Sharp divergence between wheat and barley at loci encoding novel members of the trypsin/ a-amylase inhibitors family

Pilar Sánchez de la Hoz, Atilio Castagnaro

Pilar Carbonero

Amino acid sequences for three members (CMxl, CMx2, and CMx3) of a new subfamily of trypsin/Ol-amylase inhibitors in wheat have been deduced from the nucleotide sequences of the corresponding cDNAs. A cDNA clone encoding CMx1 was selected from a wheat developing endosperm library using a probe that encoded barley trypsin inhibitor BTI-CMe at low stringency. Sequences corresponding to CMx2 and CMx3 were obtained from cDNA amplified by the polymerase chain reaction. The three CMx sequences contain a premature stop codon after 363 nt, as well as a second stop codon at the same position as in BTI-CMe (nt 439-441). Southern analysis of DNAs from diploid, tetraploid, and hexaploid wheats, as well as from aneuploid lines, indicate that there is a single *CMx* locus in each of the three genomes of hexaploid wheat, respectively associated with chromosomal arms 4AS, 4BS, and 4DL. These genes are expressed early during endosperm development and not expressed at detectable levels in other tissues. Evolutionary implications are discussed.

A number of families of plant proteinaceous inhibitors of hydrolases have been described [2, 8, 17]. One of these families includes cereal inhibitors of trypsin and of heterologous Ol-amylases, which have been implicated in plant defence because of their ability to inhibit insect digestive enzymes [4, 8, 9]. In wheat and barley, these inhibitors are encoded by a complex multigene family of over 20 members, which can be grouped into subfamilies according to inhibitory specificity, sequence similarity, and chromosomal location. There is a general parallelism in chromosomal locations of the genes and the inhibitory specificities of the different members of the family in the different cereal genomes, but there are also striking differences among closely related genomes which suggest both rapid evolutionary gene silencing and specificity changes [4, 8]. An ex-

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X75608 (pCMxl), X75609 (pCMx2) and X75610 (pCMx3).

ample of this is the apparent absence of trypsin inhibitors of this family in wheat, as judged by the absence of inhibitory activity and of proteins recognized by antibodies raised against the barley trypsin inhibitors [6, 13, 14, 15, 16]. Among these, the best characterized is BTI-CMe, which also inhibits the activated Hageman factor (β-factor XIIa of the blood coagulating cascade) and plasma kallikrein [5], but does not inhibit chymotrypsin, pepsin, papain, or α-amylases with the exception of a recently reported genetic variant with a weak activity against the α-amylase of the insect *Tenebrio molitor* [1, 13, 14]. Trypsin inhibitors of this family have also been isolated from rye [10], maize [11], and ragi [3], the last two also being active against α-amylases. Both cDNA and genomic clones corresponding to BTI-CMe have been obtained from barley [6, 16] and a cDNA clone from maize has been reported [19]. We have used a probe encoding barley BTI-CMe to screen at low stringency for possible trypsin inhibitors in hexaploid wheat (genomes AABBDD) and while no such cDNAs have been detected, a novel subset of cDNA sequences of this family, which share a premature stop codon, have been identified. These findings offer some insight into the evolution of this complex family of defence proteins.

A cDNA library from developing wheat endosperm (15 days after pollination) was constructed as previously described [12] and screened under non-stringent conditions, using as a probe a cDNA clone corresponding to BTI-CMe [16]. Clone pCMx1 was thus obtained and its nucleotide sequence had an ORF with a premature stop codon after nt 363, representing a G→U transition in the first base of a Glu codon (GAG) found in the BTI-CMe sequence. The homology between the deduced CMx1 amino acid sequence and that of BTI-CMe was maintained beyond the premature stop codon, up to nt 438, where a second stop codon appeared in the same position as that in the BTI-CMe sequence (Fig. 1). To search for the equivalent gene products of the three different genomes in hexaploid wheat, two nested sense primers, together with the antisense primer 5′-GGATCCTCTAGAGTCGACCTGCAG-3′ (Fig. 1), were used to amplify wheat

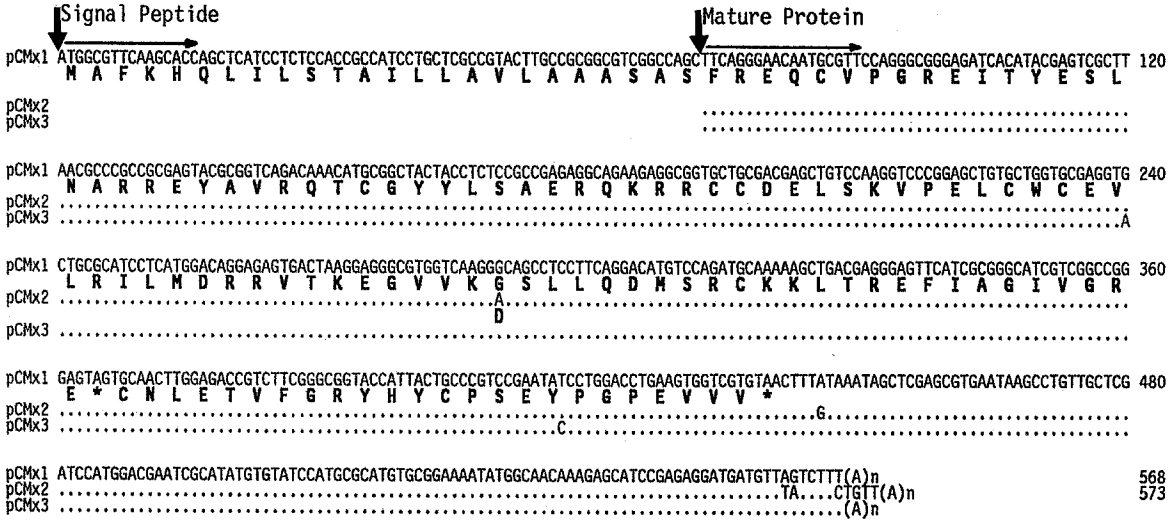

Fig. 1. Nucleotide and deduced amino acid sequences of wheat CMx cDNAs. pCMx1, clone isolated from a developing endosperm cDNA library (15 days after pollination) under non-stringent conditions (58 °C, 5× SSPE, 0.1% SDS, 5× Denhardt's solution, 100 μg/ml salmon sperm DNA) using as a probe a barley BTI-CMe cDNA (EMBL accession number X17302 [16]). pCMx2 and pCMx3, PCR-amplified sequences from a wheat endosperm cDNA. The horizontal arrows indicate the region from which the nested sense primers were derived; the antisense primer was 5′-GGATCCTCTAGAGTCGACCTGCAG-3′. Vertical arrows point the initiation of putative signal peptide and mature protein sequences. Identical nucleotides are indicated with dots; when a difference in nt leads to an aminoacid change it is also indicated. Stop codons are represented by asterisks (*).

endosperm cDNA by the polymerase chain reaction (PCR). Cloned PCR fragments consistently gave two new sequences (CMx2 and CMx3), in addition to the original CMx1 sequence, and both of them were equally interrupted by a premature stop after position 363 (Fig. 1). At least three clones of each type were sequenced, so it was unlikely that the observed differences would be the result of PCR errors. Sequence CMx2 contained one nt change relative to CMx1, at position 290 in the coding region, leading to a G→D amino acid substitution, four nt changes in the 3′ non-coding region, and a TGTT insertion before the poly(A) tail, while CMx3 contained two substitutions in the coding region that did not lead to amino acid changes. An alignment of the deduced CMx amino acid sequences with that of BTI-CMe is presented in Fig. 2A, and percentages of identity for the binary comparisons of different domains (signal peptide, mature protein, 3′ non-coding region), based both in the nucleotide and in the amino acid sequences, are given in Fig. 2B.

Available nulli-tetrasomic lines of hexaploid wheat cv. Chinese Spring (*Triticum aestivum*, genomes AABBDD) were analysed by Southern blot, using CMx1 cDNA as probe (Fig. 3). In nulli-tetrasomics, the lack of a given chromosome pair is compensated by double dosage of another chromosome of the same homeology group. The euploid gave a pattern with three bands (s, small; i, intermediate; l, large) of similar intensity, and only those aneuploids lacking or having extra doses of group 4 chromosomes showed alterations of the pattern: the s band was missing concomitantly with chromosome 4D, the l band disappeared with chromosome 4A, and the relative intensity of the i band was duplicated by doubling the dose of chromosome 4B. The three bands were further assigned to specific arms of group 4 chromosomes (4AS, 4BS, 4DL) by

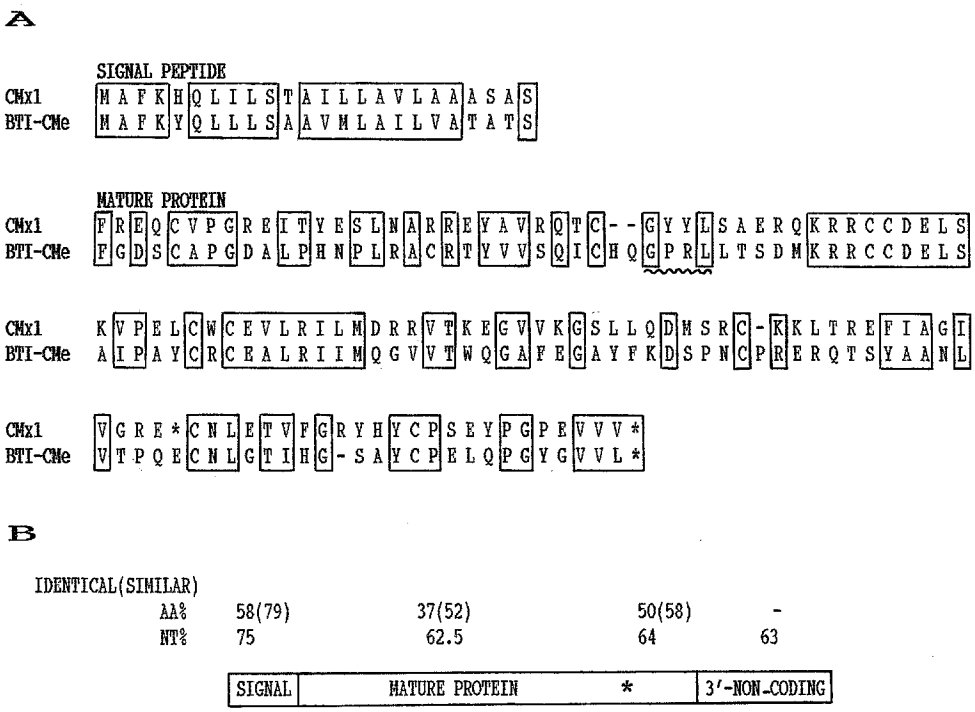

Fig. 2. A. Alignment of the amino acid sequence deduced from clone pCMx1 with that corresponding to barley trypsin inhibitor BTI-CMe. Signal peptides and mature proteins are compared. Identical or similar amino acids are boxed. The reactive site in trypsin inhibitor BTI-CMe is indicated with a wavy line. B. Percentage of identical (similar) residues in the sequences in A. Comparisons corresponding to the mature proteins are split into upstream and downstream regions of the premature stop codon in CMx1.

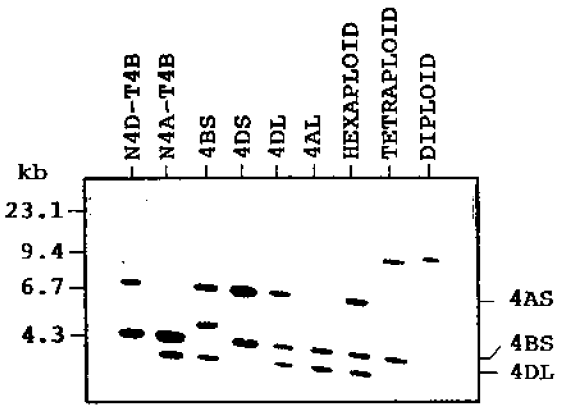

Fig. 3. Chromosomal locations of CMx genes. Southern blot hybrydization of the following genotypes: hexaploid, *Triticum aestivum* cv. Chinese Spring; tetraploid, *Triticum turgidum* cv. Senatore Capelli; diploid, *Triticum monococcum*; nulli-tetrasomic lines derived from cv. Chinese Spring: n4D-t4B lacks chromosome 4D and has 4 doses of chromosome 4B; n4A-t4B lacks chromosome 4A and has 4 doses of chromosome 4B. Ditelosomic lines 4BS, 4DS, 4DL and 4AL, lack respectively the long arm of chromosome 4B, the long arm of 4D, the short arm of 4D and the short arm of 4A. DNAs were digested with the restriction endonuclease *Bam* HI. Electrophoresis was performed in 0.8% agarose at 1 V/cm for 16 h. Clone pCMx1 was used as radioactive probe under stringent conditions.

analysis of appropriate ditelosomic lines, in which a pair of chromosome arms are deleted at a time (Fig. 3). As expected, tetraploid wheat (genomes AABB) had a two-band pattern and diploid wheat (genome AA) showed a single band. The l band, which corresponds to the A genome, was polymorphic in the hexaploid wheat with respect to the diploid and tetraploid wheats analysed (Fig. 3). Band i in the 4BS ditelosomic line was also polymorphic with respect to the euploid, probably because the 4BS chromosome arm has been incorporated from a different wheat cultivar.

Expression of the *CMx* genes, as determined by northern blot analysis, was detected during the first half of endosperm development and was not detected in coleoptiles or roots, either under illumination or in darkness (Fig. 4). Radioactive signals obtained with the CMx probe against wheat endosperm RNA were of similar intensity as those obtained with BTI-CMe probe against the barley RNA under the same conditions. The same filter

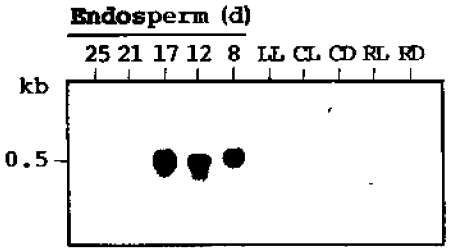

Fig. 4. Northern blot analysis of total RNAs extracted from the indicated organs of hexaploid wheat cv. Chinese Spring. RD, roots from dark-grown 7-day-old seedlings; RL, roots from light-grown 7-day-old seedlings; CD and CL, coleoptiles from dark- and light-grown 7-day-old seedlings; LL, leaves from light-grown 4-week-old plants; 8 to 25, endosperm collected at the indicated days after pollination. Clone pCMx1 was used under stringent conditions (hybridization was at 65 °C, 0.5 M Na_2HPO_4 7% SDS, 1 mM EDTA, 100 μg/ml salmon sperm DNA; washes at 65 °C, 40 mM Na_2HPO_4, 1–5% SDS, 1 mM EDTA). A ribosomal DNA probe was used to ascertain that equal amounts of total RNA were applied in each lane.

was hybridized with the BTI-CMe probe under the same conditions and no signal was detected (not shown). No CMx sequences were detected by northern blot analysis of mRNAs from different barley tissues (not shown).

Divergence of CMx sequences with respect to BTI-CMe is no greater downstream than upstream of the premature stop codon, which suggests either a recent origin of that mutation or a retention of a coding role for the downstream fragment (Fig. 2B). In this context, it is relevant to recall that a premature stop codon has also been reported for a C-hordein gene and that in that case a certain degree of amber codon suppression was observed [7 and references therein]. Signal peptide sequences are less divergent with respect to the same domain in BTI-CMe than those of the mature proteins or those of 3′ noncoding regions.

Although the CMx sequences are closer to BTI-CMe than to any other homologue, they represent a distinct subset within the family, as they are quite divergent from it, their location is in group 4 chromosomes, while the chromosome where the gene for BTI-CMe is located belongs to group 3, and they lack the highly conserved GPRL reactive site which is typical of trypsin

inhibitors from this family. This site seems to have been substituted by a GYYL site (Fig. 2A). In the case of the Bowman-Birk family of serine-protease inhibitors, a change from arginine (R) to tyrosine (Y) at the reactive site changed the specificity of inhibition from trypsin to chymotrypsin [8].

The evolution of this complex multigene family is characterized by a low intraspecific variability for its different components, as well as by sharp differences among closely-related species, which affect particular loci [4, 8, 9]. The lack of an active BTI-CMe locus in the wheat genomes and of the CMx locus in the barley genome, while most other members of the family show little divergence between the two species, indicate a rapid change in the short evolutionary period separating the barley genome from those integrated in hexaploid wheat. This would be in line with the defence role proposed for this inhibitor family, as the absence of a particular pest in the ecological niche of one of the species would favour rapid divergence and even silencing of the key inhibitor genes.

We thank L. Lamoneda and **J.** García for technical assistance and Drs F. García-Olmedo and J. Vicente-Carbajosa for stimulating discussions and critica!reading of the manuscript. This work was financed by grant Bio91-0782 from Comisión Interministerial de Ciencia y Tecnología (Spain).

Barber D, Sanchez-Monge R, Mendez E, Lazara A, García-Olmedo F, Salcedo G: New ix-amylase and trypsin inhibitors among the CM-proteins of barley *(Hordeum vulgare).* Biochim Biophys Acta 869: 115-118 (1986).

Boulter D: Insect pest control by copying nature using genetically engineered crops. Phytochemistry 34: 1453-1466 (1993).

Campos FAP, Richardson M: The complete amino acid sequence of the bifunctional ct-amylase/trypsin inhibitor from seeds of ragi (Indian finger millet; *Eleusine coracana* Goertn.). FEBS Lett 152: 300-304 (1983).

Carbonero P, Salcedo R, Sanchez-Monge R, García-Maroto F, Royo J, Gomez L, Mena M, Medina J, Diaz I: A multigene family from cereals which encades inhibitors of trypsin and heterologous ix-amylases. In: Aviles FX (ed) Innovations on Proteases and their Inhibitors, pp. 333-348. Walter de Gruyter, Berlin/New York (1993).

Chong GL, Reeck GR: Interaction of trypsin, /1-factor XIIa, and plasma kallikrein with a trypsin inhibitor isolated from barley seeds: a comparison with the corn inhibitor of activated Hageman factor. Thrombosis Res 48: 211-221 (1987).

Diaz I, Royo J, Carbonero P: The promoter of barley trypsin inhibitor BTI-CMe discriminates between wheat and barley endosperm protoplasts in transient expression assays. Plant Cell Rep 12: 698-701 (1993).

Entwistle J, Knubsen S, Müller M, Cameron-Mills V: Amber codon suppression: the *in vivo* and *in vitro* analysis of two C-hordein genes from barley. Plant Mol Biol 17: 1217-1231 (1991).

García-Olmedo F, Salcedo G, Sanchez-Monge R, Gomez L, Royo J, Carbonero P: Plant proteinaceous inhibitors of proteinases and ix-amylases. In: Miflin BJ (ed) Oxford Surveys Plant Mol Cell Biol, vol. 4, pp. 275-334. Oxford University Press (1987).

García-Olmedo F, Salcedo G, Sanchez-Monge R, Hernandez-Lucas C, Carmona MJ, Lopez-Fando JJ, Fernandez JA, Gomez L, Royo J, García-Maroto F, Castagnaro A, Carbonero P: Trypsin/ix-amylase inhibitors and thionins: possible defence proteins from barley. In: Shewry PR (ed) Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology, pp. 335-350. CAB International, Wallingford, UK (1992).

Lyons A, Richardson M, Tatham AS, Shewry PR: Characterization of homologous inhibitors of trypsin and ix-amylase from seeds of rye *(Secale cereal e* L.). Biochim Biophys Acta 915: 305-313 (1987).

Mahoney WC, Hermodson MA, Jones B, Powers DD, Corfman RS, Reeck GR: Aminoacid sequence and secondary structural analysis of the corn inhibitor of trypsin and activated Hageman factor. J Biol Chem 259: 8412-8516 (1984).

Maraña C, García-Olmedo F, Carbonero P: Linked sucrose synthase genes in group-7 chromosomes in hexaploid wheat *(Triticum aestivum* L.). Gene 63: 253-260 (1988).

Mikola J, Soulina EM: Purification and properties of a trypsin inhibitor from barley. Eur J Biochem 9: 555-560 (1969).

Moraleja MA, García-Casado G, Sanchez-Monge R, Lopez-Otin C, Romagosa I, Molina-Cano JL, Salcedo G: Genetic variants of the trypsin inhibitor from barley endosperm show different inhibitory activities. Plant Sci 89: 23-29 (1993).

Odani S, Koide T, Ono T: The complete amino acid sequence of barley trypsin inhibitor. J Biol Chem 258: 7998-8003 (1983).

Rodriguez-Palenzuela P, Royo J, Gomez L, Sanchez-Monge R, Salcedo G, Molina-Cano JL, García-Olmedo F, Carbonero P: The gene for trypsin inhibitor CMe is regulated in *trans* by the lys 3a locus in the endosperm of barley *(Hordeum vulgare* L). Mol Gen Genet 219: 474-479 (1989).

Ryan CA: Proteinase inhibitors in plants: genes for improving defenses against insects and pathogens. Annu Rev Phytopath 28: 425-449 (1990).

Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74: 5463-5467 (1977).

Wen L, Huan J-K, Zen KC, Johnson BH, Muthukrishnan S, Mackaym V, Manney TR, Manney M, Reeck GR: Nucleotide sequence of a cDNA clone that encodes the maize inhibitor of trypsin and activated Hageman factor. Plant Mol Biol 18: 813-814 (1992).