Studies of genes coding for the Triticum turgidum L. var. durum family of cystein rich proteins

Joudrier P., Lullien V., Dieryck W., Alary R., Gautier M.F.

in

Di Fonzo N. (ed.), Kaan F. (ed.), Nachit M. (ed.).
Durum wheat quality in the Mediterranean region

Zaragoza : CIHEAM
Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 22

1995
pages 263-270

Studies of genes coding for the *Triticum turgidum* L. var. *durum* family of cystein rich proteins

P. JOUDRIER
V. LULLIEN
W. DIERYCK
R. ALARY
M.F. GAUTIER
LABORATOIRE DE TECHNOLOGIE DES CEREALES
INRA
MONTPELLIER
FRANCE

SUMMARY - The proteins which make the major contribution to wheat quality are known as gliadins and glutenins. But recently, low molecular weight cystein rich proteins (LMW-CRP) have also been considered important for some quality parameters. A large number of studies have underlined the importance of SH/SS exchange reactions in the technological quality of wheats. The CRP appear to be good candidates to be implied in these reactions. In addition, it has been shown recently that a new system, NADP-dependent Thioredoxin, reduces both storage proteins and LMW-CRP and then are able to react with other components. After describing the different known members of LMW-CRP, we present results obtained with two families of CRP: the CM-proteins and the Lipid Transfer Proteins. Several cDNAs clones encoding these proteins have been isolated. All these CRP are synthesized as precursors of larger molecular weight, the molecular mass of the mature proteins range from 7 to 16 kDa and they all contain 8 to 10 cystein residues. During seed development, the pattern of accumulation of mRNAs is specific for each family. In order to obtain a substantial amount of proteins for structural analysis (such as RMN or X-ray cristallography) and to study their structure-function relationships, we used two expression systems in *E. coli*. The perspectives of this work is both to have a better knowledge of the in vivo role of these proteins and to prepare the way to transgenic cereals with improved technological properties.

Key words: Durum wheat, cDNA, expression, cystein rich proteins, LTP, CM-proteins, structure-function, technological quality.

RESUME - *"Etude des gènes codant pour des protéines riches en cystéine chez le blé dur (*Triticum durum *Desf.)". Des résultats importants ont été obtenus ces dernières années pour améliorer génétiquement la valeur d'utilisation des blés cependant des progrès sont encore possibles grâce à l'utilisation des techniques de biologie moléculaire. Il est bien établi que les protéines qui contribuent de manière prédominante à la qualité sont les protéines de réserve du grain. Cependant, récemment, il a été mis en évidence que les protéines riches en cystéine pouvaient également jouer un rôle dans la qualité technologique. Ces résultats viennent conforter les très nombreuses études qui ont souligné le rôle des réactions d'échange entre les groupements -SH et l'établissement de ponts disulfures entre les protéines du grain. De plus, des résultats encore plus récents ont mis en évidence l'existence d'un système d'oxydoréduction (Thiorédoxine h NADP-dépendente) capable d'agir à la fois sur les protéines de réserve mais également sur des petites protéines riches en cystéine (= PPRC). Dans ce travail, après avoir présenté les différentes familles de PPRC (purothionines, protéines de transfert de lipides, = LTP, CM-protéines, puroindolines) nous présentons les résultats que nous avons obtenus avec certaines d'entre elles : les LTP et CM, pour lesquelles nous avons isolé différents clones cDNA (caractéristiques et structures primaires des protéines correspondantes), l'étude de l'expression de ces gènes au cours de la maturation de la graine ainsi que la production dans des hôtes hétérologues de ces protéines en vue d'étudier les relations structure-fonction. Les perspectives de ces travaux sont de comprendre le rôle de ces protéines in vivo et dans la qualité technologique afin de préparer la voie de la transgénèse.*

Mots-clés: *Blé dur, cADN, expression, protéines riches en cystéine, LTP, CM-protéines, structure-fonction, qualité technologique.*

Introduction

Substantial improvement in wheat seeds for bread and pasta have been achieved by plant breeding but further advances should be possible from exploitation of the techniques of molecular biology (Joudrier *et al.*, 1987). The proteins which make the major contribution to quality are known as gliadins and glutenins. But, more recently, low molecular weight cysteine rich proteins have been considered as important for the technological quality (Abecassis *et al.*, 1990). A large number of studies have underlined the importance of the SH/SS exchange reaction in the technological quality of wheats. These cysteine rich proteins appear to be good candidates to be implied in such reactions. In addition, recent results have shown that these Low Molecular Weight Cysteine Rich Proteins (=LMWCRP) are reduced by the NADP Thioredoxin System (NTS=NADP-thioredoxin reductase/Thioredoxin h System) (Kobrehel *et al.*, 1991) (cf. Fig. 1). This redox system is able to reduce both storage proteins and LMWCRP.

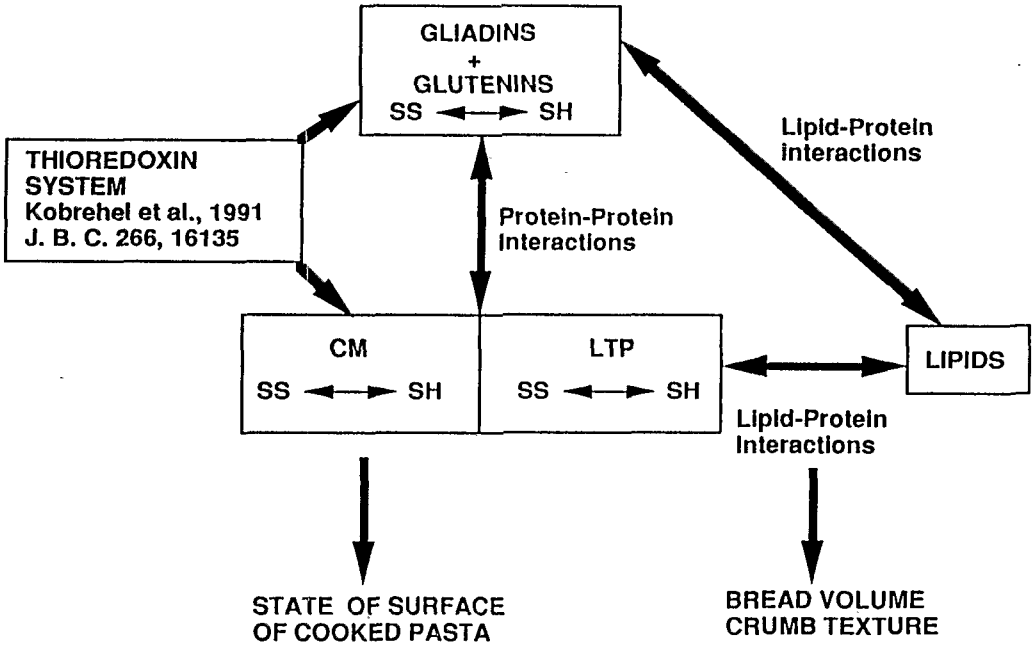

Fig. 1. Potential roles of low molecular weight cysteine rich proteins in wheat technological quality.

After modification of their sulfhydryl status, by the thioredoxin system, the sulfur rich proteins become able to react with other components such as proteins of the gluten network. In our laboratory, we have focused our interest on two families of LMWCRP: the CM proteins (CM because of their solubility in Chloroform/Methanol mixture) and the Lipid Transfer Proteins (=LTP). It has been shown that the CM proteins are involved in the state of surface of cooked pasta and that the LTP could be involved in loaf volume and crumb structure (Kobrehel and Alary, 1989a,b).

The different members of the LMWCRP family are presented in Fig. 2.

Beginning by the smallest molecular size, we find first the:

(i) Purothionins (5 to 5.5 kDa), were isolated half a century ago and their primary structure determined (for about 55 amino acid residues, they have 8 cysteins). Their biological function is not really known but *in vitro*, they have antifungal and bactericidal properties.

(ii) Lipid transfer proteins: the molecular sizes range from 7 to 9 kDa and they have 8 cystein residues. They seem to be ubiquitous and if their *in vivo* function is not known, they are able *in vitro* to transfer lipids from membranes to membranes.

PROTEINS	MOLECULAR WEIGHT		CYSTEIN NUMBER	ROLE
THIONINS	α-1 α-2	5500 Da 5500 Da	8 8	BACTERICIDE FUNGICIDE
LIPID TRANSFER PROTEINS = LTPs	LTP-7 LTP-9	7000 Da 9000 Da	8 8	LIPID TRANSFER
EXOGENOUS α–AMYLASE INHIBITORS	CM1* CM2 CM3 CM16 CM17* O.28 0.19 0.53	12000 TO 15000 Da	10 10 10 10 10 10 10 10	RESISTANCE AGAINST INSECTS
PUROINDOLINES*	a b	13000 Da 13000 Da	10 10	GRAIN FRIABILITY

* Proteins specific of *T. aestivum*

Fig. 2. Low molecular weight cystein rich proteins of wheats.

(iii) CM proteins, the molecular sizes range from 12 to 15 kDa and they have 10 cystein residues. They are known as tetrameric inhibitor against exogenous α-amylase but their role in seeds is presently unknown.

(iv) Inhibitor of α-amylase monomeric (0.28) and dimeric (0.19; 0.53). The molecular weight of the monomer is around 12 kDa and they have also 10 cystein residues.

(v) Puroindolines. The molecular weight is around 13 kDa. They have 10 cystein residues and contain a basic tryptophan rich domain. Their biological function is still unknown but they present very interesting emulsifying properties (Blochet *et al.*, 1993). These proteins were not detected in *T. durum* cultivars and PCR experiments showed that there is no puroindoline genes for all the cultivars tested (Gautier *et al.*, 1993, submitted results).

Strategy

Our general strategy is the following (Joudrier *et al.*, 1987) (Fig. 3).

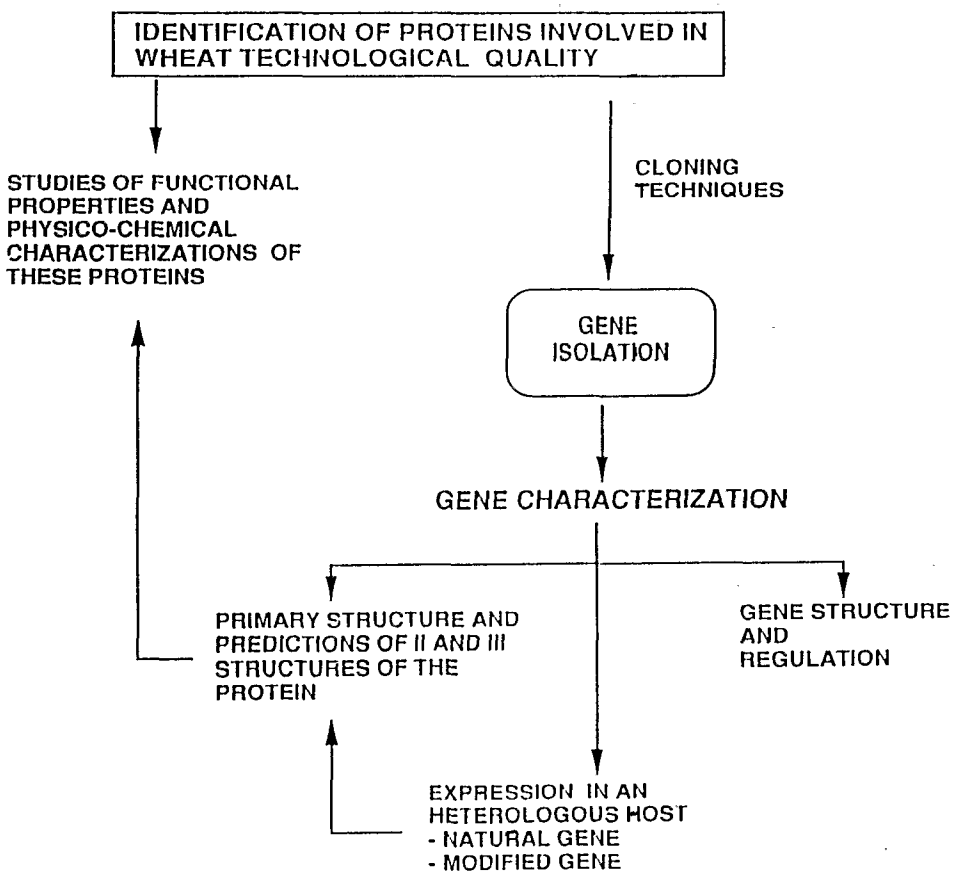

Fig. 3. Molecular biology tools to study the wheat quality.

After identification of a protein involved with a particular parameter of the technological quality of pastamaking (or breadmaking), we isolate the corresponding cDNA or genes to get:

(i) First, at the DNA level, some informations about the gene structure, the copy number, the chromosomal location, in addition we can deduce the primary structure of the corresponding protein.

(ii) Second, with the gene or cDNA we are able to produce the corresponding protein in heterologous hosts (such as bacteria cells, for example) in order to produce sufficient amount to study the tertiary structure of the protein as well as the relationships between structure and function (using site directed mutagenesis techniques).

(iii) Third, studies of the gene regulation mechanisms prepare the way to make transgenic cereals with improved qualities.

Results

The results obtained with the CM proteins and the LTPs from *T. durum* were as follows:

The CM proteins

We have focused our attention on the CM proteins because of the results obtained in our laboratory (Kobrehel and Alary, 1989a,b) showing a relationship between the state of surface of cooked pasta and the presence of a sulfur rich low molecular weight fraction. It has been shown after, that the main components of this fraction are CM proteins (Gautier *et al.*, 1989).

In *T. aestivum*, it exists 5 different CM proteins (CM1, CM2, CM3, CM16, CM17) and three in *T. durum* (CM2, CM3, CM16) (cf. Fig. 4). Similar proteins exist in other cereals like barley. Reconstitution experiments have shown that CM proteins from tetraploid wheat and the same proteins plus CM1 and CM17 from hexaploid wheat are components of the tetrameric α-amylase inhibitors (Sánchez-Monge, unpublished).

WHEAT		BARLEY	
CM1	7Ds	CMa	1H
CM2	7Bs	CMc	1H
CM3	4Bs-4Ds	CMd	4H
CM16	4Bs	CMb	4H
CM17	4Ds	CMe	3H

Fig. 4. Chromosomal locations of CM proteins genes (García-Olmedo *et al.*, 1987; García-Maroto *et al.*, 1990).

From a *T. durum* cDNA library, we have isolated clones coding for the CM proteins: CM16 (Gautier *et al.*, 1990), CM2 (Gautier *et al.*, 1991) and CM3 (Gautier *et al.*, unpublished results). From the DNA sequences, we deduced the primary structure of the corresponding proteins and showed that they contain 10 cystein residues. These proteins are synthesized with a signal peptide.

The study of the steady state level of CM mRNAs during seed maturation showed that they are temporally and spatially regulated. Using aneuploid lines, the chromosomal location of the CM genes has been made and the results are reported on Fig. 4.

Between *T. aestivum* and *T. durum*, a perfect homology, at the protein level, is observed between each kind of CM proteins (Lullien *et al.*, 1991). Assuming that the two species diverged about 8000 years ago, this high conservation during the course of evolution suggests that CM proteins could play an important biological function as known for other very conserved proteins such as histones and enzymes. Similar results have been obtained with the LTPs.

The LTPs

Lipid transfer proteins are supposed to be involved in the transportation of lipids from membranes to membranes. *In vitro*, they are able to catalyse the transfer of lipid between two membranes. Although, no physiological role is known for these proteins, they could be involved in cereal technology. They are supposed to mediate the transport of lipids at the interphase water/air of the dough alveoles. These proteins are ubiquitous, and usually, the active form is a protein of 9 kDa. In wheat, another LTP has been purified (molecular weight=7 kDa) which has a higher activity of lipid transfer.

Following the same strategy used for the CM proteins, we have isolated cDNA clones coding for both the 7 and 9 kDa LTP (Dieryck *et al.*, 1992). These proteins are synthesized as a precursor including a signal peptide. As for the CM proteins, we can observe strong homologies between LTP-9 from *T. aestivum* and *T. durum*. The study of the steady state level of LTP mRNAs during seed maturation and germination revealed that, as for the CM proteins, the expression of both LTP is temporally and spatially regulated except that they appear later during maturation. Moreover, 9 kDa and 7 kDa genes are differently expressed. In durum wheat there are at least three 9 kDa LTP genes and one additional gene is present in bread wheat. Using Chinese spring nullitetrasomic lines, two genes were localized on chromosome 5B and 5D respectively. One intron was identified in 9 kDa LTP genes.

Primary structure analysis of 9 kDa and 7 kDa wheat LTP, in comparison with known secondary and tertiary structural data of the 9 kDa wheat LTP, suggesting the existence of a lipid binding site including two amphiphatic α-helix maintained by a disulphide bridge.

Finally, is also important to note that the cystein skeleton inside and between the CM proteins and the LTPs is highly conserved, as well as for the other members of the LMWCRP, suggesting that the tertiary structure of these proteins is important for their functions. The Fig. 5 is illustrating these data.

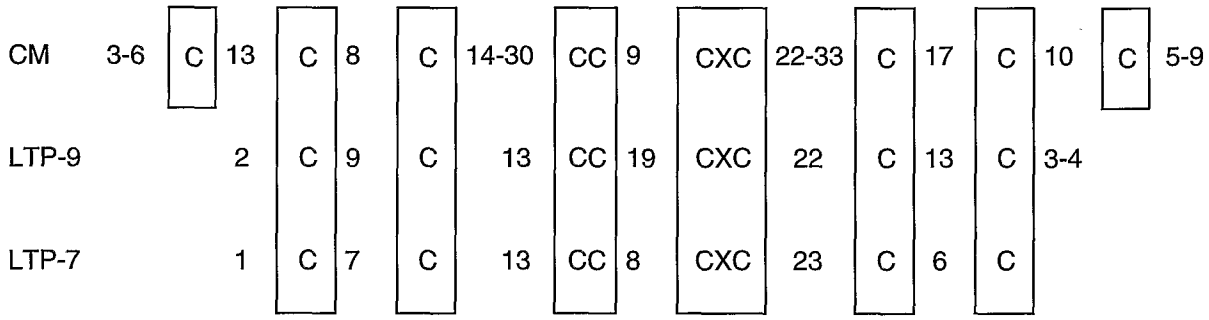

Fig. 5. Cystein skeleton of CM proteins and LTPs.

Expression

A better knowledge of the genes coding for these proteins should have several implications. Following our general strategy described before, we have started to express these proteins using expression vectors. In order to produce sufficient pure wheat proteins for physico-chemical analysis to study the structure and the relationships between structure and function, wheat genes were used to transform *E. coli*. We first started with the clones encoding the *T. durum* CM16 protein and the 9 kDa LTP. Two different systems of expression have been utilized that allow to produce the protein either as an unfused or as a fused product.

For the unfused proteins, we have cloned the cDNA part encoding the mature protein in a pET expression plasmid, under the control of a bacteriophage T7 RNA promoter that allows a high level production of the protein in *E. coli*. Upon bacteria induction with isopropyl thiogalactopyranoside, the recombinant protein accumulates in insoluble inclusion bodies. Solubilization with 6M urea containing 0.5 mM dithiothreitol, followed by slow elimination of the denaturing agents by step dialysis, results in a significant recovery of the recombinant protein in a soluble, monomeric form.

For the fused protein, we are using a system in which the CM protein or the LTP is produced fused with the Maltose Binding Protein (=MBP). This complex is soluble, allowing activity measurements for the LTP, or even structural studies (by X-Ray).

The recombinant proteins are under purification and will be used for different purposes:

(i) Tridimensional structure determination (NMR or X-Ray).

(ii) Mutants have been made (by site directed mutagenesis) to study the structure/function relationships. It is hoped that such information will lead to the design of new genes encoding even better proteins or proteins for other industrial uses (texturing or foaming proteins). These approaches are developed in collaboration with INRA Nantes, CNRS Orléans and CNRS Grenoble.

(iii) Technological assays, at a microscale, will be conducted on "wild" and "mutant" proteins.

Perspectives

(i) The most ambitious approach involves the transfer of new and useful genes to endow seed with novel properties. Medium term benefits from plant biotechnology could result from changing gene regulation or introducing genes from other species to raise the level of proteins which have the required properties (higher nutritional value, better technological properties). But, first, it is important to find out how protein synthesis is regulated during seed development if we want to manipulate protein composition by genetic engineering. To study the regulation of genes coding for these sulfur rich proteins it is necessary to isolate and characterize the corresponding genes and then to perform promoter analysis. Such studies are underway on a barley gene coding for the CMd protein.

(ii) Besides these main aims, considered as a priority for a better knowledge of the wheat technological quality, the isolated genes could be used as probes:

- For the establishment of the wheat genetic map.
- To study the genetic variability (RFLP) of different cultivars and look for "RFLP alleles" in wheat breeding programmes.

References

Abecassis, J., Gautier, M.F. and Autran, J.C. (1990). La filière blé dur - pâtes alimentaires: apports complémentaires de la technologie et de la génétique dans l'amélioration de la qualité. *Ind. Agro Alim.*, pp. 475-482.

Blochet, J.E., Chevalier, C., Forest, E., Pebay-Peyroulac, E., Gautier, M.F., Joudrier, P., Pezolet, M. and Marion, D. (1993). Complete amino acid sequence of puroindoline, a new basic and cystine-rich protein with a unique tryptophan-rich domain, isolated from wheat endosperm by Triton X114 phase partitioning. *FEBS Lett.*, (in press).

Dieryck, W., Gautier, M.F., Lullien, V. and Joudrier, P. (1992). Nucleotide sequence of a cDNA encoding a lipid transfer protein from wheat (*Triticum durum* Desf.). *Plant Mol. Biol.*, 19: 707-709.

García-Olmedo, F., Salcedo, G., Sánchez-Monge, R., Gómez, L., Royo, J. and Carbonero, P. (1987). Plant proteinaceous inhibitors of proteinases and α-amylases. *Oxford Surv. Plant Mol. Cell Biol.*, 4: 275-334.

García-Maroto, F., Marana, C., Mena, M., García-Olmedo, F. and Carbonero, P. (1990). Cloning of a cDNA and chromosomal location of genes encoding the three types of subunits of the wheat tetrameric inhibitor of insect α-amylase. *Plant Mol. Biol.*, 14: 845-853.

Gautier, M.F., Alary, R., Kobrehel, K. and Joudrier, P. (1989). Chloroform/Methanol-Soluble proteins are the main components of *Triticum durum* sulfur-rich glutenin fraction. *Cereal Chem.*, 66: 535.

Gautier, M.F., Alary, R. and Joudrier, P. (1990). Cloning and characterization of a cDNA encoding the wheat (*Triticum durum* Desf.) CM16 protein. *Plant Mol. Biol.*, 14: 313-322.

Gautier, M.F., Alary, R., Lullien, V. and Joudrier, P. (1991). Nucleotide sequence of a cDNA encoding the wheat (*Triticum durum* Desf.) CM2 protein. *Plant Mol. Biol.*, 16: 333-334.

Gautier, M.F., Aleman, M.E., Marion, D. and Joudrier, P. (1993). *Triticum aestivum* puroindolines, two basic cystine-rich seed proteins: cDNA sequence analysis and developmental gene expression. *Plant Mol. Biol.*, (submitted).

Joudrier, P., Autran, J.C. and Gautier, M.F. (1987). Pourquoi cloner les gènes des protéines de réserve du blé. *Biofutur*, 60: 46-54.

Kobrehel, K. and Alary, R. (1989a). The role of two low molecular weight glutenin fraction in the cooking quality of durum wheat pasta. *J. Sci. Food Agric.*, 47: 487-500.

Kobrehel, K. and Alary, R. (1989b). Isolation and partial characterization of two low molecular weight durum wheat glutenins. *J. Sci. Food Agric.*, 48: 441-452.

Kobrehel, K., Yee, B.C. and Buchanan, B.B. (1991). Role of the NADP/thioredoxin system in the reduction of α-amylase and trypsin inhibitor proteins. *J. Biol. Chem.*, 266: 16135-16140.

Lullien, V., Alary, R., Joudrier, P. and Gautier, M.F. (1991). Characterization of a cDNA clone encoding the *Triticum aestivum* L. CM16 protein: homology with the *Triticum durum* Desf. sequence. *Plant Mol. Biol.*, 16: 373-374.