Molecular characterization of α-gliadin genes from common wheat cultivar Zhengmai 004 and their role in quality and celiac disease

Yuge Li, Ranran Xin, Dale Zhang, Suoping Li

College of Life Science, Henan University, Kaifeng 475004, China

A total of 43 unique clones (Z4A-1 to Z4A-43 with GenBank accession numbers of HM120221, HM120222, JX828270, JN831402 to JN831406, and KC715889 to KC715923, respectively) were amplified and cloned from common wheat cultivar Zhengmai 004 using a PCR-based strategy. They included 22 full-ORF α-gliadin genes and 21 pseudogenes containing at least one in-frame stop codon. Comparative analysis of the deduced amino acid sequences showed that all the isolated genes displayed the typical structural features of α-gliadin genes and that the putative proteins of Z4A-7, Z4A-14, Z4A-17 and Z4A-20 had an extra cysteine residue in the unique domain II, while Z4A-15 lacked the second conserved cysteine residue in the unique domain I. The two fusion proteins of Z4A-15 and Z4A-20 were successfully detected by SDS-PAGE and Western blotting, although the protein level was relatively low. Based on the occurrence of the four major epitopes, as well as the lengths of the two glutamine repeats, 8, 6, and 8 genes were assigned to the Gli-2 loci on the respective chromosomes 6A, 6B, and 6D and a total of respectively 16, 0 and 23 immunogenic peptides were identified. In addition, 4 of the 5 genes with odd numbers of cysteine residues were assigned to chromosome 6D, suggesting that common wheat cultivar Zhengmai 004 has the potential to induce celiac disease (CD) and that the D genome exerts the most influence on gluten quality, but is the most deleterious for CD patients. By phylogenetic analysis, 11 exceptional α-gliadins with few or no immunogenic peptides from Triticum monococcum and Aegilops tauschii were detected, a finding that further supports the prospect of CD prevention. Finally, secondary structure prediction showed that most (98.48%) of the α-gliadins invariably contained five conserved α-helices (H1 to H5) in the two glutamine repeats and unique domains and 67.68% of the α-gliadins also contained a β-strand (S) in the C-terminal unique domains. An absent α-helix H2, 1–2 extra α-helices, or an additional β-strand (Sβ) also probably occurred in some cases. Of the 22 cloned genes in this work, 10 putative proteins contained 1–2 extra α-helices in addition to the five conserved α-helices or the additional β-strand. The observation that most of the α-helices and β-strands were present in the two unique domains and that an extra α-helix also probably occurred in the two glutamine repeats in some desirable genes strongly suggested that these two unique...
1. Introduction

Comprising approximately 50% of wheat gluten proteins, gliadins have essentially a plasticizing effect on gluten structure and mainly impart viscosity to dough [1]. Though it is generally concluded that gliadins exert mainly negative effects on overall dough strength, positive contributions of these proteins to loaf volume have also been detected [2–5]. Based on their mobility in the A-PAGE gels, as well as their different primary structures, gliadins can be divided into three groups: α, γ- and α-gliadins [6]. Among them, the α-gliadins, encoded by Gli-2 loci on the short arms of the group 6 chromosomes, are typically the storage proteins most heavily consumed by humans, being the most abundant seed storage proteins (accounting for 15%–30% in most wheat cultivars) [6–9]. They also play the largest positive role in increasing loaf volume, while showing the lowest weakening effects on dough strength [4,5]. Functional analysis in vitro [10] of such contributions to wheat flours by the α-gliadin protein subunit ACX71610 (encoded by GQ891685 and carrying an extra cysteine residue in the C-terminal unique domain II) has been confirmed. But recent advances in the study of the pathogenesis of celiac disease (CD), a T-cell-mediated chronic inflammatory disease with an incidence as high as 1% in many populations and confirmed. But recent advances in the study of the pathogenesis of celiac disease (CD), a T-cell-mediated chronic inflammatory disease with an incidence as high as 1% in many populations and caused by a permanent intolerance of dietary gluten, have also revealed that the α-gliadins are the major initiators of CD [11–14].

Based on the available literature, a variety of gluten peptides with proven in vivo or in vitro activity have been identified in gliadins as well as glutenins; however, their relative importance differs [15]. Only five peptides, one (glia-γ1: QQFPQQSFQQQ) occurring in γ-gliadins and four (glia-α9: FPFPQLFY, glia-α2: PQFPQLYPKPYQLPQ, glia-α20: FPFPQFYQPPQ, and glia-α: QGSFQPSQQ) in α-gliadins, are dominant, and are generally referred to as the immunodominant peptides. They have been shown to be recognized by T-cells from almost all CD patients, both children and adults, whereas T-cell responses to other gluten proteins are much less frequent and generally appear in young CD patients. Furthermore, they elicit a stronger T-cell response and their immune activity is designated as +++ compared to the + of the other epitopes [16–21].

Comparative analysis [13] of the deduced amino acid sequences of the full-ORF α-gliadin genes derived from several diploid wheat species representing the ancestral A (Triticum monococcum), D (Aegilops tauschii) and potentially ancestral B (Aegilops speltoides) genome of hexaploid bread wheat indicates significant differences in the average lengths of the two gluten repeats, as well as the occurrence of the four major T-cell peptides in α-gliadins, according to their genomic origin. The α-gliadins derived from the A genome almost invariably contain only glia-α9 and glia-α20 and carry a larger average number (18.8 ± 1.9) of glutamine residues in the second gluten repeat than do the A (10.2 ± 0.6) and D (9.7 ± 1.4) genomes. In contrast, the α-gliadin genes from the D genome are the most deleterious for human CD sufferers: not only can they contain all four major T-cell immunogenic peptides in variable combinations in different proteins, but some of them can even harbor a repeat of glia-α2 and form the extremely (several-fold more potent than any other known gluten peptides) immunogenic T-cell stimulator known as the 33-mer peptide: LQLQFPQFQQLPPQFPQQLPQFPQQQLPQFPQ. Thus, α-gliadin genes can be assigned to specific chromosome loci according to their marked genomic differences [12,13].

Further analysis of group 6 nulli-tetrasomic lines of Chinese Spring confirmed the reliability of such assignment methods for α-gliadin genes [23].

In conclusion, α-gliadins not only play a major role in determining gluten quality, but comprise the major source of toxicity for CD patients, given that they contain most of the main toxic components. In addition, this multigenic family encodes extensive allelic variation that has been shown to be closely associated with flour quality [24,25]. Screening of new allelic variants with specific profiles of α-gliadins from common wheat cultivars with good quality or from other valuable Triticeae species may accordingly aid in exploring gene resources both for quality improvement and potential CD prevention. The objective of the current study was to clone and characterize the novel full-ORF α-gliadin genes from common wheat cultivar Zhengmai 004, one of the major cultivars sown on a large scale in the weak-gluten wheat growing areas of China owing to its good quality and high and stable yield. To shed light on the structure—function relationships of a single α-gliadin gene, the prokaryotic expression in Escherichia coli of two genes differing in the number of cysteine residues was investigated by SDS–PAGE and Western blotting. Finally, the secondary structures of the full-ORF genes cloned in this study and other genes in the public database GenBank derived from common wheat and its relatives were predicted and the typical secondary structure of α-gliadins was summarized.

2. Materials and methods

2.1. Plant material

Seeds of Zhengmai 004 were kindly provided by Professor Hu Lin from the Wheat Research Institute of Henan Academy of Agricultural Sciences, Zhengzhou, China.

2.2. Genomic DNA extraction and gene cloning

Genomic DNA was extracted from young leaves of 10–20 wheat seedlings grown in the greenhouse, using the cetyltrimethyl ammonium bromide (CTAB) procedure. A pair of degenerate primers (F: 5′-GGA TCC ATG AAG ACC TTT CTC ATC CT-3′; R: 5′-
AAG CTT TCA GTT GGT ACC GAA GAT GCC-3') with respectively Bam H I and Hind III sites (underlined) at the 5′-end of each primer was designed according to the majority of the published open reading frame (ORF) sequences of α-gliadin genes in GenBank.

PCR was performed using LA Taq (TaKaRa, Dalian, China) with GC buffer (1 unit) in a 20-μL reaction volume containing approximately 50 ng of genomic DNA, 100 μmol L\(^{-1}\) of each dNTP, and 0.5 μmol L\(^{-1}\) of each primer. PCR cycling was at 94 °C for 4 min followed by 10 cycles of 94 °C for 30 s, 62 °C (\(T_m + 4 °C\)) for 45 s, 72 °C for 60 s, then 22 cycles of 94 °C for 30 s, 58 °C for 45 s, 72 °C for 60 s, and a final extension at 72 °C for 15 min.

PCR products were separated on 1% agarose gels and the single target fragment was purified from the gels using Gel Extraction Kit Ver 2.0 (TaKaRa, Dalian, China). The purified PCR products were cloned into the pMD-19T-sense vector (TaKaRa, Dalian, China) and transformed into E. coli (DH5α) competent cells by standard protocols. On average two recombined DNA clones for each amplified fragment were bidirectionally sequenced by the Beijing Genomics Institute (BGI, Beijing, China).

Sequence alignments were based on multiple alignments provided by the software Clustal W version 1.8 (http://www.clustal.org/), Ultraedit 3.2 (http://www.ultraedit.com/) and Bioedit 7.0 (http://www.mbio.ncsu.edu/BioEdit/). A neighbor-joining tree of the genes cloned in this study and other genes in GenBank was constructed based upon the deduced amino acid sequences without signal peptides using Mega 4.0. The identification of the four major immunogenic peptides in α-gliadins and their chromosomal locations followed Van Herpen et al. [13].

Prediction of the secondary structure of α-gliadins was performed with the latest online version (3.3) of the PSIPRED server (http://bioinf.cs.ucl.ac.uk/psipred/psiform.html).

The positive recombinant pMD-19T-α-gliadin plasmids and pET30a plasmids were digested with the enzymes Hind III and BamH I (FastDigest enzyme, Fermentas, Canada) at 37 °C for 20 min and the target fragments were purified and ligated together with the fast ligation kit of Sangon Biotech (Shanghai, China). The identity of the recombinant pET30a-α-gliadin plasmids was confirmed by PCR and DNA bidirectional sequencing (BGI, Beijing, China) and the positive recombinant plasmids were transformed into E. coli BL21 (DE3) (Novagen) competent cells. The fusion protein was induced by 1 mmol L\(^{-1}\) IPTG at 37 °C for at least 4 h. Fusion protein was extracted from the bacteria using the method described by Xu et al. [26], with some modifications. SDS-PAGE electrophoresis and Western blotting were referred to the method described by Li et al. [10].

### 3. Results

#### 3.1. Molecular characterization of the 22 full-ORF α-gliadin genes

A total of 43 unique clones, designated as Z4A-1 to Z4A-43, were isolated from common wheat cultivar Zhengmai 004 by a genomic PCR-based strategy. Among them, 22 clones (Z4A-1 to Z4A-22) were full-ORF genes that could encode protein subunits with the size of 286–312 amino acid residues. NCBI BLAST searches of their entire nucleotide sequences showed that 42 sequences had a high degree (84%–99%) of identity with the typical α-gliadin sequences in GenBank, with the exception of the complete identity of Z4A-22 with the previously submitted sequence (JX828270) that we isolated earlier from common wheat cultivar Zhengmai 9023. The 42 novel sequences were submitted to GenBank and accession numbers HM120221, HM120222, JN831402 to JN831406, and KC715889 to KC715923 were assigned.

Multiple alignment of the deduced amino acid sequences of 22 full-ORF genes and 3 typical α-gliadin genes derived from bread wheat cultivars Shan 253 (GQ891685), Chuannong 16 (DQ246448) and Gaocheng 8901 (EF561274) in GenBank showed that the 22 genes possessed typical structures of the previously characterized α-gliadin genes (Fig. 1). The size of each sequence depended principally on the length of the N-terminal repetitive region and two polyglutamine domains. Compared to other sequences, in the N-terminal repeated region, a deletion LPYPQPQ at position 82–88 was detected in Z4A-3 to Z4A-6, Z4A-8, Z4A-13, Z4A-18, Z4A-21 and Z4A-22, while an extra insertion QLPYPQP at position 100–106 was identified in Z4A-5. In the two glutamine repeats, the number of glutamine residues varied from 9 to 27 in the first and 5 to 22 in the second. In the two unique domains, six conserved cysteine residues were found in 17 genes, except that Z4A-15 lacked the second conserved cysteine residue (C\(_5\)) in the unique domain 1, Z4A-7, Z4A-14, Z4A-17 and Z4A-20 contained an extra cysteine residue created by a serine-to-cysteine residue change in the C-terminal unique domain II.

In addition to the 22 full-ORF genes, 21 pseudogenes containing at least one in-frame stop codon resulting from base transition (accounting for 80.95%) or frameshift mutations (Z4A-30, Z4A-39, Z4A-41 and Z4A-43) were identified. Of the stop codons caused by base transition, single-base C to T substitution, turning a CAA or CAG codon for glutamine residue into a TAA or TAG stop codon, accounted for 91.43% of the cases. Notably, the deduced amino sequence of Z4A-27 lacked the unique domain I compared to the other typical α-gliadin genes.

#### 3.2. Prokaryotic expression and Western blotting

To confirm authenticity and provide a useful basis for further study of structure-function relationships, two putative proteins (Z4A-15 and Z4A-20) with different numbers of cysteine residues were further constructed in the expression vector pET30a. By PCR and DNA sequencing, the positive recombinants were confirmed to have been correctly incorporated into the pET30a plasmids. The two recombinant plasmids were transformed into E. coli BL21 and the fusion proteins were induced with 1 mmol L\(^{-1}\) IPTG at...
Fig. 1 – Multiple alignment of the deduced amino acid sequences of 22 novel genes in this study and 3 typical α-gliadin genes from GenBank. Full and broken-line frames indicate the sites of conserved and extra cysteine residues, respectively.
37 °C for at least 4 h and detected by SDS-PAGE and Western blotting (Fig. 2).

SDS-PAGE electrophoresis yielded two specific protein bands of size close to that of the fusion protein at around 38 kDa (Fig. 2-a, indicated by arrows) in the induced samples of Z4A-15 and Z4A-20, though the expression levels were low compared to those of the bacterial proteins. Based on the results of Western blotting (Fig. 2-b), the induced fusion proteins of Z4A-15 and Z4A-20 extracted from E. coli were further confirmed by their strong hybridization to the anti-His Tag mouse monoclonal antibody, whereas no hybridizing signals were detected for the bacterium with the pET30a empty vector and un-induced samples.

3.3. Identification of the four major T-cell peptides and their chromosomal locations

To assign the cloned genes to specific chromosome loci and complete the characterization of the toxicity of bread wheat cultivar Zhengmai 004 for CD patients, the numbers of the four major T-cell immunogenic peptides and of glutamine residues presented in the two polyglutamine regions in the 22 α-gliadins were determined and are listed in Table 1.

As shown in Table 1, based upon the occurrence of the four major T-cell immunogenic peptides, as well as the relative lengths of the two polyglutamine domains, the deduced protein sequences of 8 genes (Z4A-3, Z4A-4, Z4A-6, Z4A-8, Z4A-13, Z4A-18, Z4A-21 and Z4A-22) that contained only glia-α9 and glia-α20 showed that the number of glutamine residues in their glutamine repeat I was relatively large, except for Z4A-22. They could accordingly be assigned to chromosome 6A based on these observations. Similarly, six other α-gliadin genes (Z4A-1, Z4A-2, Z4A-9, Z4A-11, Z4A-12 and Z4A-17) were assigned to chromosome 6B because their amino acid sequences contained none of the four major T-cell epitopes and, except for Z4A-2, carried relatively large numbers of glutamine residues in glutamine repeat II. The remaining 8 genes (Z4A-5, Z4A-7, Z4A-10, Z4A-14, Z4A-15, Z4A-16, Z4A-19 and Z4A-20) contained 2 to 4 epitopes in different combinations. Moreover, even repeats of glia-α2 were identified in the N-terminal repetitive domain of Z4A-5, resulting from an extra insertion of QLPYPQP at position 100–106. They were accordingly assigned to chromosome 6D. In total, 16, 0 and 23 epitopes were represented in 8, 6 and 8 genes located on chromosome 6A, 6B and 6D, respectively. Clearly Zhengmai 004 had full potential to induce the CD syndrome.

3.4. Phylogenetic analysis and chromosomal locations of α-gliadin genes

Based on the deduced amino acid sequences without signal peptides among the 22 cloned genes, as well as all the 95 full-ORF genes derived from three diploid wheat species (46 from T. monococcum, 12 from Ae. speltoides and 37 from Ae. tauschii) in GenBank, a phylogenetic tree was constructed, resulting in clear clustering by genomic origin (Fig. 3). Most of the sequences derived from T. monococcum and Ae. tauschii, and all the sequences derived from Ae. speltoides, formed separate clusters designated as groups 1, 3 and 2, respectively. Groups 1, 2 and 3 clearly represent the respective α-gliadin genes on the A, B and D genomes, although 11 exceptional genes originating in T. monococcum (protein IDs ACJ76933 to ACJ76938) and Ae. tauschii (protein IDs ADD1701, ABQ96115, ABQ96118, ABQ96119 and ADM96154), but clustered in group 2, were also detected. Similarly, although most of the 22 genes cloned in this work and located on chromosome 6A, 6B and 6D were clustered respectively in groups 1, 2 and 3, two (Z4A-5 and Z4A-22) exceptional genes were also found.

To determine whether the 11 above-mentioned exceptional genes were distinctive with respect to the distribution and quantity of their CD epitopes, the numbers of four major T-cell stimulatory peptides and glutamine residues harbored in the two glutamine repeats, as well as their true genome and chromosome locations obtained on the basis of the occurrence of the four major epitopes, were determined and are listed in Table 2.

As shown in Table 2, the assignments of the 11 exceptional genes based on the occurrence of the four major peptides were consistent with the clusters in the phylogenetic analysis, rather than their authentic genomes. Protein subunit ADM96154 clustered in group 1 contained only peptides glia-α9 and glia-α20, whereas the other 10 protein subunits in group 2 contained only glia-α or even lacked all four immunogenic peptides. They would accordingly be expected to be located on chromosome 6A and 6B, rather than on their actual D or A genomes, based on the quantity and distribution of the four major peptides. In addition, compared to the general number of no more than 27 glutamine residues in
the first glutamine repeat, much larger glutamine repeats I with 38 or even 66 glutamine residues were also detected in the three protein subunits ABQ96115, ABQ96118 and ABQ96119.

In summary, these findings suggest that the distribution of the four immunodominant epitopes in α-gliadins is indeed distinct for each genome in most cases, whereas the wild genetic resources of T. monococcum and Ae. tauschii harbored extensive genetic diversity and some exceptional genes.

3.5. Prediction of the secondary structure of α-gliadins

To ascertain their molecular functions, the secondary structures of the mature protein subunits of the 22 deduced α-gliadins in this study, as well as the other 176 typical α-gliadin genes derived from common wheat and its diploid or tetraploid relatives, were predicted with the latest online version (3.3) of the PSIPRED server. The results showed that the numbers of α-helices and β-strands, as well as the amino acid residues involved in each conserved α-helix and β-strand, were always variable in different proteins, though their positions and core sequences were relatively conserved. The types, positions and distributions of the α-helices and β-strands in the 198 predicted α-gliadins are displayed in Table 3. A diagram summarizing the secondary structure of typical α-gliadins on the basis of these results is given in Fig. 4.

According to the absence or presence of the relatively conserved β-strand (S) in the C-terminal unique domain II, the secondary structures of α-gliadins can be classified into types I and II, and each type can be subdivided into eight groups on the basis of the positions of the absent or extra α-helix and β-strand involved. Among them, 32.32% of the α-gliadins belonged to type I, which contained only 4-7 α-helices, whereas 67.68% of the α-gliadins formed 1-2 β-strands in addition to the 4-7 α-helices and belonged to type II (Fig. 4 and Table 3). Generally, secondary structures were infrequent (2.53%) and were found in the N-terminal repetitive domain (H1). Five conserved α-helices (H1, H2, H3, H4 and H5) were nearly always (98.48%) distributed in the glutamine repeat I (H1), repeat II (partial sequence of H4), unique domain I (H2, H3 and H4) and C-terminal unique domain II (H5). Exceptions to this rule were three genes we isolated from common wheat cultivars Zhengfeng 5 (protein ID AFX69640) and Yumai 34 (protein IDs AFX69612 and AFX69609) that lacked α-helix H2, whereas the three above-mentioned distinctive α-gliadin genes formed one (protein ID ABQ96118) or even two (protein IDs ABQ96115 and ABQ96119) distinctly larger α-helices H1. In addition, one extra α-helix H22 (11.11%), H23 (6.06%), H24 (1.52%) or two additional α-helices H23 and H22 (1.52%) also probably occurred in some cases. With regard to the other main element of the secondary structure occurring in type II, in addition to the conserved β-strand (S), an additional β-strand (S2) was detected in four protein subunits (protein IDs AFX69607, AFO17690, AFX69601 and ABS72150). Obviously, most of the α-helices and β-strands are present in the two unique domains. It is noteworthy that both the three extra α-helices H24 (protein IDs AFOQ3468, AFX69638 and ABS72143) and the four additional β-strand S2 were located around the position where an extra cysteine residue was present or had most likely occurred (protein ID AFX69601) resulting from S → C substitution.

With respect to the secondary structures of the 22 deduced α-gliadins isolated from the common wheat cultivar Zhengmai 004 in this study, considerable variation was detected. Among them, 9 deduced α-gliadins (Z4A-1, Z4A-2, Z4A-5, Z4A-9, Z4A-12, Z4A-15, Z4A-18, Z4A-21 and Z4A-22) contained only 5-7 α-helices and belonged to type I, whereas the remaining 13 deduced α-gliadins formed a β-strand (S) in the C-terminal unique domain in addition to 5-6 α-helices and belonged to type II. Five type I genes had an extra α-helix H2 (Z4A-2, Z4A-9 and Z4A-12), H43 (Z4A-22) or even two α-helices H41 and H42

Table 1 - Numbers of the four major immunogenic peptides, glutamine residues carried in the two polyglutamine domains, and chromosomal locations of the 22 cloned α-gliadin genes.

<table>
<thead>
<tr>
<th>α-Gliadin genes</th>
<th>Major T-cell peptides</th>
<th>Chromosomal location</th>
<th>Glutamines</th>
<th>Poly I</th>
<th>Poly II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Glia-α9</td>
<td>Glia-α2</td>
<td>Glia-α20</td>
<td>Glia-α</td>
<td></td>
</tr>
<tr>
<td>Z4A-3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>A</td>
</tr>
<tr>
<td>Z4A-4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>A</td>
</tr>
<tr>
<td>Z4A-6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>A</td>
</tr>
<tr>
<td>Z4A-8</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>A</td>
</tr>
<tr>
<td>Z4A-13</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>A</td>
</tr>
<tr>
<td>Z4A-18</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>A</td>
</tr>
<tr>
<td>Z4A-21</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>A</td>
</tr>
<tr>
<td>Z4A-22</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>A</td>
</tr>
<tr>
<td>Z4A-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>B</td>
</tr>
<tr>
<td>Z4A-2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>B</td>
</tr>
<tr>
<td>Z4A-9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>B</td>
</tr>
<tr>
<td>Z4A-11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>B</td>
</tr>
<tr>
<td>Z4A-12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>B</td>
</tr>
<tr>
<td>Z4A-17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>B</td>
</tr>
<tr>
<td>Z4A-5</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>D</td>
</tr>
<tr>
<td>Z4A-7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>D</td>
</tr>
<tr>
<td>Z4A-10</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>D</td>
</tr>
<tr>
<td>Z4A-14</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>D</td>
</tr>
<tr>
<td>Z4A-15</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>D</td>
</tr>
<tr>
<td>Z4A-16</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>D</td>
</tr>
<tr>
<td>Z4A-19</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>D</td>
</tr>
<tr>
<td>Z4A-20</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>D</td>
</tr>
</tbody>
</table>
(Z4A-18), and 5 type II genes possessing an extra \( \alpha \)-helix \( \text{HE}_3 \) formed by two to six glutamine residues in the glutamine repeats II, it was observed that Z4A-14 and other 3 protein subunits (Protein IDs AFX69619, ABQ52119 and ABQ52126) derived from common wheat were more similar to that of ACX71610, in which the extra \( \alpha \)-helix \( \text{HE}_3 \) consisted of five or six glutamine residues. Considering that marked positive effects on the gluten elasticity by protein subunit ACX71610 had been verified by functional analysis in vitro, it is suggested that the putative protein of Z4A-14 may also be strongly associated with the high gluten quality of bread wheat cultivar Zhengmai 004.

### 4. Discussion

#### 4.1. Variation, genomic organization and function of the \( \alpha \)-gliadin genes

Like other wheat prolamins, \( \alpha \)-gliadins are encoded by multigenic families, the copy numbers of which have been estimated to vary from 25 [27] to 150 [28] in different wheat cultivars. However, previous studies revealed that the \( \alpha \)-gliadin gene family was composed of subfamilies, with at least one subfamily consisting of pseudogenes that accounted for approximately 50\% [8] or even 87\% [13] of the total \( \alpha \)-gliadin genes and were similar in structure to the full-ORF genes. The generation of stop codons in the coding

---

Fig. 3 – Phylogenic tree based on the deduced amino acid sequences of 22 cloned \( \alpha \)-gliadin genes (without signal peptides) in this study and 95 other genes from \( T. \) monococcum, \( A e. \) speltoides and \( A e. \) tauschii. The representative genome (to left the vertical line), protein ID and species (to right the vertical line) are indicated.
sequences resulted mainly from single-base transitions, with the C to T change predominating and accounting for about 70% of these [8,13]. Additionally, and consistent with recent studies [29,30] of other wheat genomic regions, it has been shown that α-gliadin genes in the Gli-2 regions are not evenly distributed, but are clustered mainly into numerous small gene islands separated by large blocks of repetitive elements, especially retrotransposons, which are abundant (accounting for about 70% of the sequences) in these regions [7]. Thus, it has been suggested that retrotransposons contribute to the dynamic changes in these regions, including frequent gene duplications and insertions, as well as illegitimate recombination, which appears to have a major impact on increasing the number of genes [7,29,30].

The extremely high copy numbers of α-gliadin not only make it more difficult to purify a single component from a compound of related proteins, but make it more complicated to elucidate the expression and function of individual genes [31]. Heterologous expression has frequently been used to produce single pure components for studying structure–function relationships of proteins in vitro. However, heterologous expression of a protein with stable disulfide bonds in E. coli inevitably results in the formation of an inclusion-body protein, and the protein yield depends largely on the type of expressed gene. So the high-level expression of α-gliadins in vitro is still difficult [32,33], meaning that the study of structure–function relationships of single α-gliadin genes by heterologous expression, purification, and functional analysis in vitro is very limited [10].

In the present study, using a pair of degenerate primers that represent the majority of full-ORF α-gliadin genes in GenBank, 43 unique clones from Zhengmai 004 were obtained.
by comparative analysis among a total of 85 positive clones. NCBI BLAST searching of each sequence showed that 42 of them had 84%–99% identity with sequences in GenBank (except for Z4A-22 with 100% identity with JX828270, which we had previously cloned from common wheat cultivar Zhengmai 9023), suggesting that they are new members of α-gliadin gene family. In addition, consistent with previous findings, about 49% of the clones are pseudogenes, 81% of which resulted from single-base transitions, especially the C to T change that accounted for 91% of these. Of the 22 full-ORF genes, one (Z4A-15) lacked the second conserved cysteine residue in the unique domain I, while four genes (Z4A-7, Z4A-14, Z4A-17 and Z4A-20) contained an extra cysteine residue in the C-terminal unique domain II. Given the suggestion that an odd number of cysteine residues promote participation in the disulfide cross-linked gluten matrix and produce a positive effect on flour quality [33], these findings strongly suggest a close association with the high quality of Zhengmai 004. Unfortunately, SDS-PAGE and Western blotting detection of the induced α-gliadin fusion proteins expressed in E. coli confirmed that the high-level expression of α-gliadin in vitro was still difficult, although the T7 promoter induced by IPTG was a suitable promoter for inducing the expression of α-gliadin genes in E. coli. Consequently, such potential contributions to gluten quality were not successfully identified by functional analysis in vitro.

Fortunately, the functionality of a protein is determined largely by its three-dimensional structure, produced by folding secondary structures into one or several domains. Knowledge of the secondary structure of a protein may provide clues to its molecular function [34]. Generally, X-ray crystallography and nucleic magnetic resonance spectroscopy (NMR) are the two major experimental methods to determine protein structures accurately, but owing to their complexity, high cost, and time-consuming nature, progress on protein structure determination can be slow. As a result, over the last few years, computer-based automatic methods including GOR, PSIPRED, YASPIN and HNN have been developed for the rapid prediction, evaluation, and visualization of protein structures [34,35]. Of the most frequently used online software, PSIPRED is the most popular program and has several advantages over other programs including higher prediction accuracy, graphical and colored output of results, description of the confidence score values of each secondary structure element, and the facility to download results in PDF format [34,36]. However, at present, the prediction of the secondary structures of α-gliadins is still very limited. Using PSIPRED version 2.6, Xie et al. [23] predicted the secondary structures of 19 full-ORF α-gliadins that they isolated from common wheat cultivars and Aegilops tauchii accessions and found that the numbers of α-helices and β-strands were not evenly distributed in the different proteins: a high content of β-strands and most of the α-helices and β-strands were found in the two unique domains, and in particular, few or even no secondary structures were distributed in the N-terminal repetitive domain and glutamine repeat I. They accordingly inferred the C-terminal unique domain II to be the most important domain for the formation of intermolecular disulfide bonds with HMW and LMW glutenins.

To ensure the accuracy and comparability of the results, the secondary structure of a total of 198 deduced typical α-gliadins, including the 22 genes cloned in this study, as well as the abovementioned 19 full-ORF genes, were predicted in the present study. Consistent with the previous study mentioned above [23], our results also showed that numbers of α-helices...
and β-strands, as well as the amino acid residues involved in each conserved α-helix and β-strand, were always variable in different proteins, though their positions and core sequences were relatively conserved. However, our comprehensive predictions were different in some respects from those previously reported [23]. Firstly, our results demonstrated that the content of β-strands in α-gliadins was relatively low and that only 67.68% of α-gliadins contained a β-strand (S) or two β-strands (S, Sβ) in the C-terminal unique domain II; moreover, in general, only 2 to 4 amino acid residues were involved in each β-strand. Secondly, our comparative analysis revealed that more α-helices usually occurred in the unique domain I (Hα, Hα II, Hα III and Hα IV) rather than the C-terminal unique domain II (Hβ, Hβ II). Finally, though our results also indicated that the secondary structure was seldom present in the N-terminal repetitive domain, a conserved α-helix or even two α-helices were invariably present in the glutamine repeat I (H1) in all 198 predicted genes. Because the older version was not available, to our knowledge, the only explanation for these discrepancies appears to be the difference in PSIPRED versions used in the respective studies.

Generally, it has been suggested that, for the α-gliadins, a long repetitive domain, a high proportion of glutamine residues and an extra cysteine residue in the primary structure, and more α-helices and β-strands in the secondary structure, exert a positive effect on gluten quality [37–40]. Our results also support this view, not only for the above-mentioned three genes (protein IDs ABQ96115, ABQ96118 and ABQ96119) that harbor an extreme-long glutamine repeat I and could form one or even two significant longer α-helices H1 in this region, but also for some of the genes with an extra cysteine residue in the C-terminal unique domain II, which also probably formed an extra α-helix Hα II or β-strand Sβ involving the peptides precisely around the sites where an extra cysteine residue most likely occurred. Accordingly, on the basis of our comprehensive prediction, we propose that the two unique domains were the most important regions for the function of α-gliadins, whereas in some cases the glutamine repeats would also contribute. In addition, the marked influence on gluten quality of protein subunit ACX71610 identified in vitro and the marked similarity of Z4A-14 to ACX71610 in primary and secondary structure strongly suggest that Z4A-14 is closely associated with the high quality of common wheat cultivar Zhengmai 004.

4.2. The four major T-cell immunogenic peptides and their role in the determination of chromosomal location of α-gliadins and wheat quality improvement

The marked genomic differences in the occurrence of the four major T-cell immunogenic peptides and the average lengths of the two polyglutamine domains, combined with the complete amino acid sequences, make the reliable determination of chromosomal location of the α-gliadin genes feasible [23]. However, distinct genomic differences in the distributions of toxic epitopes also mean that none of the common wheat cultivars is completely safe or non-toxic for CD patients. In the present study, the number of the four T-cell immunogenic peptides and glutenamine residues occurring in the two polyglutamine domains of the 22 cloned genes were analyzed, along with their similarity to the other 95 genes originating in the three diploid species representative of the A and D genomes or the putative ancestral B genome of common wheat. In agreement with previous findings [13,15,21,23], our study confirmed that the set of epitopes, as well as the clusters formed in the phylogenetic tree, were indeed distinct for each genome. Thus, according to the distinct genomic characteristics, 6, 8 and 6 genes were assigned respectively to chromosomes 6A, 6B and 6D, and a total of 16, 0 and 23 epitopes (including a highly immunogenic 33-mer peptide present in Z4A-5) were detected. Alpha-gliadins from the A and especially the D genomes are more deleterious for CD patients, and Zhengmai 004 had the potential to cause the development of CD. However, everything has advantages and disadvantages: a study using Chinese Spring Glu-Z2 deletion lines showed that removing the α-gliadin locus from the short arm of chromosome 6D resulted in a distinct loss of technological properties, although the T-cell immunogenic epitopes decreased [41]. We also found that four of the five genes in this study that have an odd number of cysteine residues, as well as the majority of the genes in GenBank that share this characteristic, were assigned to chromosome 6D on the basis of the occurrence of the epitopes and fell into a cluster in the phylogenetic tree (data not shown). Thus, just as it has been demonstrated that the D genome contributes to many characteristics (including the effects on baking quality of HMW-GS on chromosome 1D) of common wheat [13], the Glu-Z2 locus on chromosome 6D also appears to make specific contributions to baking quality, most likely increasing loaf volume, in addition to being mainly responsible for most of the T-cell stimulatory peptides in α-gliadins.

Fortunately, however, there is evidence [42] in the literature that the amount of gluten exposure has a marked influence on the likelihood of CD development: the higher the exposure to the complex of immunogenic peptides, the higher the incidence of CD. Theoretical comparative analysis also supports this opinion [13,17]. A diet based on wheat cultivars low in T-cell stimulatory sequences may thus have high potential for CD prevention. Furthermore, given the heterogeneity of T-cell epitopes in gluten, it is possible to generate wheat varieties with few or even no toxic peptides via conventional breeding strategies [15,17]. In the phylogenetic tree we constructed, 11 exceptional α-gliadin genes originating from T. monococcum and Ae. tauschii encode few or even none immunogenic T-cell peptides. These findings further confirmed that the wild genetic resources of T. monococcum and Ae. tauschii, especially the latter, are valuable sources for wheat improvement [43,44], not only for improving the rheological properties of gluten but for decreasing the toxicity to CD patients. They also suggest identifying or generating common wheat cultivars that lack or are low in peptides harmful to CD patients, by screening primitive wheat species followed by breeding and directional selection based on the absence of specific gluten peptides.

5. Conclusions

The α-gliadins in the bread wheat cultivar Zhengmai 004 may be strongly associated with its property of weak gluten, given that important variants not only occurred in the primary structures, but were detected in their secondary structures. However, unfortunately, its full potential to cause the
development of CD was also identified. We have presented diagrams summarizing the secondary structure of typical α-gliadins, based on the comparative analysis of these structures in 198 α-gliadins, that should provide insight into structure-function relationships of the α-gliadins. Finally, considering that the α-gliadins on chromosome 6D were the most deleterious for CD patients and most closely associated with gluten quality, and further considering the identification of several distinct α-gliadins derived from Ae. tauschii lacking the four major T-cell peptides, we have confirmed the possibility and importance of screening or even producing wheat cultivars safe for CD patients.

Acknowledgments

We thank Daniel Buchan of the PSIPRED team for his prompt and detailed replies to our queries about PSIPRED. We are grateful to Professor Junmei Li of the English Department of Henan University for the language improvement. This study was supported by the National Natural Science Foundation of China (31271713) and the “Twelfth Five-Year-Plan” in National Science and Technology for Rural Development in China (2011BAD07B01 and 2012AA101105).

REFERENCES


