Cloning of cDNA and chromosomal location of genes encoding the three types of subunits of the wheat tetrameric inhibitor of insect α-amylase

Federico García-Maroto, Carmen Maraña, Montaña Mena, Francisco García-Olmedo and Pilar Carbonero

Cátedra de Bioquímica y Biología Molecular, E.T.S. Ingenieros Agrónomos-UPM, E-28040 Madrid, Spain

Key words: α-amylase tetrameric inhibitor, cDNA cloning, genetic mapping, wheat

Abstract

We have characterized three cDNA clones corresponding to proteins CM1, CM3 and CM16, which represent the three types of subunits of the wheat tetrameric inhibitor of insect α-amylases. The deduced amino acid sequences of the mature polypeptides are homologous to those of the dimeric and monomeric α-amylase inhibitors and of the trypsin inhibitors. The mature polypeptides are preceded by typical signal peptides. Southern blot analysis of appropriate aneuploids, using the cloned cDNAs as probes, has revealed the location of genes for subunits of the CM3 and of the CM16 type within a few kb of each other in chromosomes 4A, 4B and 4D, and those for the CM1 type of subunit in chromosomes 7A, 7B and 7D. Known subunits of the tetrameric inhibitor corresponding to genes from the B and D genomes have been previously characterized. No proteins of this class have been found to be encoded by the A genome in hexaploid wheat (genomes AA, BB, DD) or in diploid wheats (AA) and no anti α-amylase activity has been detected in the latter, so that the A-genome genes must be either silent (pseudogenes) or expressed at a much lower level.

Introduction

A substantial fraction of the albumins and globulins of cereal endosperm is represented by a single protein family that includes inhibitors of heterologous α-amylases and of trypsin and is encoded by genes which are dispersed over several chromosomes (for a review, see [10]). The α-amylase inhibitors from wheat and barley can be classified into monomeric, dimeric, and tetrameric forms, which show different specificity towards heterologous α-amylases from various sources [30, 35]. Attention was initially focused on the monomeric and dimeric inhibitors (see [10]) and only recently a more detailed description of the tetrameric ones has been undertaken.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X17573, pCT3; X17574, pCT1; X17575, pCT2.

A relationship between the subunits of the wheat tetrameric inhibitor and the monomeric and dimeric inhibitors was first suggested on the basis of their amino acid composition and circular dicroism spectra [5, 26, 27]. Homology was fully demonstrated both in barley and in wheat through purification of the subunits, N-terminal sequencing, and *in vitro* reconstitution of the inhibitory activity [2, 3, 11, 31, 34, 35]. Three types of subunits, with little or no activity by themselves, are integrated into the active inhibitor: wheat proteins CM1 and CM2, encoded by genes in chromosomes 7D and 7B respectively, and barley protein CMa, whose gene is in chromosome 7H, are of the first type; wheat proteins CM16 and CM17, encoded by genes in chromosomes 4B and 4D, and barley protein CMb, encoded in chromosome 4H, are of the second type; while the third type is represented by wheat proteins CM3B and CM3D, together with its barley equivalent CMd [1, 8, 11, 32, 34, 35]. The last type of subunit, two copies of which seem to be present in the tetramer, it also encoded in group 4 chromosomes, but the assignment was only tentative in the case of wheat, because proteins CM3B and CM3D could not be separated from each other by electrophoresis or by high-performance liquid chromatography [1, 8, 32]. Only N-terminal amino acid sequences have been determined for all these proteins [2, 3, 37], with the exception of protein CMd, for which the complete sequence has been deduced from the cloned cDNA [12, 29]. We have now characterized cDNA clones corresponding to the three types of subunits in wheat and present evidence that the genes corresponding to proteins CM3B and CM3D are within a few kb from those encoding proteins CM16 and CM17, respectively, in chromosomes 4B and 4D.

Materials and methods

Biological material

Hexaploid wheat *Triticum aestivum* L. (genomes AABBDD) cv. Chinese Spring and its nullitetrasomic lines, which were a gift from E.R. Sears (Columbia, MO), and tetraploid wheat *Triticum turgidum* L. (genomes AABB) cv. Senatore Capelli were used in the study.

cDNA cloning and library screening

The construction and screening of a wheat cDNA library was carried out as described previously [23]. Total polysomal RNA from developing endosperm (20 days after pollination) of cv. Chinese Spring was the source of the poly(A)$^+$ RNA used. The library was screened using as a probe the clone pUP38 [29], which was ^{32}P-labelled by nick translation. Non-stringent conditions (58 °C) were used in the screening.

Nucleotide sequencing

The complete sequence of the selected clones was determined by the dideoxy chain termination method [36]. Inserts were subcloned into M13 (mp18 and mp19) and single-stranded DNA annealed to the primer by heating at 65 °C for 2 min and slowly cooling to room temperature. Sequencing reactions used as label [α-^{35}S]dATP (> 1000 Ci/mmol) and the products were separated on 8% polyacrylamide/9 M urea 'wedge' gels. The sequences were analysed with the Beckman Microgenie software.

Southern blot analysis

Total DNA was prepared from etiolated wheat leaves after 7–10 days of germination. Leaves were ground to a powder under liquid N_2 and extracted according to Murray and Thomson [24]. DNA was restricted with the appropriate endonuclease and separated on 0.75% agarose gels. Southern blotting was performed on Nylon membranes (Hybond N, Amersham) under the manufacturer's instructions. The inserts of clones used as probes were ^{32}P-labelled by the multiprime labelling system of Feinberg and Vogelstein [7]. Hybridizations were performed under stringent conditions by standard procedures [22].

Results

Characterization of cDNA clones

A cDNA library obtained from developing wheat endosperm (20 days after pollination) was screened under non-stringent conditions with a cDNA probe (insert from clone pUP38) which corresponded to subunit CMd from the barley tetrameric inhibitor of insect α-amylase [29]. Three clones, which differed in their partial restriction maps, were selected. The nucleotide sequences of the inserts in these clones corroborated both their homology to the probe and their correspondence to the three types of subunits that make up the wheat tetrameric inhibitors. More specifically, the longest open reading frame of clone pCT1 encoded a protein whose amino acid sequence included the known N-terminal sequence of protein CM3B, as shown in Fig. 1A. Similarly, the inserts in clones pCT2 and pCT3 corresponded to proteins CM1 and CM16, respectively (Fig. 1B, C). The known N-terminal sequences of the three proteins are preceded by typical signal sequences, in agreement with previous observations concerning *in vitro* and *in vivo* synthesis of these proteins [28] and with the precursor structures of other members of the family deduced from their corresponding cDNAs [16, 17, 29, 31]. In Fig. 2, the deduced amino acid sequences of the mature proteins have been aligned with those of other members of the family for which both their *in vitro* activities and their complete sequences are known. The three sequence blocks in Fig. 2 approximately correspond to the three domains defined by Kreis *et al.* [15]. In general, the N-terminal half of each domain is more conserved than the C-terminal one among the aligned proteins. In particular, proteins CM3 and CMd have a 16-residue insertion and protein CM1 a 6-residue insertion in the boundary between the A and B domains, when compared with protein CM16. Similarly, proteins CM3 and CMd have a 9-residue insertion in the C-terminal half of the B domain, when compared with proteins CM1 and CM16. Hydropathy profiles of the three types of subunits of the tetrameric inhibitor obtained according to Hoop and Woods [13] are presented in Fig. 3.

Chromosomal locations of genes

The inserts in clones pCT1, pCT2 and pCT3 were used as probes to determine the chromosomal locations of the corresponding genes by Southern analysis of DNAs from appropriate aneuploids of cv. Chinese Spring. As expected, only aneuploids of chromosome group 4 showed pattern differences with respect to the euploid when the probes encoding proteins CM3 and CM16 were used, and only aneuploids of chromosome group 7 had altered hybridization patterns when the probe for protein CM1 was used. Two bands at 12.8 kb and 17.7 kb were detected with the pCT1 probe when euploid DNA was digested with *Eco* RI endonuclease (Fig. 4A). The 12.8 kb band corresponded to a gene in chromosome 4D because it was absent when that chromosome was absent (in nulli 4D–tetra 4A and in tetraploid wheat), whereas the 17.7 kb band was associated with chromosome 4B because the hybridization signal was markedly enhanced when the dosage of that chromosome was duplicated (in nulli 4A–tetra 4B; no stock lacking chromosome 4B is available). Additionally a weak 14.4 kb band was observed at double dosage of chromosome 4A that was not detected in the euploid, indicating that homologous sequences are also present in this chromosome, although they must be more diverged than those in the other two chromosomes from the group. Exactly the same pattern was obtained when the probe corresponding to protein CM16 was used under stringent conditions, in which no reciprocal hybridization between the CM3 (pCT1) and CM16 (pCT3) cDNAs occurred (Fig. 4B). This result, which implied that the two types of sequences are within a few kb from each other, was confirmed by using restriction enzymes *Hin*d III and *Bgl* II and by sequential hybridization of a given filter with the two probes (not shown). As shown in Fig. 4C, hybridization patterns of chromosome group 7 aneuploids with the probe corresponding to pro-

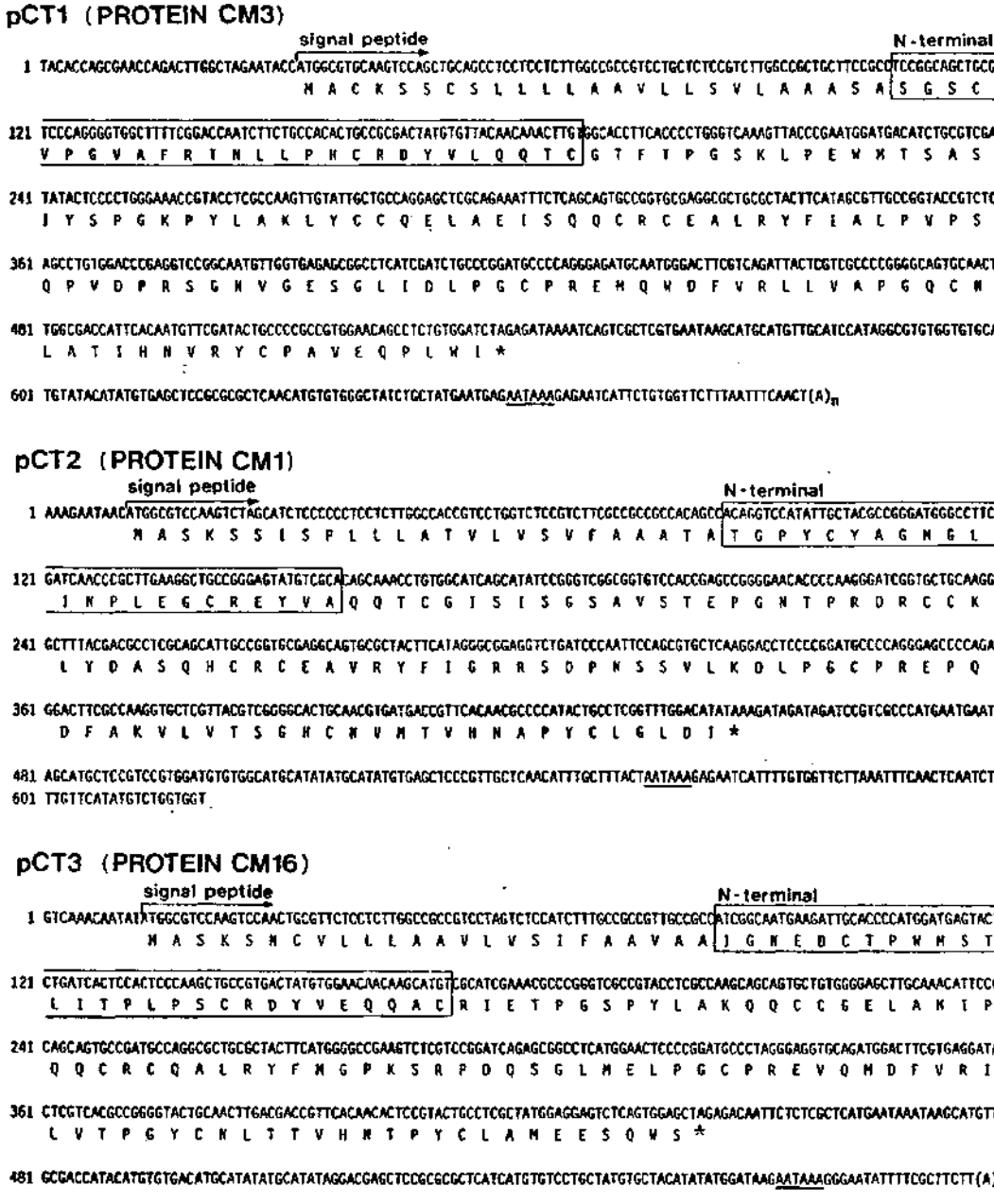

Fig. 1. Nucleotide sequences and deduced amino acid sequences of inserts in clones pCT1, pCT2, and pCT3, which respectively correspond to subunits CM3, CM1, and CM16 of the wheat tetrameric α-amylase inhibitor (indicated in parenthesis). Sequences identical to known N-terminal sequences of proteins CM3B [37] CM1 and CM16 [2] are boxed. The AATAAA polyadenilation signals are underlined.

tein CM1 indicated the existence of homologous sequences in chromosomes 7A, 7B and 7D. In this case, the band corresponding to the A genome was strong enough to be detected in the euploid.

Discussion

The deduced amino acid sequences of proteins CM1, CM3 and CM16, which represent the three types of subunits that make up the wheat

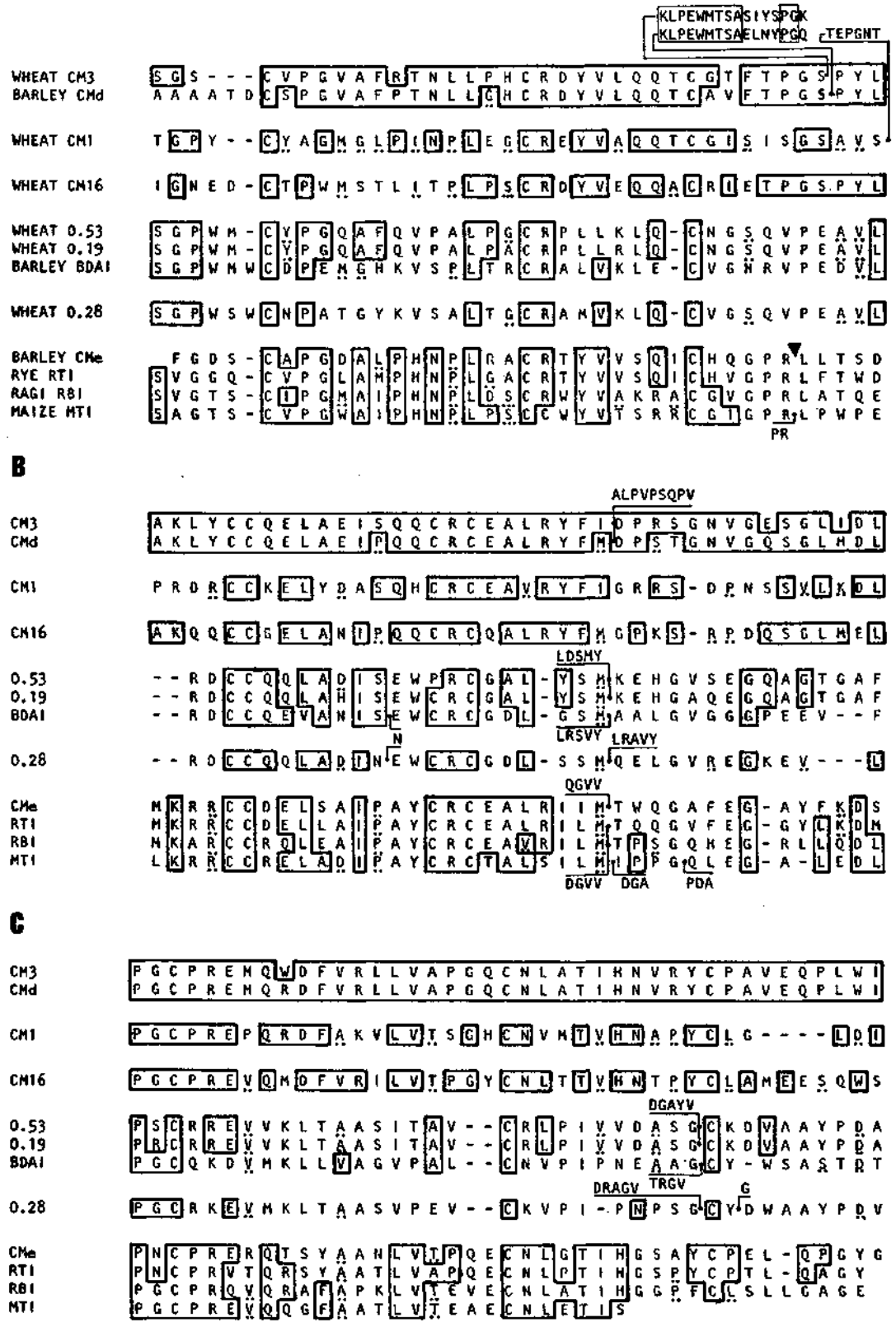

Fig. 2. Sequence alignment and comparison of subunits CM3, CM1 and CM16 of the wheat tetrameric α-amylase inhibitor with other members of the same protein family: subunit CMd from the barley tetrameric α-amylase inhibitor [29, 12]; α-amylase dimeric inhibitors 0.53 and 0.19 from wheat [20, 19] and BDAI-1 from barley [17]; α-amylase monomeric inhibitor 0.28 [14]; barley trypsin inhibitor CMe [25]; RTI, rye trypsin inhibitor [18]; RBI, bifunctional inhibitor from ragi, *Eleusine coracana* [6]; MTI, maize trypsin inhibitor [21]. The reactive sites of the trypsin inhibitors are indicated (▼). Domains A, B and C are as in Kreis *et al.* [15]. Relevant identities of residues at a given position are either boxed or underlined with two (..) or three dots (...).

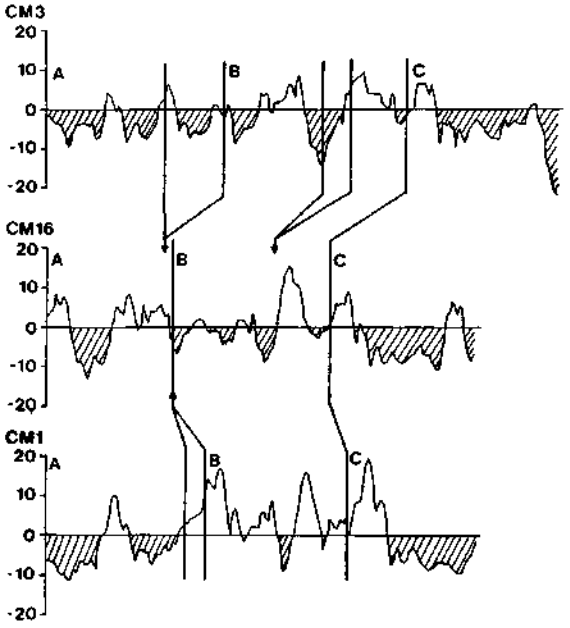

Fig. 3. Hydropathy profiles according to Hoop and Woods [13] of proteins CM3, CM16 and CM1. Regions of the profiles which correspond to insertions in proteins CM3 and CM1 with respect to protein CM16 are indicated. Domains are as in Fig. 2.

tetrameric inhibitors, fully confirm that indeed the three proteins are members of the same protein family as the monomeric and dimeric α-amylase inhibitors, the trypsin inhibitors and the bifunctional inhibitors from cereals [2, 3, 11, 34, 35, 37]. The sequences of subunits CM1, CM16, CM3 and CMd are closer to each other (47–86% identical residues) than to any of the other proteins in Fig. 2 (19–42% identical residues), and they are somewhat closer to the trypsin inhibitors (33–42% identical residues) than to the monomeric and dimeric α-amylase inhibitors (19–30% identical residues).

It has been previously suggested by Halford *et al.* [12] that the sequence TSAELXY, which is included in the 16-residue insertion of protein CMd at the end of the A domain and is also present in certain microbial α-amylases, might be responsible of the interaction with the other subunits. However, three of the six critical residues have not been conserved in the 16-residue insertion of protein CM3. The insertions in proteins CM3 and CMd create or reinforce hydro-

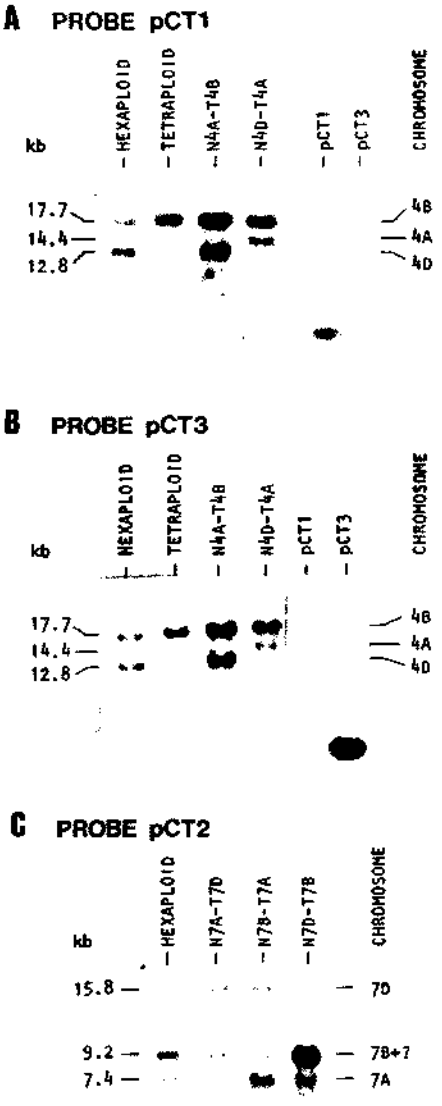

Fig. 4. Chromosomal locations of genes for the three types of subunits of the tetrameric inhibitor. Southern blot hybrydization of the following genotypes: hexaploid, *T. aestivum* cv. Chinese Spring; tetraploid, *T. turgidum* cv. Senatore Capelli; nulli-tetrasomic of cv. Chinese Spring, each of which is nullisomic (N) for a given chromosome and tetrasomic (T) for an homeologue (i.e. N4A-T4B lacks chromosome 4A and has 4 doses of chromosome 4B). A. Analysis of the indicated aneuploids using the insert in clone pCT1 (protein CM3) as probe. DNAs were digested with the *Eco* RI restriction endonuclease. B. The same digested DNAs probed with the insert in clone pCT3 (protein CM16). C. Analysis of the indicated aneuploids with the insert in clone pCT2 (protein CM1) as probe. DNAs were digested with the *Hind* III restriction endonuclease. Electrophoresis was performed in 0.8% agarose at 1 V/cm for 12 h, except for the inserts of clones pCT1 and pCT3 used as hybridization controls in A and B, which were run for 5 h.

phobic stretches in the sequences of both proteins (Fig. 3). Thus, these insertions could propitiate stabilization of the tetramer by hydrophobic interactions and would explain both the lower mobilities in SDS-PAGE and the selective solubility of these two proteins in 7:1 (v/v) chloroform/methanol mixtures in comparison with the other two types of subunits [32, 33]. The critical Arg-Leu residues that make the reactive site near the A/B boundary region of those inhibitors that show antitrypsin activity are absent from all three subunits, which is in agreement with their lack of *in vitro* activity against trypsin [2].

The probe encoding protein CM1 not only detects the corresponding gene in chromosome 7D but also a homologous sequence in chromosome 7B, which would encode the equivalent subunit CM2, and another one in chromosome 7A, whose corresponding protein has not been detected. Similarly the probe encoding protein CM16 detects homologous sequences in chromosomes 4B and 4D, which would respectively correspond to the equivalent subunits CM16 and CM17, and in chromosome 4A, to which no gene of this family has been previously assigned. Genes encoding protein CM3B and CM3D, which are undistinguishable electrophoretically, had been tentatively assigned to chromosome 4B and to an unknown chromosome of the D genome. The present results indicate that the sequences that are homologous to the probe are located in chromosomes 4A, 4B and 4D within 12–18 kb of those detected by the CM16 probe (pCT3). These findings, together with the fact that the sequences of the subunits are closer to each other than to other members of the family, allow the following hypothesis: (i) the three types of subunits of the tetrameric inhibitor were originated by a tandem duplication in chromosome 4 and by a translocation between chromosomes 4 and 7; (ii) insertions (or deletions) modified the gene for the primitive subunit as it was duplicated and translocated; (iii) the genes for the three types of subunits must have originated prior to the evolutionary branching out of barley and the diploid species that integrated into hexaploid wheat.

No proteins of this family among those extracted with either 0.1 M NaCl, 70% ethanol, or chloroform/methanol (2:1) have been found to be encoded by genes from the A genome of hexaploid wheat [1, 9, 11], and no activity against heterologous α-amylases has been found in extracts from diploid wheat species with the AA genome constitution [4, 11]. The present results indicate that there are sequences corresponding to the three types of subunits of the tetrameric inhibitor in the A genome. The fact that the corresponding proteins have not been detected suggests that these genes are either silent (pseudogenes) or expressed at a much lower level. It can be envisaged that the lack of the pertinent predators in the ecological niche of diploid wheat could have lead to the silencing of these genes.

Acknowledgments

The technical assistance of L. Lamoneda, C. Rojas and J. García is acknowledged. We thank G. Salcedo for critical reading of the manuscript. F. García-Maroto was the recipient of a postdoctoral scholarship from the Ministerio de Educación y Ciencia. M. Mena was under a contract from Centro Nacional de Biotecnología. This work was financially supported by grant Bio88-0216 from Comisión Interministerial de Ciencia y Tecnología (Spain).

References

1. Aragoncillo C, Rodriguez-Loperena MA, Carbonero P, García-Olmedo F: Chromosomal control of non-gliadin proteins from the 70% ethanol extract of wheat endosperm. Theor Appl Genet 45: 322–326 (1975).
2. Barber D, Sanchez-Monge R, García-Olmedo F, Salcedo G, Mendez E: Evolutionary implications of sequential homologies among members of the trypsin/α-amylase inhibitor family (CM-proteins) in wheat and barley. Biochim Biophys Acta 873: 147–151 (1986).
3. Barber D, Sanchez-Monge R, Mendez E, Lazaro A, García-Olmedo F, Salcedo G: New α-amylase and trypsin inhibitors among the CM-proteins of barley (*Hordeum vulgare*). Biochim Biophys Acta 869: 115–118 (1986).

4. Bedetti C, Bozzini A, Silano V, Vittozzi L: Amylase protein inhibitors and the role of Aegilops species in polyploid wheat speciation. Biochim Biophys Acta 362: 299–307 (1974).
5. Buonocore V, De Biasi M-G, Giardina P, Poerio E, Silano V: Purification and properties of an α-amylase tetrameric inhibitor from wheat kernel. Biochim Biophys Acta 831: 40–48 (1985).
6. Campos FAP, Richardson M: The complete amino acid sequence of the bifunctional α-amylase/trypsin inhibitor from seeds of ragi (Indian finger millet; *Eleusine coracana* Goertn). FEBS Lett 152: 300–304 (1983).
7. Feinberg AP, Vogelstein B: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132: 6–13 (1983).
8. García-Olmedo F, Carbonero P: Homoeologous protein synthesis controlled by homoeologous chromosomes in wheat. Phytochemistry 9: 1495–1497 (1970).
9. García-Olmedo F, Carbonero P, Jones BL: Chromosomal locations of genes that control wheat endosperm proteins. In: Pomeranz Y (ed) Advances in Cereal Science and Technology, vol 5, pp. 1–47. American Association of Cereal Chemists, St Paul, MN (1982).
10. García-Olmedo F, Salcedo G, Sanchez-Monge R, Gomez L, Royo J, Carbonero P: Plant proteinaceous inhibitors of proteinases and α-amylases. In: Miflin B (ed) Oxford Surveys of Plant Molecular and Cell Biology, vol 4, pp. 275–334. Oxford University Press, Oxford (1987).
11. Gomez L, Sanchez-Monge R, García-Olmedo F, Salcedo G: Wheat tetrameric inhibitors of insect α-amylases: alloploid heterosis at the molecular level. Proc Natl Acad Sci USA 86: 3242–3246 (1989).
12. Halford NG, Morris NA, Urwin P, Williamson MS, Kasarda DD, Lew EJ-L, Kreis M, Shewry P: Molecular cloning of the barley seed protein CMd: a variant member of the α-amylase/trypsin inhibitor family of cereals. Biochim Biophys Acta 950: 435–440 (1988).
13. Hoop TP, Woods KR: Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci USA 78: 3824–3828 (1981).
14. Kashlan N, Richardson M: The complete amino acid sequence of a major wheat protein inhibitor of α-amylase. Phytochemistry 20: 1781–1784 (1981).
15. Kreis M, Forde BG, Rahman S, Miflin BJ, Shewry PR: Molecular evolution of the seed storage proteins of barley, rye and wheat. J Mol Biol 183: 499–502 (1985).
16. Lazaro A, Rodriguez-Palenzuela P, Maraña C, Carbonero P, García-Olmedo F: Signal peptide homology between the sweet protein thaumatin II and unrelated cereal α-amylase/trypsin inhibitors. FEBS Lett 239: 147–150 (1988).
17. Lazaro A, Sanchez-Monge R, Salcedo G, Paz-Ares J, Carbonero P, García-Olmedo F: A dimeric inhibitor of insect α-amylase from barley. Cloning of the cDNA and identification of the protein. Eur J Biochem 172: 129–134 (1988).
18. Lyons A, Richardson M, Tatham AS, Shewry PR: Characterization of homologous inhibitors of trypsin and α-amylase from seeds of rye (*Secale cereale* L.). Biochim Biophys Acta 915: 305–313 (1987).
19. Maeda K, Kakabayashi S, Matsubara H: Complete amino acid sequence of an α-amylase inhibitor in wheat kernel (0.19 inhibitor). Biochim Biophys Acta 828: 213–221 (1985).
20. Maeda K, Takamori Y, Oka O: Isolation and properties of an α-amylase inhibitor (0.53) from wheat (*Triticum aestivum*). Agric Biol Chem 46: 2873–2875 (1982).
21. Mahoney WC, Hermodson MA, Jones B, Powers DD, Corfman RS, Reeck GR: Amino acid sequence and secondary structural analysis of the corn inhibitor of trypsin and activated Hageman factor. J Biol Chem 259: 8412–8416 (1984).
22. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982).
23. 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).
24. Murray MG, Thompson WF: Rapid isolation of high molecular weight plant DNA. Nucl Acids Res 8: 4321–4325 (1980).
25. Odani S, Koide T, Ono T: The complete amino acid sequence of barley trypsin inhibitor. J Biol Chem 258: 7998–8003 (1983).
26. O'Connor CM, McGeeney KF: Isolation and characterization of four inhibitors from wheat flour which display differential inhibition specificities for human salivary and human pancreatic α-amylase. Biochim Biophys Acta 658: 387–396 (1981).
27. O'Connor CM, McGeeney KF: Interaction of human α-amylases with inhibitors from wheat flour. Biochim Biophys Acta 658: 397–405 (1981).
28. Paz-Ares J, Ponz F, Aragoncillo C, Hernandez-Lucas C, Salcedo G, Carbonero P, García-Olmedo F: In vivo and in vitro synthesis of CM-proteins (A-hordeins) from barley (*Hordeum vulgare* L.). Planta 157: 74–80 (1983).
29. Paz-Ares J, Ponz F, Rodriguez-Palenzuela P, Lazaro A, Hernandez-Lucas C, García-Olmedo F, Carbonero P: Characterization of cDNA clones of the family of trypsin/α-amylase inhibitors (CM-proteins) in barley (*Hordeum vulgare* L.). Theor Appl Genet 71: 842–846 (1986).
30. Petrucci T, Tomasi M, Cantagalli P, Silano V: Comparison of wheat albumin inhibitors of α-amylase and trypsin. Phytochemistry 13: 2487–2495 (1974).
31. 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).
32. Salcedo G, Fra-Mon P, Molina-Cano JL, Aragoncillo C,

García-Olmedo F: Genetics of CM-proteins (A-hordeins) in barley. Theor Appl Genet 68: 53–59 (1984).
33. Salcedo G, Rodriguez-Loperena MA, Aragoncillo C: Relationship among low MW hydrophobic proteins from wheat endosperm. Phytochemistry 17: 1491–1494 (1978).
34. Sanchez-Monge R, Fernandez JA, Salcedo G: Subunits of tetrameric α-amylase inhibitors of *Hordeum chilense* are coded by genes located in chromosome 4Hch and 7Hch. Theor Appl Genet 74: 811–816 (1987).
35. Sanchez-Monge R, Gomez L, García-Olmedo F, Salcedo G: A tetrameric inhibitor of insect α-amylase from barley. FEBS Lett 207: 105–109 (1986).
36. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).
37. Shewry PR, Lafiandra D, Salcedo G, Aragoncillo C, García-Olmedo F, Lew EJ-L, Dietler MD, Kasarda DD: N-terminal amino acid sequences of chloroform/methanol-soluble proteins and albumins from endosperm of wheat, barley and related species. FEBS Lett 175: 359–363 (1984).