# Celiac disease : from basic insight to therapy development

Stepniak, D.T.

Submitted

# 2 Large-scale characterization of natural ligands explains the unique gluten-binding properties of HLA-DQ2

Dariusz Stepniak[1], Martina Wiesner[1], Arnoud de Ru[1, 2], Antonis K. Moustakas[3], George K. Papadopoulos[4], Peter A. van Veelen[1, 2] and Frits Koning[1]

[1] Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, the Netherlands;
[2] Centre for Medical Systems Biology, the Netherlands;
[3] Department of Organic Farming, Technological Educational Institue of Ionian Islands, Argostoli, Cephallonia, Greece;
[4] Laboratory of Biochemistry and Biophysics, Faculty of Agricultural Technology, Epirus Institute of Technology, Arta, Greece;

# Large-scale characterization of natural ligands explaines the unique gluten-binding properties of HLA-DQ2

**Dariusz Stepniak, Martina Wiesner, Arnoud H. de Ru, Antonis K. Moustakas, George K. Papadopoulos, Peter A. van Veelen, and Frits Koning**

## ABSTRACT

The association of HLA-DQ2 (DQA1*0501 / DQB1*0201) with celiac disease is extraordinarily strong: about 95% of patients express HLA-DQ2. This suggests that HLA-DQ2 has unique peptide-binding properties that allow it to present gluten-derived peptides to CD4+ T cells in the small intestine. We have now characterized peptides that are naturally bound to HLA-DQ2 molecules. In total 432 peptides were identified that represent length variants of 155 unique sequences. These sequences were aligned and the binding cores were inferred. Negatively charged amino acids were found to be favored at virtually all positions in HLA-DQ2 bound peptides, which discords with a more restricted presence of such amino acids in T cell epitopes from gluten. HLA-DQ2, however, was also found to display strong preferences or detterences for proline at several anchor and non-anchor positions that largely match with the allocation of proline in T cell stimulatory gluten peptides. Moreover, the selective presence of proline at p6 and p8 facilitates the conversion of glutamine into glutamic acid in gluten peptides by tissue transglutaminase at these preferred anchor residues. These unique peptide binding properties of HLA-DQ2 thus explain the ability of this molecule to bind a repertoire of disease-inducing gluten peptides.

## INTRODUCTION

HLA-DQ2 (DQA1*0501 / DQB1*0201) is the strongest genetic factor predisposing to celiac disease [1]. HLA-DQ2 dimers were shown to present gluten peptides to gluten-specific T cells in the intestinal lamina propria, thereby inducing and sustaining strong inflammatory responses [2,3]. The association of HLA-DQ2 with CD is extraordinarily strong: approximately 95% of CD patients carry the DQ2 phenotype wheraes in general population the frequency of this allele is about 20-30% [4,5]. This is a clear implication that the ability to present gluten-derived peptides is a unique feature of HLA-DQ2. A binding motif for HLA-DQ2 has previously been elucidated and indicates a preference for negatively charged amino acids at the p4, p6 and p7 positions and bulky hydrophobic amino acids at the p1 and p9 positions in bound peptides [6,7]. This motif, however, could not explain the binding of glutamine- and proline-rich gluten peptides. The explanation came with the finding that due to the activity of tissue transglutaminase, present in the intestinal mucosa, particular glutamine residues in gluten and similar proteins are deamidated into glutamic acid [2,3]. This enzymatic reaction thus introduces negative charges into the peptides allowing their anchoring in HLA-DQ2.

The presence of multiple proline residues in gluten peptides decreases the conformational freedom of the peptide backbone. Furthermore, the number of possible registers in which the proline-rich peptides could interact with MHC molecules, is limited since proline cannot form amide-mediated hydrogen bonds. Notwithstanding, there is a large repertoire of gluten-derived peptides that can bind to HLA-DQ2 and trigger inflammatory T cell responses. In order to gain further insight into the peptide-binding properties of HLA-DQ2 we have now characterized a large number of peptides that are naturally bound to HLA-DQ2. Alignment and analysis of the binding cores of those peptides provided unexpected insight into the peptide-binding characteristics of HLA-DQ2 that explain why this molecule is so uniquely suited to bind gluten peptides. Molecular simulation of binding between the HLA-DQ2 dimer and some of the identified peptides allowed the identification of yet unknown sites of interaction on the peptide-MHC interface. This further explicates the atypical binding preferences of HLA-DQ2.

## MATERIALS AND METHODS

### Peptide isolation from affinity-purified HLA-DQ molecules.

Human HLA-DQ2 dimers were isolated from an HLA-DQ2 homozygous Epstein-Barr Virus (EBV)-transformed B lymphoblastoid cell line. Approximately $10^{10}$ cells were grown in RPMI 1640 medium supplemented with L-glutamine and 10% FCS.

Subsequently the cells were harvested, washed with PBS and the cell pellet was stored at -80°C. The cells were lysed with 50 ml of lysis buffer (50mM Tris, 150mM NaCl, 5mM EDTA, 0.5% NP-40, 10mM iodo-acetamide, protease inhibitor mix [Complete™, Roche]). To remove the nuclei and insoluble material the lysate was centrifuged for 60 minutes at 10.000 g. It was subsequently precleared for 60 minutes with Sepharose beads and mixed with 7 ml of Sepharose beads onto which 16 mg of the HLA-DQ-specific monoclonal antibody SPV-L3 (Spits et al, 1994) was covalently linked. After 60 minutes of gentle mixing, the beads were extensively washed with lysis buffer followed by washing with 20 mM Tris-HCl, 120 mM NaCl pH 8.0, 20 mM Tris-HCl, 1 M NaCl, pH 8.0, 20 mM Tris-HCl, pH 8.0 and finally with 10 mM Tris-HCl, pH 8.0. Subsequently, the HLA-peptide complexes were eluted with 30 ml 10% acetic acid in water. All the purification steps were performed at 4°C. High molecular mass material (HLA dimers) was removed by filtration through Centriprep filtration units with a cut-off value of 10 kDa and analyzed by SDS-PAGE. The peptide fraction was freeze-dried, dissolved in 400 µL of 10% acetic acid in water and fractionated with an HPLC system (micro Smart, Pharmacia, Uppsala, Sweden) equipped with a reverse-phase C2/C18 sc 2.1/10 column (Pharmacia, Uppsala, Sweden). The material was eluted using a gradient of 0 to 100% acetonitrile (0.25% per minute) supplemented with 0.1 % TFA.

**Peptide identification by mass spectrometry**

Peptides present in the collected fractions were subsequently sequenced by tandem mass spectrometry. To this end, half of the peptide fractions from the Smart system were freeze dried and resuspended in 95/3/0.1 v/v/v water/acetonitrile/formic acid (solvent A). These resuspended fractions were analyzed by an on-line nanoHPLC-MS system. The nano HPLC system consisted of a conventional gradient HPLC system (Agilent 1100, Amstelveen, The Netherlands), the flow of which was reduced to 300 nl/min by an in house constructed splitter. Two-microliter injections were done onto a precolumn (10 mm x 100 um; AQUA-C18, 5 µm particle size, (Phenomenex, Torrance, CA, USA)) and eluted via an analytical nanoHPLC column (15 cm x 75 um; AQUA-C18, 5 um particle size. HPLC columns were packed in-house. The gradient was run from 0-50% solvent B (10/90/0.1 v/v/v water/acetonitrile/formic acid) in 90 minutes. The mass spectrometer was an HCT$^{plus}$ (Bruker Daltonics, Bremen, Germany), which was run in data-dependent MS/MS mode during peptide elution. All tandem mass spectra produced in this way were searched against the human IPI database with the database search program MASCOT v 2.1.0 (Matrixscience, London, UK). All reported hits were assessed manually, and peptides with MASCOT scores lower than 40 were usually discarded.

**SDS-PAGE and in-gel digestion**

The high molecular mass fraction of the immunoprecipitate was analyzed by standard 12% polyacrylamide gel electrophoresis. To this purpose the fraction was freeze-dried, dissolved in 1M Tris-HCl buffer pH 7.5, diluted with 4 x protein sample buffer (60% glycerol, 300 mM Tris (pH 6.8), 12 mM EDTA pH 8.0, 12 % SDS, 864 mM 2-mercaptoethanol, 0.05% bromophenol blue) and run on a 12 % SDS-PAGE gel. The proteins were visualized using Imperial Protein Stain (Pierce, Rockford IL, USA). Selected bands were cut out of the gel, extracted and digested with trypsin according to Shevchenko et al. [8]. The tryptic fragments were identified by on-line nanoHPLC-MS on an HCTplus as described above.

**Construction of databases**

A tailored database of all the known protein sequences of gluten and gluten-like proteins (gliadins, glutenins, hordeins, secalins and avenins), referred to as a gluteome, was built by extraction of relevant entries from the UniprotKB database at the European Bioinformatics Institute EBI (www.ebi.ac.uk) with the SRS system with the following search terms: organism: Triticum aestivum AND (description: gliadin OR glutenin) for wheat, Hordeum vulgare AND (description: hordein) for barley, Secale cereale AND (description: secalin) for rye, and Avena sativa AND (description: avenin) for oats.

The human IPI database was downloaded from the EBI (www.ebi.ac.uk/IPI/IPI help.html). The set of the protein sequences contained in the database is referred to as a proteome.

**Peptide binding assay**

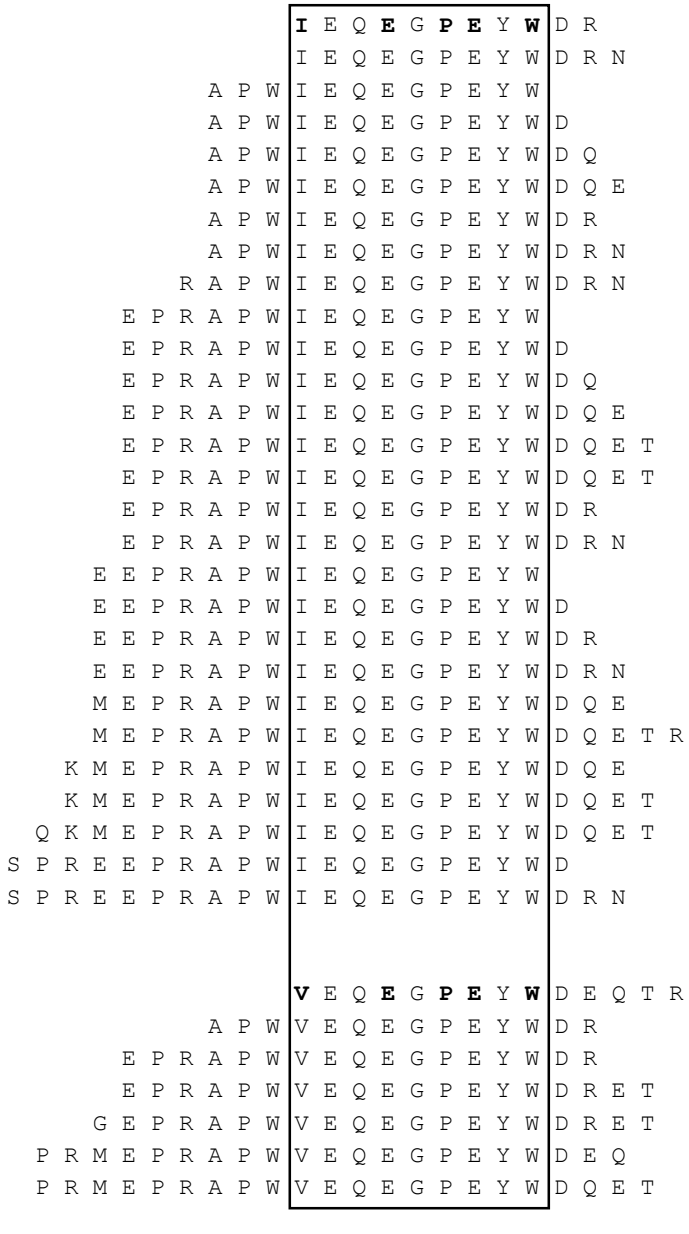
96-well FluoroNunc™ plates were coated with the HLA-DQ-specific monoclonal antibody SPV-L3, 2 µg/well in 100 µL of carbonate buffer (50 mM $Na_2CO_3$, 50 mM $NaHCO_3$, pH 9.6) for 2 h at 37°C, subsequently blocked for 3 h at 37°C with 2% FCS in PBS. HLA-DR3/DQ2 positive EBV-transformed B-cells were grown and a sample was checked for proper HLA-DQ2 expression by FACS-analysis using the SPV-L3 monoclonal antibody. Subsequently, the remainder of the cells were lysed in 20 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 1% NP-40 and protease inhibitor mix (Complete™, Roche) at 4°C, at a concentration of 4 x $10^6$ cells per 1 ml lysis buffer. Subsequently, nuclei and cell debris was removed by centrifugation (4°C, 2000g, 15 min). Next the lysate was mixed with an equal volume of ice-cold 1% solution of BSA in PBS and 100 µL aliquots were pipetted into the SPV-L3 coated wells. After overnight incubation at 4°C the plates were washed and 50 µL of binding buffer (0.1% NP-40, 0.1% Tween, 33.6 mM citric acid, 72 mM $Na_2HPO_4$, pH 5.5 and Complete™ protease inhibitor mix) was added to each well. Titration ranges of peptides to be tested (concentration range 300 – 0.2 µM) were prepared in 10% DMSO containing

a fixed amount of the biotin-labeled indicator peptides Glt-156 or MHCIα (46-63) at a concentration of 2.5 μM and 0.6 μM, respectively. Subsequently, 50 μL of the samples was applied to the SPV-L3/HLA-DQ2 coated plates. Following a 48 h incubation at 37°C each well was washed extensively. Subsequently, 100 μL of 100 nM streptavidin-europium in assay buffer (both from Wallac, Turku, Finland) was added and incubated for 45 minutes at RT. After extensive washing, 150 μL/well of enhancement solution (Wallac) was applied and the plates were read in a time resolved fluorimeter (1234, Wallac) 15-30 minutes thereafter. $IC_{50}$ values were calculated based on the observed competition between the test peptides and biotin-labeled indicator peptides and indicate the concentration of the tested peptide required for half maximal inhibition of the binding of the indicator peptide.

**Molecular simulation**

Molecular simulation of the structures of complexes between HLA-DQ2 molecules and selected peptides was performed as previously described [9]. In brief, a hybrid HLA-DQ2 molecule was constructed containing the αβ chains from the crystal structure of HLA-DQ2 [10] and the peptide backbone coordinates from HLA-DQ8 [11]. This step was necessary because the antigenic peptide in the HLA-DQ2 crystal structure has its p9 pocket (the most spacious) unoccupied and its p9Tyr residue occupies the so-called p10 pocket (shelf) [12]. Previous binding experiments indicated that this pocket is very spacious accommodating aromatic and large aliphatic amino acids, and secondarily acidic ones. This selectivity would not be exhibited by a residue occupying the p10 pocket. The peptides used were 13-mers with their nonamer core as obtained from the alignment of the nested peptides from the same family. The rotamer chosen for each peptide residue position was derived from a library of rotamers, so that there would be no steric clashes. In case a glutamate residue was to occupy pocket 6 we chose to have a rotamer nearly identical to that occupying pocket 6 in the crystal structure of the complex of DQ2 and the gliadin peptide [10]. The molecular simulation was carried out by the program Discover of Accelrys (San Diego, CA) with 1000 steps of the steepest gradient method followed by 1000 steps of the conjugate gradient method [13,14]. Numbering of the HLA-DQαβ residues follows the scheme suggested by Fremont et al. [15], as this provides structural equivalence for residues from different MHC II gene loci independent of animal species. Graphical representations of the results were per-formed with the program WebLabViewer version 3.5 of Acceelrys.

**A.**


I E Q E G P E Y W D R
I E Q E G P E Y W D R N
A P W I E Q E G P E Y W
A P W I E Q E G P E Y W D
A P W I E Q E G P E Y W D Q
A P W I E Q E G P E Y W D Q E
A P W I E Q E G P E Y W D R
A P W I E Q E G P E Y W D R N
R A P W I E Q E G P E Y W D R N
E P R A P W I E Q E G P E Y W
E P R A P W I E Q E G P E Y W D
E P R A P W I E Q E G P E Y W D Q
E P R A P W I E Q E G P E Y W D Q E
E P R A P W I E Q E G P E Y W D Q E T
E P R A P W I E Q E G P E Y W D Q E T
E P R A P W I E Q E G P E Y W D R
E P R A P W I E Q E G P E Y W D R N
E E P R A P W I E Q E G P E Y W
E E P R A P W I E Q E G P E Y W D
E E P R A P W I E Q E G P E Y W D R
E E P R A P W I E Q E G P E Y W D R N
M E P R A P W I E Q E G P E Y W D Q E
M E P R A P W I E Q E G P E Y W D Q E T R
K M E P R A P W I E Q E G P E Y W D Q E
K M E P R A P W I E Q E G P E Y W D Q E T
Q K M E P R A P W I E Q E G P E Y W D Q E T
S P R E E P R A P W I E Q E G P E Y W D
S P R E E P R A P W I E Q E G P E Y W D R N
V E Q E G P E Y W D E Q T R
A P W V E Q E G P E Y W D R
E P R A P W V E Q E G P E Y W D R
E P R A P W V E Q E G P E Y W D R E T
G E P R A P W V E Q E G P E Y W D R E T
P R M E P R A P W V E Q E G P E Y W D E Q
P R M E P R A P W V E Q E G P E Y W D Q E T

**B.**

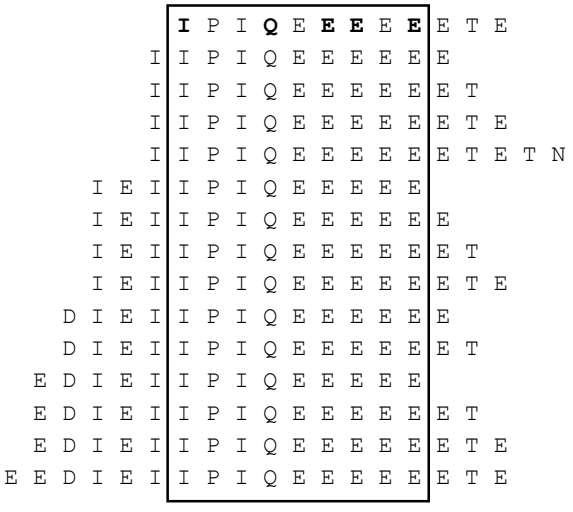



**Figure 1.** Alignment of the length variants of two MHC I-derived sequences **(A)** and a highly acidic CD20-derived peptide **(B)**. Anchor residues are given in bold.

## RESULTS

### Isolation and sequencing of HLA-DQ2 bound peptides.

Human HLA-DQ2 dimers were purified from HLA-homozygous EBV-transformed B cells and the HLA-DQ2-bound peptides were eluted with acid (see material and methods). Subsequently, filtration was used to separate peptides from protein fraction. Analysis of the high molecular weight material by SDS-PAGE revealed the presence of two bands corresponding to the predicted molecular weight of HLA-DQ2 alpha- and beta-chains (respectively 33 kDa and 28 kDa). To confirm the identity of these proteins we performed in-gel trypsin digestion of these two proteins and sequenced the generated fragments by mass spectrometry. Several protein fragments corresponding to sequences present in the HLA-DQ2 alpha- and beta-chain were identified, demonstrating that HLA-DQ2 was successfully purified (Supplementary Fig. S1A). Next, the low molecular weight material that was eluted

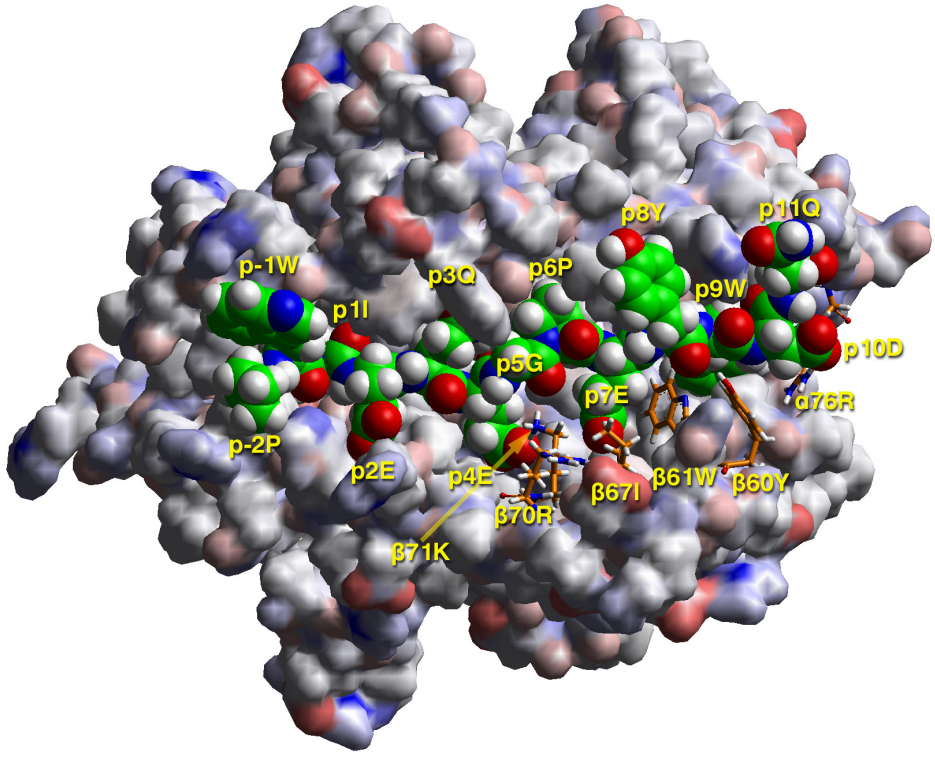



**Figure 2.** T cell receptor view of the most abundant peptide eluted from HLA-DQ2 (PW**I**EQ**E**G**P**-**E**Y**W**DR) in the groove, after energy minimization based on the crystal structure of HLA-DQ2 in complex with a gliadin peptide [10]. The peptide coordinates were replaced by those of the insulin B11-23 peptide in the crystal structure of DQ8 [11] in order to get an anchor residue into the p9 pocket. The αβ ßomain of the HLA-DQ2 molecule is depicted in van der Waals surface representation, with the surface atomic charges color-coded (blue positive; gray, neutral; red, positive), and the antigenic peptide is shown in space filling form (color code: carbon, green; oxygen, red; nitrogen, blue; hydrogen, white; sulfur, yellow). Six DQ2 residues that have particular interactions with TCR (β70Arg) and/or the antigenic peptide (α76Arg, β60Tyr, β61Trp, β67Ile, βArg and β71Lys) are shown in stick form with the same color-code as the antigenic peptide with the exception of carbon (orange). P1Ile is in the p1 pocket and p9Trp in the p9 pocket (both pointing into the plane of the paper) as expected of a peptide bound t DQ2 with high affinity.

from the HLA-DQ2 was fractionated by rpHPLC (Supplementary Fig. S1B) and the peptides present in the obtained fractions were sequenced by tandem mass spectrometry. This resulted in the identification of 432 peptides that represent length variants of 155 unique peptides.

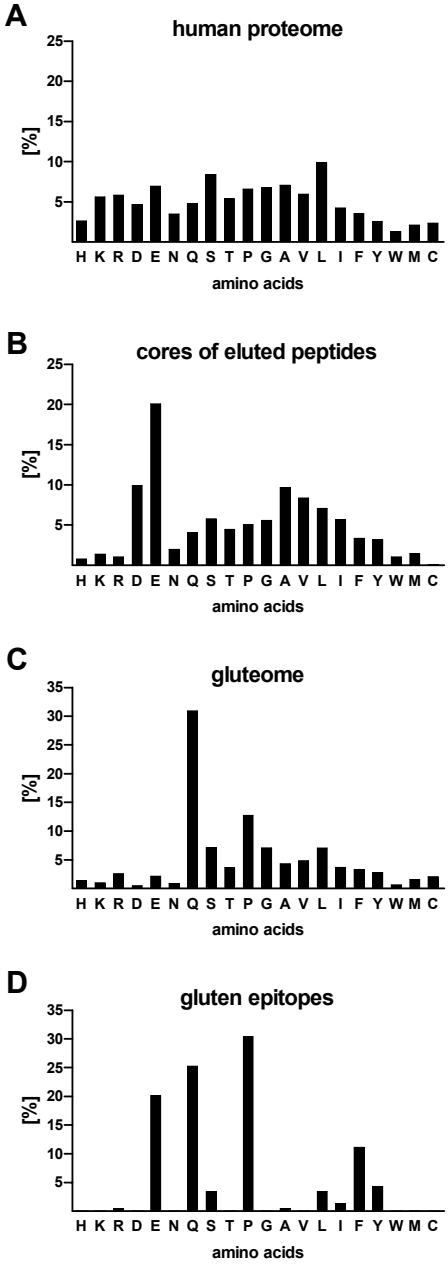
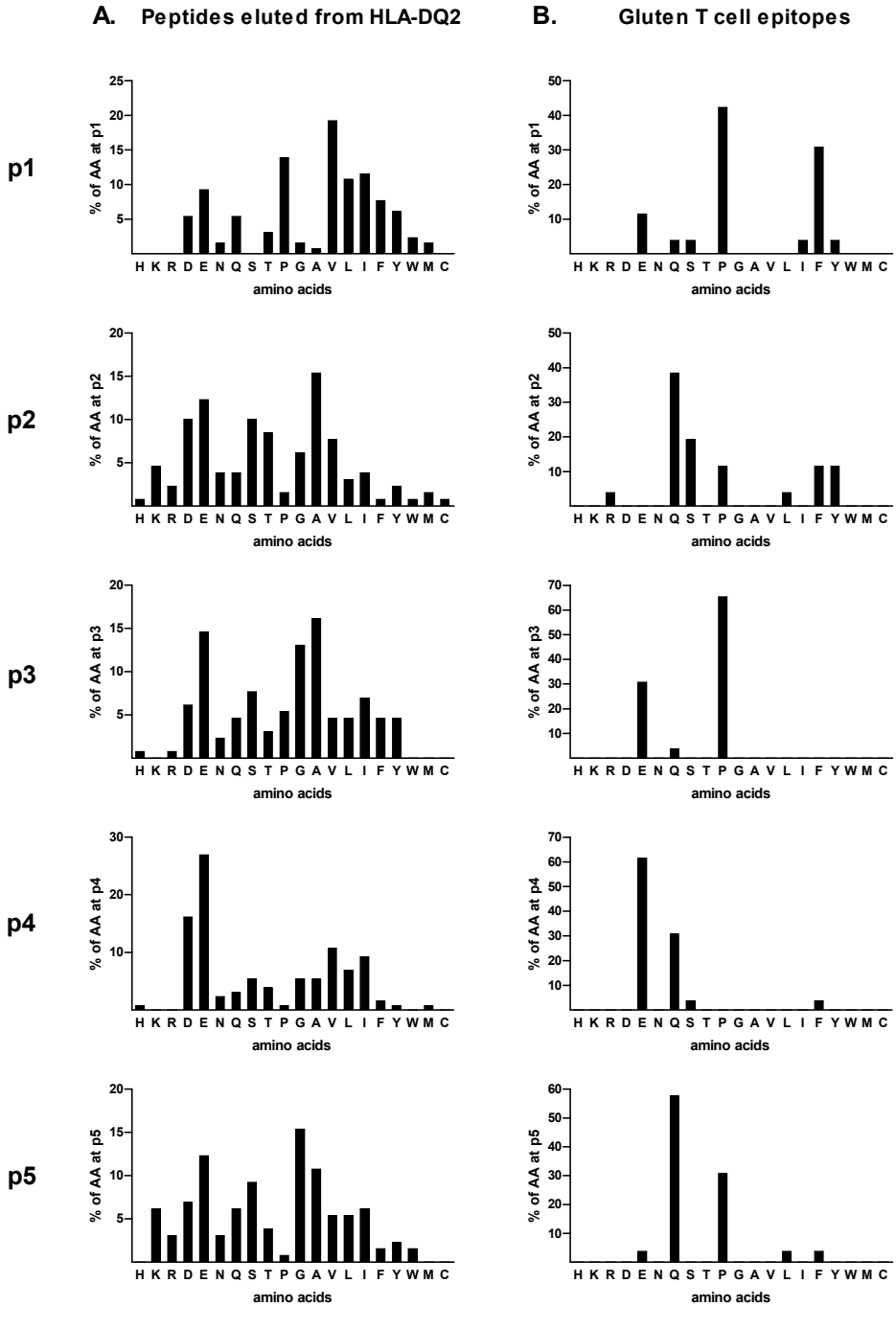



**Figure 3.** Amino acid distribution in: the human proteome **(A)**, the binding cores of peptides eluted from HLA-DQ2 **(B)**, the gluteome **(C)** and T cell stimulatory sequences from gluten and related proteins in barley, rye and oats **(D)**.


A. Peptides eluted from HLA-DQ2
B. Gluten T cell epitopes
p1
p2
p3
p4
p5
% of AA at p1
% of AA at p2
% of AA at p3
% of AA at p4
% of AA at p5
H K R D E N Q S T P G A V L I F Y W M C
amino acids

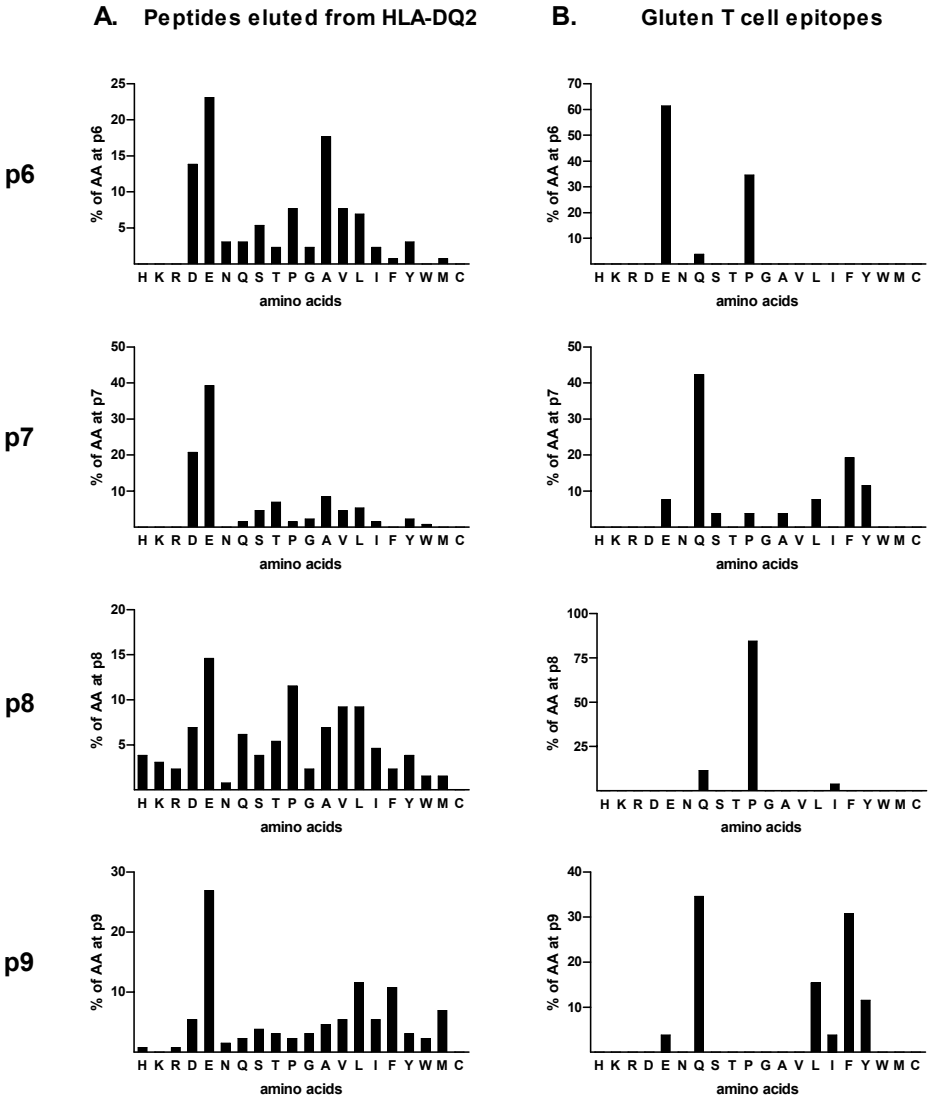



**Figure 4.** Distribution of amino acids at the p1 through p9 positions in**.** The peptides eluted from HLA-DQ2 **(A)** and T cell stimulatory sequences from gliadins, glutenins, hordeins, secalins and avenins **(B)** (see also table 1).

**Identification of the binding cores of the identified peptides**

In agreement with previous studies, the most abundant peptide identified was the HLA-class I derived peptide IEQEGPEYW. This sequence was represented by 28 length variants and 7 length variants of the homologous sequence VEQEGPEYW (Fig. 1A). Alignment of the length variants shows that the I/V-EQEGPEYW is the only 9 amino acid sequence common to all length variants (Fig. 1A), indicating that this is the binding core, which agrees with previous studies [6,7]. Molecular simulation of peptide IEQEGPEYW in the groove of HLA-DQ2 showed optimal interactions at all anchor positions (Fig. 2). Moreover, the source protein of this peptide, HLA-class I, is abundantly expressed in most cells, a factor that most likely also contributes to the dominant presence of this peptide in HLA-DQ2.

Similar to the HLA-class I peptide, the alignment of the length variants of a CD20 derived peptide indicates that the sequence IPIQEEEEE represents the binding core of this peptide (Fig. 1B). Using this method the core sequences could be determined for several other peptides. For the majority of peptides, however, fewer length variants were identified so that the minimal cores could not be determined this way. The putative minimal cores of these peptides were deduced with the aid of the previously identified HLA-DQ2 specific peptide-binding motif. This motif incorporates (large) hydrophobic amino acids at position p1 and/or p9 and negatively charged amino acids at p4, p6 and/or p7. The likelihood of the deduced binding cores was checked in HLA-DQ2 specific peptide binding studies and by molecular simulation for several peptides (data not shown). Through this procedure a single minimal core could be identified for 124 out of 155 peptides (Supplementary Table 1). For the remainder of the peptides either multiple binding frames were possible or a binding core could not be identified (Supplementary Table 2).

**HLA-DQ2 accepts negatively charged residues along the entire p1-p9 core**

Based on the alignment of the 128 minimal cores the frequency of amino acids was calculated and compared with that of the human proteome (Fig. 3A, B). Most of the amino acids in HLA-DQ2 bound peptides were present at a level that is comparable to that found in the human proteome with the notable exception of negatively and positively charged amino acids. While basic amino acids were virtually absent in HLA-DQ2-bound peptides, the vast majority of the eluted peptides were found to contain multiple acidic residues, the most extreme example being a peptide EEEEE-EAEEEEEEE derived from glucosidase II (Supplementary Table 2). Strikingly, a preference for aspartate and particularly glutamate was found not only at the p4, p6, and p7 anchor positions but at virtually all positions in the binding core including the non-anchor positions p2, p3, p5 and p8 (Fig. 4).

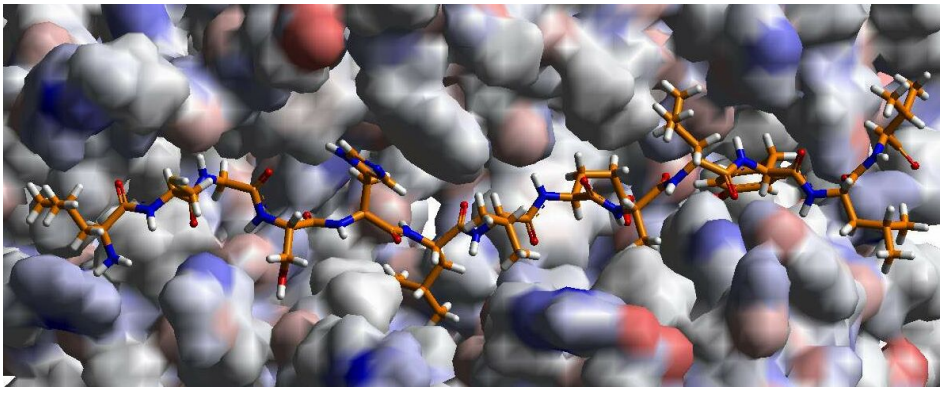


**Figure 5.** The distribution of charges on the surface of HLA-DQ2. The HLA-DQ2 molecule was modeled with an antigenic peptide in the groove and represented in the van der Waals convention. The surface is color-coded according to the atomic charges (negative, red; neutral, gray; positive, blue). Note the positive charges in and around the binding groove.

In order to explain this preference for negatively charged amino acids at both anchor and non-anchor positions we analyzed the charge distribution on the surface of the HLA-DQ2 molecule. This revealed that HLA-DQ2 has a surplus of positive charges (+8) that is larger than that found in HLA-DQ8 (+6) or HLA-DQ6 (+4) (Supplementary Table 3 and Fig. 5). As these positive charges are located around the peptide-binding groove they could facilitate the accommodation of negatively charged residues along the entire p1-p9 register.

In addition, we modeled a number of the eluted peptides in complex with HLA-DQ2. The molecular simulation of peptide AG**E**EG**E**A**GD**E**E**LP in HLA-DQ2 exposed that the p1Glu forms a salt bridge with the terminal guanidine group of α52Arg (Fig. 6A). Since glutamate occupies and covers the opening of the p1 pocket it is very likely that below this residue water molecules are present in this pocket that further promote hydrogen bonding between p1Glu and other polar residues e.g. β89Thr, β90Thr and α52Arg. Interestingly, α52Arg is also present in the p1 pocket of HLA-DQ8, which prefers acidic residues at p1, but is replaced with glycine in HLA-DQ6, which does not tolerate negatively charged amino acids at the p1 position.

The simulated structure of the complex between HLA-DQ2 and the peptide AA**V**EE**E**E**E**E**K**VAE indicates that the presence of multiple acidic residues in the binding core of the bound peptide attracts the β77Arg residue towards the p5 acidic residue (Fig. 6C). Thus, besides the residues β28Ser, β30Ser, β70Arg, β71Lys and β74Ala, that are very accommodating to acidic residues at p4, p6 and p7, β77Arg

**Table 1.** Overview of sequences derived from gluten and homologous proteins known to stimulate responses of T cells from celiac patients. Anchor positions are given in bold.

| No | Sequence | Peptide name | Reference |
|---|---|---|---|
| | - - - **1** 2 3 **4** 5 **6** **7** 8 **9** - - - | | |
| 1 | - - - **P** Q P **E** L **P** **Y** P **Q** P E L P Y | Glia-α2 (Alpha-II) | [22] |
| 2 | Q L Q **P** F P **Q** P **E** **L** P **Y** | Glia-α9 (Alpha-I) | [22] |
| 3 | P Q L **P** Y P **Q** P **E** **L** P **Y** | Glia-α9 (Alpha-III) | [23] |
| 4 | - - P **F** R P **E** Q **P** **Y** P **Q** P Q P | Glia-α20 | [24] |
| 5 | Q E Q **Q** X P **E** Q **P** **Q** Q **F** | Glu-5 | [24] |
| 6 | P E Q **P** Q Q **S** F **P** **E** Q **E** R P F | Gamma-I (Glia- γ1) | [25] |
| 7 | E G I **I** Q P **E** Q **P** **A** Q **L** | Gamma-II (Glia- γ30) | [26,27] |
| 8 | - F P **E** Q P **E** Q **P** **Y** P **Q** Q | Gamma-III | [23] |
| 9 | - - F **S** Q P **E** Q **E** **F** P **Q** P Q | Gamma-IV | [23] |
| 10 | - - P **E** Q P **F** P **E** **Q** P **Q** Q | Gamma-VI | [27] |
| 11 | - - - **P** Q P **E** Q **E** **F** P **Q** | Gamma-VII$^{A}$ | [27] |
| 12 | - - - **E** Q P **E** Q **P** **F** P **Q** | Gamma-VII$^{B}$ | [27] |
| 13 | Q P P **F** S E **E** Q **E** **S** P **F** S Q | Glt-156 | [24] |
| 14 | Q P P **F** S E **E** Q **E** **Q** P **L** P Q | Glt-17$^{A}$ | [24] |
| 15 | Q E F **P** Q P **E** Q **P** **F** P **Q** Q P | Hor-α2 | [28] |
| 16 | Q P F **P** Q P **E** Q **P** **F** P **Q** S Q | Sec-α2 | [28] |
| 17 | P Q Q **P** F P **Q** P **E** **Q** P **F** R Q | Hor-α9 | [28] |
| 18 | P Q Q **P** F P **Q** P **E** **Q** P **F** P Q | Sec-α9 | [28] |
| 19 | - Y Q **P** Y P **E** Q **E** **E** P **F** V Q | Av-α9A | [28,29] |
| 20 | - Y Q **P** Y P **E** Q **E** **Q** P **F** V Q | Av-α9B | [28,29] |
| 21 | Q P P **F** S E **E** Q **E** **Q** P **L** P | Glt-156 homologue | [24] |
| 22 | Q P P **F** S E **E** Q **E** **Q** P **I** L L | Glt-156 homologue | [24] |
| 23 | Q P P **F** S E **Q** E **Q** **P** I **L** P Q | Glt-156 homologue | [24] |
| 24 | Q Q P **F** P Q **Q** P **Q** **Q** P **F** P Q | Glia-γ2$^{A}$ | [9] |
| 25 | Q Q P **Y** P Q **Q** P **Q** **Q** P **F** P Q | Glia- γ2$^{B}$ | [9] |
| 26 | Q Q P **F** P Q **Q** P **Q** **Q** P **Y** P Q | - | [26] |

functions as an additional attractor and stabilizer of very negatively charged acidic peptides bound to HLA-DQ2.

Finally, the modeling of EI**I**PI**Q**E**E**E**E**ET demonstrated that the p9Glu forms an ionic bond with α76Arg and a hydrogen bond with α72Ser (Fig. 6B). These strong

interactions likely explain the presence of glutamate at the p9 position in roughly 30% of peptides found in HLA-DQ2.

To check whether this HLA-DQ2 binding preference matches the sequences of cereal proteins we compared the frequency of amino acids in the HLA-DQ2-eluted peptides with that found in the collection of all available gluten and gluten-like molecules from wheat, barley, rye and oat present in the public databases (referred to as the gluteome hereafter) and in the known HLA-DQ2-restricted gluten T cell stimulatory epitopes (Table 1). Large discrepancies in the abundance of certain amino acids were found between these 3 groups of sequences (Fig. 3B-D). To a large extent these can be explained by the restricted amino acid composition of gluten and gluten-like molecules, which is strongly dominated by glutamine, proline and hydrophobic amino acids. Moreover, the lack of negatively charged residues in gluten, while these are prominent in HLA-DQ2 restricted, T cell stimulatory gluten peptides is explained by the activity of tissue transglutaminase, which converts glutamine residues into glutamic acid, a prerequisite for high-affinity HLA-DQ2 binding. While HLA-DQ2 can accommodate glutamic acid all along the binding core (Fig. 7A), the occurrence of glutamic acid in the known T cell stimulatory peptides is much more restricted to certain positions (Fig. 7B). It was mostly limited to the anchor positions p4 and p6, as well as non-anchor position p3. Also, the p7 binding pocket, which has the highest preference for acidic residues, mostly accomodated glutamine and aromatic amino acids in the case of gluten epitopes (Fig. 4). Similarly, glutamate was hardly found at the p2, p5 and p8 non-anchor positions of gluten epitopes. Thus, the preference for negatively charged residues does not explain the unique binding properties of HLA-DQ2 that allows efficient presentation of gluten peptides.

**HLA-DQ2 selects for proline at anchor and non-anchor position**

A further analysis of the distribution of amino acids in the core of the HLA-DQ2 bound peptides revealed that the abundance of several amino acids is skewed at both anchor and non-anchor positions. This was observed for alanine, phenylalanine, serine, glycine and proline (Fig. 4). Although the overall presence of proline in HLA-DQ2-bound peptides is comparable to its presence in the human proteome (Fig. 3A,B), strong positive and negative effects are observed depending on the position (Fig. 7C). Most notably, proline was found to be favored at the p1 anchor position and at the p8 non-anchor position while its presence in the p3 and p6 position is comparable to that found in the human proteome. At all other positions proline is underrepresented.

It is well established that all known HLA-DQ2-restricted gluten-derived T cell stimulatory peptides contain multiple proline residues (Table 1). As noted previously, proline residues in these gluten peptides are usually found at positions

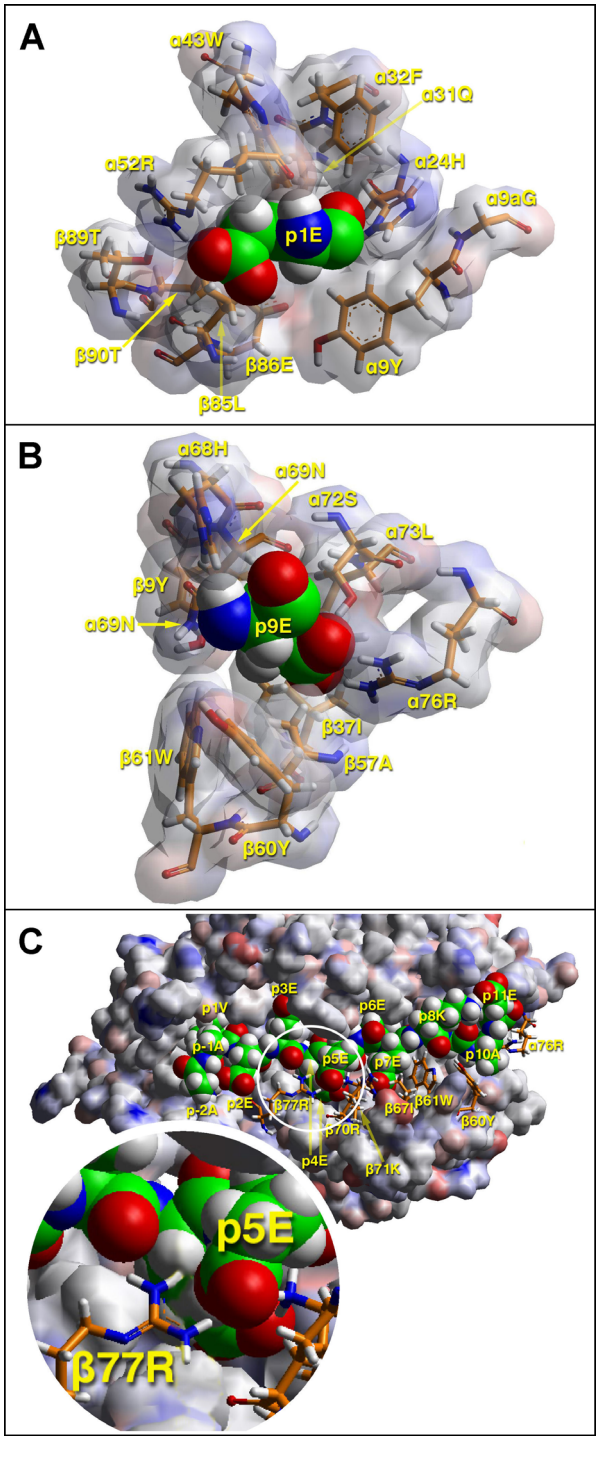

A
α43W
α32F
α31Q
α52R
α24H
α9aG
β89T
p1E
β90T
β86E
α9Y
β85L
B
α68H
α69N
α72S
α73L
β9Y
α69N
p9E
α76R
β37I
β61W
β57A
β60Y
C
p3E
p1V
p6E
p11E
p8K
p-1A
α76R
p10A
p5E
p7E
p-2A
p2E
β77R
β67I
β61W
β70R
β60Y
p4E
β71K
p5E
β77R

**Figure 6.** Molecular simulation of HLA-DQ2-peptide complexes explaining anchoring via glutamate at positions p1, p9 and p5. The antigenic peptide anchor residues are shown in space-filling mode (color code: oxygen, red; nitrogen, blue; carbon, green, hydrogen, white) while the residues from HLA-DQ2 are shown in stick form (identical atom color code except for carbon that is in orange) with transparent van der Waals surfaces and atomic charges (positive, blue; neutral, gray; negative, red). **A.** T cell receptor view of pocket 1 of the complex between peptide AG**E**EG**E**A**GD**E**E**LP (anchors in bold) and HLA-DQ2. The glutamate residue at p1 forms a salt bridge with the terminal guanidine group of α52Arg. In addition, the β86Glu forms hydrogen bonds with α31Gln (not seen because of p1Glu) and α9Tyr. The p1Glu occupies and covers the opening of this pocket; it is thus very likely that below the p1Glu there are water molecules in the pocket that further promote hydrogen bonding among polar residues (e.g. p1Glu, β89T, β90T and β52R). **B.** T cell receptor view of pocket 9 of the complex between peptide E**I**IPI**Q**E**EE**E**E**ET and HLA-DQ2. The glutamate at p9 forms a salt bridge with α76Arg, and a hy-drogen bond with α72Ser. **C.** T cell receptor view of the complex between peptide AA**V**EE**EE**E**EK****V**AE (anchors in bold) and HLA-DQ2. Note how β77Arg swings into position to form a salt bridge with p5Glu in the bound peptide.

p1, p3, p5, p6 and/or p8 (Fig. 7D), positions at which HLA-DQ2 favors or tolerates the presence of proline (Fig. 7C). Strikingly, the alignment of known T cell stimulatory peptides from gluten and gluten homologues in barley, rye and oats indicates that proline is present at position p8 in 22 out of 26 sequences, a value that is much higher than the proline content in the gluteome (Fig. 7D). Together, these results indicate that, the proline profile found in gluten peptides closely matches the preference of HLA-DQ2 for this amino acid, with the exception of the p5 position.

## DISCUSSION

The association between autoimmune disorders and specific MHC haplotypes indicates a fundamental role for antigen presentation in the development of these diseases. This is particularly clear in the case of celiac disease, which shares many similarities with other autoimmune disorders [16], and where roughly 95% of the patients carry HLA-DQ2. The capacity of these HLA molecules to bind and present gluten peptides to CD4 positive T cells in the lamina propria of the small intestine is a prerequisite for the development of celiac disease.

HLA-allele specific peptide motifs have been defined that can be used to identify potential T cell epitopes in proteins. A peptide-binding motif for HLA-DQ2 was first described in 1996 [6,17]. These and subsequent studies indicated that HLA-DQ2 preferred large hydrophobic amino acids at p1 and p9 as well as negatively charged amino acids at p4, p6 and/or p7. This motif, however, failed to clarify the role of HLA-DQ2 in celiac disease development as gluten is virtually devoid of negatively charged amino acids. This apparent discrepancy resolved with the observation that gluten could be modified by the enzyme tissue transglutaminase, which can deami-

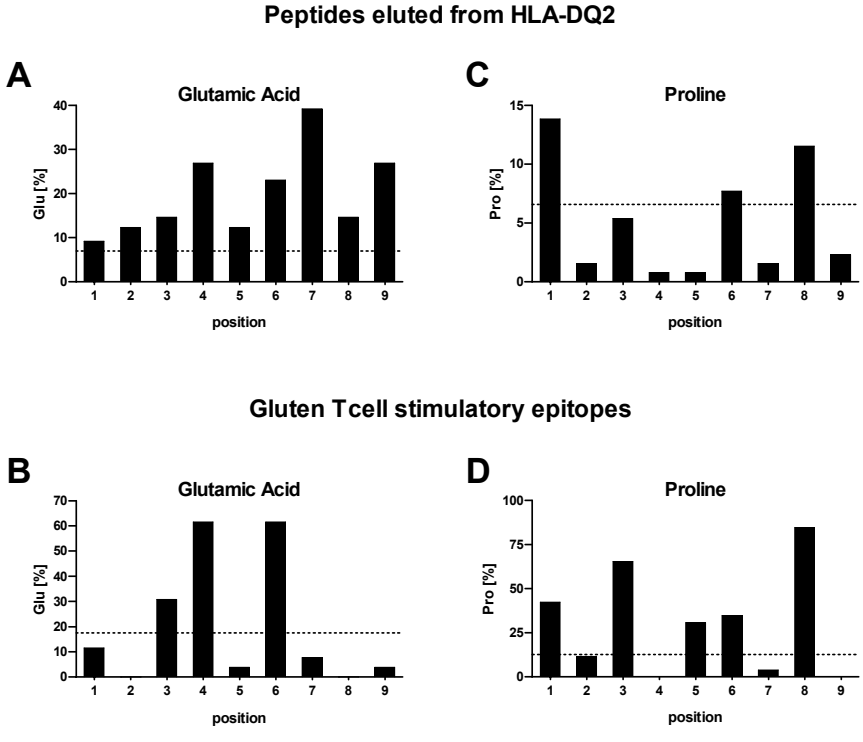



**Figure 7.** The relative distribution of glutamate and proline at the p1 through p9 positions of the HLA-DQ2 binding register. **A** and **C:** The relative abundance of the indicated amino acids in the core sequences of HLA-DQ2 eluted peptides. The dotted line indicates the abundance of these amino acids in the human proteome. **B** and **D:** The relative abundance of the indicated amino acids in the binding cores of gluten T cell stimulatory sequences from gliadins, glutenins, hordeins, secalins and avenins. The dotted line indicates the average abundance of glutamic acid in celiac disease-related T cell stimulatory epitopes or the relative abundance of proline in the gluteome.

date glutamine into glutamic acid [2,3]. Such deamidated gluten peptides were found to bind to HLA-DQ2 with high affinity and trigger gluten-specific T cell responses.

These studies also revealed that HLA-DQ2 had the ability to bind and display a large repertoire of gluten peptides which all contain multiple proline residues. Due to its peculiar nature proline restricts the flexibility in peptides and proteins. It also cannot form hydrogen bonds with mainframe amino acids in MHC-class II mole-

cules. Therefore, proline residues are poorly accepted at many positions by most MHC-class II molecules. HLA-DQ2 must thus have unique properties that allow it to bind a large repertoire of proline-rich gluten-derived peptides. In order to gain more insight into these peptide-binding properties we have now characterized a large number of peptides that are naturally bound to HLA-DQ2. This study revealed that HLA-DQ2 accepts or prefers negatively charged residues at the entire p1-p9 register. HLA-DQ2 molecule can accept acidic anchors at all five pockets because of strategically positioned polar or positively charged residues in the respective pockets (Fig. 6). In addition, analysis of the charge distribution on the surface of the HLA-DQ2 molecule indicates that HLA-DQ2, in contrast to HLA-DQ8 and HLA-DQ6 in particular, has many positive surface charges around the peptide-binding groove that facilitate the accommodation of negatively charged residues along the p1-p9 register (Fig. 5). It is important to note that these unpaired positive surface charges can attract free peptides, facilitating the docking process, but do not stabilize the peptide within the binding groove. The unique ability of HLA-DQ2 to accommodate negatively charged residues along the entire p1-p9 register, however, does not by itself explain the gluten peptide binding properties as the presence of glutamic acid residues in T cell stimulatory gluten peptides is much more restricted to certain position in these peptides. Rather, it is the unique property of HLA-DQ2 to select for a proline at the p1 anchor position and at the p8 non-anchor position that allows HLA-DQ2 to bind a large repertoire of gluten peptides. First, HLA-DQ2 is the only DQ-allele known to accept proline at p1 and roughly one-third of T cell stimulatory gluten peptides carry a proline at p1.

The uniqueness of the p1 pocket in HLA-DQ2 arises from an α53 deletion, which eliminates the requirement for a hydrogen bond from the p1 amide. There exists, however, the possibility of hydrogen bonding to the α50-52 HLA-DQ2 backbone by the amide groups of p-1 and p-2 peptide residues [9,10 and this work]. Second, a proline at p3 is accepted and present in over 50% of T cells stimulatory gluten peptides. Third, proline at p6 of DQ2 is tolerated, provided that there is no proline at p4 or p9 [18], despite the loss of one hydrogen bond to the side chain carbonyl of α62Asn. Roughly one-third of T cell stimulatory gluten peptides carry a proline at p6. The peculiar character of pocket 6 stems from the presence of β30Ser. HLA-DQ2 is the only HLA allele to have such a residue at this position (www.anthonynolan.uk.org). Finally, proline was found to be preferred at the p8 non-anchor position and 85% of T cell stimulatory gluten peptides carry proline at p8. Although selection for a particular amino acid should not be expected at non-anchor positions, the peculiar nature of proline restricts conformational freedom in peptides and may thus influence the position of the amino acid at p9, which in all cases are amino acids with a bulky side chain that can dock into the spacious p9 pocket.

Our results are in agreement with the known structure of the complex between HLA-DQ2 and a gluten peptide, which demonstrated that proline at p1/3/5/8 does not interfere with the hydrogen bond network from the polar residues of HLA-DQ2 that line the antigen-binding groove [10]. It has previously been shown that the spacing between the proline and glutamine residues determines which glutamines will be deamidated by tTG. Notably, in the very common sequence QXP in gluten, where X is any amino acid, the glutamine will almost invariably be deamidated by tTG. Consequently, the preference for a proline at p8 will lead to the deamidation of a glutamine at p6, a strong anchor residue. Likewise, a proline at p6 will often lead to the deamidation of a Q at p4, another strong anchor residue. In the same manner, the acceptance of proline at p5 and p3 can result in QXP-containing peptides that, when deamidated, will have a p1Glu and/or a p3Glu residue. While at p1 Glu is a highly favored anchor residue, at p3 it is well tolerated because of the positive surface electrostatic potential of the HLA-DQ2 molecule. Thus, HLA-DQ2 is capable of binding peptides that contain multiple negatively charged residues as well as multiple proline residues, at p1/3/5/6/8. These unique binding properties explain its capability to bind a large set of gluten peptides of diverse structure.

Gluten is a glutamine- and proline-rich family of proteins with related but distinct properties and sequences. It is well established that the high proline content renders gluten resistant to proteolytic degradation in the gastrointestinal tract. Given that the average daily gluten consumption ranges between 10 to 15 grams it is clear that in HLA-predisposed individuals HLA-DQ-gluten peptide complexes may almost continuously be formed in the small intestine. While under normal circumstances tolerance will be maintained, a perturbation of the homeostasis in the intestine may allow the development of an inflammatory T cell responses. This is in line with the observation that HLA-DQ2 homozygous individuals, expressing higher numbers of HLA-DQ2 dimers on their antigen presenting cells compared to HLA-DQ2 heterozygous individuals, have a much higher risk of developing celiac disease [19,20]. This implies that blocking of gluten presentation by HLA-DQ2 molecules might be an effective way to treat patients or prevent disease. Similarly, such compounds may be of use for the prevention of type I diabetes. The first steps towards generating such blocking compounds have already been undertaken [21] but it is clear that in order to obtain a therapeutically useful agent a substantial improvement will be required. The natural ligands for HLA-DQ2 identified in the present study provide improved insight in the peptide-binding properties, which may be of crucial importance for the development of an HLA-DQ2 blocker.

**Suplementary data for this article is available at:**
http://celiac-disease-consortium.nl/data1.htm

## ACKNOWLEDGEMENTS

We thank Bart Roep and Jeroen van Bergen for critical reading of the manuscript, as well as Jan Wouter Drijfhout and Willemien Benckhuijsen for peptide synthesis.

## GRANTS

This study was supported by the Netherlands Organization for Scientific Research (grant 912-02-028), the Celiac Disease Consortium, an Innovative Cluster approved by the Netherlands Genomics Initiative and partially funded by the Dutch Government (BSIK03009), and the Centre for Medical Systems Biology, a center of excellence approved by the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research.

## REFERENCES

1. Sollid, L. M. and E. Thorsby. 1993. HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis. *Gastroenterology* 105:910-922.
2. Molberg, O., S. N. McAdam, R. Korner, H. Quarsten, C. Kristiansen, L. Madsen, L. Fugger, H. Scott, O. Noren, P. Roepstorff, K. E. Lundin, H. Sjostrom, and L. M. Sollid. 1998. Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nat.Med.* 4:713-717.
3. van de Wal, Y., Y. Kooy, P. van Veelen, S. Pena, L. Mearin, G. Papadopoulos, and F. Koning. 1998. Selective deamidation by tissue transglutaminase strongly enhances gliadin-specific T cell reactivity. *J.Immunol.* 161:1585-1588.
4. Klitz, W., M. Maiers, S. Spellman, L. A. Baxter-Lowe, B. Schmeckpeper, T. M. Williams, and M. Fernandez-Vina. 2003. New HLA haplotype frequency reference standards: high-resolution and large sample typing of HLA DR-DQ haplotypes in a sample of European Americans. *Tissue Antigens* 62:296-307.
5. Schipper, R. F., G. M. Schreuder, J. D'Amaro, and M. Oudshoorn. 1996. HLA gene and haplotype frequencies in Dutch blood donors. *Tissue Antigens* 48:562-574.
6. van de Wal, Y., Y. M. C. Kooy, J. W. Drijfhout, R. Amons, and F. Koning. 1996. Peptide binding characteristics of the coeliac disease-associated DQ(alpha1*0501, beta1*0201) molecule. *Immunogenetics* 44:246-253.
7. Vartdal, F., B. H. Johansen, T. Friede, C. J. Thorpe, S. Stevanovic, J. E. Eriksen, K. Sletten, E. Thorsby, H. G. Rammensee, and L. M. Sollid. 1996. The peptide binding motif of the disease associated HLA-DQ (alpha 1* 0501, beta 1* 0201) molecule. *Eur.J.Immunol.* 26:2764-2772.
8. Shevchenko, A., M. Wilm, O. Vorm, and M. Mann. 1996. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal.Chem.* 68:850-858.

9. Stepniak, D., L. W. Vader, Y. Kooy, P. A. van Veelen, A. Moustakas, N. A. Papandreou, E. Eliopoulos, J. W. Drijfhout, G. K. Papadopoulos, and F. Koning. 2005. T-cell recognition of HLA-DQ2-bound gluten peptides can be influenced by an N-terminal proline at p-1. *Immunogenetics* 57:8-15.
10. Kim, C. Y., H. Quarsten, E. Bergseng, C. Khosla, and L. M. Sollid. 2004. Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. *Proc.Natl.Acad.Sci.U.S.A* 101:4175-4179.
11. Lee, K. H., K. W. Wucherpfennig, and D. C. Wiley. 2001. Structure of a human insulin peptide-HLA-DQ8 complex and susceptibility to type 1 diabetes. *Nat.Immunol.* 2:501-507.
12. Zavala-Ruiz, Z., I. Strug, M. W. Anderson, J. Gorski, and L. J. Stern. 2004. A polymorphic pocket at the P10 position contributes to peptide binding specificity in class II MHC proteins. *Chem.Biol.* 11:1395-1402.
13. De Oliveira, D. B., E. Harfouch-Hammoud, H. Otto, N. A. Papandreou, L. J. Stern, H. Cohen, B. O. Boehm, J. Bach, S. Caillat-Zucman, T. Walk, G. Jung, E. Eliopoulos, G. K. Papadopoulos, and P. M. van Endert. 2000. Structural analysis of two HLA-DR-presented autoantigenic epitopes: crucial role of peripheral but not central peptide residues for T-cell receptor recognition. *Mol.Immunol.* 37:813-825.
14. Moustakas, A. K., W. Y. van De, J. Routsias, Y. M. Kooy, P. van Veelen, J. W. Drijfhout, F. Koning, and G. K. Papadopoulos. 2000. Structure of celiac disease-associated HLA-DQ8 and non-associated HLA- DQ9 alleles in complex with two disease-specific epitopes. *Int.Immunol.* 12:1157-1166.
15. Fremont, D. H., D. Monnaie, C. A. Nelson, W. A. Hendrickson, and E. R. Unanue. 1998. Crystal structure of I-Ak in complex with a dominant epitope of lysozyme. *Immunity*. 8:305-317.
16. Sollid, L. M. and B. Jabri. 2005. Is celiac disease an autoimmune disorder? *Curr.Opin. Immunol.* 17:595-600.
17. Thorsby, E., K. E. Lundin, K. S. Ronningen, L. M. Sollid, and F. Vartdal. 1989. Molecular basis and functional importance of some disease-associated HLA polymorphisms. *Tissue Antigens* 34:39-49.
18. van de Wal, Y., Y. M. Kooy, J. W. Drijfhout, R. Amons, G. K. Papadopoulos, and F. Koning. 1997. Unique peptide binding characteristics of the disease-associated DQ(alpha 1*0501, beta 1*0201) vs the non-disease-associated DQ(alpha 1*0201, beta 1*0202) molecule. *Immunogenetics* 46:484-492.
19. Mearin, M. L., I. Biemond, A. S. Pena, I. Polanco, C. Vazquez, G. T. Schreuder, R. R. de Vries, and J. J. van Rood. 1983. HLA-DR phenotypes in Spanish coeliac children: their contribution to the understanding of the genetics of the disease. *Gut* 24:532-537.
20. Vader, W., D. Stepniak, Y. Kooy, L. Mearin, A. Thompson, J. J. van Rood, L. Spaenij, and F. Koning. 2003. The HLA-DQ2 gene dose effect in celiac disease is directly related to the magnitude and breadth of gluten-specific T cell responses. *Proc.Natl.Acad. Sci.U.S.A* 100:12390-12395.
21. Choi, K., M. Siegel, J. L. Piper, L. Yuan, E. Cho, P. Strnad, B. Omary, K. M. Rich, and C. Khosla. 2005. Chemistry and biology of dihydroisoxazole derivatives: selective inhibitors of human transglutaminase 2. *Chem.Biol.* 12:469-475.

22. Arentz-Hansen, H., R. Korner, O. Molberg, H. Quarsten, W. Vader, Y. M. Kooy, K. E. Lundin, F. Koning, P. Roepstorff, L. M. Sollid, and S. N. McAdam. 2000. The intestinal T cell response to alpha-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase. *J.Exp.Med.* 191:603-612.
23. Arentz-Hansen, H., S. N. McAdam, O. Molberg, B. Fleckenstein, K. E. Lundin, T. J. Jorgensen, G. Jung, P. Roepstorff, and L. M. Sollid. 2002. Celiac lesion T cells recognize epitopes that cluster in regions of gliadins rich in proline residues. *Gastroenterology* 123:803-809.
24. Vader, W., Y. Kooy, P. van Veelen, A. De Ru, D. Harris, W. Benckhuijsen, S. Pena, L. Mearin, J. W. Drijfhout, and F. Koning. 2002. The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. *Gastroenterology* 122:1729-1737.
25. Sjostrom, H., K. E. Lundin, O. Molberg, R. Korner, S. N. McAdam, D. Anthonsen, H. Quarsten, O. Noren, P. Roepstorff, E. Thorsby, and L. M. Sollid. 1998. Identification of a gliadin T-cell epitope in coeliac disease: general importance of gliadin deamidation for intestinal T-cell recognition. *Scand.J.Immunol.* 48:111-115.
26. Vader, L. W., A. De Ru, W. Y. van der, Y. M. Kooy, W. Benckhuijsen, M. L. Mearin, J. W. Drijfhout, P. van Veelen, and F. Koning. 2002. Specificity of tissue transglutaminase explains cereal toxicity in celiac disease. *J.Exp.Med.* 195:643-649.
27. Qiao, S. W., E. Bergseng, O. Molberg, G. Jung, B. Fleckenstein, and L. M. Sollid. 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.
28. Vader, L. W., D. T. Stepniak, E. M. Bunnik, Y. M. Kooy, W. de Haan, J. W. Drijfhout, P. A. van Veelen, and F. Koning. 2003. Characterization of cereal toxicity for celiac disease patients based on protein homology in grains. *Gastroenterology* 125:1105-1113.
29. Arentz-Hansen, H., B. Fleckenstein, O. Molberg, H. Scott, F. Koning, G. Jung, P. Roepstorff, K. E. Lundin, and L. M. Sollid. 2004. The molecular basis for oat intolerance in patients with celiac disease. *PLoS.Med.* 1:e1.