Reprinted from Nature, Vol. 286, No. 5772, pp. 520-522, July 31 1980

N-terminal amino acid sequence homology of storage protein components from barley and a diploid wheat

Peter R. Shewry*, Jean-Claude Autran†, Charles C. Nimmo, Ellen J.-L. Lew & Donald D. Kasarda

Food Proteins Research Unit, Western Regional Research Center, Agricultural Research, SEA, US Department of Agriculture, Berkeley, California 94710

Wild barley (*Hordeum spontaneum*) and the wild diploid wheat *Triticum boeoticum* were possibly the first plants cultivated by early man[1], giving rise to the domesticated forms *Hordeum vulgare L.* and *Triticum monococcum L.* In addition, *T. boeoticum* may have contributed the A genome to polyploid wheats, including common bread wheat (*Triticum aestivum*)[2] which is a hexaploid with genome composition ABD. *Hordeum* seems to be the older genus, having diverged from some common ancestor before the divergence of *Triticum* and other genera of the subtribe Triticinae[3]. Prolamins constitute the major storage protein fraction of both barley and wheat; they are located in the endosperm of the caryopsis and are soluble in alcohol–water solutions[4]. Barley and wheat prolamins (hordeins and gliadins, respectively) contain large amounts of glutamine and proline, which together make up 50–75 mol per cent of total amino acids[4,5]. The hordeins and gliadins are complex mixtures of components[6–8] that seem to be encoded by clusters of duplicated genes that have diverged to produce many distinguishable protein components. Despite the complexity of the gliadin mixture, the components retain considerable homology in their N-terminal region[9,10] and this has been reported for zeins, the prolamins of maize (*Zea mays*)[11], as well. Here, we report that a purified C-hordein component from barley is homologous in amino acid sequence with a purified ω-gliadin component from *T. monococcum* at 23 of 27 residues at the N-terminus. This result is in accord with the close relationship between the two species and indicates that, despite the propensity of prolamin genes to tolerate mutations, a significant portion of their sequences can be conserved over a period of time, which, although not accurately known, probably amounts to millions of years.

Prolamins were extracted from barley (*H. vulgare* Julia) and *T. monococcum* (from the University of Manitoba) as described previously[10,12]. The C-hordein component was obtained by ion exchange chromatography of the hordein mixture on CM-cellulose followed by gel filtration on Sephacryl S-300 (ref. 12). The ω-gliadin component from *T. monococcum* was obtained by ion exchange chromatography of the gliadin mixture on CM-cellulose according to the procedure of Booth and Ewart[13] except that the gradient ranged linearly from 5 to 43 mM in sodium acetate; this component was not purified further. Electrophoretic patterns of the purified components in aluminium lactate buffer, *p*H 3.2, where separation is based on net charge, are compared with those of the prolamin mixtures in Fig. 1. We estimated molecular weights of 57,000 for the C-hordein and 44,000 for the ω-gliadin by SDS polyacrylamide gel electrophoresis[6].

Amino acid analyses were carried out on a Durrum analyser, model D-500. Hydrolyses were for 24 h, tryptophan was not determined, and no corrections for destruction of labile amino acids were applied. Averaged results from duplicate analyses are given in Table 1. Both components were notable for their high glutamine and proline content (although glutamic acid only is shown in Table 1, this amino acid is present in C-hordein[12] and probably in the ω-gliadin[14] almost entirely in its amidated form[12,14]). The C-hordein had more proline and less glutamine than the ω-gliadin, but the sum of these two amino acids was close to 70 mol per cent for both components. Both components had about 9 mol per cent phenylalanine, only trace amounts of cystine/2, and lysine was absent from the C-hordein and present in the ω-gliadin in an amount corresponding to only about 0.5 residue on a molar basis. Low percentages of basic amino acids, combined with equivalent low percentages of free carboxyl side chains[12,14], indicate that these proteins will have few charged side chains at any *p*H.

Automatic amino acid sequencing[15] was carried out with a Beckman sequencer, model 890B, and DMAA peptide program 111374 (ω-gliadin) or 0.1 M Quadrol programs 011576 and 121178 (C-hordein). Duplicate analyses were carried out for each protein. In the first analysis of the ω-gliadin, identification of the phenylthiohydantoin (PTH) amino acids resulting from the Edman degradation[15] was by gas chromatography[16] only; this method was supplemented in the second analysis by TLC on polyamide sheets[17] and silica gel plates[18] and by hydrolysis to the free amino acid followed by analysis on the Durrum analyser. PTH amino acids from sequencing of the C-hordein were identified by HPLC[19], which clearly resolved and quantified all the PTH amino acids found in the first 28 cycles. The N-terminal sequences obtained are compared in Fig. 2. Initial yields in sequencing ranged from 50 to 76% of molar

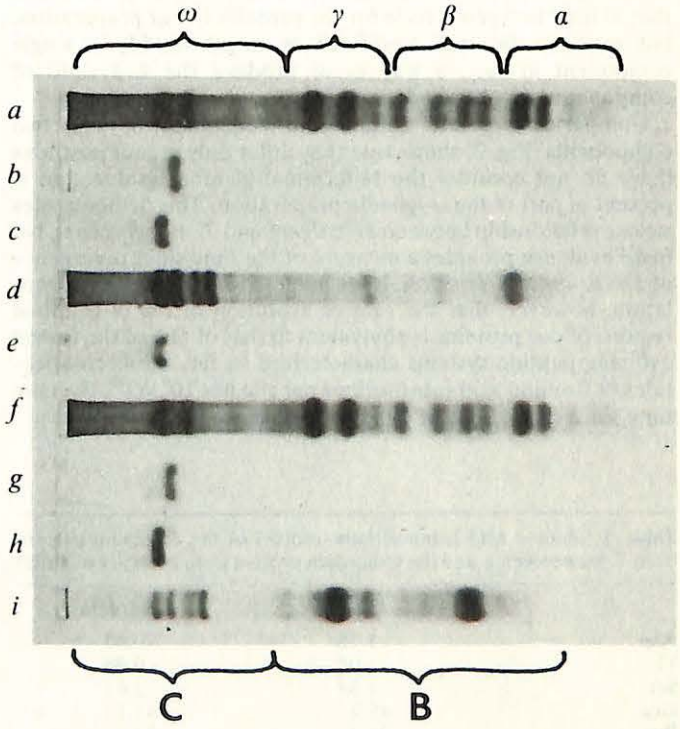

Fig. 1 Polyacrylamide gel electrophoresis[27] of purified protein components and the whole prolamin mixtures from which they were prepared, aluminium lactate buffer, *p*H 3.2, 3 M urea, migration from left (+) to right (−): *a*, prolamin mixture from *T. monococcum*; *b*, purified ω-gliadin from *T. monococcum*; *c*, purified C-hordein; *d*, prolamin mixture from barley (var. Julia); *e*, same as *c*; *f*, same as *a*; *g*, same as *b*; *h*, same as *c*; *i*, prolamin mixture from barley (Julia), but treated with 2% 2-mercaptoethanol to dissociate B-hordeins. The α-, β-, γ- and ω-regions (top) correspond to the usual assignments of electrophoretic mobilities for gliadin patterns of common wheats[5] and the B- and C-regions (bottom) correspond to the hordein patterns.

* Permanent address: Biochemistry Department, Rothamsted Experimental Station, Harpenden, Herts AL5 2JQ, UK.
† Permanent address: Institut National de la Recherche Agronomique, Laboratoire de Technologie des Cereales, 9 Place Viala, 34060 Montpellier, France.

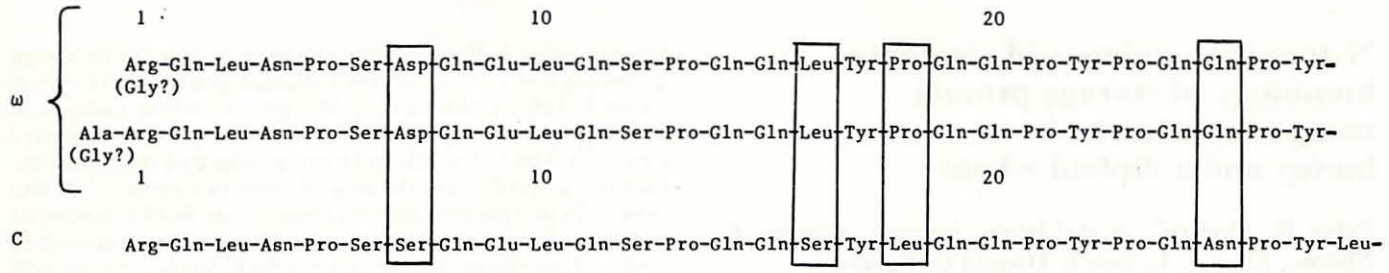

Fig. 2 N-terminal amino acid sequences of the ω-gliadin component from *T. monococcum* (ω) and the C-hordein component from barley (C). Sequences are ordered from the first arginine residue to emphasize homology. Glycine was a minor residue at cycle 1 in the sequence of the ω-gliadin. Positions where the ω-gliadin and C-hordein sequences differ are marked.

amounts applied. The highest yield of 76% for one of the C-hordein analyses indicates there was no significant N-terminal blocking.

No important minor residues were noted in sequencing the C-hordein, but three amino acids were identified in the first cycle of the ω-gliadin sequence and two amino acids at each subsequent cycle. Alanine was the major residue at cycle 1 in the ω-gliadin, but glycine and arginine were clearly identified; in subsequent cycles it became clear that the second sequence appearing at about half the level of the major sequence was the same as that of the major sequence displaced by one cycle. It evidently resulted from a protein component without the terminal alanine (or glycine) residue. Electrophoresis of the purified ω-gliadin did not show any minor components in sufficient amount to account for the second sequence; however, it is unlikely that a component differing by only one neutral amino acid would be resolved. The extra residue could mean that at least two genes code for the proteins in our preparation, but post-translational modifications might modify a single component in such a way as to produce the 2:1 ratio of components we observed.

Comparison of the amino acid sequences of our two components (Fig. 2) shows that they differ only at four positions if we do not consider the N-terminal alanine residue that is present in part of the ω-gliadin preparation. This demonstrates a close relationship between *H. vulgare* and *T. monococcum*. No fossil evidence provides a measure of the time since divergence of *Hordeum* and *Triticum*. If we assume for purposes of speculation, however, that the rate of evolution of the N-terminal regions of our proteins is equivalent to that of one of the fastest evolving peptide systems characterized so far, the fibrinopeptides (9.0 amino acid substitutions per site per 10^9 yr)[20], then the time since divergence of our two species would be 16 Myr.

Table 1 Amino acid compositions (mol%) of the ω-gliadin protein from *T. monococcum* and the C-hordein protein from barley (var. Julia)

	ω-Gliadin	*C*-hordein
Asp	1.18	0.83
Thr	1.08	0.99
Ser	4.89	2.61
Glu	45.2	41.1
Pro	26.2	31.9
Gly	0.89	0.41
Ala	1.57	0.69
Val	0.80	1.11
Met	0.13	0.22
Ile	2.22	3.02
Leu	4.70	4.31
Tyr	1.17	2.29
Phe	8.09	9.03
His	0.56	0.60
Lys	0.13	0.0
Arg	1.04	0.84
Cys/2	Trace	Trace

The similar sequences we obtained for our C-hordein and our ω-gliadin from *T. monococcum* show little homology with any other N-terminal sequences reported for α-, β- or γ-gliadins of common wheat[9,10,21,22], rye prolamins[10] or maize prolamins[11]. Our sequence is the first reported, however, for any ω-gliadin component. Autran *et al.*[10] did not note evidence of our ω-gliadin sequence in sequencing of the whole prolamin mixture from the same accession of *T. monococcum*, but this is not surprising as ω-gliadins probably constitute less than 10% of the total mixture. Our C-hordein sequence is closely similar to sequences recently obtained for the total hordein mixture[23], the C-hordein mixture[24] and a partially purified C-hordein component[25].

In common bread wheat, which is hexaploid (genomes A, B and D), gliadin proteins are encoded by genes located on homoeologous (partially homologous, but non-pairing in the polyploid) chromosomes 1 and 6 of each genome[7,26–28]. Shepherd[29] has suggested that all gliadin genes were located originally on one chromosome and that translocation of part of this chromosome to another gave rise to the common ancestor that, in turn, differentiated into the progenitors of the common wheat genomes A, B and D. This common ancestor must have already differentiated from the line that gave rise to *Hordeum* as all the prolamin genes of barley are located on chromosome 5 with the B-hordeins and C-hordeins being coded by two separate, but linked, loci[30]. Our results suggest that the C-hordein locus of barley and the ω-gliadin loci located on homoeologous chromosomes of group 1 in common wheat are homologous. It is also possible that the B-hordein locus is homologous with the gliadin loci coding for α-, β- and γ-gliadins that are located on homoeologous chromosomes of group 6 in common wheat but this is more difficult to establish as B-hordein is blocked at the N-terminus[12,25].

We thank A. Noma for amino acid analyses, B. J. Miflin for barley samples, and B. L. Jones for the sample of *T. monococcum*. P. R. Shewry and J.-C. Autran are visiting scientists in receipt of NATO postdoctoral research fellowships. Reference to a company or product name by the Department is only for information and does not imply approval or recommendation to the exclusion of others that may also be suitable.

Received 2 April; accepted 21 May 1980.

1. Harlan, J. R. in *Origins of Agriculture* (ed. Reed, C. A.) 357–383 (Mouton, The Hague, 1977).
2. Konzak, C. F. *Adv. Genet.* **19**, 407–582 (1977).
3. Smith, D. B. & Flavell, R. B. *Biochem. Genet.* **12**, 243–256 (1974).
4. Miflin, B. J. & Shewry, P. R. in *Seed protein Improvement in Cereals and Grain Legumes*, 137–158 (International Atomic Energy Agency, Vienna, 1979).
5. Kasarda, D. D., Bernardin, J. E. & Nimmo, C. C. in *Advances in Cereal Science and Technology* Vol. 1 (ed. Pomeranz, Y.) 158–236 (American Association of Cereal Chemists, St Paul, Minnesota, 1976).
6. Shewry, P. R., Ellis, J. R. S., Pratt, H. M., Miflin, B. J. *J. Sci. Fd Agric.* **29**, 433–441 (1978).
7. Wrigley, C. W. & Shepherd, K. W. *Ann. N.Y. Acad. Sci.* **209**, 154–162 (1973).
8. Mecham, D. K., Kasarda, D. D. & Qualset, C. O. *Biochem. Genet.* **16**, 831–853 (1978).
9. Bietz, J. A., Huebner, F. R., Sanderson, J. E. & Wall, J. S. *Cereal Chem.* **54**, 1070–1083 (1977).
10. Autran, J. C., Lew, E. J.-L., Nimmo, C. C. & Kasarda, D. D. *Nature* **282**, 527–529 (1979).
11. Bietz, J. A., Paulis, J. W. & Wall, J. S. *Cereal Chem.* **56**, 327–332 (1979).
12. Shewry, P. R., Field, J. M., Kirkman, M. A., Faulks, A. J. & Miflin, B. J. *J. exp. Bot.* **31** (in the press).
13. Booth, M. R. & Ewart, J. A. D. *Biochim. biophys. Acta* **181**, 226–233 (1969).
14. Charbonnier, L. *Biochim. biophys. Acta* **359**, 142–151 (1974).
15. Edman, P. & Begg, G. *Eur. J. Biochem.* **1**, 80–91 (1967).

16. Pisano, J. J., Bronzert, T. J. & Brewer, H. B. Jr *Analyt. Biochem.* **45,** 43-59 (1972).
17. Kulbe, K. D. *Analyt. Biochem.* **59,** 564–573 (1974).
18. Jeppson, J.-O. & Sjoquist, J. *Analyt. Biochem.* **18,** 264–269 (1967).
19. Bhown, A. S., Mole, J. E., Weissinger, A. & Bennett, J. C. *J. Chromatogr.* **148,** 532–535 (1978).
20. Kimura, M. *Scient. Am.* **241** (5), 98–126 (1979).
21. Kasarda, D. D., Da Roza, D. A. & Ohms, J. I. *Biochim. biophys Acta* **351,** 290–294 (1974).
22. Patey, A. L., Evans, D. J., Tiplady, R., Byfield, P. G. H. & Matthews, E. W. *Lancet* **ii,** 718 (1975).
23. Bietz, J. A. *Cereal Chem.* (in the press).
24. Shewry, P. R., March, J. M. & Miflin, B. J. *Phytochemistry* (in the press).
25. Schmitt, J. M. & Svendsen, I. *Carlsberg Res. Commun.* (in the press).
26. Shepherd, K. W. *Proc. 3rd Int. Wheat Genet. Symp.* (eds Finlay, K. W. & Shepherd, K. W.) 86–96 (Australian Acad. Sci., Canberra, 1968).
27. Kasarda, D. D., Bernardin, J. E. & Qualset, C. O. *Proc. natn. Acad. Sci. U.S.A.* **73,** 3646–3650 (1976).
28. Mitrofanova, O. P. *Tsitologiya i Genetika*–10 (3), 244–247 (1976).
29. Shepherd, K. W. in *Proc. 4th Int. Wheat Genet. Symp.* (eds Sears, E. R. & Sears, L. M. S.) 745–760 (Agric. exp. Station, University of Missouri, Columbia, 1973).
30. Shewry, P. R. *et al. Heredity* **44,** 383–389 (1980).