 1312–1314.

Hincha, D.K., Thalhammer, A., 2012. LEA proteins: IDPs with versatile functions in cellular dehydration tolerance. Biochem. Soc. Trans. 40, 1000–1003.

Hockney, R., C. 1994. Recent developments in heterologous protein production in *Escherichia coli*. Trends Biotechnol. 12, 456-463.

Hosoi, T., Uchiyama, M., Okumura, E., Saito, T., Ishiguro, K., Uchida, T., Okuyama, A., Kishimoto, T., Hisanaga, S., 1995. Evidence for cdk5 as a major activity phosphorylating Tau protein in porcine brain extract. J. Biochem. 117, 741–749.

Huang, Y., Liu, Z., 2010. Smoothing molecular interactions: The "kinetic buffer" effect of intrinsically disordered proteins. Proteins Struct. Funct. Bioinforma. 78, 3251–3259.

Iakoucheva, L.M., Kimzey, A.L., Masselon, C.D., Smith, R.D., Dunker, A.K., Ackerman, E.J., 2001. Aberrant mobility phenomena of the DNA repair protein XPA. Protein Sci. 10, 1353–1362.

Imamura, T. 2006. Surfactant-Protein Interactions. In: Dekker, M. (Eds.) Encyclopedia of Surface and Colloid Science. Taylor & Francis, New York, pp. 5230–5241.

International Grains Council. 2017. October Grain Marketing Report. GMR 482.

Ishida, T., Kinoshita, K., 2007. PrDOS: Prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res. 35, 460–464.

Ishitani, M., Xiong, L., Lee, H., Stevenson, B., Zhu, J., 1998. HOS1, a Genetic Locus Involved in Cold-Responsive Gene Expression in Arabidopsis. Plant Cell 10, 1151–1161.

Iwashita, S., Suzuki, T., Yasuda, T., Nakashima, K., Sakamoto, T., Kohno, T., Takahashi, I., Kobayashi, T., Ohno-Iwashita, Y., Imajoh-Ohmi, S., Song, S.Y., Dohmae, N., 2015. Mammalian Bcnt/Cfdp1, a potential epigenetic factor characterized by an acidic stretch in the disordered N-terminal and Ser250 phosphorylation in the conserved C-terminal regions. Biosci. Rep. 35, e00228–e00228.

Jaglo, K. R, Kleff, S., Amundsen, K. L., Zhang, X., Haake, V., Zhang, J. Z., Deit, T., Thomashow, M. F. 2001. Components of the Arabidopsis C-Repeat / Dehydration- Responsive Element Binding Factor Cold-Response Pathway Are Conserved in *Brassica napus* and Other Plant Species. Plant Physiol. 127, 910–917.

Jain, P., 2013. A study of C-repeat binding factors (CBF) associated with low temperature tolerance locus in winter wheat. PhD thesis. University of Saskatchewan, Saskatoon, Canada.

Jiang, W., Sun, D., Lu, J., Wang, Y., Wang, S., Zhang, Y., Fang, B., 2016. A cold-adapted leucine dehydrogenase from marine bacterium Alcanivorax dieselolei: Characterization and l-tertleucine production. Eng. Life Sci. 16, 283–289.

Jung, J.-H., Lee, H.-J., Park, M.-J., Park, C.-M., 2014. Beyond ubiquitination: proteolytic and non-proteolytic roles of HOS1. Trends Plant Sci. 19, 538–45.

Kalthoff, C., 2003. A novel strategy for the purification of recombinantly expressed unstructured protein domains. J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 786, 247–254.

Kang, N., Chen, D., Wang, L., Duan, L., Liu, S., Tang, L., Liu, Q., Cui, L., He, W., 2008. Rwdd1, a thymus aging related molecule, is a new member of the intrinsically unstructured protein family. Cell. Mol. Immunol. 5, 333–339.

Karan, R., Capes, M.D., DasSarma, P., DasSarma, S., 2013. Cloning, overexpression, purification, and characterization of a polyextremophilic β-galactosidase from the Antarctic haloarchaeon Halorubrum lacusprofundi. BMC Biotechnol. 13, 3.

Kawamoto, J., Kurihara, T., Kitagawa, M., Kato, I., Esaki, N., 2007. Proteomic studies of an Antarctic cold-adapted bacterium, Shewanella livingstonensis Ac10, for global identification of cold-inducible proteins. Extremophiles 11, 819–826.

Kazuoka, T., Masuda, Y., Oikawa, T., Soda, K., 2003. Thermostable aspartase from a marine psychrophile, Cytophaga sp. KUC-1: Molecular characterization and primary structure. J. Biochem. 133, 51–58.

Kazuoka, T., Oeda, K., 1992. Heat-stable COR (Cold Regulated) proteins associated with freezing tolerance in Spinach. Plant Cell Physiol. 33, 1107–1114.

Keily, J., MacGregor, D.R., Smith, R.W., Millar, A.J., Halliday, K.J., Penfield, S., 2013. Model selection reveals control of cold signalling by evening-phased components of the plant circadian clock. Plant J. 76, 247–257.

Kidokoro, S., Maruyama, K., Nakashima, K., Imura, Y., Narusaka, Y., Shinwari, Z.K., Osakabe, Y., Fujita, Y., Mizoi, J., Shinozaki, K., Yamaguchi-Shinozaki, K., 2009. The phytochrome interacting factor PIF7 negatively regulates DREB1 expression under circadian control in Arabidopsis. Plant Physiol. 151, 2046–2057.

Kim, H., Yeon, Y.J., Choi, Y.R., Song, W., Pack, S.P., Choi, Y.S., 2016. A cold-adapted tyrosinase with an abnormally high monophenolase/diphenolase activity ratio originating from the marine archaeon Candidatus Nitrosopumilus koreensis. Biotechnol. Lett. 38, 1535–1542.

Knight, H., Knight, M.R., 2000. Imaging spatial and cellular characteristics of low temperature calcium signature after cold acclimation in Arabidopsis. J. Exp. Bot. 51, 1679–1686.

Knox, A.K., Dhillon, T., Cheng, H., Tondelli, A., Pecchioni, N., Stockinger, E.J., 2010. CBF gene copy number variation at Frost Resistance - 2 is associated with levels of freezing tolerance in temperate-climate cereals Theor Appl Genet 121, 21–35.

Koornneef, M., Alonso-Blanco, C., Vreugdenhil, D., 2004. Naturally Occurring Genetic Variation in Arabidopsis thaliana. Annu. Rev. Plant Biol. 55, 141–172.

Kurepin, L. V, Dahal, K.P., Savitch, L. V, Singh, J., Bode, R., Ivanov, A.G., Hurry, V., Hüner, N.P.A., 2013. Role of CBFs as Integrators of Chloroplast Redox, Phytochrome and Plant Hormone Signaling during Cold Acclimation. Int. J. Mol. Sci. 14, 12729–12763.

Kyte, J., Doolittle, R., F. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105-132.

Laemmli, U.K. 1970. Cleavage of Structural Proteins during Assembly of Head of Bacteriophage-T4. Nature 227, 680–685.

Laudencia-Chingcuanco, Ganeshan, S., You, F., Fowler, B., Chibbar, R., Anderson, O. 2011. Genome-wide gene expression analysis supports a developmental model of low temperature gene regulation in winter wheat. BMC Genomics 12, 299.

Law, C.N., Worland, A.J., Giorgi, B., 1976. The genetic control of ear-emergence time by chromosomes 5A and 5D of wheat. Heredity (Edin). 36, 49–58.

Lee, C., Kalmar, L., Xue, B., Tompa, P., Daughdrill, G.W., Uversky, V.N., Han, K.H., 2014. Contribution of proline to the pre-structuring tendency of transient helical secondary structure elements in intrinsically disordered proteins. Biochim. Biophys. Acta - Gen. Subj. 1840, 993–1003.

Lee, H., Xiong, L., Gong, Z., Ishitani, M., Stevenson, B., Zhu, J.-K., 2001. The *Arabidopsis HOS1* gene negatively regulates cold signal transduction and encodes a RING finger protein that displays cold-regulated nucleo – cytoplasmic partitioning. Genes Dev. 15, 912–924.

Lee, W.K., Ahn, H.J., Yu, Y.G., Nam, H.W., 2014. Rhoptry protein 6 from Toxoplasma gondii is an intrinsically disordered protein. Protein Expr. Purif. 101, 146–151.

Limin, A.E., Fowler, D.B., 2006. Low-temperature tolerance and genetic potential in wheat (*Triticum aestivum* L.): Response to photoperiod, vernalization, and plant development. Planta 224, 360–366.

Linding, R., Schymkowitz, J., Rousseau, F., Diella, F., Serrano, L., 2004. A comparative study of the relationship between protein structure and β-aggregation in globular and intrinsically disordered proteins. J. Mol. Biol. 342, 345–353.

Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K., 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. Plant Cell 10, 1391–406.

Mahfoozi, S., Limin, A. E., Hayes, P. M., Hucl, P., Fowler, D. B. 2000. Influence of photoperiod response on the expression of cold hardiness in wheat and barley. Can J Plant Sci 80, 721–24.

Manning, M., Colón, W., 2004. Structural basis of protein kinetic stability: Resistance to sodium dodecyl sulfate suggests a central role for rigidity and a bias toward β-sheet structure. Biochemistry 43, 11248–11254.

Martin, A., Kather, I., Schmid, F.X., 2002. Origins of the high stability of an in vitro-selected coldshock protein. J. Mol. Biol. 318, 1341–1349.

Martin, A., Sieber, V., Schmid, F.X., 2001. In-vitro selection of highly stabilized protein variants with optimized surface. J. Mol. Biol. 309, 717–726.

Miller, A.K., Galiba, G., Dubcovsky, J., 2006. A cluster of 11 CBF transcription factors is located at the frost tolerance locus Fr-Am2 in *Triticum monococcum*. Mol. Genet. Genomics 275, 193–203.

Miura, K., Furumoto, T., 2013. Cold signaling and cold response in plants. Int. J. Mol. Sci. 14, 5312–5337.

Miura, K., Jin, J.B., Lee, J., Yoo, C.Y., Stirm, V., Miura, T., Ashworth, E.N., Bressan, R. a, Yun, D.-J., Hasegawa, P.M., 2007. SIZ1-mediated sumoylation of ICE1 controls CBF3/DREB1A expression and freezing tolerance in Arabidopsis. Plant Cell 19, 1403–1414.

Mohseni, S., Che, H., Djillali, Z., Dumont, E., Nankeu, J., Danyluk, J., 2012. Wheat CBF gene family: identification of polymorphisms in the CBF coding sequence. Genome 55, 865–881.

Nakano, T., Suzuki, K., Fujimura, T., Shinshi, H., 2006. Genome-Wide Analysis of the ERF Gene Family. Plant Physiol. 140, 411–432.

Nalam, V.J., Vales, M.I., Watson, C.J.W., Kianian, S.F., Riera-Lizarazu, O., 2006. Map-based analysis of genes affecting the brittle rachis character in tetraploid wheat (*Triticum turgidum* L.). Theor. Appl. Genet. 112, 373–381.

Novak, H.R., Sayer, C., Panning, J., Littlechild, J.A., 2013. Characterisation of an l-Haloacid Dehalogenase from the Marine Psychrophile Psychromonas ingrahamii with Potential Industrial Application. Mar. Biotechnol. 15, 695–705.

Novillo, F., Medina, J., Salinas, J. 2007. Arabidopsis CBF1 and CBF3 have a different function than CBF2 in cold acclimation and define different gene classes in the CBF regulon. Proc Natl Acad Sci USA 104, 21002–21007.

Ohshida, T., Hayashi, J., Satomura, T., Kawakami, R., Ohshima, T., Sakuraba, H., 2016. First characterization of extremely halophilic 2-deoxy-D-ribose-5-phosphate aldolase. Protein Expr. Purif. 126, 62–68.

Oikawa, T., Kazuoka, T., Soda, K., 2003. Paradoxical thermostable enzymes from psychrophile: Molecular characterization and potentiality for biotechnological application. J. Mol. Catal. B Enzym. 23, 65–70.

Preece, C., Livarda, A., Christin, P.A., Wallace, M., Martin, G., Charles, M., Jones, G., Rees, M., Osborne, C.P., 2016. How did the domestication of Fertile Crescent grain crops increase their yields? Funct. Ecol. 31, 387–397.

Pudney, P.D.A., Buckley, S.L., Sidebottom, C.M., Twigg, S.N., Sevilla, M.P., Holt, C.B., Roper, D., Telford, J.H., McArthur, A.J., Lillford, P.J., 2003. The physico-chemical characterization of a boiling stable antifreeze protein from a perennial grass (Lolium perenne). Arch. Biochem. Biophys. 410, 238–245.

Rae, B.P., Elliott, R.M., 1986. Characterization of the mutations responsible for the electrophoretic mobility differences in the NS proteins of vesicular stomatitis virus New Jersey complementation group E mutants. J. Gen. Virol. 67, 2635–2643.

Rao, L., Zhao, X., Pan, F., Li, Y., Xue, Y., Ma, Y., Lu, J.R., 2009. Solution behavior and activity of a halophilic esterase under high salt concentration. PLoS One 14, 4(9)

Rashid, M., Guangyuan, H., Guangxiao, Y., Hussain, J., Xu, Y., 2012. AP2/ERF transcription factor in rice: Genome-wide canvas and syntenic relationships between monocots and udicots. Evol. Bioinforma.8, 321–355.

Rath, A., Cunningham, F., Deber, C.M., 2013. Acrylamide concentration determines the direction and magnitude of helical membrane protein gel shifts. Proc. Natl. Acad. Sci. USA. 110, 15668–15673.

Rath, A., Deber, C.M., 2013. Correction factors for membrane protein molecular weight readouts on sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Anal. Biochem. 434, 67–72.

Rath, A., Glibowicka, M., Nadeau, V.G., Chen, G., Deber, C.M., 2009. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. Proc. Natl. Acad. Sci. USA 106, 1760–1765.

Ratnayaka, I., Båga, M., Fowler, B.D., Chibbar, R.N., 2005. Construction and Characterization of a BAC Library of a Cold-Tolerant Hexaploid Wheat Cultivar. Crop Sci. 45, 1571–1577.

Receveur-Bréhot, V., Bourhis, J.M., Uversky, V.N., Canard, B., Longhi, S., 2006. Assessing protein disorder and induced folding. Proteins Struct. Funct. Genet. 62, 24–45.

Rousseau, F., Schymkowitz, J., Serrano, L., 2006. Protein aggregation and amyloidosis: Confusion of the kinds? Curr. Opin. Struct. Biol. 16, 118–126.

Sandve, S.R., Fjellheim, S., 2010. Did gene family expansions during the Eocene-Oligocene boundary climate cooling play a role in Pooideae adaptation to cool climates? Mol. Ecol. 19, 2075–2088.

Schägger, H., 2006. Tricine–SDS-PAGE. Nat. Protoc. 1, 16–22.

Schägger, H., von Jagow, G., 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal. Biochem. 166, 368–379.

Schlegel, R. H. J. 2013. Rye: Genetics, Breeding, and Cultivation. CRC Press, Taylor and Francis Group, Florida, USA.

Schmid, F.X., 2011. Lessons about Protein Stability from in vitro Selections. Chem Bio-Chem 12, 1501–1507.

Shewry, P.R., 2009. Wheat. J. Exp. Bot. 60, 1537–1553.

Shi, Y., Mowery, R.A., Ashley, J., Hentz, M., Ramirez, A.J., Bilgicer, B., Slunt-Brown, H., Borchelt, D.R., Shaw, B.F., 2012. Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding. Protein Sci. 21, 1197–1209.

Siddiqui, K.S., Cavicchioli, R., 2006. Cold-Adapted Enzymes. Annu. Rev. Biochem. 75, 403–433.

Sieber, A.-N., Longin, C.F.H., Leiser, W.L., Würschum, T., 2016. Copy number variation of CBFA14 at the Fr-A2 locus determines frost tolerance in winter durum wheat. Theor. Appl. Genet.129, 1087-1097.

Skinner, J.S., Von Zitzewitz, J., Szucs, P., Marquez-Cedillo, L., Filichkin, T., Amundsen, K., Stockinger, E.J., Thomashow, M.F., Chen, T.H.H., Hayes, P.M., 2005. Structural, functional, and phylogenetic characterization of a large CBF gene family in barley. Plant Mol. Biol. 59, 533–551.

Snape, J.W., Law, C.N., Parker, B.B., Worland, A.J., 1985. Genetical analysis of chromosome 5A of wheat and its influence on important agronomic characters. Theor. Appl. Genet. 71, 518–526.

Somero, G.N., 1995. Proteins and Temperature. Annu. Rev. Physiol. 57, 43–68.

Stockinger, E.J., Gilmour, S.J., 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. Proc. Natl. Acad. Sci. USA. 94, 1035–40.

Struvay, C., Negro, S., Matagne, A., Feller, G., 2013. Energetics of protein stability at extreme environmental temperatures in bacterial trigger factors. Biochemistry 52, 2982–2990.

Survila, M., Heino, P., Palva, E. T. 2010. Genes and gene regulation for low-temperature tolerance. In: Jenks, M. A., Wood, A. J. (eds.), Genes for Plant Abiotic Stresses. Wiley- Blackwell, pp. 185–219.

Sutka, J. 1994. Genetic control of frost tolerance in wheat (*Triticum aestivum* L.). Euphytica 77, 277-282.

Takano, E., Maki, M., Mori, H., Hatanaka, M., Marti, T., Titani, K., Kannagi, R., Ooi, T., Murachi, T., 1988. Pig heart calpastatin: identification of repetitive domain structures and anomalous behavior in polyacrylamide gel electrophoresis. Biochemistry 27, 1964–1972.

Talbert, L.E., Smith, L.Y., Blake, M.K., 1998. More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. Genome 41, 402–407.

Tantos, A., Friedrich, P., Tompa, P., 2009. Cold stability of intrinsically disordered proteins. FEBS Lett. 583, 465–469.

Tantos, A., Han, K.-H., Tompa, P., 2012. Intrinsic disorder in cell signaling and gene transcription. Mol. Cell. Endocrinol. 348, 457–465.

Tantos, A., Szrnka, K., Szabo, B., Bokor, M., Kamasa, P., Matus, P., Bekesi, A., Tompa, K., Han, K.H., Tompa, P., 2013. Structural disorder and local order of hNopp140. Biochim. Biophys. Acta - Proteins Proteomics 1834, 342–350.

Thomashow, M.F., 1999. Plant cold acclimation: Freezing Tolerance Genes and Regulatory Mechanisms. Annu. Rev. Plant Physiol. Plant Mol. Biol. 50, 571–599.

Thomashow, M.F., 2010. Molecular basis of plant cold acclimation: insights gained from studying the CBF cold response pathway. Plant Physiol. 154, 571–577.

Tondelli A., Francia E., Barabaschi D., Pasquariello M., Pecchioni, N. 2011. Inside the CBF locus in *Poaceae*. Plant Science 180, 39–45.

Tóth, B., Galiba, G., Fehér, E., Sutka, J., Snape, J. W. 2003. Mapping genes affecting flowering time and frost resistance on chromosome 5B of wheat. Theor Appl Genet 107, 509–514.

Uemura, M., Steponkus, P.L., 1999. Cold Acclimation in Plants: Relationship Between the Lipid Composition and the Cryostability of the Plasma Membrane. J. Plant Res. 112, 245–254.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. 2012. Primer3—new capabilities and interfaces. Nucleic Acids Research, 40, e115.Uversky, V.N., 2002. What does it mean to be natively unfolded? - Uversky - 2003 - European Journal of Biochemistry - Wiley Online Library. Eur. J. Biochem. 12, 2–12.

Uversky, V.N., Dunker, A.K., 2010. Understanding Protein Non-Folding. Biochim. Biophys. Acta 1804, 1231–1264.

Vágújfalvi, A., Aprile, A., Miller, A., Dubcovsky, J., Delugu, G., Galiba, G., Cattivelli, L., 2005. The expression of several Cbf genes at the Fr-A2 locus is linked to frost resistance in wheat. Mol. Genet. Genomics 274, 506–514.

Vogel, J.T., Zarka, D.G., Van Buskirk, H.A., Fowler, S.G., Thomashow, M.F., 2005. Roles of the CBF2 and ZAT12 transcription factors in configuring the low temperature transcriptome of Arabidopsis. Plant J. 41, 195–211.

Wang, B., Wang, A., Cao, Z., Zhu, G., 2016. Characterization of a novel highly thermostable esterase from the Gram-positive soil bacterium *Streptomyces lividans* TK64. Biotechnol. Appl. Biochem.63, 334-343.

Winfield, M.O., Lu, C., Wilson, I.D., Coghill, J. a, Edwards, K.J., 2009. Cold- and light-induced changes in the transcriptome of wheat leading to phase transition from vegetative to reproductive growth. BMC Plant Biol. 9, 55.

Wise, M.J., Tunnacliffe, A., 2004. POPP the question: What do LEA proteins do? Trends Plant Sci. 9, 13–17

Worland, T, Snape J.W. 2001. Genetic basis of worldwide wheat varietal improvement. In Bonjean, A.P. and Angus, W.J.(eds) The World Wheat Book- A history of wheat Breeding. Lavoisier Publishing Paris, France. pp 59- 103.

Xiong, L., Lee, H., Ishitani, M., Tanaka, Y., Stevenson, B., Koiwa, H., Bressan, R.A., Hasegawa, P.M., Zhu, J.-K., 2002. Repression of stress-responsive genes by FIERY2, a novel transcriptional regulator in Arabidopsis. Proc. Natl. Acad. Sci. USA. 99, 10899–904.

Xue, G-P. 2003. The DNA-binding activity of an AP2 transcriptional activator *Hv*CBF2 involved in regulation of low-temperature responsive genes in barley is modulated by temperature. Plant J 33, 373–383.

Yamaguchi-Shinozaki, K., Shinozaki, K., 1994. A novel cis-acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. Plant Cell 6, 251–64.

Zhao, C., Zhang, Z., Xie, S., Si, T., Li, Y., Zhu, J-K. 2016. Mutational Evidence for the Critical Role of CBF Transcription Factors in Cold Acclimation in Arabidopsis. Plant Physiol 171, 2744–2759.

Zhen, Y., Dhakal, P., Ungerer, M.C., 2011. Fitness benefits and costs of cold acclimation in Arabidopsis thaliana. Am. Nat. 178, 44–52.

Zhen, Y., Ungerer, M.C., 2008. Relaxed selection on the CBF/DREB1 regulatory genes and reduced freezing tolerance in the southern range of Arabidopsis thaliana. Mol. Biol. Evol. 25, 2547–2555.

Zhu, J., Pearce, S., Burke, A., Robert, D., Daniel, S., Dubcovsky, J., Garland, K., 2014. Copy number and haplotype variation at the VRN‑A1 and central FR‑A2 loci are associated with frost tolerance in hexaploid wheat. Theor Appl Genet. 127, 1183–1197.

Appendix 1 List of oligonucleotide primers and their use.

Template	F/R primer	DNA sequence	Amplicon size (bp)	Annealing temp (°C)	Amplicon use
BAC clone 3149-L3-CBF12	CBF12.2F CBF12.2R	**GACGACGACAAG**ATGGACACGGGCCCGG **GAGGAGAAGCCCGGT**TCAGTGGCTCCATAGCACCG	765	60	Fragment for pET-32Ek/LIC cloning to produce pTrxHisSCBF12*, converted in this thesis to pHis-CBF12.
BAC clone 3149-L3-CBF12	L3_CBF12_1F CBF12.2R	**GACGACGACAAG**ATCCCGCCGTCCTCG **GAGGAGAAGCCCGGT**TCAGTGGCTCCATAGCACCG	726	60	Fragment for pET-32Ek/LIC cloning to produce pTrxHisSCBF12-N1, converted in this thesis to pHis-CBF12-N1.
BAC clone 3149-L3-CBF12	L3_CBF12_2F CBF12.2R	**GACGACGACAAG**ATGCCGACGTCGCCTG **GAGGAGAAGCCCGGT**TCAGTGGCTCCATAGCACCG	699	60	Fragment for pET-32Ek/LIC cloning to produce pTrxHisSCBF12-N2, converted in this thesis to pHis-CBF12-N2.
BAC clone 3149-L3-CBF12	L3_CBF12_3F CBF12.2R	**GACGACGACAAG**ATGCCGAAGCGCCCCG **GAGGAGAAGCCCGGT**TCAGTGGCTCCATAGCACCG	675	63	Fragment for pET-32Ek/LIC cloning to produce pTrxHisSCBF12-N3, converted in this thesis to pHis-CBF12-N3.
BAC clone 3149-L3-CBF12	L3_CBF12_4F CBF12.2R	**GACGACGACAAG**ATGTTCCACGGCGTGCG **GAGGAGAAGCCCGGT**TCAGTGGCTCCATAGCACCG	624	63	Fragment for pET-32Ek/LIC cloning to produce pTrxHisSCBF12-N4, converted in this thesis to pHis-CBF12-N4.
pHis-CBF12	L3_CBF12_701F L3_CBF12_701R	GGCGAGCT**A**GGAGCGCGGAGACGATGCGGCGCTATG CTCCGCGCTCC**T**AGCTCGCCCCGCTGGACG	6504	72	Q5 mutagenesis. Introduction of stop codon to generate pHis-CBF12-C1.
BAC clone 3149-L3-CBF12	CBF12.2F CBF12.2-T2R	**GACGACGACAAG**ATGGACACGGGCCCGG **GAGGAGAAGCCCGGTTA**GTACGTGCCCGC	665	60	Fragment for pET-32Ek/LIC cloning to produce pTrxHisSCBF12-T2*, converted in this thesis to pHis-CBF12-C2.
pHis-CBF12	L3_CBF13_616F L3_CBF12_616R	CACACGTCGGGC**T**AAATGGACGCGGGCACGTACTACGCG CCCGCGTCCATTT**A**GCCCGACGTGTGGAGCTCGAACATGTCG	6504	72	Q5 mutagenesis. Introduction of stop codon to generate pHis-CBF12-C3.
BAC clone 3149-L3-CBF12	CBF12.2F CBF12.2-T3R	**GACGACGACAAG**ATGGACACGGGCCCGG **GAGGAGAAGCCCGGT**CATGGCGGCTGTGA	575	60	Fragment for pET-32Ek/LIC cloning to produce pTrxHisSCBF12-T3*, converted in this thesis to pHis-CBF12-C4.
pHisCBF12	L3_CBF12_494F L3_CBF12_494R	CCGCCGCGT**A**GGCGCCAGACAATGCCAGCTCG GTCTGGCGCC**T**ACGCGGCGGAGGCGATCCC	6504	72	Q5 mutagenesis. Introduction of stop codon to generate pHis-CBF12-C5.

*Jain et al. 2013

82

Template	F/R primer	DNA sequence	Amplicon size (bp)	Annealing temp (°C)	Amplicon use
BAC clone 3149-L3-CBF12	CBF12.2F CBF12.2-T1R	**GACGACGACAAG**ATGGACACGGGCCCGG **GAGGAGAAG**CCCGGT**T**AGAGGGAGGAGGGCAC	384	60	Fragment for pET-32Ek/LIC cloning to give pTrxHisS-CBF12-T1*, converted in this thesis to pHis-CBF12-C6.
pHisCBF12	CBF12_344_FW CBF12_344_RV	ACTCGGCGTGGCTGC**G**C**TAGC**TGCCCTCCTCCCTCT AGAGGGAGGAGGGCA**GCTA**G**C**GCAGCCACGCCGAGT	6504	65	GeneArt mutagenesis. Introduction of NheI site and stop codon after CMIII-1 resulting in pHisCBF12-C7.
BAC clone 3149-L3-CBF12	CBF12.2F CBF12.2-T5R	**GACGACGACAAG**ATGGACACGGGCCCGG **GAGGAGAAGCCCGGTTA**GGGGTAGTTGAGGC	356	65	Fragment for pET-32Ek/LIC cloning to give pTrxHisS-CBF12-T5*, converted In this thesis to pHis-CBF12-C8.
BAC clone 3149-L3-CBF12	CBF12.2F CBF12.2-T6R	**GACGACGACAAG**ATGGACACGGGCCCGG **GAGGAGAAGCCCGGTTA**CTCGCCGCGCTTCC	248	60	Fragment for pET-32Ek/LIC cloning to give pTrxHisS-CBF12-T6*, converted in this thesis to pHis-CBF12-C9
BAC clone 3149-L3-CBF12	CBF12.2F CBF12.2-T7R	**GACGACGACAAG**ATGGACACGGGCCCGG **GAGGAGAAGCCCGGTTA**CGGGTGGCGCGTCTCCTT	170	60	Fragment for pET-32Ek/LIC cloning to give pTrxHisS-CBF12-T7*, converted in this thesis to pHis-CBF12-C10.
BAC clone 3149-L3-CBF12	CBF12.2F L3_CBF12_1R	**GACGACGACAAG**ATGGACACGGGCCCGG **GAGGAGAAGCCCGGTT**ACGCAGGCGACGTCG	113	60	Fragment for pET-32Ek/LIC cloning to give pTrxHisS-CBF12-R1, converted in this thesis to pHis-CBF12-C11.
pHis-CBF12	CBF12_88_FW CBF12_88_RV	ACGTCGCCTGCGTCG**G**C**TAGC**CCGAAGCGCCCCGCG CGCGGGGCGCTTCGG**GCTA**G**C**CGACGCAGGCGACGT	6504	72	GeneArt mutagenesis. Introduction of NheI site prior to CMIII-3 DNA sequence to generate MutCMIII-3. MutCMIII-3 was also used to create pHis-CBF12- DelAP2.
pHis-CBF12	CBF12_329F CBF12_329R	CTACCCCG**GG**TCGGCGTGGCTGCTCGCCGTGCCCTC GCAGCCACGCCGA**CC**CGGGGTAGTTGAGGCGC	6504	65	Q5 mutagenesis. Introduction of SmaI site at CMIII-3 to produce pHis-CBF12-MutCMIII-1.
pHisCBF12-MutCMIII-1	L3_CBF12_143F L3_CBF12_143R	CCACCCGG**G**GTTCCACGGCGTGCGCCGTC GCCGTGGAAC**C**CCGGGTGGCGCGTCTCCTTG	6504	65	Q5 mutagenesis. Introduction of SmaI site after CMIII-3. Intermediate clone for the generation of pHis-CBF12-DelCMIII-3.
pHis-CBF12	CBF12_322_FW CBF12_322_RV	GCCGCGCGCCTCAAC**GCTAG**CGACTCGGCGTGGCT AGCCACGCCGAGTCG**CTAGC**GTTGAGGCGCGCGGC	6504	65	GeneArt mutagenesis. Introduction of NheI site before CMIII-1 DNA sequence. Resulting construct was used to create pHis-CBF12-DelAP2.

*Jain et al. 2013