The PDI (Protein Disulfide Isomerase) gene family in wheat

d'Aloisio E[1], Dhanapal A[1], Tanzarella OA[2], Ciaffi M[2], Porceddu E[2]
[1] Sant'Anna School of Advanced Studies, Via Anguillarese 301, 00123 S. Maria di Galeria, Italy. [2] DABAC, Univ. of Tuscia, Via S. Camillo de Lellis, 01100 Viterbo, Italy

INTRODUCTION

The PDI (Protein Disulfide Isomerase) gene family includes several members whose products are responsible for diversified metabolic functions. PDI and PDI-like proteins differ for number and position of thioredoxin-like (TRX-like) active (***a*** type) and inactive (***b*** type) domains, for presence/absence of other domains and of the KDEL signal of retention in the endoplasmic reticulum (ER). The phylogenetic analysis of typical PDI and PDI-like protein sequences resolved them into 10 groups (1), 5 of them (I-V) had 2 TRX-like active domains, whereas the remaining ones owned only a single TRX-like active domain (VI-VIII, QSOX and APRL). In particular, QRX and APRL were not included in this study due to their putative non-isomerase enzymatic activities encoded by an additional domain. The aim of the present research was the study of the complexity and diversity of the PDI gene family in wheat, with particular focus on the genes encoding PDI-like proteins structurally similar to TaPDIL1-1 (group I), the first identified and best characterized member of the PDI family, also named typical PDI. The most important function of typical PDI is the formation and isomerization of disulfide bonds during protein folding, which are accomplished by its two active TRX-like sites sharing the characteristic tetrapeptide –CGHC-. Several studies of molecular characterization, expression analysis and cell localisation in rice and maize have suggested the involvement of typical PDI in the assembly and deposition of storage proteins in these species (2, 3, 4). The characterization and chromosome location of the three homoeologous gene sequences encoding typical PDI and of their promoter sequences have been reported previously (5).

IDENTIFICATION AND CLONING OF WHEAT PDI-LIKE GENE SEQUENCES

A cross search using rice PDI-like sequences in the wheat EST databases "TIGR wheat gene index" (TaGI, version 10) and "HarvEST Wheat" (version 1.13) fetched eight new sequences encoding PDI-like proteins in wheat, whose full length cDNAs (see Table 1) have been cloned through RACE technique from mRNA extracted from developing caryopses of *Triticum aestivum* L. cv Chinese Spring (CS) harvested eight days after anthesis (DAA).

Table 1 – Characteristics of the full-length cDNAs encoding wheat PDI-like proteins. Corresponding tentative consensus sequences (TC), identified in TIGR wheat gene index database, and orthologous rice genes have also been reported.

Clone	Length of sequence after RACE				Tentative consensus sequence (TC)	Orthologous rice gene	Protein identity
	UTR		ORF	PRO			
	5' (nt)	3' (nt)	(nt)	(aa)			
TaPDIL2-1	63	109	1764	588	TC268016	OsPDIL1-4 (AK071514)	422/588 (71,76%)
TaPDIL3-1	141	110	1623	541	TC272054	OsPDIL1-5 (AK073970)	441/541 (81,51%)
TaPDIL4-1	74	139	1101	367	TC264921	OsPDIL2-1 (NM_185280)	317/367 (86,38%)
TaPDIL5-1	55	150	1320	440	TC251262	OsPDIL2-3 (AK062254)	392/440 (89,09%)
TaPDIL6-1	13	255	447	149	TC267704	OsPDIL5-1 (AK063663)	121/149 (81,20%)
TaPDIL7-1	37	289	1239	413	TC236080	OsPDIL5-2 (AK069367)	357/413 (86,44%)
TaPDIL7-2	53	163	1254	418	TC252818	OsPDIL5-2 (AK069367)	280/418 (66,98%)
TaPDIL8-1	122	226	1455	485	TC235895	OsPDIL5-4 (AK099660)	450/485 (92,78%)

PHYLOGENETIC ANALYSIS AND DOMAIN STRUCTURE OF THE DEDUCED AMINO ACID SEQUENCES OF THE PDI-LIKE cDNAs

Based on the comparison of the amino acid sequences of 51 proteins of the PDI family, including the nine deduced PDI and PDI-like sequences of wheat and sequences of several plant species, phylogenetic analysis (Fig. not shown) was performed using the neighbour-joining (NJ) and bootstrap programs of PHYLIP package (version 3.6). The nine PDI and PDI-like sequences of wheat were assigned to the eight phylogenetic groups (I-VIII) identified in plants (1). Seven subfamilies could be split into two major clades, whereas the eighth one was considered as an out-group, due to its high diversification from the other subfamilies. The first major clade contained the I (typical PDI), II and III groups, whose genes encode proteins with two thioredoxin active domains, one at the N-terminal end and the other at the C-terminal end, as well as the VII group, whose members conserved only the N-terminal active domain. On the basis of this shared feature it is possible to put forward the hypothesis of their common evolutionary origin from duplications of a single ancestral gene, which took place in the common progenitor of monocotyledon and dicotyledon plants.

The second clade comprised the remaining three groups (IV, V, VI), whose sequences showed a much higher heterogeneity, which did not allow any inference on their origin and evolution.

The deduced amino acid sequences of the nine PDI and PDI-like cDNAs were searched for conserved motives by comparison with sequences in different protein data bases. A high level of structural similarity was found among the proteins encoded by genes of the same phylogenetic group. As mentioned before, the ***a*** type active TRX-like domains contain the active site tetra-peptide (CXXC), whereas the ***b*** type domains retain only similarities in their tri-dimensional structure from the original TRX domain, but not the active site. The signal peptide, responsible for ER translocation, was present in all proteins, except TaPDIL8-1. TaPDIL1-1 (the typical PDI) had a multidomain structure comprising 4 sequential TRX-like domains ***a-b-b'-a'*** and carried both the N-terminal signal peptide as well as the C-terminal KDEL signal (5). TaPDIL2-1 and TaPDIL3-1, whose multidomain organization was similar to that of typical PDI, had an additional N-terminal domain, rich of acidic residues, which showed a putative low-affinity, high capacity calcium-binding site. Moreover, TaPDIL3-1 showed two non-characteristic tetra-peptide active sites (-CERS- and -CVDC-), which probably affect its redox potential and consequently its function. Also the two paralog genes of group VII preserved the same multidomain organization, but lacked the ***a'*** C-terminal domain as well as the KDEL signal. Moreover, the members of group VII were characterized by the presence of a transmembrane segment, which, in the absence of the KDEL signal, could retain the protein in the ER by its anchoring to the membrane. All the known transmembrane PDIs have been shown to be involved in development control. In the nematode *Caenorhabditis elegans*, for example, it has been demonstrated that both the transmembrane region and the active TRX catalytic site are essential for the determination of the body shape (length) and ray morphology (6). TaPDIL4-1 had the signal peptide but lacked the KDEL signal and had the following modular domain organization: ***a°-a-D***, where ***D*** stands for a C-terminal α-helical domain of about 100 aa, whose function is unknown. Recently it has been demonstrated that the C-terminal part of the ***D*** domain is responsible for the ER retention of the PDI-D protein in the amoeba *Dictyostelium discoideum* (7). The modular domain organization of TaPDI5-1 was ***a°-a-b***. TaPDIL5-1 contained the signal peptide and a modified NDEL signal for retention in the ER. TaPDIL6-1 was the smallest of the PDI-like proteins, presenting only the signal peptide, the **a°** domain with a mutated active site tetrapeptide (CKHC) and a modified signal of ER retention EDEL. TaPDIL8-1 contained an active ***a°*** type domain with a mutated tetrapeptide sequence (CYWS), two transmembrane regions, one in the N-terminus and one in the C-terminus, and a C-terminal DUF1692 domain, which is present in several proteins of unknown function, one of which, ERGIC-32 (8), has partially been characterised.

GENOMIC ORGANIZATION AND CHROMOSOME LOCATION

Genomic sequences from *T. aestivum* cv CS were amplified using the same three combinations of primer pairs, designed in the UTR, which had previously been used to amplify the cDNA sequences. The amplified genomic fragments corresponding to the cDNAs TaPDIL2-1 (1939 bp), TaPDIL4-1 (1317 bp), and TaPDIL5-1 (1531 bp) were of 4619 bp, 3834 bp and 5341 bp, respectively; all three genomic fragments were cloned and sequenced. Alignment of the genomic and cDNA sequences of the three genes revealed a perfect nucleotide identity in the overlapping coding regions and the presence of 12 exons in TaPDIL2-1, 11 in TaPDIL4-1 and 9 in TaPDIL5-1. In figure 1, as an example, the genomic structures of TaPDIL4-1 and of its orthologous genes in rice and *Arabidopsis* are represented from ATG to the stop codon. The comparison of the genomic organisations of the three wheat PDI-like sequences (TaPDIL2-1, TaPDIL4-1 and TaPDIL5-1) with their orthologous genes of rice and *Arabidopsis* showed a high level of conservation of their structural features (exon/intron structure, exon length and position of the active sites) among the members of the same phylogenetic group. Most likely such conservation reflects the essential functional role of their encoded proteins.

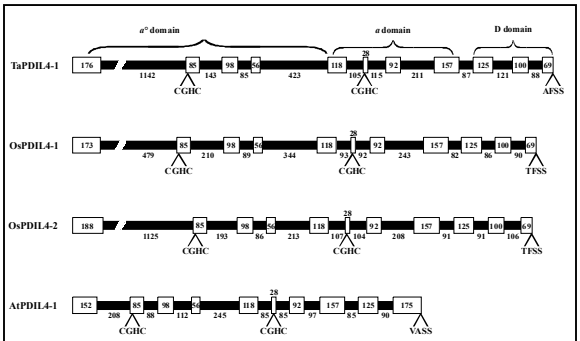

Figure 1 – Intron-exon structure of genes belonging to the IV group of plant PDI gene family. The white boxes indicate exons and the black bars introns, numbers represent exon and intron sizes (bp). The position of the thioredoxin-like catalytic sites are indicated in all the sequences. Exons encoding the two active and inactive thioredoxin domains as well as the D domain in TaPDIL4-1 are labelled by brackets.

The chromosome location of the genes encoding two wheat PDI-like proteins was determined through Southern analyses. The sequence TaPDIL4-1 was located in the short arms of homoeologous group 1 chromosomes, TaPDIL5-1 in the long arms of homoeologous group 5 chromosomes (data not shown).

EXPRESSION ANALYSIS

The expression patterns of three wheat PDI-like sequences (TaPDIL2-1, TaPDIL4-1, TaPDIL5-1) were studied by northern analysis of RNAs from different tissues: coleoptiles and young roots of seedlings (20 days after germination); flag leaves of plants with fully emerged spikes; developing spikes at 5-7 days intervals (from 1-1.5 cm long spikes until full heading); developing caryopses from 4 to 36 DAA. The transcription level of the cloned PDI-like sequences during plant development was compared with that of the typical PDI, previously characterized in wheat (Figure 2). As expected, the mRNA of the typical PDI (TaPDIL1-1) was constitutively expressed in all analysed tissues, with much higher level of transcription in immature caryopses, in particular between 8 and 12 DAA. The expression of typical PDI was relatively high also in the middle stages of spike development, whereas the lowest amount of transcripts was detected in vegetative tissues, such as mature leaves and coleoptiles. Also TaPDIL4-1 was constitutively expressed in all analysed tissues, but the highest amount of transcripts was detected during spike development and in caryopses collected 4-8 DAA. In the vegetative tissues, TaPDIL4-1 resulted, as for typical PDI, slightly more expressed in roots than in flag leaves and coleoptiles. TaPDIL5-1 showed an expression profile similar to TaPDIL4-1, although the overall transcript level resulted lower. TaPDIL5-1 transcripts were absent in coleoptiles and flag leaves, whereas its expression in roots was similar to that in spikes and caryopses. The expression profile of TaPDIL2-1 was similar to that of typical PDI. During the caryopse development the overall level of expression of TaPDIL2-1 was much lower than that of TaPDIL1-1, although the maximum level was between 8 and 12 DAA, like in typical PDI.

CONCLUSIONS AND PERSPECTIVES

Further characterization of the cloned PDI-like sequences will include their quantitative expression analyses by real-time PCR in different plant tissues, including developing spikes and caryopses, young seedlings (20 days old) grown at low (5°C) and high (35°C) temperatures for 24 and 48 h. Evaluation of the copy number per genome and chromosome location by Southern analysis will be extended to the remaining six genes. An exhaustive knowledge of the structural features and regulation of the PDI family genes will be useful to design the most suitable strategies for their functional characterization, in particular for the silencing of single genes or groups of genes through RNA interference (RNAi) technology. The effect on the characteristics of seed storage proteins produced by the progressive knock-out of PDI and PDI-like genes will allow the understanding of their role in the formation of protein aggregates and will highlight possible functional redundancies. The study of the gene family encoding these proteins in wheat is important from the applied viewpoint, for their involvement in determining the technological properties of wheat flour, and it would also be very important for understanding the molecular evolution of this multigenic family in a polyploid context.

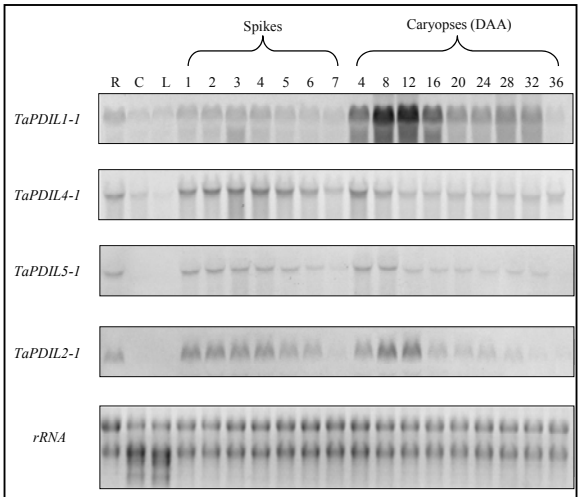

Figure 2 – Northern blot analysis of *TaPDIL1-1*, *TaPDIL4-1*, *TaPDIL5-1* e *TaPDIL2-1* in different wheat tissues: R=roots, C=coleoptiles, L=mature leaves, Spikes at different development stages (1-7) and caryopses collected between 6 and 36 days after anthesis (DAA). The same amount of total RNA (10 μg) was loaded in each track, as shown in the gel stained with ethidium bromide.

REFERENCES

1. Houston NL, Fan C, Xiang QY, Schulze JM, Jung R, Boston RS (2005) Plant Physiol 137:762-778.
2. Li CP, Larkins BA (1996) Plant Mol Biol 30: 873-882.
3. Shimoni Y, Zhu X, Levanoy H, Segal G, Galili G (1995) Plant Physiol 108: 327-335.
4. Takemoto Y, Coughlan SJ, Okita TW, Satoh H, Ogawa M, Kumamaru T (2002) Plant Physiol 128:1212-1222.
5. Ciaffi M, Paolacci AR, d'Aloisio E, Tanzarella OA, Porceddu E (2006) Gene 366: 209-218.
6. Clissold PM, Bicknell R (2003) BioEssays 25: 603-611.
7. Monnat J, Neuhaus E, Pop MS, Ferrari DM, Kramer B, Soldati T (2000) Mol Biol Cell 11:3469-3484.
8. Breuza L, Halbeisen R, Jenö P, Otte S, Barlowe C, Hong W, Hauri HP (2004) J Biol Chem 279(45): 47242-47253.