N-terminal amino acid sequences of chloroform/methanol-soluble proteins and albumins from endosperms of wheat, barley and related species

Homology with inhibitors of α-amylase and trypsin and with 2 S storage globulins

Peter R. Shewry, Domenico Lafiandra*†, Gabriel Salcedo+, Cipriano Aragoncillo+, Francisco Garcia-Olmedo+, Ellen J.-L. Lew*, Mary D. Dietler* and Donald D. Kasarda*

*Biochemistry Department, Rothamsted Experimental Station, Harpenden, Herts. AL5 2JQ, England, *Food Proteins Research Unit, Western Regional Research Centre, ARS United States Deparment of Agriculture, Berkeley, CA 94710, USA,* and +*Departamento de Bioquimica, E.T.S. Ingenieros Agronomos, Madrid 3, Spain*

Received 27 July 1984

The N-terminal amino acid sequences of two chloroform/methanol soluble globulins from barley and one form wheat are reported. They are homologous with N-terminal sequences previously reported for α-amylase and trypsin inhibitors from cereals and 2 S storage proteins from castor bean and rape. Three albumins were also purified from *Aegilops squarrosa* and *Triticum monococcum*. These had N-terminal amino acid sequences most closely related to the α-amylase and trypsin inhibitors. The relationships of this superfamily of seed proteins are discussed.

Barley *Wheat* *Amino acid sequence* *Protein homology* *Seed protein*

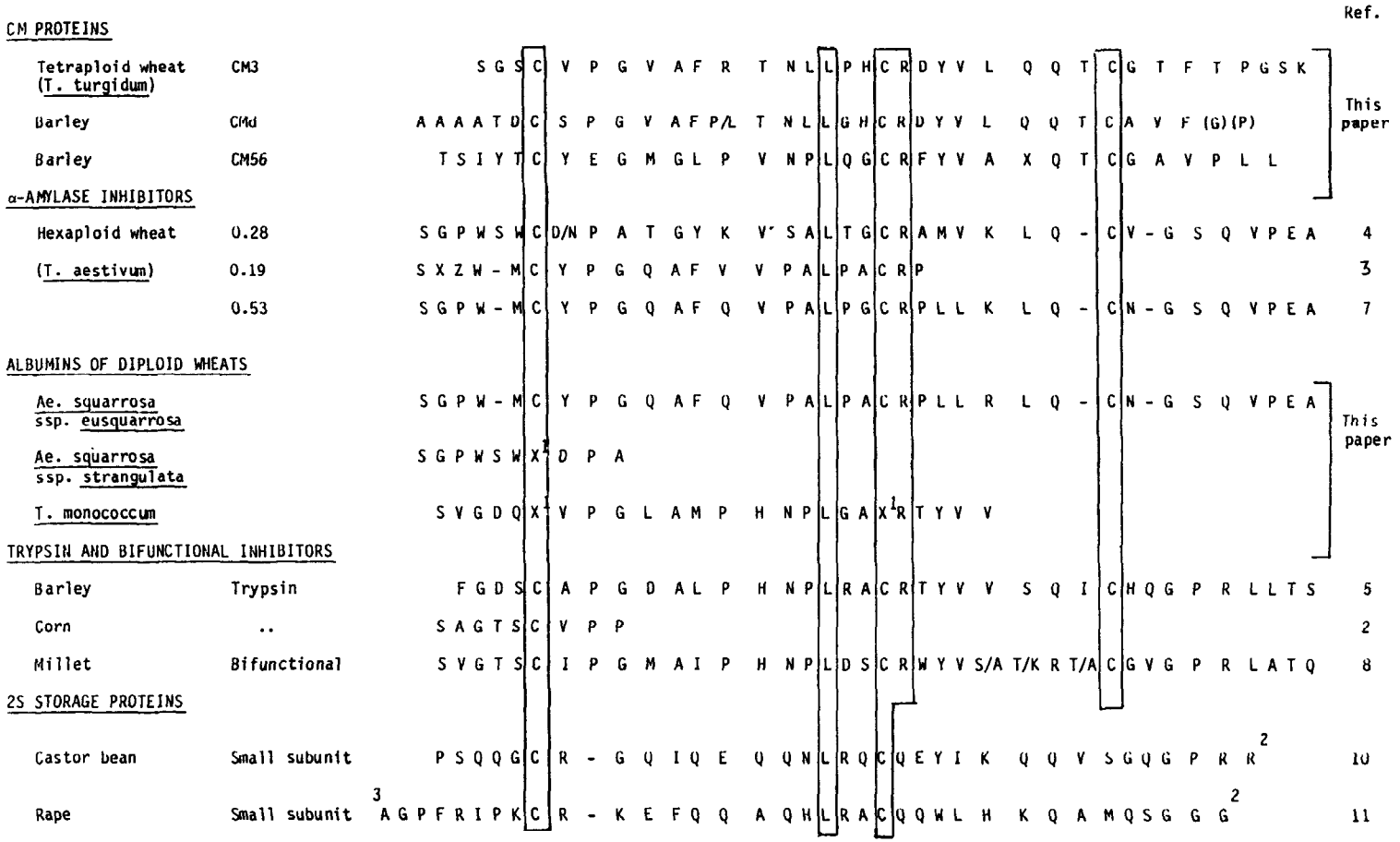
1. INTRODUCTION

The increasing sensitivity of methods for direct protein sequencing, and the availability of protein sequences deduced from nucleotide sequencing of cloned cDNA and genomic DNA, have facilitated the comparison of proteins of diverse origin. In some cases this has revealed the presence of sequence homology in proteins with no known functional relationships (see [1]), indicating an origin from a single ancestral protein. One such grouping, or family, contains a number of seed proteins, including inhibitors of α-amylase and trypsin (at least one of which is bifunctional) from cereals [2–9] and storage globulins with sedimentation coefficients ($s_{20,w}$) of about 2 from castor bean [10,11] and oil seed rape [12].

A further group of low M_r salt-soluble proteins from cereal seeds are characterized by solubility in mixtures of chloroform and methanol, and have accordingly been called CM proteins [13–21]. Their function is not known, although they do not appear to be stored in protein bodies [22]. In the present paper we show that CM proteins from wheat and barley and several albumins from diploid wheats have N-terminal amino acid sequences homologous to the α-amylase/trypsin inhibitors and 2 S storage proteins. These proteins together constitute a superfamily with limited sequence homology as defined in [1].

† Present address: Istituto de Germoplasma, CNR, Via Amendola 165/A, Bari, Italy

Published by Elsevier Science Publishers B.V.
00145793/84/$3.00



[1] These proteins were not alkylated, and this residue may correspond to cysteine.
[2] Carboxyterminus of the mature subunit.
[3] Tentative N-terminus after cleavage of signal peptide.

The notation of the residues follows standard single letter codes: A, alanine, C, cysteine: D, aspartic acid; E, glutamic acid, F, phenyl alanine, G, glycine, H, histidine, I, isoleucine, K, lysine, L, leucine, M, methionine, N, asparagine, P, proline, R, arginine, S, serine, T, threonine, V, valine, W, tryptophan, Y, tryosine, Z, glutamate or glutamine, X, unidentified. Residues in parentheses are provisional.

Fig.1. N-terminal amino sequences of the protein superfamily comprising CM proteins, α-amylase inhibitors, trypsin inhibitors and 2 S storage proteins. The sequences are aligned for maximum homology. Conserved residues are boxed.

2. MATERIALS AND METHODS

CMd and CM3 were prepared from barley (*Hordeum vulgare* L.) cv. Zephyr and tetraploid wheat (*Triticum turgidum*) cv. Senatore Capelli respectively as described previously [17,19]. CMd was also prepared, as was CM56, by ion-exchange chromatography and gel filtration of a reduced and pyridylethylated total hordein fraction from barley mutant Risø 56 [23]. Albumin fractions were prepared from the diploid species *T. monococcum* (accession MG4273), *Ae. squarrosa* ssp. *eusquarrosa* (accession G3489) and *Ae. squarrosa* ssp. *strangulata* (G1276). An acetic acid (0.01 M) extract of flour was mixed with 3 volumes of 1.5% (w/v) NaCl. The precipitated proteins were removed by centrifugation, redissolved in 0.01 M acetic acid, dialysed and lyophilized. They were then separated by gel filtration on Biogel P100 in 0.01 M acetic acid, and in the case of the component from *eusquarrosa*, followed by ion-exchange chromatography on carboxymethyl cellulose in ammonium acetate buffer (pH 5.2) with a linear gradient of 0.1 to 0.2 M.

Proteins prepared from *T. turgidum*, *T. monococcum*, *H. vulgare* cv. Zephyr and *Ae. squarrosa* ssp. *eusquarrosa* were reduced and carboxymethylated [24]. Electrophoresis in aluminium lactate buffer at low pH and SDS–PAGE were as described previously [25,26]. N-terminal amino acid sequences were determined with a Beckman 890B automatic sequencer and Beckman 1M Quadrol program 112078. Phenylthiohydantoin (PTH) amino acids were determined by high performance liquid chromatography [27], identification of PTH valine and PTH methionine were checked by thin-layer chromatography [28].

3. RESULTS AND DISCUSSION

Three main CM proteins (designated CM1 to CM3) are present in wheat and five (designated CMa to CMe) in barley [17,19,21]. They have similar M_r values (about 12000–16000 by SDS–PAGE) and amino acid compositions. Comparison of the latter by the compositional difference index of Cornish-Bowden [29] indicates sequence homology, notably between CMd of barley and CM3 of wheat [20]. These two proteins are also unusual in being soluble in 7:1 (v/v) chloroform/methanol. Their homology is confirmed by comparison of N-terminal amino acid sequences (fig.1). The main difference is at the N-terminus, where four alanine residues are present in CMd. This unusual sequence was confirmed by the analysis of CMd preparations from two different genotypes (Zephyr and Risø 56). A second CM protein, with an M_r by SDS–PAGE of about 13000, was also prepared from Risø 56. This preparation, called CM56 in fig.1, may correspond to CMc as described by Salcedo et al. [19]. The N-terminal sequence was less closely related to CM3, and again had a unique N-terminus.

Three α-amylase inhibitors, called 0.19, 0.28 and 0.53 according to their electrophoretic mobility at high pH, have been previously purified from hexaploid bread wheat (*T. aestivum*) [30] and their partial or complete amino acid sequences determined [3,4,7] (fig.1). These have been shown to have homologous amino acid sequences to trypsin inhibitors from barley [5] and maize [2] and to a bifunctional α-amylase/trypsin inhibitor from finger millet (*Eleusine coracana*) [8] (fig.1). In the course of studies of the phylogeny of wheats we also purified three albumin proteins from *T. monococcum* and *Ae. squarrosa*, diploid species thought to be related to the progenitors of the A and D genomes respectively of hexaploid bread wheats [31]. One preparation from *Ae. squarrosa* ssp. *eusquarrosa* had identical mobility on electrophoresis at pH 3.1 to the 0.19 α-amylase inhibitor of *T. aestivum*, while a second from *Ae. squarrosa* ssp. *strangulata* had a similar mobility to the 0.28 inhibitor. The component from *T. monococcum* was faster than the 0.19 and 0.28 α-amylase inhibitors (results not shown). The N-terminal amino acid sequences of these components (fig.1) are related to those reported for the α-amylase/trypsin inhibitors. The albumin from *Ae. squarrosa* ssp. *eusquarrosa* differed in only two residues out of 36 from the 0.53 α-amylase inhibitor, indicating that it may also have been an α-amylase inhibitor. The albumin from *T. monococcum* was more closely related to the trypsin inhibitors and the bifunctional inhibitor. Only small amounts of these albumins were prepared, and we did not test their inhibitory properties.

The α-amylase/trypsin inhibitors are also homologous with the storage proteins (with $s_{20,w}$

values of about 2) of castor bean (*Ricinus communis*) [10,11] and oilseed rape (*Brassica napus*) [12] (see fig.1). Both proteins are composed of two subunits which are associated by disulphide bonds [10,32]. The small subunits (M_r values about 4000) are homologous with the N-terminal regions of the α-amylase/trypsin inhibitors (see fig.1), and the larger subunits (M_r values 7000 and 9000) with the C-terminal regions. It has been shown that the two subunits of the rape protein are synthesised as a single precursor protein which is probably processed to the mature subunits by the loss of a leader sequence and a short-linker peptide [12]. The CM proteins from barley have been shown to be synthesised as larger precursors which seem to be co-translationally processed into the mature proteins [22].

Comparison of the sequences in fig.1 indicates that the CM proteins are also related to this group. The region between residues 9 and 24 is the most conserved, with cysteines at 9 and 23 and leucine at 20 in all the proteins. Within this region CM3 of wheat has 13 out of 16 residues in common with CMd of barley, 9 with the millet bifunctional and wheat 0.19 α-amylase inhibitors but only four and three with the small subunits of the castor bean and rape proteins respectively. The CM proteins, albumins and α-amylase/trypsin inhibitors also have a third cysteine residue at 33 and arginine at 24. The extreme N-terminal region is the least conserved, with variation in the number and sequence of the residues preceding the first cysteine. The three CM proteins are not related to each other or to the other proteins in this region.

The extreme variability at the N-terminus may indicate little or no functional role for this region. Conversely the conservation of the number and positions of the cysteine residues probably indicates an important function. It has been noted previously that protease inhibitors are often rich in cysteine residues [33,34] and that the positions of these are conserved [35]. It has therefore been suggested that disulphide bonds are important in stabilising the active conformation [36]. Their conservation in the CM proteins may indicate that this group of proteins has a specific function, which requires a precise and stable conformation.

Thus it appears that all the proteins discussed in this paper constitute a protein superfamily [1], with limited sequence homology (below 50% in some cases) and diverse functions. Although they are present in monocotyledonous (grasses) and dicotyledonous (castor bean, rape) plants, they appear to be restricted to seeds where they may be synthesised in the endosperm (cereals, castor bean) or embryo (rape). It is of interest that although their structural genes have diverged to encode proteins with different amino acid sequences and functions, they share the property of being expressed only in the storage tissues of developing seeds.

ACKNOWLEDGEMENTS

P.R.S. wishes to thank Ms S. Parmar for technical assistance and his colleagues at Rothamsted for valuable discussions. The work at Rothamsted was supported by EEC grant no.470. P.R.S. and D.D.K. acknowledge NATO grant No.0324/82. The work in Madrid, was supported by a grant from the 'Comision Asesora de Investigacion Cientifica y Tecnica'. D.D.K. and D.L. thank Ms N. Fulrath for assistance with protein separations and Dr J.G. Waines (University of California, Riverside) for supplying *Ae. squarrosa*. *T. monococcum* was supplied by the Istituto de Germoplasma, CNR, Bari, Italy. Reference to a company or product does not imply approval or recommendation by the US Department of Agriculture to the exclusion of others that may be suitable.

REFERENCES

[1] Dayhoff, M.O., Barker, W.C., Hunt, L.T. and Schwartz, R.M. (1978) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O. ed.) vol.5, suppl.3, National Biomedical Research Foundation, Washington, DC.

[2] Hochstrasser, K., Illchmann, K. and Werle, E. (1970) Hoppe-Seyler's Z. Physiol. Chem. 351, 721–728.

[3] Redman, D.G. (1976) Biochem. J. 155, 193–195.

[4] Kashlan, N. and Richardson, M. (1981) Phytochemistry 20, 1781–1784.

[5] Odani, S., Koide, T. and Ono, T. (1982) FEBS Lett. 141, 279–282.

[6] Odani, S., Moide, T. and Ono, T. (1983) J. Biol. Chem. 258, 7998–8003.

[7] Maede, K., Hase, T. and Matsubara, H. (1983) Biochim. Biophys. Acta 743, 52–57.

[8] Campos, F.A.P. and Richardson, M. (1983) FEBS Lett. 155, 300–304.
[9] Petrucci, T., Sannia, G., Parlamenti, R. and Silano, V. (1978) Biochem. J. 173, 229–235.
[10] Sharief, F.S. and Li, S.S.-L. (1982) J. Biol. Chem. 257, 14753–14759.
[11] Odani, S, Koide, T., Ono, T. and Ohnishi, K. (1983) Biochem. J. 213, 543–545.
[12] Crouch, M., Tenbarge, K.M., Simon, A.E. and Ferl, R. (1983) J. Mol. Appl. Genet. 2, 273–283.
[13] Garcia-Olmedo, F. and Garcia-Faure, R. (1969) Lebensm. Wiss. Technol. 2, 94–96.
[14] Garcia-Olmedo, F. and Carbonero, P. (1970) Phytochemistry 9, 1495–1497.
[15] Rodriguez-Loperena, M.A., Aragoncillo, C., Carbonero, P. and Garcia-Olmedo, F. (1975) Phytochemistry 14, 1219–1223.
[16] Aragoncillo, C., Rodriguez-Loperena, M.A., Carbonero, P. and Garcia-Olmedo, F. (1975) Theor. Appl. Genet. 45, 322–326.
[17] Salcedo, G., Rodriguez-Loperena, M.A. and Aragoncillo, C. (1978) Phytochemistry 17, 1491–1494.
[18] Sacedo, G., Sanchez-Monge, R., Argamenteria, A. and Aragoncillo, C. (1980) Plant Sci. Lett. 19, 109–119.
[19] Salcedo, G., Sanchez-Monge, R. and Aragoncillo, C. (1982) J. Exp. Bot. 33, 1325–1331.
[20] Paz-Ares, J., Hernandez-Lucas, C., Salcedo, G., Aragoncillo, C., Ponz, F. and Garcia-Olmedo, F. (1983) J. Exp. Bot. 34, 388–395.
[21] Salcedo, G., Fra-Mon, P., Molina-Cano, J.L., Aragoncillo, C. and Garcia-Olmedo, F. (1984) Theor. Appl. Genet. 68, 53–59.
[22] Paz-Ares, J., Ponz, F., Aragoncillo, C., Hernandez-Lucas, C., Salcedo, G., Carbonero, P. and Garcia-Olmedo, F. (1983) Planta 157, 74–80.
[23] Kreis, M., Shewry, P.R., Forde, B.G., Rahman, S. and Miflin, B.J. (1983) Cell 34, 161–167.
[24] Crestfield, A.M., Moore, S. and Stein, W.H. (1963) J. Biol. Chem. 238, 623–627.
[25] Lafiandre, D. and Kasarda, D.D. (1984) Analyt. Biochem., in press.
[26] Shewry, P.R., Lew, E.J.-L. and Kasarda, D.D. (1981) Planta 153, 246–253.
[27] Bhown, A.S., Mole, J.E., Weissenger, A. and Bennett, J.C. (1978) J. Chromatogr. 178, 532–535.
[28] Jeppson, J.-O. and Sjoquist, J. (1967) Anal. Biochem. 18, 264–269.
[29] Cornish-Bowden, A. (1979) J. Theor. Biol. 76, 369–386.
[30] Buonocore, V., Petrucci, T. and Silano, V. (1977) Phytochemistry 16, 811–820.
[31] Feldman, M. (1973) in: Evolution of Crop Plants (Simmonds, N. ed.) pp.120–128, Longmans, London.
[32] Reddy, M.N., Kreim, P.S., Heinrikson, R.L. and Kezdy, F.J. (1975) J. Biol. Chem. 250, 1741–1750.
[33] Richardson, M. (1977) Phytochemistry 16, 159–169.
[34] Laskowski, M. and Kavo, I. (1980) Ann. Rev. Biochem. 49, 593–626.
[35] Lonnerdal, B. and Janson, J.-C. (1972) Biochim. Biophys. Acta 278, 175–183.
[36] Ryan, C.A. (1979) in: Herbivores: Their Interaction with Secondary Plant Metabolites (Rosenthal, G.A. and Janzen, D.H. eds) pp.559–618, Academic Press, London.