A Universal Approach to Eliminate Antigenic Properties of Alpha-Gliadin Peptides in Celiac Disease

Cristina Mitea1*, Elma M. J. Salentijn2‡, Peter van Veelen1, Svetlana V. Goryunova2, Ingrid M. van der Meer2, Hetty C. van den Broeck2, Jorge R. Mujico1†, Veronica Monserrat1, Luud J. W. J. Gilissen2,3, Jan Wouter Drijfhout1, Liesbeth Dekking1, Frits Koning1,3⁎, Marinus J. M. Smulders2,3⁎

1 Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, The Netherlands, 2 Plant Research International, Wageningen UR, Wageningen, The Netherlands, 3 Allergy Centre Wageningen, Wageningen, The Netherlands

Abstract
Celiac disease is caused by an uncontrolled immune response to gluten, a heterogeneous mixture of wheat storage proteins, including the α-gliadins. It has been shown that α-gliadins harbor several major epitopes involved in the disease pathogenesis. A major step towards elimination of gluten toxicity for celiac disease patients would thus be the elimination of such epitopes from α-gliadins. We have analyzed over 3,000 expressed α-gliadin sequences from 11 bread wheat cultivars to determine whether they encode for peptides potentially involved in celiac disease. All identified epitope variants were synthesized as peptides and tested for binding to the disease-associated HLA-DQ2 and HLA-DQ8 molecules and for recognition by patient-derived α-gliadin specific T cell clones. Several specific naturally occurring amino acid substitutions were identified for each of the α-gliadin derived peptides involved in celiac disease that eliminate the antigenic properties of the epitope variants. Finally, we provide proof of principle at the peptide level that through the systematic introduction of such naturally occurring variations α-gliadin genes can be generated that no longer encode antigenic peptides. This forms a crucial step in the development of strategies to modify gluten genes in wheat so that it becomes safe for celiac disease patients. It also provides the information to design and introduce safe gluten genes in other cereals, which would exhibit improved quality while remaining safe for consumption by celiac disease patients.

Introduction
Celiac Disease (CD) is an intestinal T cell-mediated disease caused by the gluten fraction of wheat or the homologous proteins from barley or rye. CD has prevalence between 0.5 and 2% in human populations [1] and is characterized by a chronic intestinal inflammation upon ingestion of gluten proteins. Recently, the molecular aspects have been comprehensively addressed in several review papers [2–5]. In short, in CD patients CD4+ T cells are induced by gluten [7]. The α-gliadins are the only gluten molecules that harbor the p31-43/49 peptide (LQLQPFPQPQLPYPQPQLPY) that has very potent T cell stimulatory properties [13]. It harbours the p56–75 peptide (LQLQPFPQPQLPYPQPQLPY) was identified that encodes six partially overlapping T cell epitopes and trigger T cell responses [2–5]. In addition to the adaptive CD4+ T cell response to gluten, CD is characterized by the upregulation of IL-13, an intraepithelial T cell infiltrate expressing the NKG2D receptor, and the overexpression of a ligand for NKG2D (MICA) [6,7].

Many gluten peptides with T cell stimulatory properties have now been identified. Such peptides have been found in wheat α-, γ-, and ω-gliadins as well as in low molecular weight (LMW) and high molecular weight (HMW) glutenins [8–14]. Several studies have demonstrated that peptides derived from α-gliadins induce strong T cell responses in the large majority of patients, while responses to the other peptides are less frequently found [8–10,13]. An α-gliadin derived 33-mer peptide (amino acid sequence LQLQPFPQPQLPYPQPQLPY) was identified that encodes six partially overlapping T cell epitopes and has very potent T cell stimulatory properties [13]. It harbours the p56–75 peptide (LQLQPFPQPQLPYPQPQLPY) that has been identified as the dominant gluten epitope [9,10]. Furthermore, α-gliadins are the only gluten molecules that harbor the p31-43/49 peptide that has been implicated in the innate immune response induced by gluten [7].

The α-gliadins are a gene family encoded by the Gli-2 loci, Gli-A2, Gli-B2 and Gli-D2, located on the short arm of three homoeologous chromosomes (6AS, 6BS and 6DS) of hexaploid bread wheat (Triticum aestivum L.). These loci can contain from 25–
35 to even 150 α-gliadin genes per haploid genome [15–17], although most of these (72–95%) are presumably pseudogenes [16,17]. Sequencing of genomic α-gliadin clones from hexaploid bread wheat enabled to differentiate the sequences according to their loci Gli-A2, Gli-B2 and Gli-D2 based on genome-specific SNPs [16,17]. Relevant for CD, the occurrence and frequency of the HLA-DQ2 epitopes DQ2-Glia-ζ1, DQ2-Glia-ζ2 and DQ2-Glia-ζ3 (previously designated Glia-ζ2, ref. 12) and the HLA-DQ8 T-cell epitope DQ8-Glia-ζ1 also differs between the loci [17]. This was corroborated by the observation that T cell clones specific for the DQ2-Glia-ζ2 epitope did not recognize gluten derived from diploid species carrying the S-genome, ancestrally related to the B genome of bread wheat, while gluten derived from diploid A- and D-genome species was recognized [18]. Variation in T cell stimulatory capacity of cereal-derived gluten was observed with other T cell clones as well [19–21]. Indeed, differences have been observed in the T cell stimulatory capacity of pasta and bread wheat varieties [22,23], but none were safe for CD patients.

Given the overall importance of the α-gliadins in CD we set out to determine the naturally existing sequence variation in CD epitopes as deduced from α-gliadin transcripts from developing wheat grains. The immunogenic potential of these epitope variants was subsequently tested in T-cell proliferation assays. This produced insight in which key amino acid changes are sufficient to abolish T-cell recognition. Finally, we verified that the observed differences in antigenicity of α-gliadin peptides derived from diploid species corresponded to differences in the antigenicity of the gluten from these species, ancestrally related to bread wheat. Altogether the results offer a molecular basis for differential CD reactions to gluten from these species, ancestrally related to bread wheat.

### Results

**Genetic variation in α-gliadin (Gli-2) transcripts**

In total 3022 expressed α-gliadin sequences (expressed sequence tags and mRNA sequences from NCBI and UniGene) originating from 11 different T. aestivum L. varieties were analyzed. These α-gliadin transcripts were grouped into 35 contigs with at least 90% sequence homology. Forty per cent of the α-gliadin transcripts clustered with 22 genomic sequences and were attributed to locus Gli-A2, 35% originated from Gli-D2 and only 25% came from Gli-B2 [Fig. S1]. After tracing all non-synonymous DNA polymorphisms 83 different transcript contigs were obtained for the 3′ region of the gene that each contained at least four sequence equivalents. This indicates a high sequence diversity among expressed α-gliadin sequences in these 11 T. aestivum L. varieties.

**Variants of T cell stimulatory and innate stimulatory sequences**

The N-terminal region of α-gliadins contains the p31-43 epitope implicated in the innate immune response, and the immunodominant DQ2-Glia-ζ1 and DQ2-Glia-ζ2 epitopes as well as the DQ2-Glia-ζ3 T cell epitope. The carboxyl-terminal part encodes the immunodominant DQ8-Glia-ζ1 T cell epitope (Table 1). To obtain information on the immunogenic potential of the various α-gliadins, the translated amino acid sequences for each locus were checked for the presence of canonical T cell epitopes and variants thereof. Tables 2, 3, 4, 5 present the most frequently expressed epitope variants (>5 transcripts each).

![Table 1](https://example.com/table1.png)  
**Table 1. Amino acid sequences of α-gliadin derived peptides.**

Table 2 lists HLA-DQ2-Glia-ζ1 epitope variants expressed in bread wheat.

![Table 2](https://example.com/table2.png)  
**Table 2. HLA-DQ2-Glia-ζ1 epitope variants expressed in bread wheat.**

Five antigenic peptides derived from α-gliadins are known to be involved in celiac disease. For the peptides that can provoke an adaptive immune response in CD patients (HLA-DQ2 restricted or HLA-DQ8 restricted) the minimal 8-mer “canonical” epitope cores are shown. One peptide, p31-43, is known to be involved in an innate immune response observed in CD. For each of the epitopes is specified the name, which immune response it evokes, the restriction element and the amino acid sequence.

<ref>doi:10.1371/journal.pone.0015637.t002</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>

<ref>doi:10.1371/journal.pone.0015637.t001</ref>
and Gli-B2 loci. The latter expressed substantially truncated versions of the 33-mer (Fig. S2).

T cell stimulatory capacity of α-gliadin derived peptides

Several of the amino acid variants that we found in the α-gliadin transcriptome have never been described before while some have been described but were never tested for their T cell stimulatory capacity. In order to determine which variants are capable of inducing T cell responses, the DQ2-Glia-α1, DQ2-Glia-α2, DQ2-Glia-α3, and DQ8-Glia-α1 variants were synthesized as 15- or 16-mer peptides and tested for their capacity to bind to HLA-DQ2 and induce in vitro proliferation of HLA DQ2- or DQ8-restricted T cell clones.

**DQ2-Glia-α1 variants.** Virtually all peptides that carry the canonical 9-mer Glia-α1 epitope core P1[F/Y]2P3Q4P5E6L7P8Y9 (in which the glutamic acid at p6 is introduced by TG2-deamidation of the original glutamine) were able to stimulate DQ2-Glia-α1 T cells. Only an arginine residue at the position preceding the epitope core diminished T cell stimulation. In the core sequence, several amino acid substitutions diminished or abolished the T cell stimulatory capacity, such as a proline to serine substitution at p3 or p8 and a proline to alanine substitution at p5 (Table 6, no. 3). Peptides in which an amino acid was deleted at p3 or p4 were not causing any proliferation of the T cells. Strikingly, such safe peptides are all from locus Gli-B2 (Table 6, Fig. S3).

**DQ2-Glia-α2 variants.** Full responses of DQ2-Glia-α2 T cells were only observed against peptides that carry the core P1:F2Q3P4E5L6P7Y8Q9. A deletion of the glutamine at p2, indicative for α-gliadins from locus Gli-B2, or substitution of this glutamine by histidine diminished or abolished the stimulatory capacity. Furthermore, a single substitution of the proline for an alanine residue at either p6 or p8 abolished the T cell stimulating capacity. The latter substitution is found in Gli-B2, Gli-D2 or Gli-α2 variants are from locus Gli-A2, Gli-B2 or Gli-D2. N = total transcript count per variant. In bold: amino acid variation.

doi:10.1371/journal.pone.0015637.t005

| Table 5. DQ8-Glia-α1 epitope variants expressed in bread wheat. |
|-----------------------|-------------------|-------------------|-------------------|-------------------|
|                      | Gli-A2 | Gli-B2 | Gli-D2 | N    |
| QQSEPQPSQQN           | 186    | 381    | 567    |
| QQSEPQPSQQN           | 409    | 409    |
| QQFYPQPSQQN           | 114    | 114    |
| QQSEPQPSQQN           | 103    | 103    |
| QQFPFSQQN             | 112    | 112    |
| QQSEPQPSQQN           | 111    | 111    |
| QQSEPQPSQQN           | 4      | 27     | 31    |
| QQSEPQPSQQN           | 6      | 6      |
| Total                 | 512    | 413    | 528    | 1453  |

Expressed variants of the DQ8-Glia-α1 epitope, represented by ≥5 transcripts and the number of transcripts per epitope variants per chromosomal locus (Gli-A2, Gli-B2 or Gli-D2). N = total transcript count per variant. In bold: amino acid variation.

doi:10.1371/journal.pone.0015637.t005

Table 6, Fig. S3.

Table 5. DQ8-Glia-α1 epitope variants expressed in bread wheat.

| Table 6. DQ2-Glia-α3 epitope variants expressed in bread wheat. |
|-----------------------|-------------------|-------------------|-------------------|-------------------|
|                      | Gli-A2 | Gli-B2 | Gli-D2 | N    |
| PQPQLYPQQ            | 607    | 607    |        |
| POPQLYPSQ            | 431    | 431    |
| PPPQLPPYQ           | 382    | 382    |
| LOPQLPSPQ           | 106    | 106    |
| FPLQLPFQP           | 31     | 31     |
| GQPQLYSPQ           | 85     | 85     |
| POPQLYSEP           | 30     | 30     |
| PHGPQYPYQ           | 29     | 29     |
| POPQLYSPQ           | 7      | 7      |
| POPQLYSEP           | 6      | 6      |
| POPQLYGSQ           | 6      | 6      |
| Total               | 629    | 419    | 726    | 1774  |

Expressed variants of the DQ2-Glia-α3 epitope represented by ≥5 transcripts and the number of transcripts per epitope variants per chromosomal locus (Gli-A2, Gli-B2 or Gli-D2). N = total transcript count per variant. In bold: amino acid variation.

doi:10.1371/journal.pone.0015637.t003
T cell stimulatory capacity of diploid wheat accessions and lack the capacity to stimulate Glia-α2 T cell clones, whereas α-gliadins encoded by Glia-B2 carry a deletion that prevents Glia-α1 T cell clone stimulation. Alpha-gliadins with the two intact epitopes, Glia-α1 and Glia-α2, are encoded by Gl-D2 (Tables 2 and 3).

### T cell stimulatory capacity of diploid wheat accessions

Previously, differential reactivity of α-gliadin specific T cell clones against gluten extracts from diploid wheat accessions has been reported [18,19]. In order to link such differential reactivity to the presence of specific epitope variants, a panel of diploid wheat accessions containing either the A, S or D genome was tested for their reactivity of an α-gliadin specific monoclonal antibody (mAb) and T cells.

Pepsin-trypsin digests of gluten extracts from kernels of 29 diploid Triticum and Aegilops accessions were prepared and tested in a competition ELISA with a mAb specific for a sequence partially overlapping with the DQ2-Glia-α1 and DQ2-Glia-α2 epitopes (Fig. 1A) and, after treatment with TG2, with T cell clones specific for either the DQ2-Glia-α1 and DQ2-Glia-α2 epitope (Fig. 1B and 1C). The results indicate that the mAb reacts strongly with all extracts except three derived from diploids expressing the S genome (Fig. 1A). Similarly, and in agreement with previous results [18], extracts from A, S and D origin were capable of stimulating the DQ2-Glia-α1 specific T cell clone (Fig. 1B) while the extracts of the diploids expressing the A genome failed to stimulate the DQ2-Glia-α2 specific T cell clone (Fig. 1C).

Based on our observation that the α-gliadins expressed from locus Gl-D2 of the T. aestivum A genome carry a variant DQ2-Glia-α2 epitope in which the proline at p8 has been replaced by a serine, and our experimental result that introducing this substitution in a peptide leads to loss of DQ2-Glia-α2 T cell stimulatory properties (Table 6), we wanted to determine if this amino acid substitution was indeed the cause of loss of immunogenicity. For this purpose we sequenced the α-gliadin transcripts from T. monococcum (A genome) accessions.

| No. | Peptide | locus | IC50 DQ2-Glia-α1 | Glia-α1 T cell clone | IC50 DQ2-Glia-α2 | Glia-α2 T cell clone |
|-----|---------|-------|-----------------|----------------------|-----------------|----------------------|
| 1   | QLQFPFQPELYFQPE | Gli-D2 | 5               | +                    | 18              | +                    |
| 2   | QLQFPFQPELYFQPE | Gli-D2 | 5               | +                    | 34              | +                    |
| 3   | QLQFPFQPELYFQPE | Gli-D2 | 8               | ±                    | 11              | -                    |
| 4   | QLQFPFQPELYFQPE | Gli-D2 | 12              | -                    | 24              | -                    |
| 5   | QLQFPFQPELYFQPE | Gli-D2 | 14              | ±                    | 19              | +                    |
| 6   | QLQFPFQPELYFQPE | Gli-D2 | 21              | +                    | 21              | -                    |
| 7   | QLQFPFQPELYFQPE | Gli-D2 | 42              | +                    | 41              | ±                    |
| 8   | QLQFPFQPELYFQPE | Gli-A2 | 7               | ±                    | 32              | -                    |
| 9   | QLQFPFQPELYFQPE | Gli-A2 | 57              | +                    | 41              | -                    |
| 10  | QLQFPFQPELYFQPE | Gli-A2 | 67              | +                    | 24              | -                    |
| 11  | QPQFEL-PELYFQPE | Gli-B2 | 4               | -                    | 4               | ±                    |
| 12  | QPQFEL-PELYFQPE | Gli-B2 | 8               | -                    | 8               | +                    |
| 13  | QPQFEL-PELYFQPE | Gli-B2 | 45              | -                    | 45              | -                    |
| 14  | QPQFEL-PELYFQPE | Gli-B2 | 86              | -                    | 86              | -                    |
| 15  | QPQFEL-PELYFQPE | Gli-B2 | nd              | -                    | nd              | -                    |
| 16  | QPQFEL-PELYFQPE | Gli-A2/D2 | 35          | +                    |                 |                      |
| 17  | QPQFEL-PELYFQPE | Gli-A2/D2/B2 | 3080        | -                    |                 |                      |
| 18  | LGEPFQPSQENP | Gli-D2/B2 | 16         | +                    |                 |                      |
| 19  | LGEPFQPSQENP | Gli-A2 | 33              | -                    |                 |                      |
| 20  | LGEPFQPSQENP | Gli-D2 | nd              | -                    |                 |                      |

Variants of the DQ2-Glia-α1 and DQ2-Glia-α2 epitopes as encoded by the α-gliadin transcriptome were synthesized as deamidated 14- to 17-mer peptides (column 1, underlined: DQ2-Glia-α1/α2 epitope region) and tested for stimulation of DQ2-Glia-α1 and DQ2-Glia-α2 specific T cell clones in a proliferation assay. = 100 times reduced T cell stimulation compared to the ‘canonical’ epitope; = 1000 times reduced T cell stimulation compared to the ‘canonical’ epitope. For each epitope shorter versions of the peptide variant, including the putative epitope core flanked by at least two amino acids, were tested in a cell free in vitro peptide binding assay for binding to HLA-DQ2 antigen presenting cells (lysates from HLA-DR3/DQ2 positive EBV-transfected B-cells). IC50 DQ2-Glia-α1/α2 = mean value of the half maximal inhibitory concentration (IC50) returned by the binding assays for respectively DQ2-Glia-α1 and DQ2-Glia-α2 epitope variants. IC50 values were calculated based on the observed competition between the tested peptides and biotin-labelled indicator peptides and indicate the concentration of the tested peptide required for half maximal inhibition of the binding of the indicator peptide.

doi:10.1371/journal.pone.0015637.t006
monococcum contain only a single form of DQ2-Glia-\(\alpha_1\) and DQ2-Glia-\(\alpha_2\). Moreover, in the DQ2-Glia-\(\alpha_2\) epitope the proline at p8 was consistently replaced by a serine (Table S1). Together these results establish that this naturally occurring single amino acid substitution is sufficient to completely eliminate the T cell stimulatory properties of the DQ2-Glia-\(\alpha_2\) epitope in gluten.

Differential reactivity of the DQ2-Glia-\(\alpha_3\) specific T cells towards the extracts of the diploids correlated with an arginine to proline replacement at p2 in \(\alpha\)-gliadins derived from the S-genome, FRPQQPYPQ\(\rightarrow\) FPQQPYPQ (Table 6).

Elimination of \(\alpha\)-gliadin toxicity by a naturally occurring single amino acid substitution

The large majority of known antigenic peptides derived from wheat gluten, as well as homologous peptides derived from the hordeins from barley and the secalins from rye, contain a proline at p8 [24]. Based on our observation that a single proline to serine replacement at p2 in \(\alpha\)-gliadins derived from the S-genome, FRPQQPYPQ\(\rightarrow\) FPQQPYPQ (Table 6).

We also analysed the effects of proline to serine substitutions in the 33-mer \(\alpha\)-gliadin derived peptide that encodes 6 partially overlapping antigenic DQ2-Glia-\(\alpha_1\) and -\(\alpha_2\) sequences (Fig. 2B) as well as in an elongated version of the 33-mer which also encodes the DQ2-Glia-\(\alpha_3\) epitope (Table S2). In all cases, the proline to serine substitutions completely abrogated the response of DQ2-Glia-\(\alpha_1\), DQ2-Glia-\(\alpha_2\) and DQ2-Glia-\(\alpha_3\) specific T cell clones (Fig. 2B, Table S2).

Finally, we tested the effect of systematic introduction of single, double, triple and quadruple P to S substitutions in the DQ2-Glia-\(\alpha_1\) peptide with five T cell clones isolated from small intestinal biopsies of 3 CD patients. T cell responses of single T cell clones were only observed for two of the single P to S substitutions at p3 and p5. In contrast, with a single P to S substitution at p8 and in all but one of the other cases the substitutions completely eliminated the T cell response of all five T cell clones (Fig. 3).

Thus, the naturally occurring proline to serine substitution constitutes a universal approach to remove the antigenic properties of HLA-DQ2 restricted \(\alpha\)-gliadin peptides.
Discussion

Although T cell responses to peptides derived from α-, γ- and ω-gliadins as well as from HMW- and LMW-glutenins have been described, various studies have indicated that the α-gliadins are among the most immunogenic regarding CD [8–10,13,25]. A crucial step towards the elimination of gluten toxicity would thus be the elimination of T cell stimulatory α-gliadin sequences. Our extensive genetic analysis of over 3000 α-gliadin transcripts from different bread wheat accessions showed a high heterogeneity of

Figure 2. Amino acid substitution eliminates toxicity of known α-gliadin epitopes. The DQ2-Glia-α3 epitope and the known 33-mer were synthesized in deamidated form either as the original sequence or after substitution in each epitope of the prolines at position 8 with serine. These peptides were tested in T cell proliferation assays. A: T cell proliferation assay using a DQ2-Glia-α3 T cell clone. B: T cell proliferation assay using a DQ2-Glia-α1 T cell clone. Positive control: synthetic peptide encoding the specific minimal T cell epitope. Stim. index: stimulation index defined as the specific proliferation of a sample divided by the background proliferation.

doi:10.1371/journal.pone.0015637.g002

Figure 3. Response of DQ2-Glia-α1 epitope specific T cell clones against single and multiple P to S substituted peptides. Five T cell clones derived from 3 CD patients were tested against the deamidated form of the DQ2-Glia-α1 peptide (sequence PFPQPELPY) and variants thereof in which prolines at position 3, 5, 8 and 10 were systematically substituted for serine, both as single substitutions and in all possible combinations. Shown is the response to the substituted peptides relative to unsubstituted DQ2-Glia-α1 epitope. The introduced substitutions are underlined. The most C-terminal substituted proline at position 10 lies outside the 9 amino acid core of the T cell stimulatory peptide.

doi:10.1371/journal.pone.0015637.g003
the α-gliadins genes and considerable differences in the number of T cell stimulatory sequences encoded by the various α-gliadin genes. We identified three major factors determining these differences: i) the length of tandem repeats of antigeic sequences; ii) natural amino acid substitutions that affect the antigenicity of T cell epitopes, and iii) amino acid deletions that eliminate the antigenicity of T cell epitopes.

The α-gliadins from locus Gli-D2 generally encode several copies of both the DQ2-Glia-x1 and DQ2-Glia-x2 epitopes in addition to the DQ2-Glia-x3 and DQ8-Glia-x1 epitopes. Gli-A2 genes usually encode only the DQ2-Glia-x1 and DQ2-Glia-x3 epitopes while the Gli-B2 genes encode no or at most one DQ2-Glia-x1 epitope, next to the DQ8-Glia-x1 epitopes. The 33-mer sequence with 6 T cell stimulatory sequences [13] was only found in a minority of the α-gliadins analyzed, all of which are expressed from Gli-D2. Many natural occurring amino acid substitutions affecting the antigenicity of the canonical α-gliadin peptides were identified. Typical examples are the proline to serine substitution at p8 in the DQ2-Glia-x2 epitope and the arginine to proline substitution at p2 in the DQ2-Glia-x3 epitope, that both completely eliminate the T cell stimulatory properties of these peptides.

The analysis of gluten extracts from the diploid wheat varieties underscores these observations and provides a molecular basis for the previous observation that A-genome diploid T. monococcum varieties lack the DQ2-Glia-x2 epitope [18]. All α-gliadins from this genome encode an altered version of the DQ2-Glia-x2 epitope with a serine at p8 that fails to induce T cell responses. We observed that a similar substitution eliminates the T cell stimulatory properties of the DQ2-Glia-x1 and DQ2-Glia-x3 epitopes. The more variable reaction pattern of T-cells and antibodies towards gluten extracts of the S-genomes reflects the higher level of genetic variation in these outcrossing species. These α-gliadins contain another variant of the DQ2-Glia-x1 and DQ2-Glia-x2 region that does not incude the A-genome specific serine at p8. Furthermore, amino acid deletions in the canonical α-gliadin peptides prohibit binding to HLA-DQ2 and hence T cell recognition. A typical example is the deletion of the glutamate at p4 in the DQ2-Glia-x1 epitope which generates a peptide that no longer binds to HLA-DQ2, presumably due to defective docking of the anchor residues into their respective pockets in the HLA-DQ2 molecule. Our results confirm previous observations [18,19] that the α-gliadin locus Gli-D2 encodes the most toxic α-gliadins while substantially less toxicity is associated with those from Gli-A2 and Gli-B2, but also provide the molecular basis for these differences.

Unfortunately, due to the complexity of both the Gli-2 gene family and the wheat genome, it will be a difficult task to generate tetraploid pasta and hexaploid bread wheat that is entirely safe for consumption by all CD patients by conventional breeding methods. Our results now provide a rationale for an alternative approach as we demonstrate that by the introduction of naturally occurring amino acid substitutions the toxicity of all four T cell epitopes in α-gliadins can be eliminated. Using novel methods such as zinc finger nucleases [26–28] we can introduce the underlying SNPs as specific mutations into the α-gliadin genes of wheat to eliminate toxicity completely. Technically this will not be easy, but our results indicate precisely the three complementary sets of actions that need to be performed.

i) In the Gli-B2-derived α-gliadins analyzed, none of the DQ2 epitopes are present due to a single amino acid deletion in the region encoding the DQ2-Glia-x1 and DQ2-Glia-x2 epitopes, which generates a peptide that has decreased binding affinity for HLA-DQ2 and is no longer recognized by T cells. Therefore, a single amino acid substitution in the DQ2-Glia-x3 epitope will result in a peptide that has completely lost HLA-DQ2 binding properties and T cell stimulatory properties. To eliminate the remaining DQ8-Glia-x1 epitope in some Gli-B2 α-gliadins, a single glutamine to arginine substitution, which naturally occurs in the α-gliadins from the A-genome, would suffice. Such a minimally genetically modified B-genome α-gliadin gene would thus no longer encode any T cell stimulatory peptides. Moreover, by starting with an α-gliadin gene in which the sequence of the p31-43 peptide is naturally altered (example the gene encoding protein no. 9 in SI 2), the chance of innate immune stimulation by a protein derived from such a gene would also be minimized.

ii) For Gli-A2 α-gliadins, the approach to eliminate toxicity would be to introduce two proline to serine substitutions at p8 in the DQ2-Glia-x1 and DQ2-Glia-x3 epitopes present. As these α-gliadins encode a shorter version of the immunodominant 33-mer in which the DQ2-Glia-x2 epitope is already non-functional, and contain a version of the DQ8-Glia-x1 epitope that has no T cell stimulatory properties, these two substitutions would completely remove toxicity in proteins encoded by such modified genes.

iii) Regarding the Gli-D2 α-gliadins, we found that the proline to serine substitutions at p8 completely abrogated the in vitro T cell stimulatory properties of the 33-mer peptide and of an elongated version of the 33-mer that also encodes the DQ2-Glia-x3 epitope. This result underscores our previous observation that most T cell stimulatory gluten peptides have a proline at p8 [24]. However, to render α-gliadins from the D-genome non-toxic, up to seven substitutions need to be introduced in a single gene.

An alternative approach is to design safe α-gliadin genes that can subsequently be introduced into celiac disease safe cereals such as rice or maize, for the production of gluten proteins. Such modified proteins will be very similar to existing α-gliadins they will most likely have indistinguishable technological properties. Thus, such gluten proteins could enhance the baking properties of these cereal crops, or they could be extracted from these crops and used as an ingredient to generate novel high quality foods for celiac disease patients. For the generation of high quality cereals that can replace wheat-based products the simple introduction of detoxified α-gliadins is unlikely to be sufficient, as baking quality is mostly determined by the HMW and LMW glutenin proteins. Therefore, additional studies will have to investigate how other gluten proteins can be detoxified as there is substantial evidence that these contain T cell stimulatory peptides as well. In previous studies we provided evidence that such epitopes can be found in the γ-gliadins as well as in the LMW- and HMW-glutenins [11,12] and others have extended these observations [14,25]. In particular, we found that T cell responses to LMW-glutenins were found in children while these are much less frequent in adults [12,25]. Moreover, in a recent study a highly antigenic α-gliadin peptide was described [25] that is identical to an antigenic hordein-derived peptide reported by us earlier [20]. In essence this hordein/omega peptide is a sequence variant of the DQ2-Glia-x1 peptide and also carries a proline at the p8 position [20]. It is therefore feasible that the toxicity of this peptide can be eliminated by a proline to serine substitution at p8 as well. Preliminary results show that amino acid substitutions similar to those that destroy the T cell stimulatory properties in α-gliadins might also be effective for γ-gliadin derived epitopes (Salentijn et al, in prep), but it is likely that for the LMW- and HMW-glutenins other approaches will be required. This will be the subject of further studies.

In conclusion, we have demonstrated that by utilizing naturally occurring amino acid substitutions the toxicity of the four T cell epitopes in α-gliadins can be eliminated. Such modified proteins
will most likely display indistinguishable technological properties. Thus, our results provide a rational approach to eliminate CD related toxicity from \( \alpha \)-gliadins, which represents a first but crucial step towards the realization of safe gluten containing food products for CD patients.

Materials and Methods

Analysis of \( \alpha \)-gliadin transcripts from diploid wheat varieties

_**T. monococcum**_ accessions CGN10500, CGN2035 and CGN10555 (CGN, Wageningen, The Netherlands) were used for cloning and sequencing of \( \alpha \)-gliadin transcripts. Plants were grown under greenhouse conditions. Developing green kernels of single plants were harvested and used for RNA isolation according to Doyle and Doyle [29] but with 1% (w/v) poly(vinylpyrrolidone)-10 in the extraction buffer. For the production of first strand cDNA 1 \( \mu \)g of total RNA was treated with DNase I (Invitrogen, amplification grade; 18068-015) followed by RT PCR (Invitrogen SuperScript III First-Strand Synthesis System for RT-PCR; 18080-051) using random hexamer primers in a final reaction volume of 20 \( \mu \)l. Primers specific for \( \alpha \)-gliadin genes, located on the conserved sequences at the 5' and 3' end of the coding region of the \( \alpha \)-gliadin, were used to amplify \( \alpha \)-gliadin transcripts from the cDNA samples (\( \alpha F1: 5'\text{agtgaaraCmmtteYcate and} \alpha 5'R: 5'\text{gttagtacngaNgatg} \)). The PCR conditions were: 5 min. at 94°C, 30 cycles (94°C for 1 min., 49°C for 1 min. and 72°C for 2 min), 72°C for 10 min, 25 \( \mu \)l reaction volume. The PCR products were cloned and sequenced.

Characterization of expressed \( \alpha \)-gliadin sequences

Over 3200 _T. aestivum_ expressed sequence tags (ESTs) and mRNAs designated as \( \alpha \)-gliadin or \( \alpha/\beta \)-gliadin were downloaded from the NCBI UniGene library (Ta. 152968, Ta. 23792, Ta. 24084, Ta. 25210, Ta. 27702, Ta. 28482) (http://www.ncbi.nlm.nih.gov/UniGene) on 13 April 2007. The sequences were from _T. aestivum_ libraries of various tissues, treatments and cultivars (Chinese Spring 34.2%; Gleneal 20.8%; Cheyenne 10.4%; Recital 7.8%; Mercia 7.2%; unknown cultivar 8.9%; Wyna 3.2%; HILine 2.9%; Butte 86, 2.4%; Hartog 1.5%; Soleil 0.6%; Nostar 0.1%; Novobirskaya 67, 0.1%). The DNA sequences were aligned using the SeqMan II (DNAstar) and first assembled at a minimum number of sites = 750, in MEGA 4)(Fig. S1).

Sequence variation in epitope regions

To analyze all sequence variation, the 55 \( \alpha \)-gliadin EST contigs, now assigned to a specific chromosome, were reassembled one by one at 99–100% match (SeqMan II, Lasergene, DNAStar). This yielded 717 different allelic variants. The consensus nucleotide sequences were translated (MEGA 4) and explored for epitopes and surrounding sequence regions using a text explorer program (PatternResearch, in house developed) after which the output file was analysed in Excel.

T cell clones, T cell proliferation and HLA-DQ2 binding assays

Gluten specific T cell clones were generated from small intestinal biopsies of celiac disease patients as described before [8,11,12]. All patients signed an informed consent form which was approved by the hospital ethics committee. Proliferation assays were performed in triplicate in 150 \( \mu \)l Iscove’s Modified Dulbecco’s Medium (Bio Whittaker, Verviers, Belgium) with 10% pooled normal human serum in 96 well flat-bottom plates using 2 \( \times \)104 gluten specific T cells stimulated with 105 irradiated HLA-DQ2-matched allogeneric peripheral blood mononuclear cells (3000 rad) in the presence or absence of antigen (1–10 \( \mu \)g/ml) [8,11,12]. After 2 days \(^3\)H-thymidine (0.5 \( \mu \)Ci/well) was added to the cultures, and 18–20 hours thereafter the cells were harvested. \(^3\)H-thymidine incorporation in the T cell DNA was determined with a liquid scintillation counter (1205 Betaplate Liquid Scintillation Counter, LKB Instruments, Gaithersburg, MD). A binding assay was performed as described previously [24].

Supporting Information

**Figure S1 Phylogenetic analysis of \( \alpha \)-gliadin sequences.** A neighbor-joining tree was made with 55 EST consensus nucleotide sequences from hexaploid bread wheat together with 56 genomic DNA sequences derived from the diploid wheat species _T. monococcum_ (A genome, green dots), _T. spelta_ (S/B genome, blue dots) and _Aegilops tauschii_ (D genome, red dots) after alignment using Clustal W. The EST sequences (black dots) can be assigned to their loci in hexaploid wheat as they cluster into the same three groups as the sequences from the diploid species A (genome = locus Gli-A2, S/B genome = locus Gli-B2, D genome = locus Gli-D2 [17]).

**Figure S2 Sequence variation in the N-terminal and C-terminal regions of \( \alpha \)-gliadin genes from hexaploid wheat.** The 23 most frequently found expressed sequence tag (EST) contigs were translated, the amino acid sequences were aligned and grouped per chromosomal location (the Gli-A2 locus on chromosome 6AS, Gli-B2 on 6BS, and Gli-D2 on 6DS) to present the variation in various gluten epitope regions. Note the large differences in the number of times (N) each sequence was present in the set of ESTs. In red: amino acid variation in the sequence. In back: chymotrypsin or trypsin sites (>72% affinity).
Figure S3 DQ2-Glia-a1 variants. The effect of various amino acid substitutions on the ability of the DQ2-Glia-a1 peptide to stimulate DQ2-Glia-a1 T cells.

(TIF)

Table S1 Amino acid sequences of the DQ2-Glia-a1/-a2/-a3 region of α-gliadin cDNA transcripts from three diploid T. monococcum accessions (diploid, AA genome).
The gene sequences have been submitted as accession numbers HQ317881-HQ317890.

(DOC)

References
1. Rewers M (2005) Epidemiology of celiac disease: what are the prevalence, incidence, and progression of celiac disease. Gastroenterology 128: 847–851.
2. Koning F (2005) Celiac Disease: caught between a rock and a hard place. Gastroenterology 129: 1294–1301.
3. Solidi LM (2002) Celiac disease: dissecting a complex inflammatory disorder. Nat Rev Immunol 2: 647–655.
4. Stepniak D, Koning F (2006) Celiac Disease: sandwiched between innate and adaptive immunity. Immunol Rev 117: 41–49.
5. Meresse B, Chen Z, Ciszewski C, Tretiakova M, Bhagat G, et al. (2004) Coordinated induction by IL15 of a TCR-independent NKG2D signaling pathway converts CTL into lymphokine-activated killer cells in celiac disease. Immunity 21: 357–366.
6. Huë S, Mention J-J, Monteiro RC, Zhang SL, Cellier C, et al. (2004) A direct role for NKG2D/MICA interaction in villous atrophy during celiac disease. Immunity 21: 367–377.
7. van de Wal Y, Kooy Y, van Veelen P, Pena S, Mearin L, et al. (1998) Small intestinal cells of celiac disease patients recognize a natural pepsin fragment of gliadin. Proc Natl Acad Sci USA 95: 10050–10054.
8. Arentz-Hansen H, Korner R, Mollerberg O, Quasten H, Vader W, et al. (2000) The intestinal T cell response to α-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase. J Exp Med 191: 603–612.
9. Anderson RP, Domano P, Godkin AJ, Jewell DP, Hill AV (2000) In vivo antigen challenge in celiac disease identifies a single transglutaminase-modified peptide as the dominant α-gliadin T cell epitope. Nat Med 6: 337–342.
10. van de Wal Y, Kooy YM, van veelen P, Vader W, August SA, et al. (2000) Glutenin is involved in the gluten-driven mucosal T cell response. Eur J Immunol 30: 3133–3139.
11. Vader W, Kooy Y, van veelen P, de ru A, Harris D, Benckhuijsen W, et al. (2002) The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. Gastroenterology 122: 1729–1737.
12. Shan L, Mollerberg O, Parrot I, Hausch F, Filiz F, et al. (2002) Structural basis for gluten intolerance in celiac sprue. Science 297: 2275–2279.
13. Qiao S-W, Bergsgt E, Mollerberg O, Jung G, Fleckenstein B, et al. (2005) Refining the rules of gliadin T cell epitope binding to the disease-associated DQ2 molecule in celiac disease: importance of proline spacing and glutamine deamidation. J Immunol 175: 254–261.
14. Okita TW, Chessbrov H, Reeves CD (1985) Evolution and heterogeneity of the α/β-type and γ-type gliadin DNA sequences. J Biol Chem 260: 8203–8213.
15. Anderson OD, Linl JC, Greene FC (1997) The α-gliadin gene family I. Characterization of ten new wheat α-gliadin genomic clones, evidence for limited sequence conservation of flanking DNA, and Southern analysis of the gene family. Theor Appl Genet 95: 56–58.
16. van Herpen TWJM, Goryunova SV, van der Schoot J, Mitreve M, Salenijn EMJ, et al. (2006) Alpha-gliadin genes from the A, B, and D genomes of wheat contain different sets of celiac disease epitopes. BMC Genomics 7: 1.
17. Molberg O, Uhlen AK, Jensen T, Farte NS, Fleckenstein B, et al. (2005) Mapping of gluten T-cell epitopes in the bread wheat ancestors: implications for celiac disease. Gastroenterology 128: 393–401.
18. Spaenij-Dekking I, Kooy-Winkel A, van veelen P, Drijfhout JW, Jonker H, et al. (2005) Natural variation in toxicity of wheat: potential for selection of nontoxic varieties for celiac disease patients. Gastroenterology 129: 797–806.
19. Vader LW, Stepniak DT, Ronnik EM, Kooy YM, De Haan W, et al. (2003) Characterization of cereal toxicity for celiac disease patients based on protein homology in grains. Gastroenterology 123: 1105–1113.
20. Vader LW, de ru A, van de Wal Y, Kooy YM, Benckhuijsen W, et al. (2002) Specificity of tissue transglutaminase explains cereal toxicity in celiac disease. J Exp Med 193: 643–649.
21. van den Broeck HC, de Jong HC, Salenijn EMJ, Dekking L, Bosch D, et al. (2010) Presence of celiac disease epitopes in modern and old hexaploid wheat varieties. Wheat breeding may have contributed to increased prevalence of celiac disease. Theor Appl Genet 121: 1527–1539.
22. Van den Broeck HC, Chen H, Lacoze X, Deautour J-C, Gilissen L, et al. (2010) In search of tetraploid wheat accessions reduced in celiac disease-related gluten epitopes. Mol Biol Systems 6: 2206–2213.
23. Stepniak D, Wiesner M, de ru AH, Moustanakis AK, Drijfhout JW, et al. (2008) Large-scale characterization of natural ligands explains the unique gluten-binding properties of HLA-DQβ2. J Immunol 180: 3268–3278.
24. Tye-Din JA, Stewart JA, Dromey JA, Beisharh T, van Heel DA, et al. (2010) Comprehensive, quantitative mapping of T cell epitopes in gluten in celiac disease. Sci Transl Med 2: 64ra51.
25. Jin L, Hsiu A-P, Schnable PS (2007) Recent advances in plant recombination. Curr Opin Plant Biol 10: 131–135.
26. Lloyd A, Plaisier CL, Carroll D, Drews GN (2005) Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. Proc Natl Acad Sci USA 102: 2232–2237.
27. de Pater S, Neuteboom LW, Pina JE, Hooykaas PJ, van der Zaal BJ (2009) ZFN-induced mutagenesis and gene-targeting in Arabidopsis through Agrobacterium-mediated floral dip transformation. Plant Biotechnol J 7: 821–835.
28. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure from small quantities of fresh leaf tissues. Phytochem Bull 19: 11–15.
29. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596–1599.
30. Kagnoff MF (2007) Celiac disease: pathogenesis of a model immunogenetic disorder. Gastroenterology 132: 1789–1799.
31. Koning F (2005) Celiac Disease: caught between a rock and a hard place. Gastroenterology 129: 1294–1301.