Rothamsted Repository Download

A - Papers appearing in refereed journals

Shewry, P. R. 2019. What is gluten- why is it special? *Frontiers in Nutrition.*

The publisher's version can be accessed at:

- https://dx.doi.org/10.3389/fnut.2019.00101

The output can be accessed at: https://repository.rothamsted.ac.uk/item/8ww33.

© 20 June 2019, Please contact library@rothamsted.ac.uk for copyright queries.

What is gluten- why is it special?

Peter R. Shewry[1*]

[1]Rothamsted Research (BBSRC), United Kingdom

Conflict of interest statement

Author contribution statement

Keywords

Abstract

Word count: 196

Wheat gluten has an immense impact on human nutrition as it largely determines the processing properties of wheat flour, and in particular the ability to make leavened breads, other baked products, pasta and noodles. However, there has been increasing interest in wheat gluten over the past two decades because of its well-established role in triggering coeliac disease, and its perceived role in other adverse reactions to wheat. The literature on wheat gluten is vast and extends back over two centuries, with most studies focusing on the structures of gluten proteins and their role in determining the functional properties of wheat flour and dough. This article provides a concise account of wheat gluten, focusing on properties and features which are relevant to its role in triggering coeliac disease and, to a lesser extent, other gluten-related disorders. It includes descriptions of the biological role of the gluten proteins, the structures and relationships of gluten protein families, and the presence of related types of protein which may also contribute to functional properties and impacts on health. It therefore provides an understanding of the gluten protein system at the level required by those focusing on its impact on human health

Contribution to the field

There has been increasing interest in wheat gluten over the past two decades because of its well-established role in triggering coeliac disease, and its perceived role in other adverse reactions to wheat. The literature on wheat gluten is vast and extends back over two centuries, making it difficult for those interested in gluten-related effects on health to gain a broad interest of the structures and relationships of gluten and associated proteins. This article presents a concise account of wheat gluten, focusing on properties and features which are relevant to its role in triggering coeliac disease and, to a lesser extent, other gluten-related disorders. It includes descriptions of the biological role of the gluten proteins, the structures and relationships of gluten protein families, and the presence of related types of protein which may also contribute to functional properties and impacts on health. It therefore provides an understanding of the gluten protein system at the level required by those focusing on its impact on human health

Funding statement

Data availability statement

Generated Statement: All datasets generated for this study are included in the manuscript and the supplementary files.

What is gluten- why is it special?

2 **Peter Shewry**

3 Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, UK

4

5 Correspondence:
6 Peter Shewry peter.shewry@rothamsted.ac.uk

7

8 **Abstract**

9 Wheat gluten has an immense impact on human nutrition as it largely determines the processing
10 properties of wheat flour, and in particular the ability to make leavened breads, other baked
11 products, pasta and noodles. However, there has been increasing interest in wheat gluten over
12 the past two decades because of its well-established role in triggering coeliac disease, and its
13 perceived role in other adverse reactions to wheat. The literature on wheat gluten is vast and
14 extends back over two centuries, with most studies focusing on the structures of gluten proteins
15 and their role in determining the functional properties of wheat flour and dough. This article
16 provides a concise account of wheat gluten, focusing on properties and features which are
17 relevant to its role in triggering coeliac disease and, to a lesser extent, other gluten-related
18 disorders. It includes descriptions of the biological role of the gluten proteins, the structures
19 and relationships of gluten protein families, and the presence of related types of protein which
20 may also contribute to functional properties and impacts on health. It therefore provides an
21 understanding of the gluten protein system at the level required by those focusing on its impact
22 on human health

23 **Key words**

24 Wheat, gluten, coeliac disease, protein, prolamin, gliadin, gluten, ATI

25 **Running title**

26 What is gluten?

27

28 **INTRODUCTION**

29 Wheat gluten was one of the earliest proteins to be studied scientifically, by Jacopo Beccari
30 (Professor of Chemistry at the University of Bologna) in his article "De Frumento"
31 (Concerning Grain) in 1745 (Beccari, 1745; Bailey, 1941). It has since been studied in great
32 detail by cereal chemists, because of its role in underpinning the ability to make leavened bread,
33 other baked goods, pasta and noodles. These properties are only shared to a very limited extent
34 by related cereals (barley and rye). Hence, gluten underpins the production of staple foods for
35 a substantial proportion of the global population, particularly in temperate zones.

36 Although gluten was identified as the trigger for coeliac disease almost 70 years ago (Dicke,
37 1950), interest in gluten outside the scientific community was limited to those unfortunate
38 enough to suffer from coeliac disease until early in the present century, which has seen an
39 explosion of interest, particularly in the popular press and social media. As an example, a
40 "Google" search carried out in December 2018 gave almost 400 million hits in less than a
41 minute. This interest relates, of course, to the proposed role of gluten in triggering a range of
42 adverse reactions, with substantial proportions of the population in many countries choosing to

43 adopt a gluten-free, or low-gluten, diet. However, despite this massive interest few people have
44 a clear understanding of gluten itself: what is it, what is the origin, why is it special?

45 This article, which forms part of the Special Research Topic "Gluten, from Plant to Plate:
46 Implications for People with Celiac Disease", therefore provides a broad account of wheat
47 gluten including its synthesis and deposition in the developing grain, the structures and
48 evolutionary relationships of its component proteins, and its unique properties which are
49 exploited in grain processing, focusing on features which are relevant to its role in triggering
50 coeliac disease. It does not cover other impacts of wheat proteins on human health, notably
51 allergy and non-coeliac gluten sensitivity (NCGS) which are discussed in other recent review
52 articles (Catassi et al., 2017; Juhasz et al., 2018).

53 **WHAT IS GLUTEN?**

54 **Gluten is defined based on its origin and solubility**

55 Gluten is classically defined as the largely proteinaceous mass which remains when a dough
56 made from wheat flour and water is gently washed in an excess of water or dilute salt solution
57 to remove most of the starch and soluble material (Wrigley and Bietz, 1988). The remaining
58 material, which has been described as "rubbery", comprises about 75-80% protein on a dry
59 matter basis, depending on how well the material is washed. Hence "gluten proteins" are
60 defined as those present in this mass and, because similar material cannot be isolated from
61 doughs made with flours from other cereals, gluten proteins are restricted to the grain of wheat
62 (species of the genus *Triticum*). However, related proteins are present in other cereals (as
63 discussed below) and these are frequently referred to as gluten in the non-specialist literature
64 and the wider popular media.

65 More correctly, gluten and related proteins from other cereals are classified as "prolamins".
66 This name was coined by T.B. Osborne, the father of plant protein chemistry who worked at
67 the Connecticut Agricultural experiment station from 1886 till 1928. During this period he
68 published some 250 papers, including studies of seed proteins from 32 species. This allowed
69 him to develop a broad classification of proteins based on their extraction in a series of solvents
70 (Osborne, 1924). This extraction is often performed sequentially (and called "Osborne
71 fractionation") with the four Osborne fractions being called albumins (soluble in water),
72 globulins (soluble in dilute saline), prolamins (soluble in 60-70% alcohol) and glutelins
73 (insoluble in the other solvents but may be extracted in alkali). The first two fractions are
74 readily distinguished and the names are still in use, while prolamins were recognised as a
75 defined group present only in cereal grains with the name being based on their high contents
76 of proline and amide nitrogen (now known to be derived from glutamine). This fraction is given
77 specific names in different cereal species: gliadin in wheat, hordein in barley, secalin in rye,
78 zein in maize etc.

79 However, the final fraction (glutelin) is more difficult to define, as it effectively comprises all
80 proteins which are insoluble in the three previous solvents but can be solubilised under
81 conditions of extreme pH. In fact, glutelins are now known to comprise a mixture of unrelated
82 proteins, including insoluble structural and metabolic proteins such as those bound to
83 membranes and cell walls. However, these proteins are only present in small amounts and in
84 wheat (and most other cereals) the major glutelin components are in fact prolamin subunits
85 which are not extractable with alcohol/water mixtures due their presence as high molecular
86 mass polymers stabilised by inter-chain disulphide bonds. In wheat these proteins are called
87 glutenin and are present in about equal amounts to the alcohol-soluble gliadins, the two groups
88 comprising gluten.

Gluten proteins are the major storage protein fraction

Gluten proteins are the major group of proteins which are stored in the grain to support germination and seedling development. They are restricted in distribution to the starchy endosperm cells of the grain, and have not been detected in any other tissues of the grain or plant. Their pathway and mechanisms of synthesis and deposition have been studied in detail (see Tosi, 2012) but two points are particularly relevant here. Firstly, they are initially deposited in discrete protein bodies, which fuse during the later stages of grain development to form a continuous matrix surrounding the starch granules (Figure 1 A). This matrix forms a continuous protein network within the cell, which can be revealed when the starch is removed from a flour particle by enzyme digestion (Figure 1 B). It is easy to envisage how the protein networks present in the individual cells can be brought together during dough mixing to form the continuous gluten network in dough.

The second important point is that gluten proteins are not uniformly distributed in the starchy endosperm cells, but enriched in the outer 2 to 3 layers of cells (which are called the sub-aleurone cells). This is illustrated in Figure 1C, which shows a section of the starchy endosperm cells and outer layers from the lobe of the grain at a late stage of development stained with toluidine blue to show protein. In fact, Kent (1966) calculated that the protein content of the cells of the starchy endosperm varies by over 4-fold, from 45% in the sub-aleurone cells to 8% in the central region. Furthermore, the gluten protein composition also varies, with the percentage of high molecular weight glutenin subunits (HMW subunits) increasing and the proportion of low molecular weight (LMW) subunits and gliadins (except for ⍵-gliadins) decreasing (these protein types are discussed below) (He et al., 2013). These gradients in composition are reflected to some extent in the contents and compositions of gluten proteins in the flour streams produced by commercial roller milling, meaning that these fractions may also vary in their impact on health (Tosi et al., 2018).

Implications for coeliac disease

Fractionation by conventional milling combined with pearling (abrasion) or peeling (friction) could lead to flour streams that are enriched or depleted in coeliac-active proteins. The use of vital gluten (which is produced commercially for fortification of food products) also has implications. This will contain all of the gluten proteins present in the flour of origin, but may also contain other biologically active proteins as "co-passengers".

GLUTEN PROTEINS

Gluten comprises several related families of proteins encoded by multigene families

The gluten protein fraction comprises a complex mixture of components which can be separated into groups by electrophoresis. Electrophoresis of the gliadins at low pH separates four groups of bands, called (in terms of decreasing mobility) α-gliadins, β-gliadins, ɣ-gliadins and ⍵-gliadins. However, comparisons of amino acid sequences show that the α- and β-gliadins form a single group, sometimes called α-type gliadins.

The glutenin polymers are too big to be separated by conventional electrophoresis, but reduction of the inter-chain disulphide bonds that stabilise the polymers allows the subunits to be separated by sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE) into two groups of bands, called the high molecular weight (HMW) and low molecular weight (LMW) subunits. The latter group can be further sub-divided into a major group of components (B-type LMW subunits) and two minor groups (C-type and D-type).

Comparisons of amino acid sequences of these groups of gluten protein components clarifies their relationships, showing that whereas the HMW subunits and ꞷ-gliadins form discrete groups, with the α-gliadins, ɣ-gliadins and B-type LMW subunits forming a third group. The minor groups of C-type and D-type LMW subunits appear to be modified forms of gliadins in which mutations to form cysteine residues allow their incorporation into glutenin polymers, with the C-type LMW subunits being modified α-gliadins or ɣ-gliadins and the D-type modified ꞷ-gliadins. This classification is summarised in Table 1, which also shows their relative amounts and summarises their characteristics (molecular masses and partial amino acid compositions).

Table 1 also groups the types of gluten proteins discussed above into three "families" (the HMW, sulphur(S)-rich and S-poor prolamins), which were defined about 30 years ago based on emerging sequence data (Shewry et al, 1986). This classification remains valid despite the vast increase in our knowledge of gluten protein sequences over the past few decades. For example, in May 2015 Bromilow et al. (2017a) retrieved over 24,000 sequences related to gluten proteins from the UniProt database. Removal of redundant, partial and mis-assigned sequences allowed the assembly of a curated database of 630 sequences.

The retrieval of over 600 sequences of gluten proteins does not, of course, mean that individual wheat genotypes contain this number of gluten proteins. Although the precise number of gluten proteins present in mature seed has not been determined, examination of two-dimensional (2D) electrophoretic separations indicates that the number of gluten proteins present in detectable amounts is probably between 50 and 100. This is consistent with the recent study of Bromilow et al. (2017b), who identified 63 gluten proteins in a single cultivar, using mass spectroscopy and a curated sequence database (Bromilow et al., 2017a). However, this study identified eight individual HMW subunit proteins, which is twice the number known to be present in the cultivar studied. This highlights the problems inherent in identifying gluten proteins based on short peptide sequences.

Although the prolamin groups discussed above undoubtedly account for the vast majority of the gluten proteins, recent work has shown that small amounts of a further are present. These have been defined as δ-gliadins, although sequence comparisons indicate that they form part of the wider family of ɣ-prolamins (being closest in sequence to the ɣ3-hordeins of barley) (Anderson et al, 2012, Guo et al, 2018a). Proteomic analysis indicates that they account for 1.2% of the total normalised spot volume in grain of Chinese Spring wheat (Altenbach et al., 2019).

Molecular basis for gluten protein polymorphism

The large numbers of individual gluten proteins present in single genotypes, and the 10-fold greater number of sequences in databases, arises from three factors: the presence of multigene families, the high level of polymorphism between genotypes and, to a more limited extent, post-translational modification. It is therefore necessary to consider these factors in turn.

Common wheat (*Triticum aestivum*), which includes modern bread wheat and spelt, is a hexaploid species, with three genomes (called A, B and D) derived from related wild grasses. Only two of these genomes (A and B) are present in the tetraploid durum (pasta) wheat and emmer (forms of *Triticum turgidum*) while einkorn (*Triticum monococcum*) is diploid with only the A genome. Gluten proteins are encoded by loci on the group 1 and group 6 chromosomes of all three genomes, meaning that the gluten fraction can be expected to comprise more individual protein components in common wheat than in the other. A detailed discussion of the genetics of gluten proteins is outside the scope of this article, but the reader can refer to Shewry et al (2003) for a detailed account.

180
181 Furthermore, all of the gluten protein loci comprise multiple genes. The simplest loci are the
182 *Glu-1* loci which are located on the long arms of the group 1 chromosomes. Each of these loci
183 comprises two genes which encode two types of HMW subunit of glutenin (called x-type and
184 y-type). However, because not all of the *Glu-1* genes are expressed in all genotypes, the number
185 of HMW subunit proteins in cultivars of bread wheat vary from 3 to 5 (Payne, 1987). Because
186 of the simple genetic system, and the fact that the HMW subunits have been studied in more
187 detail than most groups of gluten proteins, it is possible to define alleles at all three loci. Thus,
188 the widely occurring pairs of subunits called 1Dx2+1Dy12 and 1Dx5+1Dy10 are alleles, while
189 the pairs of subunits called 1Dx2+1Dy12 and 1Bx7+1By9 are homeoalleles (alleles on
190 different genomes). The greater complexity of other gluten protein loci makes it much more
191 difficult to recognise allelic forms of genes and proteins, although detailed analyses of allelic
192 variation in LMW subunits have been reported (reviewed by Juhász and Gianibelli, 2006).
193 However, whereas the individual HMW subunits can be assigned to sequenced genes, this is
194 very difficult, if not impossible, for many other gluten proteins because of the complexity of
195 the loci. For example, Huo et al (2018b) assembled sequences of the α-gliadin loci on the three
196 genomes of bread wheat, showing a total of 47 genes of which 26 encoded intact full-length
197 protein products. Similarly, Qi et al (2009) reported the sequences of 29 putatively functional
198 γ-gliadin genes (encoded by genes at the *Gli-1 loci* on the short arms of the group 1
199 chromosomes) in a single cultivar. Further information on the structures of the gluten protein
200 multigenic loci are being provided by genome analysis (see, for example, Huo et al., 2018a,
201 Juhasz et al., 2018, Clavijo et al., 2019).
202 It is also likely that the numbers of expressed genes vary between genotypes. Thus, the high
203 polymorphism in gluten protein composition observed between genotypes may arise both from
204 variation in the numbers of expressed genes, and variation in the sequences of the encoded
205 proteins.
206 A third factor which may contribute to protein polymorphism is post-translational
207 modification. Gluten proteins contain between about 20 mol % and 50 mol % of glutamine
208 residues so post-translational deamidation has long been recognised as a possibility. It may, for
209 example, account for the fact that HMW subunits often form "trains" of spots in 2D
210 electrophoresis, while Dupont et al (2011) reported the presence of HMW subunit sequences
211 in 43 spots separated on 2D gels. However, the extent of deamidation has never been quantified.
212 Other proposed modifications, such as glycosylation (Tilley et al, 1993) and phosphorylation
213 (Tilley and Schofield, 1995) have not been substantiated by further studies. Other types of post-
214 translational modification may include cyclisation of N-terminal glutamine to give
215 pyroglutamate (which is likely to be responsible for many gluten proteins having "blocked" N-
216 termini), differential processing of the signal peptide (Masci et al., 1998) and proteolysis by
217 legumain-like asparaginyl endoproteinase (DuPont et al, 2004).
218 Finally, the proportions of gluten proteins may also be affected by the environment, including
219 temperature during grain development and availability of nutrients (nitrogen and sulphur)
220 (reviewed by Dupont and Altenbach, 2003; Altenbach, 2012). In particular, increases in the
221 proportions of gliadins occur under high nitrogen availability and of ⍵-gliadins when nitrogen
222 availability is high but sulphur is limiting.
223
224 *Implications for coeliac disease*
225 Protein polymorphism is clearly a challenge for attempts to eliminate "toxic" proteins and to
226 develop coeliac-safe wheats, whether by exploiting natural variation or by genetic
227 engineering/genome editing.
228 Effects of environment on gluten protein composition will also have impacts on the abundances
229 of specific coeliac disease epitopes.

Gluten proteins contain unique repetitive domains

230   The most important characteristic of wheat gluten proteins in relation to their role in coeliac
231   disease is the presence of protein domains comprising repetitive sequences. The domains vary
232   in extent, but generally account for between about 30% and 50% of the protein sequence in S-
233   rich gliadins and LMW subunits, between 75% and 85% in HMW subunits, and almost the
234   whole protein in ⍵-gliadins (reviewed by Shewry et al., 2009). They comprise tandem repeats
235   of short peptides comprising between three and nine amino acid residues, and may be based on
236   tandem repeats of one motif or tandem and interspersed repeats of two or more motifs.

237   The most widely studied repetitive sequences are those present in the HMW subunits of
238   glutenin. These comprise repeats based on three motifs: the hexapeptide PGQGQQ, the
239   nonapeptide GYYPTSPQQ or GYYPTSLQQ and, in x-type subunits only, a tripeptide GQQ
240   (P, proline; G, glycine; Q, glutamine, Y, tyrosine; P. proline; T, threonine, S, serine; L, leucine)
241   (Shewry et al., 2009). The motifs present in the other groups of gluten proteins are generally
242   less well conserved and the identification of consensus motifs is more subjective than in the
243   HMW subunits, but all are rich in proline and glutamine, for example, PQQPFPQQ (F, phenyl
244   alanine) in ɣ-gliadins. It should be noted that these sequences are responsible for the
245   characteristic amino acid compositions of the whole proteins, notably the high contents of
246   glutamine (35-55 mol%) and proline (10-25 mol%) in all groups of prolamins, high glycine in
247   HMW subunits (11-12 mol%) and high phenyl alanine (about 11 mol%) in ⍵-gliadins
248   (reviewed by Shewry et al., 2009).

249   The repeated sequences may also be responsible for the unusual solubility properties of gluten
250   proteins. Although glutamine is a hydrophilic amino acid, the regularly repeated glutamine
251   residues in gluten proteins are considered to form protein:protein hydrogen bonds resulting in
252   insolubility in water (as discussed by Belton (1999) for HMW subunits). However, in most
253   gluten proteins, all of the cysteine residues, which may form interchain or intrachain disulphide
254   bonds, are located in the non-repetitive domains.

255   The repetitive sequences also play a crucial role in triggering coeliac disease. In fact, all of the
256   31 "coeliac disease relevant T-cell epitopes" listed by Sollid et al (2012) are present in the
257   repetitive domains of wheat or related cereals (barley, oats, rye) and all groups of gluten
258   proteins (gliadins and glutenins) contain epitopes. Nevertheless, some individual proteins
259   within these groups may lack recognised coeliac epitopes (although the current list of epitopes
260   is considered to be incomplete). This is illustrated by Figure 2 (Shewry and Tatham, 2016) and
261   discussed in detail by Gilissen et al (2014), Shewry and Tatham (2016) and Juhasz et al. (2018).

263   *Implications for coeliac disease*

264   As discussed above, all of the coeliac-toxic epitopes in wheat gluten proteins are present in the
265   repeated sequences, with multiple epitopes present in some repetitive domains. This clearly
266   poses a significant challenge for attempts to "remove" epitopes by transgenesis or gene editing.

267   **THE PROLAMIN SUPERFAMILY**

268   The prolamins, including wheat gluten proteins, were historically defined as a unique class of
269   proteins restricted to the grain of cereals and related grass species, based on their unusual amino
270   acid compositions and solubility properties (Osborne, 1924) and this dogma was not questioned
271   until the increasing availability of protein sequence data allowed wider comparisons to be
272   made. The first report that prolamins were related to a wider range of proteins was in 1985,
273   when Kreis et al (1985) showed the sequences present in the cysteine-rich non-repetitive
274   regions of prolamins were related to sequences in two other groups of seed proteins: cereal

275 inhibitors of α-amylase and trypsin (now called ATIs) and 2S albumin storage proteins of
276 dicotyledonous seeds. Although these groups of proteins have little sequence identity with each
277 other or with prolamins, the homology was based on very high conservation in the numbers
278 and spacing of cysteine residues. Further comparisons exploiting the vast increase in sequence
279 data have since identified several other groups of related proteins, which are together referred
280 to as the "prolamin superfamily".

281 The prolamin superfamily includes proteins which are not restricted to cereals and grasses, and
282 present in tissues other than seeds (Shewry et al, 2004). However, several types are present in
283 wheat grain, and may contribute to the functional properties and role in diet and health (Shewry
284 et al., 2009). They are therefore briefly discussed here and summarised in Table 2.

285 **Farinins and purinins**

286 It has been known for many years that wheat flour contains proteins with molecular masses
287 below 30 kDa which are related to gluten proteins, including types described as globulins, low
288 molecular weight gliadins and avenin-like proteins. Kasarda et al (2013) have recently
289 discussed the relationships of these proteins and suggested that they should be classified into
290 two types, which they termed farinins and purinins. Both are more closely related to gliadins
291 than the other protein types discussed below, but lack the repeated sequences which are typical
292 of gliadins. Hence they have been classed as globulins based on solubility. The farinins
293 correspond to the avenin-like proteins (defined based on homology with the avenin proteins of
294 oats) with two types called a (which correspond to LMW gliadins) and b (Kan et al, 2006).
295 These groups differ in that the b-type proteins contain a duplicated sequence of about 120
296 residues, resulting in a higher molecular weight (about 30kDa compared with 17kDa). The b-
297 type proteins are associated with the surface of the starch granule and are post-translationally
298 cleaved to give two subunits (11kDa and 19kDa) linked by a single disulphide bond (Kasarda
299 et al, 2006). Ma et al (2013) showed that over-expression of a transgene encoding a b-type
300 protein resulted in improved flour mixing properties and an increased proportion of large
301 glutenin polymers, presumably due to their ability to form inter-chain disulphide bonds.
302 The low molecular weight gliadins/purinins have masses of about 17kDa to 19kDa (Salcedo et
303 al., 1977) and are more closely related to the ɣ-gliadins in sequence (Clarke et al., 2003;
304 Kasarda et al., 2013). They may, perhaps, be considered to be similar to the "ancestral"
305 prolamin proteins, before they diverged due to development and amplification of the repetitive
306 sequence domains. Mixing of heterologously expressed proteins into dough showed similar
307 effects to the incorporation of gliadins (Clarke et al., 2003).

308 **Puroindolines (Pins) and grain softness protein (GSP).**

309 Hardness is one of the major characteristics used to divide wheat into end use classes. It is
310 determined by the *Hardness* (*Ha*) locus on the short arm of chromosome 5D of bread wheat,
311 although the name is misleading because the encoded genes actually determine softness. This
312 locus is not present in durum wheat which is therefore ultrahard. The *Ha* locus comprises three
313 genes (Chantrey et al, 2005), encoding proteins called puroindoline a (Pin a), puroindoline b
314 (Pin b) and grain softness protein (GSP). The mature Pin a and Pin b proteins comprise about
315 120 amino acid residues including 10 cysteine residues which form inter-chain disulphide
316 bonds. They also contain five (in Pin a) or three (in Pin b) tryptophan residues which are
317 grouped together in the sequences. Comparison of wholemeal flours of 40 wheat cultivars (19
318 soft and 21 hard) grown on four French sites showed 0.029-0.060 % dry wt of Pin a and 0.004-
319 0.031% dry wt of Pin b (Igrejas et al., 2001). Differences in the expression of these proteins,
320 and/or their amino acid sequences, account for about 75% of the variation in grain hardness in
321 bread wheat (Turner et al., 2004).

The third gene at the *Ha* locus encodes a protein which is cleaved post-translationally, probably in the vacuole by a similar legumain-type asparaginyl endoproteinase to the enzyme(s) responsible for proteolysis of gluten proteins (as discussed above). This releases a 15 residue peptide from the N-terminus (Van de Bulck et al., 2002). This peptide contains three proline residues which are hydroxylated to give hydroxyprolines and then *o*-glycosylated with arabinogalactan chains to give a mass of about 23 kDa (Van der Bulck et al., 2005). The resulting "arabinogalactan peptide" (AGP) accounts for about 0.39% of the dry weight of white flour (Van der Bulck et al., 2005) and is readily fermented by the colonic microflora (Harris et al., 2019). The remaining part of the protein, termed "grain softness protein" (GSP), may contribute to hardness to a limited extent (by about 10 units measured by the Perten Single Kernal Characterisation System (SKCS)) (Wilkinson et al, 2017), but the biological roles of AGP and GSP are not known.

Non-specific lipid transfer proteins (LTPs)

Unlike the other proteins discussed here, LTPs are not restricted to seed tissues, or to cereals and other grass species. Although they were initially defined on their ability to transfer phospholipids between liposomes and membranes *in vitro*, their true physiological role is unknow with one possible function being to contribute to defence to biotic stresses. They occur in two classes, with masses of about 9 kDa (LTP1) and 7 kDa (LTP2) and are concentrated in the aleurone layer and embryo of the wheat grain (reviewed by Marion et al., 2004). Many LTPs have been identified as allergens, in seeds, fruit and pollen (Marion et al, 2004), with LTP1 of wheat contributing to both food allergy and Bakers' asthma (respiratory allergy to wheat flour) (Pastorello et al., 2007; Palacin et al., 2007).

α-Amylase/trypsin inhibitors

Wheat inhibitors of α-amylase and trypsin have been studied for over 40 years, resulting in an extensive and somewhat confusing literature. This results partly from the complexity of the fraction but also from use of different nomenclatures, based on relative electrophoretic mobilities (the major components being called 0.19, 0.28 and 0.53), solubility in chloroform:methanol (called CM1 to CM17) and subunit structure (monomeric, dimeric and tetrameric forms occurring) (Carbonero and Garcia-Olmedo, 1999). Dupont et al. (2011) used mass spectroscopy of proteins separated by 2D electrophoresis to identify two spots corresponding to forms of the putative monomeric trypsin inhibitor(s) CM1/3, two related to the monomeric amylase inhibitor WMAI, two related to the homodimeric amylase inhibitor WDAI1 and nine related to subunits of the heterotetrameric amylase inhibitor WTAI (1x CM1, 2 x CM2, 2x CM3, 2 x CM16 and 2 x CM17). More recently, Geisslitz et al (2018) have used targeted LC-MS to quantify the amounts of the major ATIs (WDAI/0.19+0.53; WMAI1/0.28, CM2, CM3, CM16 and CM17), showing that they together accounted for 3.4-4.1 mg/g in wholemeal flour of bread wheat.

Wheat ATIs are well characterised as wheat allegens, particularly in Bakers' asthma but also on ingestion of food (reviewed by Salcedo et al., 2004). In addition, they have been studied widely over the past few years because of putative roles in other adverse reactions to wheat consumption, including coeliac disease and non-coeliac wheat/gluten sensitivity (as discussed in other contributions to this special section).

ATIs have also been reported to contribute to the cooking quality of pasta, where they were initially reported to be glutenin components (called durum sulphur-rich glutenin, DSG) (Kobrehel and Alary, 1989a ,b; Gautier et al, 1989).

Implications for coeliac disease

8

368 Wheat grain contains many other proteins including other families of protease and amylase
369 inhibitors, thionins, ribosome-inactivating proteins, and putative defence-related proteins with
370 unknown functions (reviewed by Shewry et al., 2009). All of these may be present in food
371 products, present either in flours or as "contaminants" in vital gluten. However, the proteins
372 discussed above share some properties which may be particularly relevant. Firstly, most are
373 small globular proteins which are tightly folded and stabilized by multiple interchain disulphide
374 bonds. Hence, they are particularly stable to heating during food processing and to degradation
375 in the gastro-intestinal tract: although proteolysis may occur, the proteins will not disintegrate
376 because the fragments are held together by the disulphide bonds. Secondly, they may interact
377 strongly with gluten proteins and hence be present in vital gluten. These interactions may be
378 stabilised by non-covalent forces, such as the LMW gliadins/purinins, or by disulphide bonds
379 formed either during grain development and maturation or re-arrangements during processing.
380 Irrespective of the mechanism, the fact that they may be present in "gluten protein" fractions
381 shows that they must be considered when interpreting studies carried out on human responses
382 to wheat proteins.

383 **GLUTEN PROTEINS HAVE UNIQUE BIOPHYSICAL PROPERTIES WHICH**
384 **UNDERPIN GRAIN PROCESSING**

385 Several factors have contributed to the global success of wheat, one being its wide adaptability.
386 However, the main reason why it is grown in preference to other cereal crops in many countries
387 is the functional properties of wheat flour. As discussed above, wheat is the only cereal which
388 can be baked to give leavened bread and other baked products, as well as pasta and noodles.
389 The quality for these end uses is determined largely by the gluten proteins, which form a
390 continuous network in dough. This network provides the cohesiveness required for making
391 products such as pasta as well as the visco-elasticity required for breadmaking.

392 Despite a massive literature the molecular basis for the biophysical properties of gluten is still
393 not completely understood, and it is not possible to provide a detailed discussion here.
394 However, two points are particularly relevant. Firstly, the properties depend on the
395 contributions of both the gliadins and glutenins, with the glutenin subunits forming large three
396 dimensional networks stabilised by inter-chain disulphide bonds which interact with gliadins,
397 and with other glutenin networks, by non-covalent forces, particularly hydrogen bonds.
398 Secondly, the polymers are stabilised by a combination of forces. The importance of disulphide
399 bonds is readily demonstrated as these can be disrupted using reducing agents, with
400 catastrophic effects on functionality. The importance of hydrogen bonds is less easy to
401 demonstrate, but Belton (1999) has proposed that hydrogen bonds are particularly important in
402 developing optimal protein interactions during dough mixing.

403 *Implications for coeliac disease*

404 The clearest implication for coeliac disease is that any drastic modification to the composition
405 of the gluten protein fraction and/or to the sequences of the individual subunits are likely to
406 have effects on functionality. Although these effects are not easy to predict, that fact that bread
407 making wheats have been selected for functional properties for almost a century suggests that
408 most modifications will be detrimental. Thus, although it may be possible to produce
409 "acceptable" loaves from modified lines of wheat in the laboratory and in small scale systems
410 (see, for example, Gil-Humanes et al., 2014a, b) , this is a much greater challenge for large
411 scale commercial production where profit margins are narrow and small differences in
412 parameters such as loaf height, crumb texture, colour and shelf life will affect the quality of the
413 product and hence acceptability by consumers.

414 **CONCLUSION**

Wheat gluten fulfils an essential biological role as the major grain storage protein fraction, and is the major determinant of the functional (processing) properties of the grain. It is a highly complex mixture of proteins, encoded by multigene families at multiple loci on the three genomes of bread wheat, with a high degree of polymorphism between genotypes. The individual proteins also have unusual structures, including extensive domains of repetitive sequences. In addition, a range of related proteins are present in the grain and may be present in isolated gluten fractions. All of these factors must be considered when studying the role of gluten in coeliac disease and other adverse responses to wheat consumption, and in designing strategies to develop safe types of wheat and wheat products.

FUNDING

ACKNOWLEDGEMENT

REFERENCES

Altenbach SB. New insights into the effects of high temperature, drought and post-anthesis fertiliser on wheat grain development. J Cereal Sci (2012) 56: 39-50.

Altenbach SB, Chang H-C, Simon-Buss A, Mohr T, Hou N, Gu YQ. Exploiting the reference genome sequence of hexaploid wheat; a proteomic study of flour proteins from the cultivar Chinese Spring. Functional and Integrative Genomics (2019) doi:10.1007/s10142-019-00694-z

Amend T, Beauvais F. (1995) Der mechanismus der Teigbildung: Vorstoß in den molekularen Strukturbereich. Getreide Mehl Und Brot 49:359-362.

Anderson D, Dong L, Huo N, Gu, YQ. A new class of wheat gliadin genes and proteins. PLOS ONE (2012) 7:12 e52139

Bailey CH. A translation of Beccari's "Concerning grain" 1728. Cereal Chem. (1941) 18:555-561.

Beccari JB. De Frumento. De Bononiensi Scientifiarurm et Atrium atque Academia Commentarii Tomi Secundi, Bononia (1745).

Belton PS. On the elasticity of wheat gluten. J. Cereal Sci. (1999) 29:103-107.

Bromilow S, Gethings LA, Buckley M, Bromley M, Shewry PR, Langridge JI, Mills ENC. A curated gluten protein sequence database to support development of proteomics methods for determination of gluten in gluten-free foods. J Proteomics (2017a) 163:67-75

Bromilow SNL, Gethings LA, Langridge JI, Shewry PR, Buckley M, Bromley ML, Mills ENC. Comprehensive proteomic profiling of wheat gluten using a combination of data-independent and data-dependent acquisition. Frontiers in Plant Science, (2017b) 7, 2020. 10.3389/fpls.2016.02020

Carbonero P, García-Olmedo F "A multigene family of trypsin/α-amylase inhibitors from cereals" in PR Shewry and R Casey, editors. Seed Proteins. Dordrecht Germany: Kluwer Academic Publishers (1999). p 617-633.

Catassi C, Alaedini A, Bojarski C, Bonaz B, Bouma G, Carroccio A et al. The overlapping area of non-celiac gluten sensitivity (NCGS) and wheat-sensitive irritable bowel syndrome (IBS): an update. Nutrients (2017) 9: 1268; doi:10.3390

Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, Dubois I, Dossat C, Sourdille P, Joudrier P, Gautier M.-F, Cattolico L, Beckert M, Aubourg S, Weissenbach J, Caboche, M, Bernard M, Leroy P, Chalhoub B. Molecular basis of evolutionary events that shaped the *Hardness* locus in diploid and polyploidy wheat species (*Triticum* and *Aegilops*). The Plant Cell (2005) 17 :1033-1045.

Clarke BC, Phongkham T, Gianibelli MC, Beasley H, Békés F. The characterisation and mapping of a family of LMW-gliadin genes: effects on dough properties and bread volume. Theor Appl Genet (2003) 106:629-635.

Clavijo BJ, Venturini L, Schudoma C, Accinelli GG, Kaithakottil G, Wright J et al. An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provided gemonic evidence for chromosomal teanslocations. Genome Research (2019) 27: 885-896.

Dicke W.K. Coeliakie. MD Thesis University of Utrecht (1950).

DuPont FM, Altenbach SB. Molecular and biochemical impacts of environmental factors op wheat grain development and protein synthesis. J Cereal Sci (2002) 38: 133-146.

DuPont FM, Vensel WH, Chan R, Kasarda DD. Similarities of omega gliadins from *Triticum urartu* to those encoded on chromosome 1A of hexaploid wheat and evidence for their post-translational processing. Theor Appl Genet (2004) 108:1299-1308.

Dupont FM, Vensel WH, Tanaka CK, Hurkman WJ, Altenback SB. Deciphering the complexities of the wheat flour proteome using quantitative two-dimensional electrophoresis, three proteases and tandem mass spectrometry. Proteome Science (2011). 9, 10. doi.org/10.1186%2F1477-5956-9-10

Gautier M-F, Alary R, Kobrehel K, Joudrier P. Chloroform methanol-soluble proteins are the main components of *Triticum durum* sulfur-rich glutenin fractions. Cereal Chem (1989) 66:535-535.

Geisslitz S, Ludwig C, Scherf KA, Koehler P. Targeted LC–MS/MS Reveals Similar Contents of α-Amylase/Trypsin-Inhibitors as Putative Triggers of Nonceliac Gluten Sensitivity in All Wheat Species except Einkorn. J Agric Food Chem (2018) 66:12395-12403.

Gil-Humanes J, Piston F, Barro F, Rosell C. The shutdown of celiac disease-related gliadin epitopes in bread wheat provides flours with increased stability and better tolerance to over-mixing) PLOS ONE (2014a) https://doi.org/10.1371/journal.pone.0091931

Gil-Humanes, J, Piston F, Altamirano-Fortoul R, Real A, Comino S, Sousa C et al. Reduced-gliadin wheat bread: an alternative to the gluten-free diet for consumers suffering gluten-related pathologies. PLOS ONE (2014b) https://doi.org/10.1371/journal.pone.0090898

Gilissen LJWJ, van der Meer IM, Smulders MJM. Reducing the incidence of allergy and intolerance to cereals J Cereal Sci (2014) 59:337-353.

Harris S, Powers S, Monteagudo-Mera A, Kosik O, Lovegrove A, Shewry PR, Charalampopoulos, D. Determination of the prebiotic activity of wheat arabinogalactan peptide (AGP) using batch culture fermentation. European Journal of Nutrition (2019) in press.

He J, Penson S, Powers S, Hawes C, Shewry PR, Tosi, P. Spatial patterns of gluten protein and polymer distribution in wheat grain. J Agric Food Chem. (2013) 61:6207-6215.

Huo N, Zhang S., Zhu T., Dong L., Wang Y, Mohr T. et al. Gene duplication and evolution dynamics in the homeologous regions harbouring multiple prolamin and disease resistance gene families in hexaploidy wheat. Front Plant Sci (2018a) https://doi.org/10.3389/fpls.2018.00673

Huo N, Zhu T, Altenbach S, Dong L, Wang Y, Mohr T et al. Dynamic evolution of α-gliadin prolamin gene families in homeologous genomes of hexaploidy wheat. Scientific Reports (2018a) 8:5181.

Igrejas G, Gaborit T, Oury F-X, Chiron H, Marion D, Branlard G. Genetic and environmental effects on puroindoline-a and puroindoline-b content and their relationship to technological properties in French bread wheats. J Cereal Sci (2001) 34:37–47.

Juhász, A, and Gianibelli, M.C. (2006). "Low-Molecular-Weight Glutenin Subunits: Insights into This Abundant Subunit Group Present in Glutenin Polymers" in Wrigley CWF, Békés F, Bushuk W. Gliadin and Glutenin. The Unique Balance of Wheat Quality. St Paul Mn: AACC (2006). p .71-212.

Juhász, A, Belova T, Florides C, Maulis C, Fischer I, Gell G et al. Genome mapping of seed-borne allergens and immunoresponsive proteins in wheat. Science Advances (2018) 4: eaar8602

Kasarda DD, Adalsteins E, Lew EJ-L, Lazo GR, Altenbach B. Farinin: characterisation of a novel wheat endosperm protein belonging to the prolamin superfamily. J Agric Food Chem (2013) 61:2407-2417.

Kan Y, Wan Y, Mitchell RAC, Beaudoin F, Leader DJ, Edwards KJ, Poole R, Wang D, Shewry PR. Transcriptome analysis reveals novel storage protein transcripts in seeds of *Aegilops* and wheat. J Cereal Sci (2006) 44:75-85.

Kent NL Subaleurone endosperm cells of high protein content. Cereal Chem (1966) 43:585-601.

Kobrehel K, Alary R. Isolation and partial characterisation of two low-molecular weight durum-wheat (*Triticum durum*) glutenins. J Sci Food Agric (1989a) 48:441-452.

Kobrehel K, Alary R. The role of a low-molecular weight glutenin fraction in the cooking quality of durum-wheat pasta. J Sci Food Agric (1989b) 47:487-500.

Kreis M, Forde BG, Rahman S, Miflin BJ, Shewry PR. Molecular evolution of the seed storage proteins of barley, rye and wheat. J Mol Biol (1985) 183:499-502

Ma F, Li M, Yu L, Li Y, Liu Y, Li T, Liu W, Wang H, Zheng Q, Li K, Chang J, Yang G, Wang, Y, Guang Y. Transformation of common wheat (*Triticum aestivum* L.) with *avenin-like b* gene improves flour mixing properties. Mol Breed (2013) 32:853-865.

Marion D, Douliez J-P, Gautier M-F, Elmorjani K. "Plant lipid transfer proteins: relationships between allergenicity, and structural, biological and technological properties" in Mills ENC, Shewry PR. Plant Food Allergens, editors. Oxford UK: Blackwell Science (2004). p. 57-86.

Masci S, D'Ovidio R, Lafiandra D, Kasarda D. Characterization of a low-molecular-weight glutenin subunit gene from bread wheat and the corresponding protein that represents a major subunit of the glutenin polymer. Plant Physiol (1998) 118:1147-1158.

Osborne TB. The Vegetable Proteins. 2nd Edition. London UK: Longmans, Green and Co (1924). p.154.

Palacin A, Quirce S, Armentaria A, Fernandez-Nieto M, Pacios LF, Asencio T, Sastre J, Diaz-Perales A, Salcedo G. Wheat lipid transfer protein is a major allergen associated with bakers' asthma. J Allergy Clin Immunol (2007) 120, 1132:1138

Pastorello EA, Farioli L, Conti A, Pravettoni V, Bonomi, S, Iametti S, Fortunato D, Scibilia J, Bindsley-Jensen C, Ballmer, Weber B, Robino AM Ortolani C. Wheat IgE-mediated food allergy in European patients: a-amylase inhibitors, lipid transfer proteins and low molecular weight glutenins. Allergenic molecules revealed by double-blind, placebo-controlled food challenges. Int Arch Allergy Immunol (2007) 144:10–22.

Payne PI. Genetics of wheat storage proteins and the effect of allelic variation on breadmaking quality. Ann Rev Pl Physiol (1987) 38:141-153.

Qi P-F, Wei Y-M, Ouellet T, Chen Q, Tan X, Zheng Y-L. The γ-gliadin multigene family in common wheat (*Triticum aestivum*) and its closely related species. BMC Genomics (2009) 10: 168. doi 10.1186/1471-2164-10-168

Salcedo G, Prada J, Aragoncillo C. Low MW gliadin-like proteins from wheat endosperm. Phytochem (1979) 18:725-727.

Salcedo G, Sanchez-Monge R, Garcia-Casado G, Armentaria A, Gomez L, Barber D. "The cereal α-amylase/trypsin inhibitor family associated with Bakers' asthma and food allergy". In Mills ENC, Shewry PR editors. Plant Food Allergens. Oxford UK: Blackwell Science (2004). p.70-86.

Shewry PR, Tatham AS. Improving wheat to remove coeliac epitopes but retain functionality. J. Cereal. Sci. (2016) 67:12-21.

Shewry PR, Jenkins J, Beaudoin F, Mills ENC. "The classification, functions and evolutionary relationships of plant proteins in relation to food allergens". In Mills ENC, Shewry PR editors. Plant Food Allergens. Oxford UK: Blackwell Science (2004). p.24-41.

Shewry PR, Tatham AS, Forde J, Kreis M, Miflin BJ. The classification and nomenclature of wheat gluten proteins: a reassessment. J Cereal Sci (1986) 4, 97:106.

Shewry PR, Halford MG. Cereal seed storage proteins: structures, properties and role in grain utilisation. J Exp Bot (2002) 53:947-958.

Shewry PR, Halford NG, Lafiandra D. "The genetics of wheat gluten proteins". In Hall JC, Dunlap JC, Friedman T, editors. Advances in Genetics Vol. 49. London: Academic Press (2003). p.111-184.

Shewry, P.R., Tatham, A.S., Barro, F., Barcelo, P., Lazzeri, P. (1995) Biotechnology of breadmaking: unraveling and manipulating the multi-protein gluten complex. Biotechnology 13, 1185-1190.

Shewry PR, D'Ovidio R, Lafiandra D, Jenkins JA, Mills ENC, and Bekes F. "Wheat grain proteins". In Khan K, Shewry PR, editors. Wheat: Chemistry and Technology 4th edition. St Paul Mn: AACC (2009). p. 223-298.

Shewry PR, Jenkins J, Beaudoin F, Mills ENC. "The classification, functions and evolutionary relationships of plant proteins in relation to food allergens" In Mills ENC, Shewry PR editors. Plant Food Allergens. Oxford UK: Blackwell Science (2004). p.24-41.

Sollid LM, Qiao S-W, Anderson RP, Gianfrani C, Konig F. Nomenclature and listings of celiac disease relevant gluten T-cell epotopies restricted by HLA-DQ molecules. Immunogenetics (2012) 64: 455-460.

Tilley KA, Schofield JD. Detection of phosphorylation in the high Mr subunits of wheat glutenin. J Cereal Sci (1995) 22:17-19

Tilley KA, Lookhart GL, Hoseney RC, Mawhinney TP. Evidence for glycosylation of the high molecular weight glutenin subunits 2, 7, 8 and 12 from Chinese Spring and TAM 105 wheats. Cereal Chem (1993) 70: 602-606.

Tosi P. Trafficking and deposition of prolamins in wheat. J Cereal Sci (2012) 56:81-90

Tosi P, He J, Lovegrove A, Gonzalez-Thuillier I, Penson S, Shewry PR. Gradients in compositions in the starchy endosperm of wheat have implications for milling and processing. TIFST (2018) 82:1–7.

Turner AS, Bradburne RP, Fish L, Snape JW. New quantitative trait loci influencing grain texture and protein content in bread wheat. J Cereal Sci (2004) 40:51-60.

Van den Bulck K, Loosveld A-MA, Courtin CM, Proost P, Van Damme J, Robben J, Mort A, Delcour JA. Amino acid sequence of wheat flour arabinogalactan-peptide, identical to part of grain softness protein GSP-1, leads to improved structural model. Cereal Chem (2002) 79:329-331.

Van den Bulck, K, Swennen K, Loosveld A-MA, Courtin CM, Brijs K, Proost P, Van Damme J, Van Campenhout S, Mort A, Delcour JA. Isolation of cereal arabinogalactan-peptides and structural comparison of their carbohydrate and peptide moieties. J Cereal Sci (2005) 41:59-67.

609 Wilkinson MD, Tosi P, Lovegrove A, Corol DI, Ward JL, Palmer R, Powers S, Passmore D,
610   Webster G, Marcus SE, Knox JP, Shewry PR. The *Gsp-1* genes encode the wheat
611   arabinogalactan peptide. J Cereal Sci (2017) 74:155-164.
612 Wrigley CW, Bietz JA. "Proteins and amino acids". In: Pomeranz Y, editor. Wheat Chemistry
613   and Technology, Third Edition. St Paul Mn: AACC (1988). p.159-275.

614

615

616 **Conflict of interest statement**

617 The author has no conflicts of interest

618

619

620

621

622

623 Table 1. Summary of the types and characteristics of wheat gluten proteins (based on Shewry
624   and Halford, 2002).

625

Gluten protein type	Molecular mass	% total gluten fraction	Polymers or monomers?	Partial amino acid composition (mol %)
HMW prolamins				
HMW subunits	65-90,000	6-10	polymers	30-35% glutamine, 10-16% proline, 15-20% glycine, 0.5-1.5% cysteine, 0.7-1.4% lysine
S-rich prolamins				
α-gliadins			monomers	30-40% glutamine, 15-20% proline, 2-3% cysteine, less than 1% lysine
γ-gliadins	30-45,000	70-80		
B-type and C-type LMW subunits			polymers	
S-poor prolamins				

			monomers	40-50% glutamine, 20-30% proline, 0-0.5% phenyl alanine, 0-0.5% lysine, 0 cysteine (1 cysteine residue in D-type LMW subunits
ω-gliadins	30-75,000	10-20		
D-type LMW subunits			polymers	

626

627

628

629

Table 2. Wheat grain proteins of the prolamin superfamily (based on literature discussed in the text)

Protein group	Molecular mass	Characteristics	Abundance	Functional properties/impact on health
farinins	17,000 to 30,000	Correspond to avenin-like proteins and LMW gliadins	Not determined	Transgenic expression results in improved mixing properties
purinins (low molecular weight gliadins)	17,000-19,000	Possibly correspond to "ancestral" type of prolamin	Not determined	Behave like gliadins in dough
puroindolines a and b	13,000	Tryptophan-rich loop region which may be involved in binding to starch granule surface	0.029-0.060 % dry wt of Pin a and 0.004-0.031 % dry wt. of Pin b in wholemeal flour	Determine about 75% of the variation in softness in European wheats
grain softness protein (GSP) +	Approx. 15,000	Associated with the starch granule surface	Not determined	Small effect on grain softness.
arabinogalactan peptide (AGP)	23,000	15 residue peptide *o*-glycosylated with arabinogalactan chains at 3 hydroxyproline residues.	0.39 % dry wt. white flour	Prebiotic properties in vitro.
non-specific lipid-transfer proteins (LTP)	9,000 (LTP1) +	Bind and transport lipids *in vitro*. Concentrated in	Not determined	LTP1 is a food and respiratory allergen.

	7,000 (LTP2)	aleurone layer and embryo.		
α-amylase/trypsin inhibitors (ATIs)	12,000 to 16,000	Monomeric, dimeric and tetrameric forms, some subunits inhibit trypsin or α-amylase	0.34-0.41 % dry wt. of wholemeal flour	Include respiratory and food allergens, putative links to coeliac disease, NCWS and other adverse reactions to wheat. Contribute to pasta-making quality.

632

633

634

Figure Captions

Figure 1. The origin of wheat gluten

    a.   transmission electron microscopy of starchy endosperm cells at a late stage of grain development (46 days after anthesis) shows that the individual protein bodies have fused to form a continuous proteinaceous matrix. Taken from Shewry et al. (1995) with permission, provided by Dr. M. Parker (IFR, Norwich, UK).

    b.   digestion of a flour particle to remove starch reveals a continuous proteinaceous network. Taken from Amend (1995) with permission.

    c.   Transverse section of the lobe region of a developing wheat grain stained with Toluidine Blue to show the tissue structure and deposited protein (in blue). Figure kindly provided by Cristina Sanchis Gritsch and Paola Tosi (Rothamsted Research).

Figure 2. The distribution of T-cell epitopes (shown as red bars) in representative wheat gluten proteins (identified by GenBank accession codes). The epitopes are based on Sollid et al. (2012).

α-gliadin P18573: DQ2.5-glia-α1a, DQ2.5-glia-α1b, DQ2.5-glia-α2 & DQ8-glia-α1. **γ-gliadin** AAK84774: DQ2.5-glia-ω1/hor-1/sec-1, DQ8-glia-γ1a, DQ8-glia-γ2, DQ8-glia-γ4c & DQ8-glia-γ5. **ω-gliadin (A/D)** AAT74547: DQ2.5-glia-γ5, DQ8-glia-γ1a, DQ2.5-glia-ω1/hor-1/sec-1, DQ8-glia- γ1b & DQ2.5-glia- γ3. **ω-gliadin (B)** AB181300 no coeliac toxic epitopes present. **LMW subunit** AAS66085:DQ2.5-glut-L1. **HMW Subunit (1Bx17)** BAE96560: DQ8.5-glut-H1. **HMW Subunit (1Dy10)** AAU04841: DQ8.5-glut-H1.
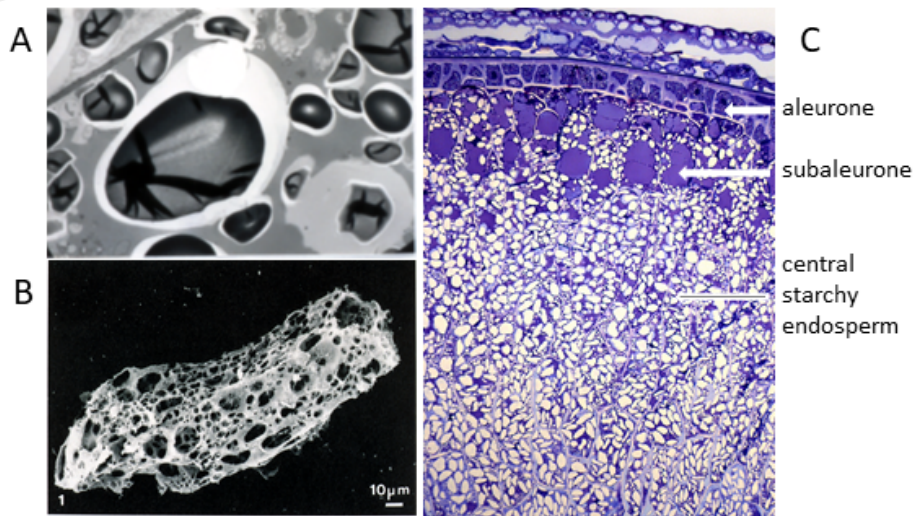
Modified from Shewry and Tatham (2015).

Figure 1.TIF

Figure 2.JPEG

α-gliadin P18573

γ-gliadin AAK84774

ω-gliadin (A/D) AAT74547

ω-gliadin (B) AB181300

LMW glutenin AAS66085

HMW Subunit (1Bx17) BAE96560

HMW Subunit (1Dy10) AAU04841