Discovery of a number of novel and known human genes whose protein products bear striking similarity to two or more wheat gliadin domains raised the possibility that human intestinal non-HLA peptides homologous to celiac T-cell epitopes could play a role in non-HLA gene specification in celiac disease. Database searching of the entire human genome identified only 11 gut-expressed proteins with high T-cell epitope homology, particularly to the DQ2-α-I-gliadin epitope (i.e. TFIIA, FOXJ2 and IgD; mean BestFit quality score = 40 versus random value of 24). Others were similar to DQ2-α-I-gliadin (i.e. PAX9; BestFit quality 46 versus 20 for random), or DQ2-α-II-gliadin (PHLDA1, known in mice as the T-cell death-associated gene; BestFit quality 43 versus 30 for random) epitopes. Among proteins previously screened for gliadin homology, noteworthy was achaete scute homologous protein (DQ2-α-I-gliadin; BestFit quality 41 versus 22 for random). With the exception of IgD, all are nuclear factors. Paying particular attention to the position of potential major histocompatibility complex (MHC) anchor residues, several were selected for testing in a DQ2-γ-I-gliadin-restricted T-cell system. All native 10-mer peptides were inactive, even when deamidated, but V96F substitution of deamidated TFIIA amino acid residues 91–100 stimulated IL-2 release at levels exceeding the wheat gliadin positive control. Also active, but only slightly, was L1009F substitution of AIB3 amino acid residues 1004–1013. PlotSimilarity alignment of TFIIA from eight species revealed sub-threshold similarity score in the peptide region, in contrast to the highly conserved amino and carboxy termini. Molecular modeling of TFIIA[V96F] peptide points to an important juxtaposition of an upwardly projecting phenylalanine residue at peptide position 6 that likely contacts a receptor complementarity-determining region, and a downwardly projecting glutamic acid residue that fits into the shallow MHC P7 pocket. These observations tentatively point to a new multi-gene hypothesis for the initiation of celiac disease in which deamidated free human peptides with T-cell epitope homology (particularly those made more homologous by mutation) escape negative selection, as per deamidation of the HEL48 – 62 peptide in the hen egg lysozyme model of autoimmunity. Deamidation following peptide release due to injury triggers inflammation, thereafter repeatedly provoked by dietary gliadin immunodominant peptides concentrated in the proximal small intestine.

Celiac disease is a strongly heritable immune disease affecting at least 0.2–0.5% of the population,1 presenting classically in childhood, although it is now understood that the disease may be diagnosed at any age. Characterized by
hypersensitivity to dietary wheat gliadins, patients suffer from acute loss of proximal intestinal absorptive capacity as small intestinal villi atrophy and crypts expand leading to diarrhea and pediatric failure to thrive. Immune involvement is indicated by high serum antibody titer to gliadin and tissue transglutaminase, gliadin-dependent activation of expanded intraepithelial and submucosal T-cell populations, and by the very high prevalence of DQ2 (≥90%) among affected individuals. A smaller fraction are DQ8 positive. Despite strong linkage to the HLA locus on chromosome 6, no mutated HLA haplotypes have been detected and most DQ or DQ8 individuals in the general population are normal. Recent estimates attributing more than 39% of the disease risk to unknown genes outside the MHC region lend support to the suggestion that celiac disease is a complex genetic disease initiated by non-HLA genes, as well as sequential human genomic genes, as well as sequential human genomic

<table>
<thead>
<tr>
<th>Name/symbol</th>
<th>Access no.</th>
<th>Function</th>
<th>Intra-cell</th>
<th>Epitope homology</th>
<th>BestFit score</th>
<th>Express</th>
<th>Native sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAGH26</td>
<td>U80739</td>
<td>?</td>
<td>Yes</td>
<td>DQ2-γ-1-gliadin</td>
<td>41 ± 27</td>
<td>n</td>
<td>PQQQQQQQ</td>
</tr>
<tr>
<td>FL20259</td>
<td>AK000266</td>
<td>?</td>
<td>Yes</td>
<td>DQ2-γ-1-gliadin</td>
<td>39 ± 22</td>
<td>st, c</td>
<td>PQQQQQQQ</td>
</tr>
<tr>
<td>FOXJ2</td>
<td>AF135132</td>
<td>Transcr factor</td>
<td>Yes</td>
<td>DQ2-γ-1-gliadin</td>
<td>41 ± 23</td>
<td>c, n</td>
<td>PQQQQQQQ</td>
</tr>
<tr>
<td>IGHD</td>
<td>K02882</td>
<td>Antibody</td>
<td>No</td>
<td>DQ2-α-I-gliadin</td>
<td>41 ± 20</td>
<td>c, c</td>
<td>PQQRTFRIQ</td>
</tr>
<tr>
<td>NRL</td>
<td>M81840</td>
<td>Leucine zip</td>
<td>Yes</td>
<td>DQ8-I-gluten</td>
<td>41 ± 20</td>
<td>st</td>
<td>HGYGPSPEET</td>
</tr>
<tr>
<td>PAX9</td>
<td>P55771</td>
<td>Transcr factor</td>
<td>Yes</td>
<td>DQ2-γ-1-gliadin</td>
<td>46 ± 20</td>
<td>st</td>
<td>QPTQFPALPY</td>
</tr>
<tr>
<td>SSXT</td>
<td>X79201</td>
<td>Translocation</td>
<td>Yes</td>
<td>DQ2-γ-1-gliadin</td>
<td>37 ± 30</td>
<td>st, c</td>
<td>PQQGYPQQQ</td>
</tr>
<tr>
<td>TFIIA or GTF2A1</td>
<td>P52655</td>
<td>Transcr factor</td>
<td>Yes</td>
<td>DQ2-γ-1-gliadin</td>
<td>39 ± 23</td>
<td>c, ct, n</td>
<td>PQQQTVQQQA</td>
</tr>
</tbody>
</table>

M95929, whose BestFit score did not differ from random, was excluded.

Suggested by Unigene and annotation.

Presence/absence of an endoplasmic signal peptide determined by SignalP and/or SPScan.

Given is quality score versus (n) value ± SD obtained when the probing 10 to 14-mer epitope sequence was randomised ten times. BestFit default settings were utilized.

Likely expression suggested by Unigene with occasional contribution from hybridization data. Stomach (st), small intestine (si), colon (c), connective tissue (ct) and/or nerve fibers (n); or no apparent gastrointestinal expression (–). The actual expression may be more extensive.

Homologous sequence in native human protein. Bold type indicates amino acid residues that differ from non-deamidated DQ2-γ-1-gliad.

<table>
<thead>
<tr>
<th>Name/symbol</th>
<th>Access no.</th>
<th>Function</th>
<th>Intra-cell</th>
<th>Epitope homology</th>
<th>BestFit score</th>
<th>Express</th>
<th>Native sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAGH26</td>
<td>U80739</td>
<td>?</td>
<td>Yes</td>
<td>DQ2-γ-1-gliadin</td>
<td>41 ± 27</td>
<td>n</td>
<td>PQQQQQQQ</td>
</tr>
<tr>
<td>FL20259</td>
<td>AK000266</td>
<td>?</td>
<td>Yes</td>
<td>DQ2-γ-1-gliadin</td>
<td>39 ± 22</td>
<td>st, c</td>
<td>PQQQQQQQ</td>
</tr>
<tr>
<td>FOXJ2</td>
<td>AF135132</td>
<td>Transcr factor</td>
<td>Yes</td>
<td>DQ2-γ-1-gliadin</td>
<td>41 ± 23</td>
<td>c, n</td>
<td>PQQQQQQQ</td>
</tr>
<tr>
<td>IGHD</td>
<td>K02882</td>
<td>Antibody</td>
<td>No</td>
<td>DQ2-α-I-gliadin</td>
<td>41 ± 20</td>
<td>c, c</td>
<td>PQQRTFRIQ</td>
</tr>
<tr>
<td>NRL</td>
<td>M81840</td>
<td>Leucine zip</td>
<td>Yes</td>
<td>DQ8-I-gluten</td>
<td>41 ± 20</td>
<td>st</td>
<td>HGYGPSPEET</td>
</tr>
<tr>
<td>PAX9</td>
<td>P55771</td>
<td>Transcr factor</td>
<td>Yes</td>
<td>DQ2-γ-1-gliadin</td>
<td>46 ± 20</td>
<td>st</td>
<td>QPTQFPALPY</td>
</tr>
<tr>
<td>SSXT</td>
<td>X79201</td>
<td>Translocation</td>
<td>Yes</td>
<td>DQ2-γ-1-gliadin</td>
<td>37 ± 30</td>
<td>st, c</td>
<td>PQQGYPQQQ</td>
</tr>
<tr>
<td>TFIIA or GTF2A1</td>
<td>P52655</td>
<td>Transcr factor</td>
<td>Yes</td>
<td>DQ2-γ-1-gliadin</td>
<td>39 ± 23</td>
<td>c, ct, n</td>
<td>PQQQTVQQQA</td>
</tr>
</tbody>
</table>

Table 1. Human proteins identified in FASTA searches (E ≤ 1.0) of the primate database using celiac T-cell epitopes DQ2-γ-1-gliadin, DQ2-α-II-gliadin, DQ2-α-I-gliadin, DQ8-α-I-gliadin and DQ8-I-glutenin as probes

Human T-cell-like Epitopes
Table 2. Human proteins identified in TFASTX (“TFTX”) searches (E ≤ 1.0) of the primate database with probing full length α-gliadin sequence and via library screening using full length α-gliadin or α-gliadin oligonucleotide as probe

<table>
<thead>
<tr>
<th>Name/symbol</th>
<th>Access no.</th>
<th>Meth. Id.</th>
<th>Function*</th>
<th>Intra-cell?b</th>
<th>Epitope homology</th>
<th>BestFit scorec</th>
<th>Expressd</th>
<th>Native sequencee</th>
</tr>
</thead>
<tbody>
<tr>
<td>'AIB3</td>
<td>D80003</td>
<td>TFTX</td>
<td>Nucl. coat</td>
<td>Yes</td>
<td>DQ2-α-II-gliadin</td>
<td>38 ± 3 2</td>
<td>st/c</td>
<td>POPQPLPOQQQ</td>
</tr>
<tr>
<td>ASCL1</td>
<td>L08424</td>
<td>Oligo</td>
<td>Transcr.</td>
<td>Yes</td>
<td>DQ2-α-II-gliadin</td>
<td>35 ± 2 2</td>
<td>n!</td>
<td>POPQPLFLP9</td>
</tr>
<tr>
<td>KIAA1048</td>
<td>AB028971</td>
<td>TFTX</td>
<td>?</td>
<td>Yes</td>
<td>DQ2-α-II-gliadin</td>
<td>41 ± 2 4</td>
<td>c</td>
<td>POPQPOQPF</td>
</tr>
<tr>
<td>PHLD1</td>
<td>ZZ5194</td>
<td>TFTX</td>
<td>Apoptosis</td>
<td>Yes</td>
<td>DQ2-α-II-gliadin</td>
<td>43 ± 3 4</td>
<td>st/c</td>
<td>POPQOLQPQPQ</td>
</tr>
<tr>
<td>SMARCF1</td>
<td>AB001895</td>
<td>TFT</td>
<td>Chrom. reg.</td>
<td>Yes</td>
<td>DQ2-α-II-gliadin</td>
<td>38 ± 3 6</td>
<td>st/c</td>
<td>PQSOPPYSQP</td>
</tr>
</tbody>
</table>

Excluded were AF010404, AF238868, U91316, U91935 and AJ251016 whose BestFit scores did not differ from random.

* Suggested by Unigene and annotation.

b Presence/absence of an endoplasmic signal peptide determined by SignalP and/or SPScan.

c Given is quality score versus (c) value ± SD obtained when the probing 10 to 14-mer epitope sequence was randomised ten times. BestFit default settings utilized.

d Expression suggested by Unigene with occasional contribution from hybridization data. Stomach (st), small intestine (si), colon (c), connective tissue (ct) and/or nerve fibers (n); or no apparent gastrointestinal expression (–). Actual expression may be more extensive.

e Homologous sequence in native human protein. Bold indicates amino acids that differ from non-deamidated DQ2-γ-I-gliadin (QPQQSFQPSQQQ) or DQ2-α-II-gliadin (QPFPQPQLPY). BestFit and Gap (AF010404, not shown; BestFit quality ¼ 155027) were again prevalent (Table 2) and included the apoptosis-associated protein PHLD1 whose mouse homologue derives from the T-cell death-associated gene. Sequences approximating the DQ2-α-II-gliadin epitope (55–90% identity; mean 69±12%) were most common, with PHLD1 (90% identical) differing in only two residues. In BestFit and Gap analyses, all but two (retina-derived POU-domain factor-1 [U91935] and calcium-activated potassium channel SK3 [A251016]; quality = 28 versus 26 ± 3; quality = 23 versus 23 ± 4, respectively) differed from random. Expression was apparent in small intestine, colon, stomach, and connective tissue (Table 2).

Search for epitope homologues

We screened the primate database by FASTA for peptide homology to gliadin immunodominant peptides in deamidated [QPQQSFQPSQQQ] or native [QPQQFLP9] or (QPFPQPQLPY, SGQGSFQPSQQN) forms. Searches were performed also using the glutenin immunodominant peptide SGQGSFQPSQQN (DQ8-I-glutenin). At E ≤ 1.0, nine proteins were identified. With the exception of IgD and two proteins of unknown function, all were nuclear factors involved, for the most part, in transcriptional control (Table 1). Although the level of homology varied (36–80% identity; mean 62±13%), linear recapitulation of DQ2-γ-I-gliadin epitope-like sequences predominated. Moderately striking DQ2-α-I-gliadin (80% identity; Pax 9) and DQ8-I-glutenin (55% identity; neural retina leucine zipper [NRL] epitope-like sequences were also apparent. To sort out the glutamine-rich and proline-rich nature of these epitopes, advantage was taken of BestFit (Table 1) and Gap (not shown) analyses in which the alignment quality can be compared quantitatively to the average random quality, as determined by randomizing the probing epitope sequence multiple times. All differed from random, with the exception of homeobox protein M95929 (not shown; BestFit quality = 35 versus random quality of 31 ± 5). Unigene provided an approximation of organ distribution. All except CAGH26 and neural retinal-specific leucine zipper protein (NRL) appear to be expressed in the gastrointestinal tract in small intestine, colon, stomach, connective tissue and/or nerve fibers (Table 1).

Six more human proteins were identified in repeated TFASTX screens (E ≤ 1.0) using the complete α-gliadin protein sequence (M10092) as probe. Each was subsequently scrutinized for epitope homology by BestFit and Gap. Nuclear factors were again prevalent (Table 2) and included the apoptosis-associated protein PHLD1, whose mouse homologue derives from the T-cell death-associated gene. Sequences approximating the DQ2-α-II-gliadin epitope (55–90% identity; mean 69±12%) were most common, with PHLD1 (90% identical) differing in only two residues. In BestFit and Gap analyses, all but two (retina-derived POU-domain factor-1 [U91935] and calcium-activated potassium channel SK3 [A251016]; quality = 28 versus 26 ± 3; quality = 23 versus 23 ± 4, respectively) differed from random. Expression was apparent in small intestine, colon, stomach, and connective tissue (Table 2).

Similarly analyzed were four proteins (40–70% epitope identity; mean 50±14% identified in α-gliadin oligonucleotide or cDNA screens of human cDNA and genomic libraries. Lacking significant T-cell epitope BestFit scores were: (i) ALR (AF010404, not shown; BestFit quality = 37 versus random quality of 36 ± 3) which displays strong regional gliadin-like homology at the nucleic acid and protein level (BestFit quality = 179 versus random quality of 77 ± 10); (ii) the Genescan-predicted protein from gliadin-like genomic clone 2 (GLSH-1; AF238868; BestFit quality = 23 versus 19 ± 4, and 16 versus 16 ± 3, both compared to DQ2-α-II-gliadin), and (iii) acyl CoA hydrolase (U91316; BestFit quality = 18 versus 15 ± 5). In contrast, the transcription factor achaete scute homologous protein (ASCL1; L08424) approximates both DQ2-α-I- and DQ2-α-II-gliadin epitopes (70% and 60% identical, respectively)
and, when removed from the germline, severely affects the development of small intestinal ganglia. Relevance of the latter to celiac disease is, however, considered unlikely.

**Activation of celiac T-cells by TFIIA peptide**

To test whether linear epitope homologues may cause disease (Figure 1), we took advantage of a well-characterized DQ2-gliadin-restricted T-cell system derived from intestinal T-cells of celiac patients by transfection. This system consisted of: (i) gliadin-responsive (acid-treated SWP50 gliadin) clonal cells generated by stably transfecting the Va and Vb gene segments from the DQ2-restricted, gliadin-specific and human gut-derived T-cell clone 4.32 into murine T-cell hybridoma BW 58a-/b- cells, and (ii) EBV-transformed B-lymphoblastoid cells expressing MHC chains DQa1p0501 and DQb1p0201 as antigen presenters. IL2-release was measured using time-resolved fluorometry (Delfia reagents; Wallac, Finland) 18 hours later.

![Figure 1](image-url) Test of human peptides homologous to the gliadin DQ2-γ-I T-cell epitope in a celiac-specific T-cell activation assay. Peptides were identified in FASTA searches of the primate database and synthesized with the third last glutamine residue as glutamic acid (arrow), as per the sensitivity of this glutamine in epitope DQ2-γ-I to deamidation. Several peptides were prepared with a single amino acid substitution. T-cell activation by DQ2-γ-I and TFIIA[V96F] was robust, in contrast to slight AIB3[L1009F] activity and a completely negative response from all native human peptides. TFIIA[V96F] comprises residues 91–100 of general transcription factor II A (PS2655), whereas AIB3[L1009F] derives from residues 1004–1013 of transcriptional coactivator AIB3 (AF177388) also known as nuclear receptor coactivator RAP250, peroxisome proliferator-activated receptor interacting protein or thyroid hormone receptor-binding protein. Peptides were synthesized, purified by HPLC, checked by mass spectrometry, and preincubated (five, ten or 20 µm) overnight with EBV-transformed B-lymphoblastoid cells (5 x 10⁴/well) expressing MHC chains DQa1p0501 and DQb1p0201, as described. Antigen presenters were then added to gliadin-responsive clonal cells (2 x 10⁴/well) of a previously characterized line derived by stably transfecting the Va and Vb gene segments from the DQ2-restricted, gliadin-specific and human gut-derived T-cell clone 4.32 into murine T-cell hybridoma BW 58a-/b- cells. IL2-release was measured using time-resolved fluorometry (Delfia reagents; Wallac, Finland) 18 hours later.
DQ2-γ-1-gliadin homology (as per the test cell system), BestFit score, codon usage of residues differing from DQ2-γ-1-gliadin, and the relative position of potential anchor residues at P1 (“P”), P4 (“S”), P6 (“P”), P7 (“E”) and P9 (“Q”) that together mediate DQ2-γ-1-gliadin binding to MHC chains. Attention was also paid to likely in vivo deamination by tissue transglutaminase following extracellular release as a consequence of infection or tissue damage. Accordingly each selected 10-mer was synthesized with glutamic acid rather than glutamine at position 8, the site of tissue transglutaminase-dependent deamination in the DQ2-γ-1-gliadin epitope. Finally, several synthesized with a single amino acid substitution explored the possibility that gene mutation may be activating. Thus, serine was substituted for threonine (BLOSSUM 62 [BL] value 26) at positions 52 or 95 of IGHD and TFIIA, respectively, and for glutamine (BL value 0) at position 1008 of AIB3. Both were chosen to engage the MHC P4 pocket more readily, as confirmed later in modeling studies (see Figure 3). Glutamine was substituted for proline (BL value −1) at position 1007 of AIB3 to relieve proline-related structural constraints. Finally, phenylalanine was substituted for valine (BL value −1) or leucine (BL value 0) at positions 96 or 1009 of TFIIA or AIB3, respectively, thereby presenting a benzene ring to the T-cell receptor surface (see Figure 3) as per the immunodominant DQ2-γ-1-gliadin epitope.

No native human peptides stimulated IL-2 release (Figure 1). However, substitution of phenylalanine for valine-96 in TFIIA was activating at levels greatly exceeding the DQ2-γ-1-gliadin positive control. Slightly but reproducibly stimulatory at 20 μM was AIB3[L1009F], notably with the same non-polar hydrophobic to bulky hydrophobic alteration. No other substitution was activating, including IGHD[T52S], TFIIA[T95S], AIB3[P1007Q] or AIB3 [Q1008S], each of which was chosen to generate structural changes aimed at improving MHC and T-cell receptor engagement. The inactivity of IGHD draws attention to the apparent negative influence of a charged residue at position 3 (arginine) and/or a hydrophobic at position 8 (isoleucine), despite residues appropriate for almost (IGHD) or all (IGHD[T52S]) MHC pockets. Not inhibitory was position 5 threonine (TFIIA) in place of serine (DQ2-γ-1-gliadin), in keeping with the conservative nature of this alteration. Prior DQ2-γ-1-gliadin substitution studies15,18,27–29 have been of a limited nature.

Should inherited substitution underlie disease initiation, compatibility with normal protein function would be expected for maintenance in the germline. TFIIA, for example, is an important transcription initiator of class II genes. Functional compatibility was considered by PlotSimilarity alignment of TFIIAs from eight different species including Homo sapiens, Rattus norvegicus, Mus musculus, Drosophila melanaster, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Arabidopsis thaliana (Figure 2). Although conservation between human and rat or mouse is remarkably high (rat, 99% identity over 377 amino acid residues; mouse, 97% over 378 amino acid residues), quantitative alignment of all eight TFIIAs reveals low similarity in the DQ2-γ-1-gliadin homologous region (Figure 2; asterisk), suggesting that mutation in this region may be inconsequential. Indeed, deletion studies in yeast31 point to amino acid residues 55–240 as simply linking and spatially separating the conserved amino and carboxy-terminal domains in a sequence-independent manner. Suggested cross-species maintenance of charge35 is not apparent upon examination by Isoelectric with linkage region pI values ranging from 4.14 (S. cerevisiae) to 7.8 (C. elegans); and H. sapiens, R. norvegicus, M. musculus all at 7.09. Currently, 19 single nucleotide polymorphisms have been discovered in the human TFIIA gene (NCBI) from analyses of a mean of 22 chromosomes. Of these, 17 are in introns, and two are in the proximal upstream sequence.

**Molecular modeling of peptides in HLA-DQ2**

Molecular modeling was applied to TFIIA[V96F] and AIB3[L1009F] (Figure 3) to glean structural clues underlying differences in activity. Lacking the HLA-DQ2 crystal structure, we modeled DQα1*0501, DQβ1*0201 and peptides into the
Homology between MHC extracellular domains is moderately high (65% similarity/56% identity and 74% similarity/69% identity, for DQα1'0501 and DRβ1'0101). Thus required amino acid substitutions were not extensive. Orientation of TFIIA[V96F] in the HLA-DQ2 binding groove (Figure 3(a)) overlaps perfectly with DQ2-γ1-gliadin (Figure 3(a) and (b)). HLA-DQ2 pockets are filled with downwardly projecting proline (P1) and threonine/serine (TFIIA/DQ2-γ1; P4), a shallow proline (P6) and a downwardly protruding alanine/glutamine (TFIIA/DQ2-γ1; P9). Critical glutamic acid at position 7 is more superficially arranged. Jutting upward or somewhat sideways, are glutamine (P-1, P2, P8) and phenylalanine (P5), all of which are potentially available for contact with the T-cell receptor. The less active AIB3[L1009F] peptide differs in the length of the polar side-chain carbon backbone at P4 (Q/S), the lack of an upwardly projecting glutamine residue at position 3, and in the orientation of both the phenylalanine ring (position 5) and position 8 glutamine. Similar subtle differences explain the lack of activity of the other peptides. Glutamine substituted for serine (FLJ20259 and AIB3) may not fit into the P4 pocket, unlike a threonine for serine substitution (IGHD, TFIIA); and IGHD’s position 3 arginine residue is predicted (not shown) to project to the T-cell surface in a manner that could sterically inhibit position 5 phenylalanine. A planar position 5 phenylalanine residue appears to be particularly important, and cannot be substituted by the polar amino acid residue threonine (FLJ20259) or the smaller hydrophobic amino acid residues valine (TFIIA) or leucine (AIB3). As noted previously, glutamic acid in the P7 pocket is essential. 2 An alanine for glutamine substitution (TFIIA) has no apparent effect. Molecular modeling studies using FLEXIDOCK (Sybyl; Tripos Inc., St. Louis MO) suggest seven putative hydrogen bonds that may help immobilize TFIIA[V96F] and DQ2-γ1-gliadin peptides in the HLA-DQ2 binding groove. Noteworthy is the P2 glutamine residue, which is associated with hydrogen bonds from the protein backbone nitrogen atom to the side-chain oxygen atom of DQβ1'0201 asparagine residue 82 [1FYT B chain]. Also predicted is a hydrogen bond from the protein backbone oxygen atom to a side-chain nitrogen atom of the same DQβ1'0201 asparagine residue 82; and a third from the side-chain nitrogen atom to the protein backbone oxygen atom of DQβ1'0201 arginine residue 77. Single hydrogen bonds are associated with P3 glutamine (from the side-chain nitrogen atom to the protein backbone oxygen atom of DQα1'0501 phenylalanine residue 58), P7 glutamic acid (from the protein backbone oxygen atom to the side-chain nitrogen atom of

Figure 3. Conformation models of the TFIIA[V96F], gliadin DQ2-γ1 and AIB3[L1009F] peptides. (a) Overlapping TFIIA[V96F] and DQ2-γ1 peptides (white/red/blue) bound to the DQα1'0501/DQβ1'0201 heterodimer (green). The two peptides differ only at position 5 (threonine [TFIIA] versus serine [DQ2-γ1]; “T/S”). HLA-DQ2 binding pockets at positions 1 (“P1”), 4 (“P4”), 6 (“P6”), 7 (“P7”) and 9 (“P9”) are occupied with proline, threonine/serine, proline, glutamic acid and alanine/glutamine, respectively. Noted is the planar orientation of phenylalanine at position 6 (+), deep to which lies the P4 pocket in this angle view. Only HLA-DQ2 residues within 6 Å of the peptides are shown. (b) Side-view stick model of overlapping TFIIA[V96F] and gliadin DQ2-γ1 peptides. Readily visualized are amino acid residues projecting down to HLA-DQ2 pockets P1, P4, P6, P7 and P9, and up to the planar T-cell receptor. (c) Similar overlap of AIB3[L1009F] and DQ2-γ1 peptides. The phenylalanine residue in AIB3 is oriented more vertically, and the glutamine residue at position 8 is shifted slightly. Models were constructed in Sybyl Composer from the recently published crystal structure of complexed HLA-DR1/αβ T-cell receptor/hemagglutinin (1FYT).39 Homology between MHC extracellular domains is moderately high (65% similarity/56% identity and 74% similarity/69% identity, respectively for DQα1'0501 and DRβ1'0101; and DQβ1'0201 and DRβ1'0101), thus required amino acid substitutions were not extensive. All substitutions were verified by Gap.
DQα1’0501 asparagine residue 69), P8 glutamine (from the protein backbone oxygen atom to the side-chain nitrogen atom of DQβ1’0201 tryptophan residue 61), and P9 alanine (from the protein backbone nitrogen atom to the side-chain oxygen atom of DQα1’0501 asparagine residue 69).

Potential multi-gene susceptibility

Building on the hypothesis that DQ2-γ-I-gliadin-, DQ2-α-II-gliadin- or DQ2-α-I-gliadin-homologous regions of human non-HLA genes (possibly made more homologous by mutation) could help explain the genetic complexity and geographic variability of celiac disease, a screening strategy was developed. Sequential database searching and in vitro analysis with DQ2-γ-I-gliadin-restricted T-cell transfectants identified a peptide in the human TFIIA non-conserved region that, upon valine to phenylalanine substitution, promoted dose-dependent IL-2 release at levels that exceeded the gliadin positive control. An AIB3 peptide was slightly positive; and several other promising peptides await functional investigation in DQ2-α-II-gliadin and DQ2-α-I-gliadin systems.

Although conformation and charge figure highly in MHC and T-cell receptor binding, advantage could be taken of the unique glutamine-rich and proline-rich nature of gliadin T-cell epitopes that together make up 80% of residues. Moreover, human homologous peptides lend themselves to quantitative BestFit or Gap analyses interpreted in the context of ten randomized alignments of the shuffled test sequence. Chance FASTA (E = 1.0) alignments are thereby eliminated. Emerging from these analyses are short, precisely defined genetic regions that are particularly well suited for PCR-based heteroduplex mutational analysis of patient DNA whose sensitivity extends to single-nucleotide substitutions when coupled with capillary electrophoresis. Such an approach is a subject for future investigation in which positive control templates with substitutions are required for optimization of heteroduplex formation. Aided by efforts to document single nucleotide polymorphisms (SNPs) among different geographical groups, the probability of TFIIA (V96F; using “gta” for valine (SNPs) among different geographical groups, the probability of TFIIA (V96F; using “gta” for valine (SNPs) among different geographical groups, the probability of TFIIA (V96F; using “gta” for valine (SNPs) among different geographical groups, the probability of TFIIA (V96F; using “gta” for valine
general marker of tissue degeneration augmented in Alzheimer’s and Huntington’s disease, and a major autoantigen in celiac disease. Although celiac disease-associated neuropathology has been noted, the relationship remains unclear. Nonetheless, the possibility that TFIIA and other nuclear proteins are non-HLA risk factors in celiac disease is intriguing. Recently, positional cloning identified a novel mutable gene coding for a nuclear transcription factor-like protein as the disease gene in autoimmune disease polyendocrinopathy–candidiasis–ectodermal dystrophy. A similar approach led to the discovery that X-linked autoimmunity-allergic dysregulation syndrome is attributable to mutations in the gene for fork head domain-containing protein JM2. Application of positional cloning to celiac disease does not appear practical. Although familial risk attributable to the HLA locus (approximately 3.6-fold to fivefold) differs greatly from observed (20 to 60-fold), linkage to non-HLA regions is diverse, an observation suggesting the involvement of several non-HLA genes, the identity of which may be dependent on geographical origin. Zhong et al. highlighted 12 potential linkage regions (3q27, 5q33.3, 6p23, 6p21.3, 6p12, 7q31.3, 11p11, 15q26, 19p13.3, 19q13.1, 19q13.4, and 22cen) in a 15 family Irish group; however, two independent replicative attempts based on UK family groups of slightly larger size failed to reproduce these results. One possible exception was a suggested weak association at 15q26. An Italian whole genome analysis involving 105 families revealed a potential risk factor in 5qter, and a second 11qter factor (also seen in a recent Swedish study together possibly with 2qter), but no other associations were apparent. A recent Finnish study displayed ambiguous non-HLA results. Limited attempts to link celiac diseases with known genes such as T-cell receptor delta, TNF 1-3, CD28/CTLA4 were unsuccessful or lacked specificity.

In summary, the availability of the complete human genome is revolutionizing our approach to the genetics of autoimmune disease. Computer-based screening coupled with in vitro testing offers a rapid and efficient first step in the potential identification of human proteins from which deamidated inflammatory peptides may derive. Coupled with new advances in molecular modeling and testing in other celiac T-cell-restricted systems, it is expected that multiple candidate disease peptides will be identified.

Acknowledgments

The authors gratefully appreciate the laboratory of Drs Hans Sjöström and Ove Nörn (The Panum Institute, Copenhagen, Denmark) for T-cell assays. This work was supported by NIH grant EY09747 (to G.W.L.).

References

25. Lundin, K. E., Scott, H., Hansen, T., Paulsen, G.,
22. Lipman, D. J., Wilbur, W. J., Smith, T. F. & Waterman,
21. Grogan, J. L., Kramer, A., Nogai, A., Dong, L., Ohde,
20. Quaratino, S., Thorpe, C. J., Travers, P. J. & Londei,
18. van de Wal, Y., Kooy, Y. M., van Veelen, P. A., Pena,
29. van de Wal, Y., Kooy, Y. M., Drijfhout, J. W., Amons,
17. Anderson, R. P., Degano, P., Godkin, A. J., Jewell,
27. Johansen, B. H., Gjertsen, H. A., Vartdal, F., Buus, S.,
28. Vartdal, F., Johansen, B. H., Friede, T., Thorpe, C. J.,
16. van de Wal, Y., Kooy, Y. M., van Veelen, P. A., Pena,
30. Averof, M., Rokas, A., Wolfe, K. H. & Sharp, P. M.
32. Yokomori, K., Zeidler, M. P., Chen, J. L., Verrijzer,


*Edited by J. Karn*

*(Received 19 November 2001; received in revised form 8 March 2002; accepted 12 March 2002)*