Evolutionary implications of sequential homologies among members of the trypsin/α-amylase inhibitor family (CM-proteins) in wheat and barley

Domingo Barber [a], Rosa Sanchez-Monge [a], Francisco García-Olmedo [a], Gabriel Salcedo [b] and Enrique Méndez

Servicio de Endocrinología, Centro Ramón y Cajal, Carretera de Colmenar Km 9.100, 28032 Madrid, and Departamento de Bioquímica, E.T.S. Ingenieros Agrónomos, Universidad Politécnica, Ciudad Universitaria, 28040 Madrid (Spain)

Key words: Amino acid sequence homology; Multi-gene family; Trypsin inhibitor; α-Amylase inhibitor; (Cereal endosperm)

The N-terminal amino acid sequence of four members of the trypsin/α-amylase inhibitor family in wheat, CM1, CM2, CM16 and CM17, has been investigated for 27–29 cycles by automated sequencing procedure. None of the proteins showed inhibitory activity against trypsin or α-amylases from different sources. The N-terminal sequences of these four proteins present a high degree of homology to each other as well as to those reported for other members of the same family in wheat and barley. Such homology is higher between a given protein and a second one associated with a different genome than between that protein and any other encoded in the same genome, indicating that most of the dispersion of the corresponding multi-gene family over several chromosomes took place before the wheat/barley evolutionary branching-out.

It is becoming increasingly evident that a substantial fraction of the albumins and globulins of wheat and barley endosperm is represented by a group of homologous components that includes inhibitors of trypsin and α-amylase, as well as members with no known in vitro activity [1–3]. Most of these proteins are soluble in chloroform/methanol mixtures (CM-proteins; Refs. 4, 5). The CM-proteins in wheat and barley are encoded by multi-gene families which are dispersed over several chromosomes [6,7]. Considerable information about sequences and in vitro activities of these proteins is available in barley, where at least six different components have been identified by N-terminal sequencing and cDNA cloning [2,3,8]. We have now determined N-terminal sequences of four wheat proteins; CM1 and CM2, encoded by genes located in chromosomes 7D and 7B, respectively, and CM16 and CM17, whose genes are respectively associated with chromosomes 4A and 4D [6], and have gained new insights into the evolution of this protein family in plants. In addition we have investigated the inhibitory properties of these four proteins.

Mature endosperm was obtained and milled from kernels of *Triticum aestivum* cv. Candeal by standard procedures.

Proteins CM1 and CM2 were purified by selective extraction with chloroform/methanol and gel filtration. In addition, CM1 was further purified by CM-cellulose ion-exchange chromatography followed by reversed-phase HPLC in a Vydac column as reported [9], and CM2 by preparative electrofocusing as previously described [10]. CM16 and CM17 were isolated by extraction with NaCl,

followed by extraction of the salt-soluble proteins by chloroform/methanol and preparative electrofocusing [11]. CM17 was further purified by HPLC as in the case of CM1. Purity of proteins was determined by two-dimensional electrophoresis [1]. For chemical analysis proteins were reduced and [^{14}C]carboxymethylated as in Ref. 11.

The N-terminal amino acid sequence of the carboxymethylated proteins were obtained by automatic Edman degradation as described [11], except that the phenylthiohydantoin derivatives were identified by HPLC. Phenylthiohydantoin (PTH) derivatives were separated in a Nova-Pak C_{18} (3.9 $\times$ 150 mm) column. Two buffers were employed: A, 35 mM sodium acetate (pH 5.0)/acetonitrile (5:1); B, 2-propanol/water (3:2). The following Waters program was used: 1 min at 0% B; 3.5 min from 0% to 40% B with a hyperbolic step (curve 4 in a M-680 Waters controller) followed by an isocratic step of 3 min at 40% B. The column was maintained at 38°C and a flow rate of 1 ml/min was used.

Amino acid analyses of proteins were carried out as previously reported [11]. The amino acids liberated on hydrolysis of the PTH-derivatives were analysed as in Ref. 12.

Enzyme inhibition tests were carried out as in a previous report [3].

Proteins CM1, CM2, CM16 and CM17 were purified essentially as previously described [9], except that their purity was refined through additional purification steps as indicated in the methods. The purity of proteins was determined by two-dimensional electrophoresis and by sequencing the N-terminal regions. No evidence of contaminants in any of the sequenced residues was observed. The amino acid composition of the four proteins adjusted to their molecular weight [9] is shown in Table I. The N-terminal sequence of the [^{14}C]carboxymethylated CM1, CM2, CM16 and CM17 obtained by automatic Edman degradation of the first 27–29 residues is shown in Fig. 1. A high degree of homology is observed between CM16 and CM17 (73%) as well as between CM1 and CM2 (86%). The N-terminal regions of these proteins have been aligned and compared with those previously reported for members of the same family in wheat [2,10,13] and in barley [2,3,11,14], as shown in Fig. 1. Gaps

TABLE I

AMINO ACID COMPOSITION OF PURIFIED WHEAT CM-PROTEINS

Compositions are expressed in residues per mole of protein based on the molecular weight. Cys was determined as CM-Cys. Trp was not determined. Numbers in parenthesis are the assumed numbers of residues

Amino acid	CM1	CM2	CM16	CM17
Cys	10.4 (10)	9.7 (10)	7.6 (8)	7.9 (8)
Asp	11.8 (12)	10.2 (10)	6.9 (7)	7.1 (7)
Thr	5.7 (6)	6.9 (7)	5.1 (5)	5.5 (6)
Ser	8.1 (8)	5.8 (6)	4.8 (5)	4.7 (5)
Glu	13.1 (13)	10.2 (10)	17.6 (18)	17.7 (18)
Pro	9.2 (9)	13.1 (13)	11.4 (11)	12.5 (13)
Gly	15.5 (16)	11.0 (11)	7.4 (7)	8.2 (8)
Ala	6.7 (7)	4.2 (4)	5.6 (6)	5.2 (5)
Val	6.7 (7)	7.9 (8)	5.8 (6)	5.7 (6)
Met	2.0 (2)	1.0 (1)	2.8 (3)	2.2 (2)
Ile	3.6 (4)	2.7 (3)	4.0 (4)	3.4 (3)
Leu	5.9 (6)	8.4 (8)	9.5 (10)	9.6 (10)
Tyr	5.1 (5)	5.1 (5)	4.5 (5)	4.4 (4)
Phe	1.9 (2)	1.8 (2)	1.9 (2)	1.6 (2)
Lys	1.5 (2)	3.8 (4)	2.7 (3)	2.5 (3)
His	2.7 (3)	3.0 (3)	1.7 (2)	1.9 (2)
Arg	7.5 (8)	8.5 (9)	6.4 (6)	6.5 (7)
M_r	12600	12450	11800	11850

have been introduced to achieve maximum homology, so that a total of 30 positions have been considered in all binary sequence comparisons, for each of which the percent of identical residues has been recorded in Table II, together with the significance levels of their respective compositional divergence indexes, based on the overall amino acid compositions [15]. No in vitro inhibitory activity was found for any of the native proteins when tested against trypsin, or α-amylases (salivary, pancreatic, *tenebrio molitor, Aspergillus oryzae* and barley), at concentrations of up to 10 μg per assay.

Based on the structural and compositional data in Table II, the CM-proteins can be grouped into clusters of closely related components which, together with the available genetic information, allow some conclusions with respect to the evolution of this family. Thus, CM16/CM17/CMb and CM3/CMd seem to correspond to two distinct loci associated with group 4 chromosomes. The

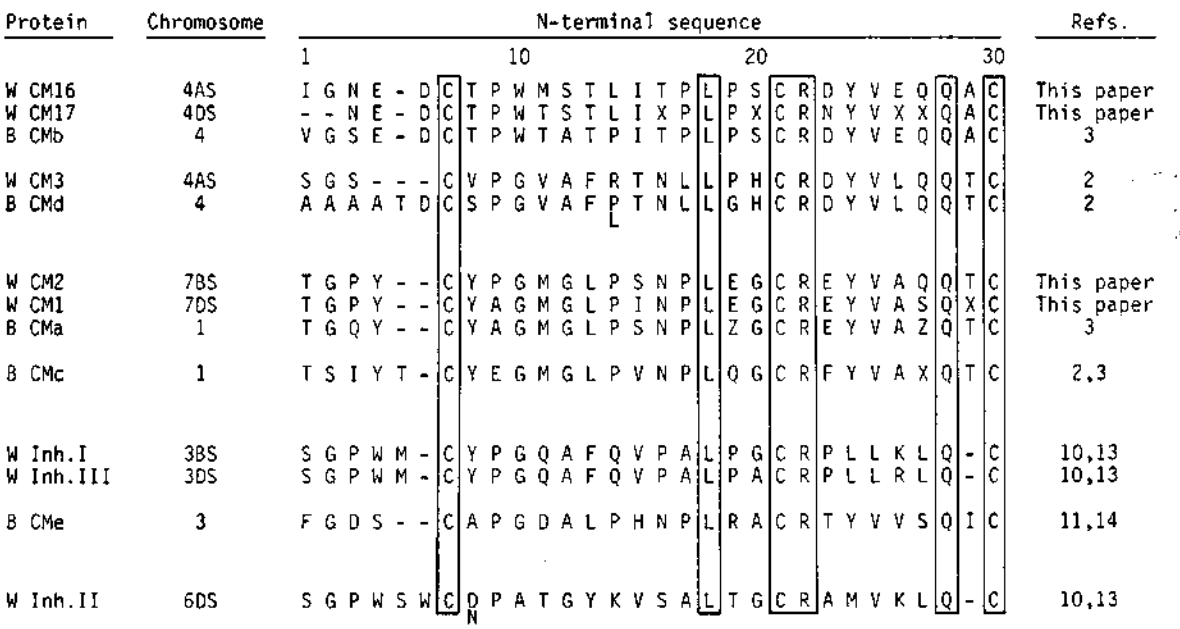

Fig. 1. Comparison of N-terminal amino acid sequences of wheat and barley proteins from the trypsin/α-amylase inhibitor family (CM-proteins). Proteins have been grouped according to the chromosomal locations of their structural genes, and within each chromosome group according to their locus. Invariant positions are boxed and gaps introduced for maximum alignment are indicated (–). W = wheat and B = barley. The following identities have been proposed based on N-terminal sequences and inhibitory activity data: W Inh.I = 0.53 α-amylase inhibitor, W Inh.II = 0.28 α-amylase inhibitor and W Inh.III = 0.19 α-amylase inhibitor (see Refs. 10 and 13); B CMe = barley trypsin inhibitor (see Refs. 11 and 14). Chromosome numbers are followed by one letter indicating the genome, when needed, and a second letter (S = short) indicating chromosome arm, when known. Barley chromosome 1 has been shown to be homologous to chromosome group 7 of wheat. The standard IUPAC-IUB single-letter designations are used [24].

presence of the two types of locus in chromosome 4A of wheat (CM16/CM3) and in chromosome 4 of barley (CMb/CMd) indicates that the intra-chromosomal duplication was already present in their common ancestor. There is a very high degree of homology between wheat proteins CM1 and CM2, respectively encoded by genes located in the short arms of chromosomes 7D and 7B, and between these proteins and protein CMa, whose gene is located in chromosome 1 of barley (equivalent to group 7 of wheat). In fact, CMa is closer to the wheat proteins than to a very similar protein, CMc, which is associated with the same barley chromosome. With respect to chromosome group 3, two types of proteins have been reported: wheat α-amylase inhibitors I and III, which have almost identical N-terminal sequences and are respectively associated with the short arms of chromosomes 3B and 3D [10], and barley trypsin inhibitor CMe (chromosome 3), rather diverged from them and closer to the chromosome group 7 proteins. Wheat α-amylase inhibitor II, whose gene is in the short arm of chromosome 6D [10], is closely related to inhibitors I and III and less so to the rest of the CM-proteins.

The present results not only confirm the predicted homology for the protein pairs CM1-CM2 [16] and CM16-CM17 [17] but further support the proposed homology of chromosome 1 of barley with group 7 of wheat and of chromosome 4 of barley with group 4 of wheat [18,19].

In general, binary comparisons of the wheat and barley proteins studied here indicate that there is greater similarity between a given protein from one genome and the appropriate one from a different genome than between that protein and any other encoded in the same genome, which suggests that most of the dispersion of this multigene family must have occurred prior to the branching-out of the barley genome from the di-

TABLE II

BINARY COMPARISONS OF SEQUENCED MEMBERS OF THE TRYPSIN/α-AMYLASE INHIBITOR FAMILY

In each case, the figure above the diagonal indicates the percent of identical positions in the N-terminal strech shown in Fig. 1 (30 positions), and the stars (*) below the diagonal mark those binary comparisons for which the compositional divergence index based on the overall amino acid compositions [15] indicates homology with 95% confidence (in all other binary comparisons the indexes indicated homology with 90% confidence). Comparisons within a chromosome group are boxed.

	Chromosome group												
	4					7				3			6
	CM16	CM17	CMb	CM3	CMd	CM2	CM1	CMa	CMc	Inh.I	Inh.III	CMe	Inh.II
CM16	–	73	83	46	43	46	43	40	33	30	30	40	30
CM17	*	–	66	36	36	36	36	33	30	26	26	36	30
CMb	*	*	–	53	46	46	43	40	33	33	33	46	33
CM3	*	*	*	–	70	53	43	46	40	46	46	50	33
CMd			*	*	–	46	36	40	43	33	33	43	26
CM2						–	86	86	73	43	40	56	40
CM1						*	–	83	70	40	36	56	36
CMa	*	*				*	*	–	73	36	33	53	33
CMc				*				*	–	36	33	46	33
Inh.I								*		–	93	36	56
Inh.III										*	–	40	50
CMe						*	*	*		*	*	–	30
Inh.II								*		*	*		

ploid genomes included in allohexaploid wheat.

Although the list in Fig. 1 includes the most prominent members of this protein family in wheat and barley, it is probable that additional components remain to be identified and characterized; for example, cDNA clones have been obtained in barley that encode additional CM-proteins [8].

No obvious correlation can be found between the sequences displayed in Fig. 1 and the in vitro inhibitory properties of the corresponding proteins. Thus although the wheat α-amylase inhibitors (chromosomes 3B, 3D, 6D) can be clearly segregated from the rest of wheat and barley CM-proteins, it is possible to find within a cluster of highly homologous proteins, such as the CM1/CM2/CMa/CMc group, proteins which are active against α-amylase (CMa), or trypsin (CMc), together with inactive proteins (CM1, CM2). Recent data on the tetrameric α-amylase inhibitors from wheat [20] suggest that at least some of the CM-proteins with unknown in vitro activity might be subunits of this type of inhibitor, although there are other possible explanations, such as inactivation during purification or specificity for other enzymes, which have not been tested.

It should be also pointed out that there are weak homologies between CM-proteins and certain domains of B-hordeins and α-gliadins [8,21], whose genes are located in chromosome groups 1 and 6 (1 in wheat equivalent to 5 in barley), which would extend the dispersion of this type of sequence over 5 out of the 7 chromosome homology groups in wheat and barley.

Finally, proteins clearly belonging to this family have been described in other plant species. Thus, the corn trypsin inhibitor [22] and the trypsin/α-amylase bifunctional inhibitor from *Eleusine coracana* [23] can be grouped with trypsin inhibitor CMe, with which they respectively share 17 and 16 positions out of 30. Weaker but significant

homologies exist between this group of proteins and the 2S reserve proteins of some dicotyledonous plants (castor bean and rape; Ref. 2).

We acknowledge the technical assistance of F. Soriano. This work has been supported by Comisión Asesora de Investigación Científica y Técnica grant No. 2022/83. D.B. is a fellow of the Fondo de Investigaciones Sanitarias.

References

1 Fra-Mon, P., Salcedo, G., Aragoncillo, C. and García-Olmedo, F. (1984) Theor. Appl. Genet. 69, 167–172
2 Shewry, P., Lafiandra, D., Salcedo, G., Aragoncillo, C., García-Olmedo, F., Lew, E.J.-L., Dietler, M.D. and Kasarda, D.D. (1984) FEBS Lett. 175, 359–363
3 Barber, D., Sanchez-Monge, R., Mendez, E., Lazaro, A., García-Olmedo, F. and Salcedo, G. (1986) Biochim. Biophys. Acta 869, 115–118
4 Rodriguez-Loperena, M.A., Aragoncillo, C., Carbonero, P. and García-Olmedo, F. (1975) Phytochemistry 14, 1219–1223
5 Salcedo, G., Sanchez-Monge, R., Argamenteria, A. and Aragoncillo, C. (1980) Plant Sci. Lett. 19, 109–119
6 Aragoncillo, C., Rodriguez-Loperena, M.A., Carbonero, P. and García-Olmedo, F. (1975) Theor. Appl. Genet. 45, 322–326
7 Salcedo, G., Fra-Mon, P., Molina-Cano, J.L., Aragoncillo, C. and García-Olmedo, F. (1984) Theor. Appl. Genet. 68, 53–59
8 Paz-Ares, J., Ponz, F., Rodriguez-Palenzuela, P., Lazaro, A., Hernandez-Lucas, C., García-Olmedo, F. and Carbonero, P. (1986) Theor. Appl. Genet. 71, 842–846
9 Sanchez-Monge, R., Barber, D., Mendez, E., García-Olmedo, F. and Salcedo, G. (1986) Theor. Appl. Genet. 72, 108–113
10 Salcedo, G., Rodriguez-Loperena, M.A. and Aragoncillo, C. (1978) Phytochemistry 17, 1491–1494
11 Lazaro, A., Barber, D., Salcedo, G., Mendez, E. and García-Olmedo, F. (1985) Eur. J. Biochem. 149, 617–623
12 Mendez, E. and Lai, C.Y. (1975) Anal. Biochem. 68, 47–53
13 Maeda, K., Kakabayashi, S. and Matsubara, H. (1985) Biochim. Biophys. Acta 828, 213–221
14 Odani, S., Koide, T. and Ono, T. (1983) J. Biol. Chem. 258, 7998–8003
15 Cornish-Bowden, A. (1980) Anal. Biochem. 105, 233–238
16 García-Olmedo, F. and Carbonero, P. (1970) Phytochemistry 9, 1495–1497
17 Aragoncillo, C., Rodriguez-Loperena, M.A., Salcedo, G., Carbonero, P. and García-Olmedo, F. (1978) Proc. Natl. Acad. Sci. USA 75, 1446–1450
18 Hart, G.E., Islam, A.K.M.R. and Shepherd, K.W. (1980) Genet. Res. 36, 311–325
19 Powling, A., Islam, A.K.M.R. and Shepherd, K.W. (1981) Biochem. Genet. 19, 237–254
20 Buonocore, V., De Biasi, M.-G., Giardina, P., Peorio, E. and Silano, V. (1985) Biochim. Biophys. Acta 831, 40–48
21 Kreis, M., Forde, B.G., Rahman, S., Miflin, B.J. and Shewry, P.R. (1985) J. Mol. Biol. 183, 499–502
22 Mahoney, W.C., Hermodson, M.A., Jones, B., Powers, D.D., Corfman, R.S. and Reeck, G.R. (1984) J. Biol. Chem. 259, 8412–8416
23 Campos, F.A.P. and Richardson, M. (1984) FEBS Lett. 167, 221–223
24 IUPAC-IUB JCBN (1984) Eur. J. Biochem. 138, 9–37