Molecular characterization of *Cyclophilin B* genes and promoter sequences in wheat and rice

Wu H, Bhave M
Environment and Biotechnology Centre, Faculty of Life and Social Sciences, Swinburne University of Technology, VIC 3122, Australia

INTRODUCTION

The storage proteins (prolamins) of wheat endosperm are commonly grouped into the monomeric gliadins and polymeric glutenins, the latter comprised of low and high molecular weight (LMW and HMW) subunits[1, 2]. This forms the highly structured protein network (gluten) that determines the viscoelastic properties of the flour which affect its end-uses[1, 2, 3]. The storage protein bonding, folding, assortment and deposition processes in the endosperm imply significant roles for ER-localised chaperones and 'foldase' enzymes, and three groups of proteins are likely involved: the enzymes protein disulphide isomerases (PDI), peptidyl-prolyl *cis-trans* isomerases (PPIases), and the Binding Protein (BiP)/heat shock protein (Hsp70) chaperones[4, 5]. The cyclophilins (CyPs), belonging to PPIase class, are of particular interest as preliminary evidence shows their involvement in seed storage protein processes[6, 7]. The ER-localised Cyps (commonly called CypB) are of particular relevance due to the storage protein folding and assortment processes for both transport pathways initiating and/or occurring in the ER. However, little genetic information exists on genes encoding this 'foldase' enzyme, and on the flanking regulatory sequences which might affect the timing, tissue specificity or transcription rates. This work reports on the *CypB* genes and their promoters in wheat.

MATERIAL AND METHODS

Various primers were designed based on the Tentative Consensus (TC264488) sequence of a putative *CypB* gene (Table 1, Fig 1a). Three overlapping gene sections were amplified from the genomic DNA (gDNA) of *T. aestivum* cv. Rosella using the primer pairs WC1-CYPBR5, WC4-CYPBR1, WC1-CYPBR9. Fragment WC1-CYPBR5 from *T. urartu*, WC1-CYPBR9 from *Ae. tauschii*, and WC1-CYPBR9 from cDNA of *T. aestivum* cv. Cranbrook were also amplified. In order to optimize the possibility of isolating all homoeologous copies, 6-10 clones of each ligation were selected randomly and their plasmids prepared. The inserts were amplified by vector-based primers T7 and SP6 and subjected to preliminary detection of restriction fragment length polymorphisms (RFLP) by *Hae*III. Clones representing various RFLP types were sequenced and sequence analysis was conducted with Bioedit sequence alignment editor v7.0.5 (http://mbio.ncsu.edu/BioEdit). The flanking regions of *CypB* genes were amplified by IPCR. The enzyme (*SacI*) was chosen for digestion of gDNA of *T. uratu* and *Ae. tauschii*, as it had no sites in the *CypB* gene of *T. urartu* and one site in an intron of the C*ypB* gene of *Ae. tauschii* (this work; data not shown). IPCR amplification was carried out with CYPBF3/CYPBIPCR-R1 (Table 1) from *SacI* digested and ligated gDNA. The PCR products were cloned and sequenced.

The bacterial artificial chromosome (BAC) clone containing the putative *CypB* gene of rice was identified by a BLASTn search of TIGR (The Institute for Genome Research) Rice Genome Annotation Database Assembly 2006 (http://www.tigr.org/tdb/e2k1/osa1/) using TC264488 as query sequence. The structure of the rice *CypB* gene was then analysed by alignment with the experimentally isolated *CypB* genes of common wheat. The 'Rice Genome Browser' interface of the TIGR database was used to find flanking region upstream of ATG as the putative promoter sequence.

Table 1: Primer pairs used for PCR or IPCR amplification of *CypB* gene and its promoter

Forward primer	Reverse primer
WC1 5'ATACGATCCAAG**ATG**GCG 3'	CYPBR5 5'CTGTGGAACGAACTCCCCTT3'
WC4 5'CGCACAAGGTCTACTTCGA3'	CYPBR1 5'CTGTCGGCAATGACAACC3'
WC1 5'ATACGATCCAAG**ATG**GCG 3'	CYPBR9 5' ACGGTGGTGTGTGATGCC3'
CYPBF3 5'ACAGGTGAGAAAGGCATGGGC3'	CYPBIPCR-R1 5'CCACGCCCTCATCGCCAT3'

RESULTS AND DISCUSSION

PCR amplification of genomic copies of *CypB* genes from *T. aestivum* cv. Rosella with WC1-CYPBR5 led to a product of size approximately 1.8 kilo base pairs (kb) (data not shown), much larger than the corresponding region of TC264488 (317 base pairs; bp), suggesting the presence of introns. Six plasmid clones (TaA1-TaA6) of this fragment were selected randomly and RFLP analysis with *HaeIII* showed three patterns (data not shown). One representative clone of each (TaA1, TaA4, TaA5) was sequenced. In order to get longer sections, an overlapping fragment (WC4-CYPBR1) was generated from *T. aestivum cv*. Rosella and led to products of ~2.0kb. PCR products of ten clones of this fragment (TaB1-TaB10) were analysis by RFLP (*HaeIII*) and the

representative clones of these types (TaB2, TaB6) were sequenced. Once the sequences of these five clones provided clues to the gene structures, primer CYPBR9 located at the 3'end of the gene, 196bp after the stop codon, was used with WC1 (114bp to 960bp section on TC264488) for amplification of a longer section of *T. aestivum* cv. Rosella. The PCR product was ~2.6kb, and its ten randomly selected clones all showed one identical RFLP type with *HaeIII*. One of these clones (TaC1) was sequenced. In the overlapping sections, TaA1 had 100% identity with TaB6 and TaA4 had 100% identity with TaB2 (data not shown). Two contigs of C*ypB* genes in *T. aestivum* cv. Rosella could thus be generated, with total lengths of 2386bp (TaA1-TaB6) and 2307bp (TaA4-TaB2). The insert of clone TaC1 of fragment WC1-CYPBR9 was 2606bp long and exhibited a 100% identity with contig TaA1-TaB6 in the overlapping area. Three genomic copies of *CypB* gene in *T. aestivum* cv. Rosella were thus identified, represented by (i) clone TaC1; (ii) contig TaA4-TaB2; and (iii) clone TaA5. The cDNA clone (TaCD4) of section WC1-CYPBR9 of cv. Cranbrook was sequenced with length of 851bp.

Fragment WC1-CYPBR5 was amplified from gDNA of *T. urartu* and its sequence showed a length of 1786bp (data not shown) and a high identity to sequences of clone TaA4. Contig TaA4-TaB2 were thus deduced to be from the A genome (*TaCypB-A*). Fragment WC1-CYPBR9 from gDNA *Ae. tauschii* also yielded ~2.6kb products. One clone of it (Aet1) was sequenced, showing a length of 2560bp, and high identity to TaA5; TaA5 was deduced to be from D genome (*TaCypB-D).* Thus the sequence of TaC1 was predicted to be from B genome (*TaCypB-B*).

The sequence of the isolated *CypB* cDNA (TaCD4) was compared with the above three *CypB* genomic genes/ contigs for elucidation of the exon/intron structure. *TaCypB-B* and *TaCypB-A* both contained 7 exons and 6 introns, while *TaCypB-D* was partial (Fig. 1). The three genomic genes in common wheat showed exon conservations with no gaps and a few SNPs in exon II, V and VII (data not shown). The major variations were in introns I, II, and IV. Intron I of *TaCypB-B* (242bp) was longer than in *TaCypB-A* and *TaCypB-D* respectively; intron II in *TaCypB-B* (369bp) than that in *TaCypB-A* and *TaCypB-D* respectively, and intron IV of *TaCypB-B* (851bp) was longer than that in *TaCypB-A* and *TaCypB-D*. Intron III was least variable in length. Variation in introns of the three common wheat genes resulted in certain diagnostic restriction sites (Fig. 1b).

The exons of *TaCypB-B* were joined together to create exon contigs for translation purpose. The deduced CypB protein exhibited a length of 213 amino acids, with an estimated molecular weight of ~23kDa (data not shown). It contained the putative N-terminal signal peptide (SP), the seven-amino-acid insertion characteristic plant Cyps, three essential residues (R, F and H) for PPIase catalytic activity, conserved W for CsA binding, and putative unusual C-terminal ER-localization signal EVPL (Fig 1b).

To our knowledge, this is the first complete sequence of a wheat *CypB* gene. The results show a complete retention of all residues responsible for the catalytic activity and immunosuppressant drug binding as reported from human cytosolic CypA and ER-localized CypB[8, 9]. The major variations noticed in the intron sequences provided the basis for physical mapping of the genes, and could be exploited for genetic mapping purposes, or as a marker to test association with grain quality or other QTLs. From the lack of success to amplify the full-length *TaCypB-D* gene of common wheat with the primer CYPBR9, we can deduce that this gene may be partial in common wheat or the sequence in this area may be divergent. The hypothesis can be tested by amplification of 3'- and 5' flanking sequences or by inverse PCR.

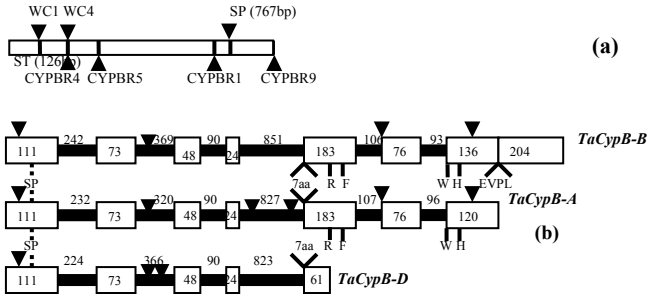

Fig 1: Gene structures of the *CypB* genes isolated from *T. aestivum* cv. Rosella and the *CypB* gene identified in rice.
(a) The putative *TaCypB* cDNA used for designing primers for amplification of *CypB* genes. Arrows pointing down or up indicate forward or reverse primers, respectively.
(b) Schematic representation of the gene structures of the three *TaCypB* genes from *T. aestivum* cv. Rosella. Boxes indicate exons, solid lines between them indicate introns, and the numbers show their lengths in base pair (bp). Arrows indicate restriction sites for *HaeIII*. The signal peptide (SP), the seven-amino-acid insertion, three essential residues (R, F and H) for PPIase catalytic activity, conserved W for CsA binding, and putative unusual ER-localization signal sequence EVPL at the C-terminal of the putative protein product are indicated.

The putative rice *CypB* gene was identified by searching the Rice Genome Annotation Database Assembly 2006 (http://www.tigr.org/tdb/e2k1/osa1/) by BLASTn, using TC264488 as query sequence. The BAC (LOC_Os06g49480 of *Oryza sativa* ssp *japonica cv. Nipponbare*) containing a highly similar sequence on chromosome 6 of the rice genomic assembly was identified. Alignment of *TaCypB-B* with this putative rice *CypB* gene and its coding sequence (CDS) showed a seven-exon structure, with 54.7% overall DNA

sequence identity, with a much higher identity over exons (83%) than introns (30%), and the putative proteins of rice (220 amino acids) and wheat (213 amino acids) share 83.2% identity (data not shown). The high degree of conservation suggests this to be the orthologous rice *CypB* gene.

The 'Rice Genome Browser' interface of TIGR Rice Genome Annotation (http://www.tigr.org/tigr-scripts/osa1_web/gbrowse/rice/; accessed 03/2008) was used to identify sequence upstream of the rice C*ypB* gene. The promoters of *TuCypB-A* in *T. urartu* and *AetCypB-D* in *Ae. tauschii* were amplified by inverse PCR (IPCR), and led to fragments of sizes of ~2.5kb and ~3kb respectively, which were cloned and sequenced. Alignments of the 500bp promoter sequences of *TuCypB-A* and *AetCypB-D* showed 75.9% identity over 500bp region and 88.5% identity over the 370bp region upstream of the start codon. The main regulatory elements were identified in rice and wheat *CypB* promoter sequences. The TATA-box motif (TAAATAAA) for determination of transcription start site at 484bp upstream of the start codon, two CAAT-box motifs downstream of the TATA box, as well as CGCAAT (-276) and CTCAAT (-245) boxes was identified. No TATA-box and CAAT box were identified in promoters of *TuCypB-A* and *AetCypB-D,* whereas the GC-rich motifs critical to transcription initiation were present at -114 in rice *CypB*, -180 in *TuCypB-A*, and -183 in *AetCypB-D*. In addition, the E-boxes (CACCTG) being responsible for tissue specific expression were presented at -169 (rice *CypB*), -133 (*TuCypB-A*), -133 (*AetCypB-D*).

Motifs similar to the AT-rich motif (ATATTTAT/AA/T) were identified at -343 and -489 in rice *CypB* promoter, -314 in *TuCypB-A* promoter and -318 in *AetCypB-D* promoter. The Dof cores (AAAG) for light regulation and tissue specific gene expression appeared frequently, and motifs similar to the endosperm consensus motif (TG(T/A/C)AAA(A/G)(G/T)) considered responsible for the endosperm-specific expression also occurred in all three promoters, e.g. -368 (rice *CypB*), -273 and -205 (*TuCypB-A*), -276 and -208 (*AetCypB-D*). The dyad repeats (CAAN2-9TTG) related to endosperm specific expression were located at -468 (rice *CypB*), and a similar sequence (CAAN12TGG) at position -495 (*TuCypB-A*) and (CAAN11TGG) at position -201 (*AetCypB-D*). The CT-leader boxes (CTCTCTCT) influencing the gene expression quantitatively were present at -49 in rice *CypB*, -34 in *TuCypB-A*, and -44 in *AetCypB-D*. ER stress responsive element (ERSE-II) motif (ATTGG-N-CC(A/C)N(C/G/T) was also identified at position -954. The sequence up to -370bp upstream of start codon thus appears sufficient for transcription initiation of the genes.

The promoter sequences of the putative rice and wheat *CypB* genes contain conserved regulatory sequences such as GC-rich motif critical for initiation of transcription, E-box for tissues specific gene expression, Endosperm motif and dyad repeats for endosperm specific expression, AT-rich motif for seed-specific enhancer and Dof for core for tissues specific expression. These support the tissue-specific up-regulation of *Cyp* genes noted in developing wheat endosperm[6]. The high levels of storage protein expression are suggested to be an internal stress[10], and presence of the ER stress responsive element (ERSE-II) in the rice promoter also supports the suggested role of ER-Cyps in stress response[11, 12].

REFERENCES

1. Shewry PR, Halford NG, Tatham AS, Popineau Y, Lafiandra D, Belton, PS (2003) Adv Food Nutr Res. 45: 219-230.
2. Shewry PR, Nigel GH, Peter SB, Arthur ST (2002) Phil Trans R Soc Lond. B 357: 133-142.
3. Shewry PR, Halford NG (2002) J Exp Bot. 53: 947-958.
4. Galat A (2003) Curr. Top Med Chem 3: 1315-1347.
5. Galili G, Sengupta-Gopalan C, Ceriotti A (1998) Plant Mol Biol 38: 1-29.
6. Grimwade B, Tatham AS, Freedman RB, Shewry PR, Napier JA (1996) Plant Mol Biol 30: 1067-1073.
7. Sheldon PS, Venis MA (1996) Biochem J. 315: 965-970.
8. Pflügl G, Kallen J, Schirmer T, Jansonius JN, Zurini MGM, Walkinshaw MD. (1993) Nature 361: 91-94.
9. Price ER, Zydowsky LD, Jin MJ, Baker CH, McKeon FD, Walsh CT (1991) Proc Natl Acad Sci USA. 88: 1903–1907.
10. Hara-Nishimura I, Matsushima R, Shimada T, Nishimura M (2004) Plant Physiol. 136: 3435-3439.
11. Derkx P, Madrid S (2001) Mol Genet Genom 266: 537-545.
12. Joseph JD, Heitman J, Means AR (1999) Fungal Genet Biol. 27:55-66.