Amino acid sequence alignment of cereal storage proteins

Gerald R. Reeck and Charles Hedgcoth

Department of Biochemistry, Willard Hall, Kansas State University, Manhattan, KS 66506, USA

Received 14 November 1984

An alignment is presented of portions of the amino acid sequences of two gliadins and a glutenin from wheat and of a barley hordein. The two gliadins exhibit similarity over much of their sequences. The glutenin is similar in sequence to the gliadins only over a restricted region. Our analysis of the aligned sequences leads us to suggest the word 'modular' to describe the architecture of these proteins. The term is intended to connote the joining together, in the course of evolution, of several units (modules) of distinctive character, under a set of rules that allows considerable flexibility in the arrangement of modules within a molecule.

Gliadin *Glutenin* *Hordein* *Amino acid sequence* *Wheat* *Barley* *Homology*

1. INTRODUCTION

The wheat storage proteins are a large group of physically and chemically similar molecules [1]. They have been largely intractable towards efforts of protein chemists to obtain and study individual polypeptides. Recently, a good deal of amino acid sequence information for several wheat storage proteins has been inferred from nucleotide sequences of genomic or cDNA clones [2–5]. One of the more interesting issues that can be addressed with amino acid sequences of related proteins is the evolutionary history of the molecules. We have found that the reported sequences of two gliadins, a high-M_r glutenin, and a B-hordein from barley are similar and can be readily aligned. This observation indicates that the 4 proteins are members of a family of related (homologous) proteins and allows us to begin to understand the construction of the molecules in the course of evolution.

2. MATERIALS AND METHODS

The amino acid sequences that we have aligned were deduced from nucleotide sequences of a genomic clone for a wheat α/β-gliadin [2] and of cDNA clones for a wheat γ-gliadin [3], a high-M_r glutenin from wheat [4], and a B-hordein from barley [5].

The similarities among the sequences were strong enough that the sequences could be aligned by visual inspection. In comparing alternate alignments, we used the scheme of McLachlan [6] to assign a score to each pairing of amino acid residues. The rules and data of Chou and Fasman [7] were used to predict secondary structure.

3. RESULTS AND DISCUSSION

The complete sequence of a wheat gliadin of the α/β subfamily is available from the nucleotide sequence of the genomic clone [2]; the partial sequences inferred from cDNA sequences correspond to the C-terminal portions of the other 3 proteins. The complete amino acid sequence of the α/β-gliadin is the basis of our numbering scheme. Alignment of the longest (γ-gliadin) of the 3 partial sequences starts after the continuous stretch of 18 glutamine residues at positions 96–113 of the α/β-gliadin. Optimal alignment of the two gliadin sequences requires few deletions, and the sequences exhibit rather strong similarity. The aligned sequences (fig.1) reveal 3 distinct regions. Two regions, 114–181 (particularly 126–181) and 219–266, contain hydrophobic residues at levels typical of folded proteins. The greatest similarity between the two gliadins is in these regions. The sequence linking these two regions (182–218) is quite

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

rich in glutamine, contains few hydrophobic residues, and has a fairly high proline content. Glutamine matches dominate the alignment in this region. Other amino acids are frequently matched with quite dissimilar residues. In other words, in this region the sequences are not highly conserved on a residue-by-residue basis, but the overall glutamine-rich character of the region is preserved.

The shorter partial sequences from a barley B-hordein and a wheat high-M_r glutenin polypeptide align readily with the C-terminal portion of the α/β-gliadin and γ-gliadin sequences. Note, however, that some known sequence of the high-M_r glutenin is not aligned. This lies on the N-terminal side of that portion of the high-M_r glutenin sequence shown in the alignment. A convenient starting point in making the alignments was a Met-Cys sequence that occurs in all 4 sequences (positions 246 and 247 in fig.1). There is rather strong conservation on each side of this Met-Cys sequence.

Large portions of the α/β-gliadin and γ-gliadin sequences are similar. From this, we infer that they are homologous proteins, i.e., that at least those portions of the two sequences that we have aligned have derived from a common ancestral polypeptide. Similarly, at least those portions of the B-hordein and high-M_r glutenin for which sequences are available are homologous to the two gliadins. Thus, the 4 sequences constitute a family of related polypeptides, which appear to have been derived (over at least a portion of their sequences) from a common ancestral polypeptide.

The B-hordein sequence is somewhat more similar to the γ-gliadin than to the α/β gliadin. This suggests that the B-hordeins are the barley analogs of wheat γ-gliadins. Although it is much larger than the gliadins, the high-M_r glutenin (M_r ~100000 from SDS gel electrophoresis [8]) apparently is also related evolutionarily to the gliadins, the majority of which have M_r values of 30000–50000 from SDS gel electrophoresis. We then would tentatively divide the family of 4 sequences into 3 subfamilies: the α/β-gliadin group; a group that contains the γ-gliadin and the B-hordein; and the high-M_r glutenin group.

Our alignment suggests that primary structure in this family of cereal storage proteins is 'modular'; i.e., in each protein several units of distinctive character are linked together. We prefer to call the

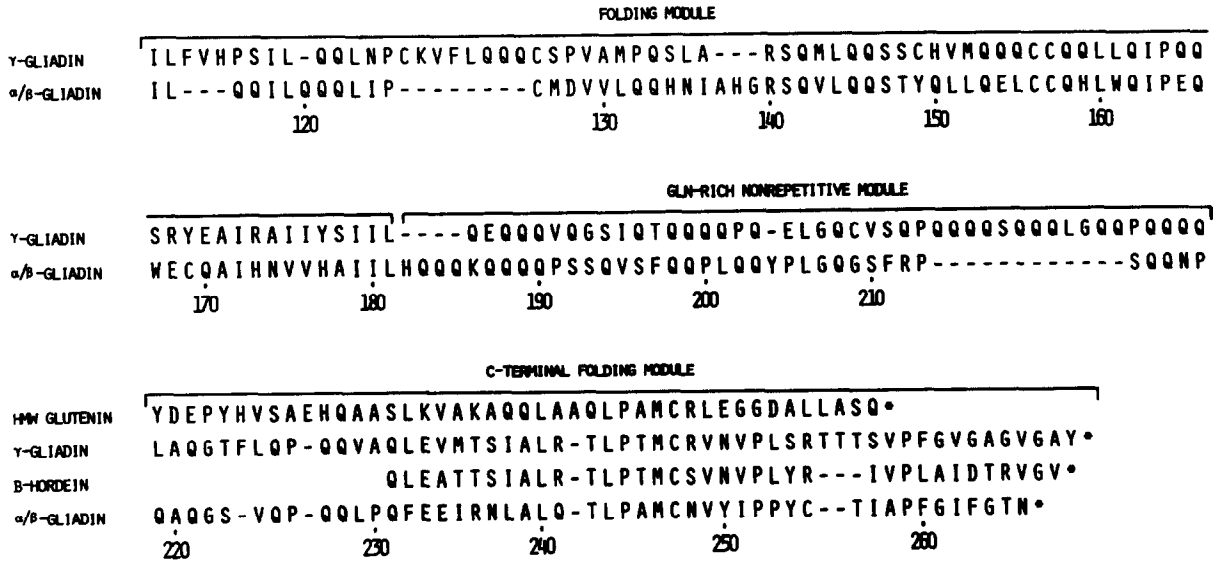

Fig.1. Alignment of sequences from 4 cereal storage proteins. Residue numbers are indicated for the α/β-gliadin. Sequence information is available to the N-terminal side of the sequences shown for the gliadins and the glutenin. That information is not shown because there are no homologies among the 3 proteins in those sequences (see text). The C-terminal sequence of each protein is complete, as indicated by *.

units modules rather than domains since their identification is entirely at the level of primary structure, and the term 'domain' ordinarily carries connotations of tertiary structure. Furthermore, some of the modules are quite short and more likely to be linkers or bridges between domains rather than domains.

We can identify 3 types of modules. One is composed of repeats (with some variations) of a sequence of 12–20 residues that is glutamine-rich and, to a lesser extent, proline-rich. A second type of module is also glutamine-rich and rather deficient in hydrophobic residues, but does not contain repeats. An extreme form of this type of module is a continuous stretch of glutamine residues. A third type of module contains hydrophobic residues at a level typical of globular proteins.

Glutamine/proline-rich repetitive modules have been noted in an α/β-gliadin [2], a γ-gliadin [3], and in high-M_r glutenin polypeptides [4,8]. In each case, these occur immediately to the N-terminal side of the sequence shown in fig.1. Although this type of module has a characteristic amino acid composition, there is no apparent similarity in the repeat sequences of the 3 subfamilies. Thus, the portions of the proteins comprised of this sort of module apparently are not homologous from subfamily to subfamily, although they might be homologous within a subfamily. Indeed, the nature of the repeating sequence within these modules could be a diagnostic feature of a subfamily within the wheat storage protein family. The repeats in primary structure in this type of module could be expected to give rise to repeats in 3 dimensions, i.e., to produce a helix.

The second type of module, the glutamine-rich nonrepetitive regions, is represented by residues 182–218 of the α/β-gliadin and by the corresponding portion of the γ-gliadin. Although this region is not highly conserved, it does appear that these two modules are homologous, that is, that they have likely evolved from a stretch of common ancestral polypeptide. Lacking repeats and being relatively deficient in hydrophobic residues, this type of module is unlikely to be structured in a 3-dimensional sense and may resemble a random coil. We consider the continuous stretch of glutamine residues (positions 96–113) in the α/β-gliadin to be an extreme form of this type of module although one could argue that it is a distinct type of module since polyglutamine might have a tendency to form an α-helix [7–9].

The third type of module we shall call a folding module. This type occurs in the α/β-gliadin at positions 114–181 and 219–266 and in the corresponding positions of the γ-gliadin. The C-terminal portions of B-hordein and the high-M_r glutenin are also representatives of this type of module (fig.1), which contains hydrophobic residues at a level typical of globular proteins. This region contains all the cysteine residues in the α/β-gliadin and all but one of the cysteine residues in the portion of the γ-gliadin for which the sequence has been reported. It is these regions in which the highest level of sequence conservation is apparent between the α/β- and γ-gliadins. Sequences commonly are conserved in proteins to preserve tertiary structure (i.e., a folding pattern). This, along with amino acid composition, suggests that this sort of module in the wheat storage proteins is folded into a specific 3-dimensional structure; hence the term 'folding module'. The two folding modules of the α/β-gliadin sequence have high helix and β-sheet potential under the predictive scheme of Chou and Fasman [7]. Physical measurements indicate that gliadins contain about 25% helix [1]. Our analysis suggests that the helical regions are at least largely within the folding modules. Intramolecular or intermolecular disulfide bonds could form between folding modules. Their relatively high conservation suggests that the folding modules may be the key structural features in this family of cereal storage proteins.

There appears to have been considerable flexibility in combining modules in the evolution of wheat storage proteins. Each of the proteins studied so far contains a folding module at the C-terminus. In the α/β- and γ-gliadins this is preceded by a glutamine-rich structureless module. In the high-M_r glutenin, on the other hand, the C-terminal folding module is preceded by a large glutamine/proline-rich repetitive module [4]. (It will be most interesting to learn whether a folding region homologous to residues 114–181 of the α/β-gliadin occurs in the high-M_r glutenin on the N-terminal side of the glutamine/proline-rich repetitive module.) Another example of different arrangements of modules is this: in the α/β-gliadin a continuous stretch of glutamine immediately

precedes the first folding module. There is no analogous glutamine-rich structureless module in the γ-gliadin. Instead, in this molecule, a glutamine/proline-rich repetitive module immediately precedes the first folding module. The apparent flexibility, which is probably limited by rules we do not yet understand, in the evolutionary construction in this family of proteins is part of the reason for our choice of the term 'modular'.

In the case of the gliadins, it appears that after duplication of a gene coding for 3 modules (two folding modules separated by a glutamine-rich nonrepetitive module) different repetitive modules were attached to the N-terminal regions, forming molecules homologous over only part of their sequence (the two folding modules and the linking glutamine-rich nonrepetitive module). In subsequent evolution of the homologous regions, constraints on the structures of the folding module maintained greater sequence similarity than in the module that joins the folding modules.

We note that an alternate view of organization of sequence units in wheat gliadins appeared recently [10], after our manuscript had been written. The data presented there are nicely accommodated by the approach outlined in this paper.

ACKNOWLEDGEMENTS

This work was supported by the Kansas Agricultural Experiment Station (publication no.85-172-j) and by USDA grant 84-CRCR-1-1514 to C.H.

REFERENCES

[1] Kasarda, D.D., Nimmo, C.C. and Kohler, G.O. (1979) in: Wheat: Chemistry and Technology (Pomeranz, Y. ed.) 2nd edn, pp.227–299, Am. Assoc. Cereal Chemists, St. Paul, MN.

[2] Rafalski, J.A., Scheets, K., Metzler, M., Peterson, D.M., Hedgcoth, C. and Söll, D.G. (1984) EMBO J. 3, 1409–1415.

[3] Bartels, D. and Thompson, R.D. (1983) Nucleic Acids Res. 11, 2961–2977.

[4] Forde, J., Forde, B.G., Fry, R.P., Kreis, M., Shewry, P.R. and Miflin, B.J. (1983) FEBS Lett. 162, 360–366.

[5] Morde, B.G., Kreis, M., Bahramian, M.B., Matthews, J.A. and Miflin, B.J. (1981) Nucleic Acids Res. 9, 6689–6707.

[6] McLachlan, A.D. (1971) J. Mol. Biol. 61, 409–424.

[7] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 45–148.

[8] Thompson, R.D., Bartels, D., Harberd, N.P. and Flavell, R.B. (1983) Theor. Appl. Genet. 67, 87–96.

[9] Lewis, P.N., Go, N., Go, M., Kotelchuck, D. and Sheraga, H.A. (1970) Proc. Natl. Acad. Sci. USA 65, 810–815.

[10] Kasarda, D.D., Okita, T.W., Bernardin, J.E., Baecker, P.A., Nimmo, C.C., Lew, E.J.-L., Dietler, M.D. and Greene, F.C. (1984) Proc. Natl. Acad. Sci. USA 81, 4712–4716.