

# APPLY NOW

2021 Microbial Pathogenesis in Alzheimer's Disease Grant  
Applications Close September 21

✓ Up to \$1.7 million in grant funding

✓ Individual grants of up to \$250,000

IDSA 20  
EXCELLENCE



## Large-Scale Characterization of Natural Ligands Explains the Unique Gluten-Binding Properties of HLA-DQ2

This information is current as of September 20, 2021.

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

*J Immunol* 2008; 180:3268-3278; ;  
doi: 10.4049/jimmunol.180.5.3268  
<http://www.jimmunol.org/content/180/5/3268>

**References** This article **cites 33 articles**, 10 of which you can access for free at:  
<http://www.jimmunol.org/content/180/5/3268.full#ref-list-1>

Why *The JI*? [Submit online.](#)

- **Rapid Reviews! 30 days\*** from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Fast Publication!** 4 weeks from acceptance to publication

*\*average*

**Subscription** Information about subscribing to *The Journal of Immunology* is online at:  
<http://jimmunol.org/subscription>

**Permissions** Submit copyright permission requests at:  
<http://www.aai.org/About/Publications/JI/copyright.html>

**Email Alerts** Receive free email-alerts when new articles cite this article. Sign up at:  
<http://jimmunol.org/alerts>

*The Journal of Immunology* is published twice each month by  
The American Association of Immunologists, Inc.,  
1451 Rockville Pike, Suite 650, Rockville, MD 20852  
Copyright © 2008 by The American Association of  
Immunologists All rights reserved.  
Print ISSN: 0022-1767 Online ISSN: 1550-6606.



# Large-Scale Characterization of Natural Ligands Explains the Unique Gluten-Binding Properties of HLA-DQ2<sup>1</sup>

Dariusz Stepniak,\* Martina Wiesner,\* Arnaud H. de Ru,\*<sup>†</sup> Antonis K. Moustakas,<sup>‡</sup> Jan Wouter Drijfhout,\* George K. Papadopoulos,<sup>§</sup> Peter A. van Veelen,\*<sup>†</sup> and Frits Koning<sup>2\*</sup>

Celiac disease is an enteropathy caused by intolerance to dietary gluten. The disorder is strongly associated with DQA1\*0501/DQB1\*0201 (HLA-DQ2) as ~95% of celiac patients express this molecule. HLA-DQ2 has unique Ag-binding properties that allow it to present a diverse set of gluten peptides to gluten-reactive CD4<sup>+</sup> T cells so instigating an inflammatory reaction. Previous work has indicated that the presence of negatively charged amino acids within gluten peptides is required for specific binding. This, however, only partly explains the scale of the interaction. We have now characterized 432 natural ligands of HLA-DQ2 representing length variants of 155 distinct sequences. The sequences were aligned and the binding cores were inferred. Analysis of the amino acid distribution of these cores demonstrated that negatively charged residues in HLA-DQ2-bound peptides are favored at virtually all positions. This contrasts with a more restricted presence of such amino acids in T cell epitopes from gluten. Yet, HLA-DQ2 was also found to display a strong preference for proline at several anchor and nonanchor positions that largely match the position of proline in gluten T cell epitopes. Consequently, the bias for proline at p6 and p8 facilitates the enzymatic conversion of glutamine into glutamic acid in gluten peptides at p4 and p6, two important anchor sites. These observations provide new insights in the unique ability of HLA-DQ2 to bind a large repertoire of glutamine- and proline-rich gluten peptides. This knowledge may be an important asset in the development of future treatment strategies. *The Journal of Immunology*, 2008, 180: 3268–3278.

Referred to as HLA-DQ2, DQA1\*0501/DQB1\*0201 is the strongest genetic factor predisposing to celiac disease (1). HLA-DQ2 dimers were shown to present gluten-derived peptides to gluten-specific CD4<sup>+</sup> T cells in the intestinal lamina propria, thereby inducing and sustaining strong inflammatory responses (2, 3). It is well-established that HLA-DQ2 can bind and present a relatively large and diverse repertoire of gluten-derived peptides. The association of HLA-DQ2 with celiac disease is extraordinarily strong: ~95% of celiac disease patients express HLA-DQ2, whereas in the general population, the frequency of this allele is between 20 and 30% (4). Virtually all HLA-DQ2-negative patients are HLA-DQ8 positive and it has been described that several gluten-derived peptides can bind to HLA-DQ8 and elicit T cell responses (5, 6). Together, these data indicate that the ability to present gluten-derived peptides is a unique feature of

HLA-DQ2 and HLA-DQ8 molecules. 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 (7, 8). Gluten, however, hardly contains negatively charged amino acids. This discrepancy was solved when it became clear that the glutamine- and proline-rich gluten molecules are substrates for the enzyme tissue transglutaminase. This enzyme can convert glutamine into glutamic acid, thereby introducing the negative charges required for binding to HLA-DQ2 (2, 3). A preference for hydrophobic and negatively charged anchor residues, however, is not unique to HLA-DQ2. HLA-DR3, for example, is known to preferentially bind peptides with an isoleucine and glutamic acid at the p1 and p4 positions, respectively. We therefore sought to further investigate the peptide-binding properties of HLA-DQ2 to understand its capacity to bind and present a large repertoire of gluten-derived peptides. To this end, 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 clarifying why this molecule is so uniquely suited to bind gluten peptides.

\*Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, The Netherlands; <sup>†</sup>Centre for Medical Systems Biology, Leiden, The Netherlands; <sup>‡</sup>Department of Organic Farming, Technological Educational Institute of Ionian Islands, Argostoli, Cephalonia, Greece; and <sup>§</sup>Laboratory of Biochemistry and Biophysics, Faculty of Agricultural Technology, Epirus Institute of Technology, Arta, Greece

Received for publication September 18, 2007. Accepted for publication December 6, 2007.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>1</sup> This work 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 was 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.

<sup>2</sup> Address correspondence and reprint requests to Dr. Frits Koning, Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, P.O. Box 9600, 2300 RC Leiden, The Netherlands. E-mail address: f.koning@lumc.nl

Copyright © 2008 by The American Association of Immunologists, Inc. 0022-1767/08/\$2.00



Table I. Overview of peptides eluted from HLA-DQ2<sup>a</sup>

Peptide Source	Sequence	Length Variants
Actin	KEITALAPSTMKIK	3
$\alpha$ tubulin	APVLSAEKAYHEQ	2
$\alpha$ tubulin	DYEEVGDADGEDEG	1
$\beta_2$ -microglobulin	HPSDIEVLLKN	2
$\beta$ tubulin	DATAEEEEDFGEEAEEE	4
$\beta$ tubulin	DPTGTYHGDSDLQLDR	2
$\beta$ -actin	EEETAAALVVDN	4
$\beta$ -actin	YVALDFEQEMA	3
$\beta$ -actin	VPIYEGYALPHAI	1
Calcium-binding protein (stromal cell-derived factor 4)	DGIVTAEELLESYMDP	3
Calmodulin	LQDMINEVDADGN	4
CD20	IEIPIQEEEE	15
CD20	IKEEVVGLTETSSQPK	1
CD74	MRMATPLLMQALPM	22
Desmocollin type 4	ADGYSADLPLPLPIR	2
Endoplasmic precursor	ESDDEAAVEEEEEK	1
Eukaryotic translation initiation factor 4A	DIETFYNTTVEEMP	1
Glyceraldehyde 3-phosphate dehydrogenase	GKVDIVAINDPFIID	2
Glyceraldehyde 3-phosphate dehydrogenase	DIQVVAINDPFIID	1
Golgi phosphoprotein	EFEEAQVREENLPDE	1
Golgi-localized phosphoprotein 130 kDa	VDEQYQEEAEEVQED	1
GP73 Golgi phosphoprotein 2	AALSVSQENPEMEGPE	2
Heat shock cognate 71-kDa protein	FDVSIILTIEDGIFE	1
Heparan sulphate 2 sulphotransferase	GPRQDADLDEEDM	5
HLA DM $\alpha$ -chain	ADWAQEQGDAPAILFDK	2
HSP-90	DPTADDTSAAVTEEMPP	2
HSP-90	IDEDEVAEEPN	3
HSP-70	KSINPDEAVAYG	1
HSP-70	TPSYVAFDTERLIG	2
Hyaluronan-binding protein	YPEDEVGQEDAEESD	1
Hypoxanthine-guanine phosphoribosyltransferase	SPGVVISDDEPGYD	2
Ig $\kappa$ -chain V-III region POM	TLTISSLQSEDFAV	1
IgE FcR 2	NLNLGLADLSSFK	7
IgG	YPSDIAVEWESNGQPE	1
IgG	FTLTISSLQSEDF	1
Importin $\beta$ -1 subunit	EADVADQEEPAT	1
Initiation factor 4A-I	GPDGMEPEGVIESNWNEIV	1
Integral membrane protein 2A	DSDPAAIITHDFE	1
Integrin $\beta_1$	DDLENVKSGLGTDLMNEM	1
IFN- $\gamma$ receptor $\alpha$	APTSTFGYDKPHVLVDL	1
IL-21 receptor	GSPLAGLDMDTFDG	1
IL-2 receptor $\beta$ -chain precursor	SPFPSSSFSPGGLAPEISPLE	1
IL-4 receptor $\alpha$ chain precursor	APVECEEEEEVEEE	1
ITM2A	REDLVAVEEIRD	1
ITM2C protein	LPQTYIIQEEEMV	1
LckBP1 p75; HCLS1; HS1	DEPEGDYEEVLEPE	1
LDL receptor	DIQAPDGLAVDWIHSN	1
Legumain	DYTGEDVTPQN	5
Legumain	GDAEAVKIGSGKVL	1
Low-affinity IgE FcR	ISQLEELRAEQQR	1
Major vault protein	GDKVVAGDEWLFEGPG	2
Mannose-6-phosphate receptor	VPAAYRGGVDDQLGE	2
MAPK-activating protein	SDTEVYGEFYVPVPP	2
MHC I	APWLEQEGPEYWDQ	28
MHC I	APWVEQEGPEYWR	7
MHC I	DTQFVRFDSDAASQ	16
MHC I	GKDYIALNEDLRS	9
MHC I	TAADTAAQITQ	11
MHC I	EPAADLIDMGPDPA	5
MHC I	FISVGYVDDTQFV	7
MHC I	LTWKRGDGEDOTQD	9
MHC II	AEYWNSQKDLLEQKR	6
MHC II	DVGEFRAVTELG	4
MHC II	SDVGEYRAVTELG	2
MHC II	DFDFGDEIFH	13
MHC II	NPDQSGEFMDFDGEFFH	2
MHC II	GKPVTTGVSETVFLPR	1
MHC II	GPSGOYTHEFD	1
MHC II	DPLKVYPPPLKGSFPENLRHL	1
MHC II	DQEETAGVVSTPLIR	1
Microsomal stress 70 protein ATPase core	DVYVGYESVELADSNPQ	1

(Table continues)

Table I. (Continued)

Peptide Source	Sequence	Length Variants
Microtubule-associated protein 1A	<u>APAKAENE</u> EAAANPAW	1
Moesin	YPEDVSEELIQDIT	1
MOP-4	APDIDGDEDLPGPP	1
Na/K-transporting ATPase	FASIVTGVVEEGR	1
NAP-1-related protein	DLDDVVEEVEEETG	5
Natural killer cell-specific antigen KLIP1	EEFVDADAADAAAA	1
NEDD4-2 minus WW2	VPEFWETISEEVN	2
Nef-associated factor 1	TSSEFEVVTPEEQ	2
Neutral amino acid transporter B	GPAGDATVASEKESVM	1
Notch 1 (TAN-1)	IEAVQSETVEPPP	1
Nucleoporin	FGTLPSPSLGNSSILK	1
Occludin	ELDDYREESEEVY	1
Phosphatidylinositol 4-kinase type-II $\beta$	EEGEAGDEELPLPPG	6
Phosphoglycerate kinase 1	LPVDFVTADKFD	2
Putative amyloid protein	GDEVVEEAEEPEYE	2
Putative NF- $\kappa$ B activating protein	SDTEVYGFYFVPPPY	1
Ras GTPase-activating-like protein	QDILNEIAKDIRNQR	1
Ras related protein Rap-2c	VPLILLVGNKVDLEPER	1
Ras-related protein Ral-B	DGEEVQIDILDTAGQE	1
Regulator of G-protein signaling 14	GGNEQKALVLD	1
Rho-GDP-dissociation inhibitor S12121	LAQIAAENEDEHHSV	1
Saposin	IVDSYLPVILDIK	5
Saposin	<u>FVAEYEPVLI</u> EIL	2
Saposin	SSLISILLEEVSP	3
Saposin	VAPFMANIPLLLYP	2
Secretory granule proteoglycan core protein	<u>DLGQHGLEEDFM</u>	13
Sequestome	EPPEAEAAAGPGP	3
Sequestome 1	MPESEGPSLLDPSQEGP	2
Sialidase 1	DGDVPDGLNLGAVVSDVETG	1
Sialyl transferase	WDILQEI <sup>a</sup> SP <sup>a</sup> E <sup>a</sup> I <sup>a</sup> Q <sup>a</sup> PNPP	1
Similar to calyculin-3 precursor	EISLVGDDLDPERE	1
Snf7 homolog associated with Alix3	EDELMAELELEEQ	1
Sodium/potassium-transporting ATPase $\alpha$ -1 chain precursor	TPIAAEIEHFIH	1
Sodium/potassium-transporting ATPase $\alpha$ -1 chain precursor	<u>SSLTGESE</u> PQTRSPD	1
Sortilin-related receptor	HPINEYYIADASEDQ	2
Sorting nexin 1	KPQPTYEELEEEQED	2
STAM	EPEPEPAFIDEDKMDQ	1
TGF- $\beta$ 1	DRVAGESAEPEPEPE	5
TNF- $\alpha$	LPDGPAGSWEQLIQER	1
TRAF 6-binding protein	KPSPSAAEADFD	3
Transferrin receptor A chain	IQGAANALSGDVWDIDN	1
Transferrin receptor A chain	AAAEVAGQFVIKLTHD	1
Transferrin receptor A chain	IPVQTI <sup>a</sup> SR <sup>a</sup> AAEKLF	2
Transferrin receptor A chain	SAGDFGSV <sup>a</sup> GATEWL	3
Transformation/transcription domain associated protein	EIDIELAPGDQSTSP	1
Translation initiation factor IF1H	GPDDDEIQFDDIGDDDED	1
Tripeptidyl-peptidase I precursor	EPFLITN <sup>a</sup> EIVDYLSGG	2
Tripeptidyl-peptidase I precursor	SELVQAVSDPSSP	1
Ubiquitin and ribosomal protein	GKTI <sup>a</sup> TLEVEPSDT	4
Ubiquitin-conjugating enzyme E2 N	EPVPGIKAE <sup>a</sup> PDES <sup>a</sup> N	2
Vacuolar ATP synthase subunit E	DVDVQIDQESYLPE	1
Vacuolar protein sorting protein	LPDEGEPTDEETDGDISDSMD	1
WB syndrome protein G	GEPDYVNGEVAATEA	1

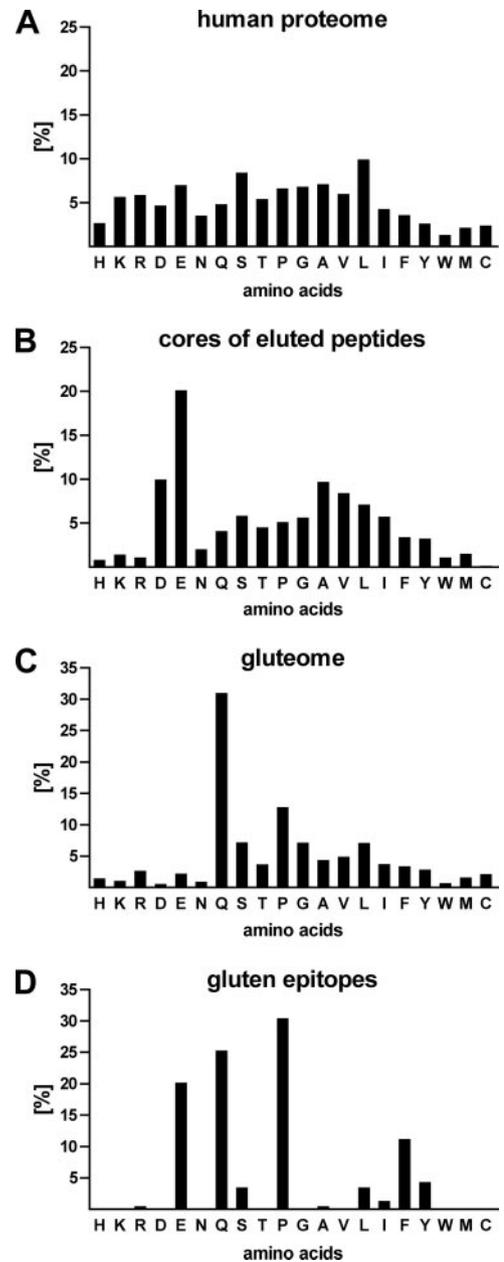
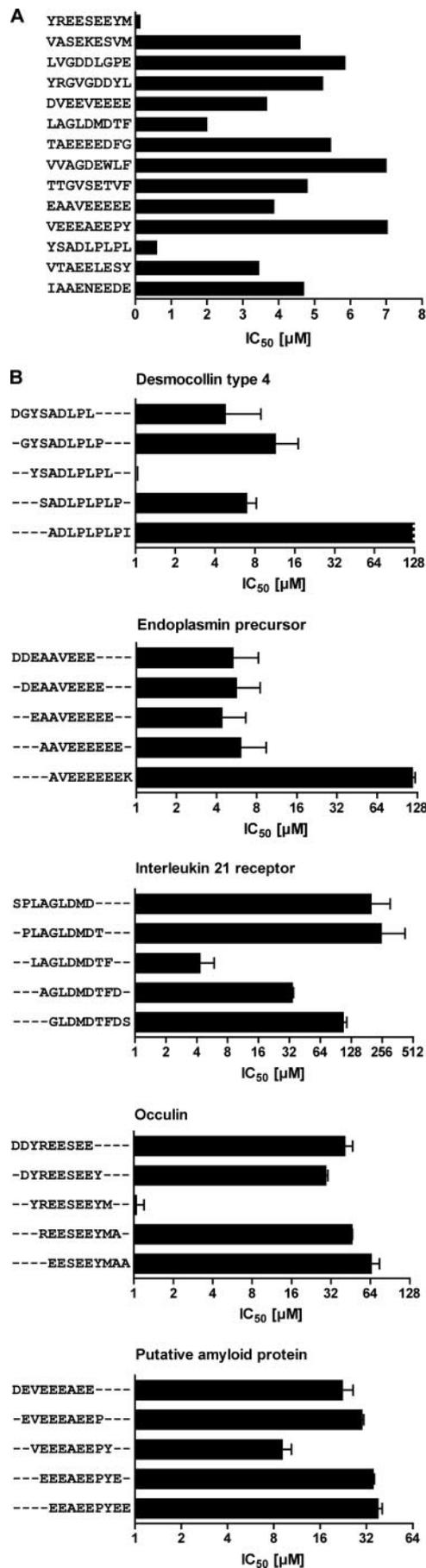
<sup>a</sup> The putative binding cores are underlined. The average pI of the 9-mer core sequences was 4.1 (SD  $\pm$  0.9).

$\mu$ l/well of enhancement solution (Wallac) was applied and the plates were read in a time-resolved fluorometer (1234; Wallac) 15–30 min thereafter. IC<sub>50</sub> 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 (11). In brief, a hybrid HLA-DQ2 molecule was constructed containing the  $\alpha\beta$ -chains from the crystal structure of HLA-DQ2 (12) and the peptide backbone coordinates from HLA-DQ8 (13). This step was necessary be-

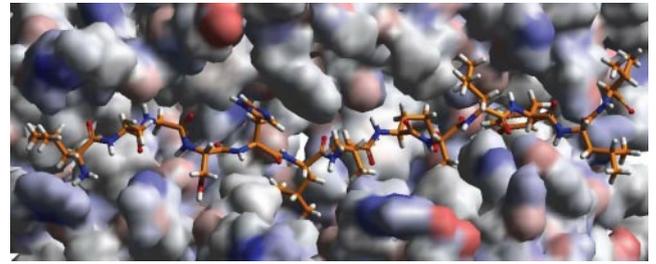
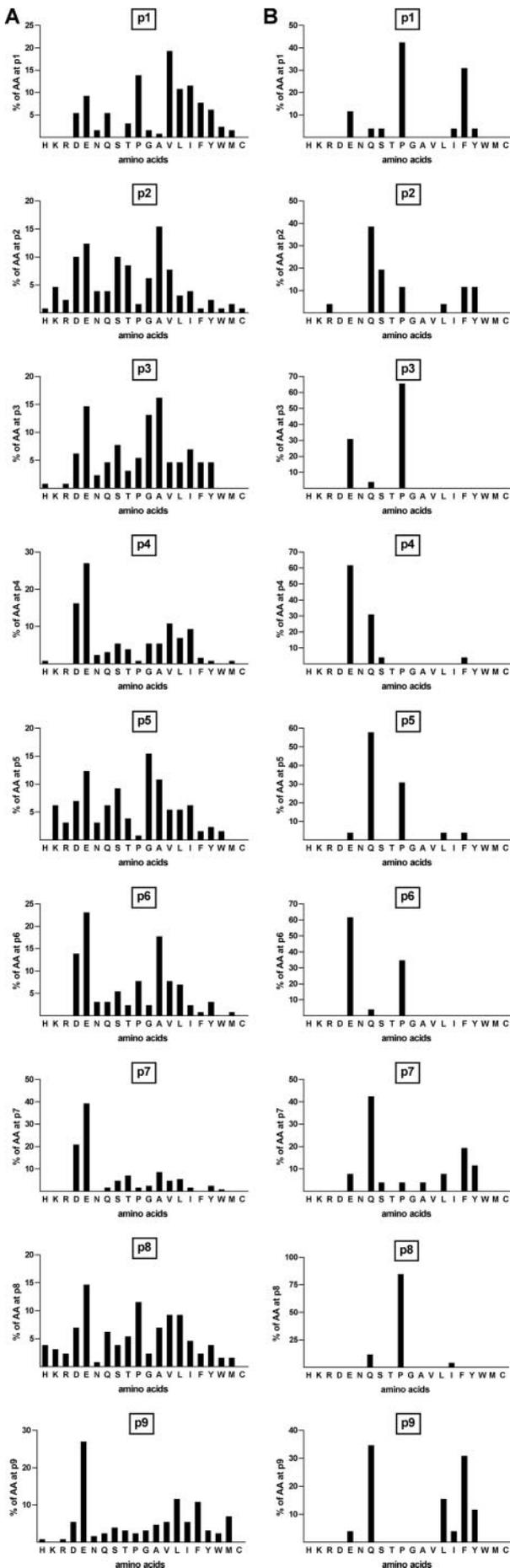
cause the antigenic peptide in the HLA-DQ2 crystal structure has its p9 pocket (the most spacious) unoccupied and its p9Tyr residue occupies the so-called p10 pocket (shelf) (14). Previous binding experiments indicated that this pocket is very spacious, accommodating aromatic and large aliphatic amino acids and secondarily acidic ones. This selectivity would not be exhibited by a residue occupying the p10 pocket. The peptides used were 13-mer 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 with that occupying pocket 6 in the crystal structure of the complex of DQ2 and the gliadin peptide (12). The molecular simulation was conducted by the program Discover of Accelrys



**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).

with 1000 steps of the steepest gradient method followed by 1000 steps of the conjugate gradient method (15, 16). Numbering of the HLA-DQ $\alpha\beta$  residues follows the scheme suggested by Fremont et al. (17), as this provides structural equivalence for residues from different MHC class II (*MHC II*) gene loci independent of animal species. Graphical representations of the results were performed with the program WebLabViewer (version 3.5; Accelrys).

**FIGURE 2.** Binding of 9-aa core sequences to HLA-DQ2. The ability of 14 predicted core sequences to bind to HLA-DQ2 (A). The correctness of the binding core prediction as tested in the binding studies of the truncated peptides (B). Mean values of three independent experiments are shown. The IC<sub>50</sub> values were calculated based on the observed competition between the biotin-labeled indicator peptide (QLQFPQPELPY) and the test peptide. IC<sub>50</sub> represents the concentration of the test peptide that is required for 50% inhibition of the binding of the indicator peptide.



**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; and positive, blue). Note the positive charges in and around the binding groove.

**Results**

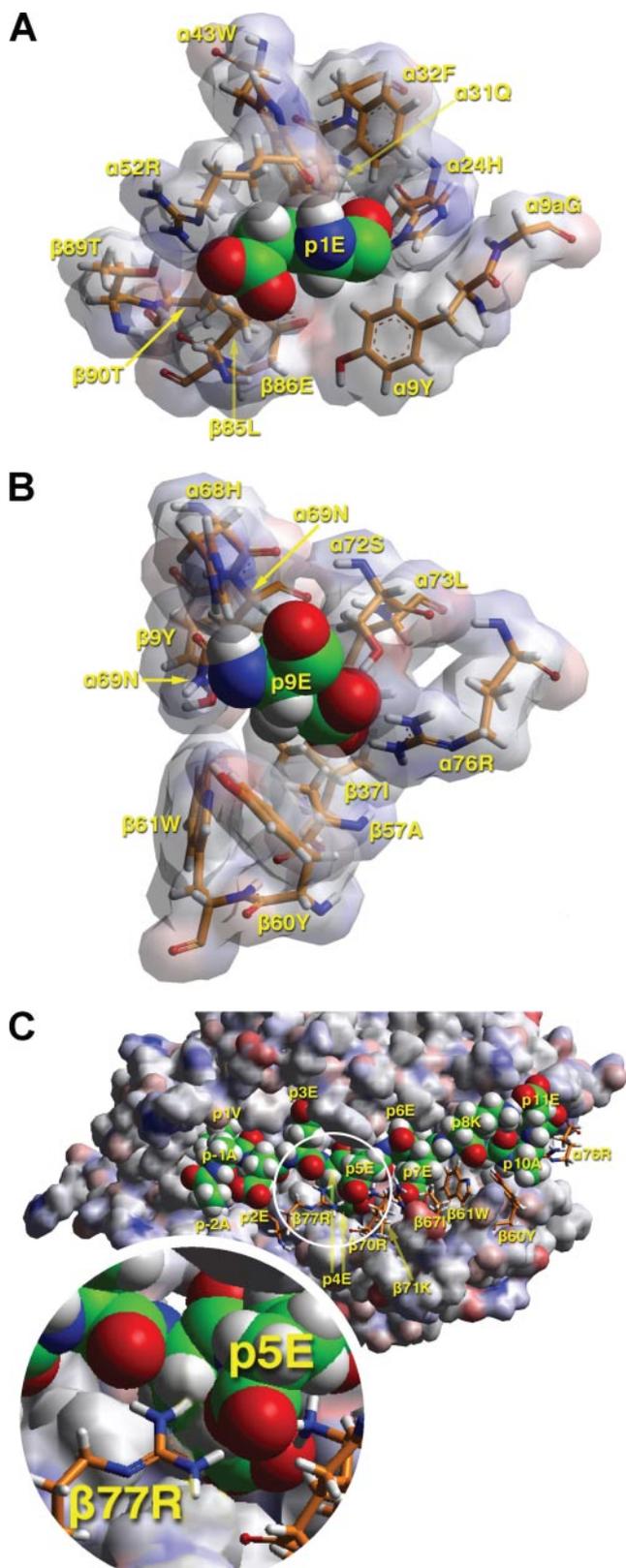
*Isolation and sequencing of HLA-DQ2-bound peptides*

Human HLA-DQ2 dimers were purified from HLA-homozygous EBV-transformed B cells and the HLA-DQ2-bound peptides were eluted with acid (see *Materials and Methods*). Subsequently, filtration was used to separate peptides from the protein fraction. Analysis of the high molecular mass material by SDS-PAGE revealed the presence of two bands corresponding to the predicted molecular mass of HLA-DQ2  $\alpha$ - and  $\beta$ -chains (respectively, 33 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 MS. Several protein fragments corresponding to sequences present in the HLA-DQ2  $\alpha$ - and  $\beta$ -chain were identified, demonstrating that HLA-DQ2 was successfully purified (data not shown). Next, the low molecular mass material that was eluted from the HLA-DQ2 was fractionated by reverse-phase HPLC and the peptides present in the obtained fractions were sequenced by tandem MS. This resulted in the identification of 432 peptides that represent length variants of 155 unique peptides.

*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-aa sequence common to all length variants (Fig. 1A), indicating that this is the binding core. This is in line with previous studies (7, 8). In a similar way, the alignment of the length variants of a CD20-derived peptide implies that the sequence IPIQEEEEEE represents the binding core (Fig. 1B). Using this method, the core sequences could be determined for several other peptides. In most cases, however, fewer length variants were identified so that the minimal cores could not be determined this way. The putative minimal cores of those 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. Through this procedure, a single minimal core could be identified for 124 of 155 peptides (Table I). For the remainder of the

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



**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; and 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*, TCR view of pocket 1 of the complex between peptide AGEEGEAGDEELP (anchors in bold) and HLA-DQ2. The glutamate residue at p1 forms a salt

bridge with the terminal guanidine group of  $\alpha 52$ Arg. In addition, the  $\beta 86$ Glu forms hydrogen bonds with  $\alpha 31$ Gln (not seen because of p1Glu) and  $\alpha 9$ Tyr. 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,  $\beta 89$ T,  $\beta 90$ T, and  $\alpha 52$ R). *B*, TCR view of pocket 9 of the complex between peptide EIIPQEEEEET and HLA-DQ2. The glutamate at p9 forms a salt bridge with  $\alpha 76$ Arg and a hydrogen bond with  $\alpha 72$ Ser. *C*, TCR view of the complex between peptide AAVEEEEEKVAE (anchors in bold) and HLA-DQ2. *Inset*, Note how  $\beta 77$ Arg swings into position to form a salt bridge with p5Glu in the bound peptide.

#### Peptide binding to HLA-DQ2 of identified core sequences

To verify that the above approach correctly identified the 9-aa binding cores, we selected 14 9-aa core sequences and tested their ability to bind to HLA-DQ2 in a cell-free binding assay (Fig. 2A). For all those peptides,  $IC_{50}$  values ranged from 0.2 to 7  $\mu$ M, which is consistent with the values previously found for other HLA-DQ2 ligands (7, 18). To further ensure the correctness of the identified 9-aa cores we tested the ability of N- and C-terminally truncated versions of the 9-aa cores to bind to HLA-DQ2 (Fig. 2B). In four of five cases, this explicitly confirmed that the identified cores represent the only binding frame. The fifth peptide (endoplasmic precursor), with the highly acidic sequence DDEAAVEEEEE, appeared to bind also in frames other than the predicted one. This binding can be explained by the fact that HLA-DQ2 accepts negatively charged amino acids along the entire p1-p9 binding core (see below). The results of the binding studies were used for further refinement of our core prediction algorithm. Further confirmation of the accuracy of the approach was obtained by molecular simulation of HLA-DQ2 binding with several other ligand peptides (data not shown).

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

Based on the alignment of the 124 minimal cores, the frequency of amino acids was calculated and compared with that of the human proteome (Fig. 3, A and 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. Although 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 EEEEEEEAEEEEEE derived from glucosidase II. Strikingly, a preference for aspartate and particularly glutamate was found not only at the p4, p6, and p7 anchor positions, but also at virtually all positions in the binding core including the non-anchor positions p2, p3, p5, and p8 (Fig. 4).

To explain this preference for negatively charged amino acids at both anchor and nonanchor 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) (Fig. 5 and data not shown). 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 to analyze the molecular basis for the acceptance of a negatively charged amino acid at p1, p5, and p9,

bridge with the terminal guanidine group of  $\alpha 52$ Arg. In addition, the  $\beta 86$ Glu forms hydrogen bonds with  $\alpha 31$ Gln (not seen because of p1Glu) and  $\alpha 9$ Tyr. 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,  $\beta 89$ T,  $\beta 90$ T, and  $\alpha 52$ R). *B*, TCR view of pocket 9 of the complex between peptide EIIPQEEEEET and HLA-DQ2. The glutamate at p9 forms a salt bridge with  $\alpha 76$ Arg and a hydrogen bond with  $\alpha 72$ Ser. *C*, TCR view of the complex between peptide AAVEEEEEKVAE (anchors in bold) and HLA-DQ2. *Inset*, Note how  $\beta 77$ Arg swings into position to form a salt bridge with p5Glu in the bound peptide.

Table II. Overview of sequences derived from gluten and homologous proteins known to stimulate responses of T cells from celiac patients

No	Sequence <sup>a</sup>											Peptide Name					Reference		
				<b>1</b>	2	3	<b>4</b>	5	<b>6</b>	<b>7</b>	8	<b>9</b>							
1				<b>P</b>	Q	P	<b>E</b>	L	<b>P</b>	<b>Y</b>	P	<b>Q</b>	P	E	L	P	Y	Glia- $\alpha$ 2 ( $\alpha$ -II)	28
2	Q	L	Q	<b>P</b>	F	P	<b>Q</b>	P	<b>E</b>	<b>L</b>	P	<b>Y</b>						Glia- $\alpha$ 9 ( $\alpha$ -I)	28
3	P	Q	L	<b>P</b>	Y	P	<b>Q</b>	P	<b>E</b>	<b>L</b>	P	<b>Y</b>						Glia- $\alpha$ 9 ( $\alpha$ -III)	29
4				<b>F</b>	R	P	<b>E</b>	Q	<b>P</b>	<b>Y</b>	P	<b>Q</b>	P	Q	P			Glia- $\alpha$ 20	24
5	Q	E	Q	<b>Q</b>	X	P	<b>E</b>	Q	<b>P</b>	<b>Q</b>	Q	<b>F</b>						Glu-5	24
6	P	E	Q	<b>P</b>	Q	Q	<b>S</b>	F	<b>P</b>	<b>E</b>	Q	<b>E</b>	R	P	F			$\gamma$ -I (Glia- $\gamma$ 1)	30
7	E	G	I	<b>I</b>	Q	P	<b>E</b>	Q	<b>P</b>	<b>A</b>	Q	<b>L</b>						$\gamma$ -II (Glia- $\gamma$ 30)	23, 31
8		F	P	<b>E</b>	Q	P	<b>E</b>	Q	<b>P</b>	<b>Y</b>	P	<b>Q</b>	Q					$\gamma$ -III	29
9			F	<b>S</b>	Q	P	<b>E</b>	Q	<b>E</b>	<b>F</b>	P	<b>Q</b>	P	Q				$\gamma$ -IV	29
10			P	<b>E</b>	Q	P	<b>F</b>	P	<b>E</b>	<b>Q</b>	P	<b>Q</b>	Q					$\gamma$ -VI	23
11				<b>P</b>	Q	P	<b>E</b>	Q	<b>E</b>	<b>F</b>	P	<b>Q</b>						$\gamma$ -VII <sup>A</sup>	23
12				<b>E</b>	Q	P	<b>E</b>	Q	<b>P</b>	<b>F</b>	P	<b>Q</b>						$\gamma$ -VII <sup>B</sup>	23
13	Q	P	P	<b>F</b>	S	E	<b>E</b>	Q	<b>E</b>	<b>S</b>	P	<b>F</b>	S	Q				Glt-156	24
14	Q	P	P	<b>F</b>	S	E	<b>E</b>	Q	<b>E</b>	<b>Q</b>	P	<b>L</b>	P	Q				Glt-17 <sup>A</sup>	24
15	Q	E	F	<b>P</b>	Q	P	<b>E</b>	Q	<b>P</b>	<b>F</b>	P	<b>Q</b>	Q	P				Hor- $\alpha$ 2	32
16	Q	P	F	<b>P</b>	Q	P	<b>E</b>	Q	<b>P</b>	<b>F</b>	P	<b>Q</b>	S	Q				Sec- $\alpha$ 2	32
17	P	Q	Q	<b>P</b>	F	P	<b>Q</b>	P	<b>E</b>	<b>Q</b>	P	<b>F</b>	R	Q				Hor- $\alpha$ 9	32
18	P	Q	Q	<b>P</b>	F	P	<b>Q</b>	P	<b>E</b>	<b>Q</b>	P	<b>F</b>	P	Q				Sec- $\alpha$ 9	32
19		Y	Q	<b>P</b>	Y	P	<b>E</b>	Q	<b>E</b>	<b>E</b>	P	<b>F</b>	V	Q				Av- $\alpha$ 9A	32, 33
20		Y	Q	<b>P</b>	Y	P	<b>E</b>	Q	<b>E</b>	<b>Q</b>	P	<b>F</b>	V	Q				Av- $\alpha$ 9B	32, 33
21	Q	P	P	<b>F</b>	S	E	<b>E</b>	Q	<b>E</b>	<b>Q</b>	P	<b>L</b>	P					Glt-156 homolog	24
22	Q	P	P	<b>F</b>	S	E	<b>E</b>	Q	<b>E</b>	<b>Q</b>	P	<b>I</b>	L	L				Glt-156 homolog	24
23	Q	P	P	<b>F</b>	S	E	<b>Q</b>	E	<b>Q</b>	<b>P</b>	I	<b>L</b>	P	Q				Glt-156 homolog	24
24	Q	Q	P	<b>F</b>	P	Q	<b>Q</b>	P	<b>Q</b>	<b>Q</b>	P	<b>F</b>	P	Q				Glia- $\gamma$ 2 <sup>A</sup>	11
25	Q	Q	P	<b>Y</b>	P	Q	<b>Q</b>	P	<b>Q</b>	<b>Q</b>	P	<b>F</b>	P	Q				Glia- $\gamma$ 2 <sup>B</sup>	11
26	Q	Q	P	<b>F</b>	P	Q	<b>Q</b>	P	<b>Q</b>	<b>Q</b>	P	<b>Y</b>	P	Q				—	31

<sup>a</sup> Anchor positions are given in bold.

three positions previously unknown to be involved in harboring such amino acids. The molecular simulation of peptide AGEEGEAGDEELP in HLA-DQ2 indicates that the p1Glu forms a salt bridge with the terminal guanidine group of  $\alpha$ 52Arg (Fig. 6A). Because 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.,  $\beta$ 89Thr,  $\beta$ 90Thr, and  $\alpha$ 52Arg. Interestingly,  $\alpha$ 52Arg 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 AAVEEEEEKVAE indicates that the presence of multiple acidic residues in the binding core of the bound peptide attracts the  $\beta$ 77Arg residue toward the p5 acidic residue (Fig. 6C). Thus, besides the residues  $\beta$ 28Ser,  $\beta$ 30Ser,  $\beta$ 70Arg,  $\beta$ 71Lys, and  $\beta$ 74Ala, that are very accommodating to acidic residues at p4, p6, and p7,  $\beta$ 77Arg functions as an additional attractor and stabilizer of very negatively charged acidic peptides bound to HLA-DQ2.

Finally, the modeling of EIPIQEEEEET demonstrated that the p9Glu forms an ionic bond with  $\alpha$ 76Arg and a hydrogen bond with  $\alpha$ 72Ser (Fig. 6B). These strong interactions likely explain the presence of glutamate at the p9 position in roughly 30% of the peptides found in HLA-DQ2.

To determine whether the preference of HLA-DQ2 for negatively charged amino acids 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 II). Large discrepancies in the abundance of certain amino acids were found between these three groups of sequences (Fig. 3, B–D). To a large extent, this can be

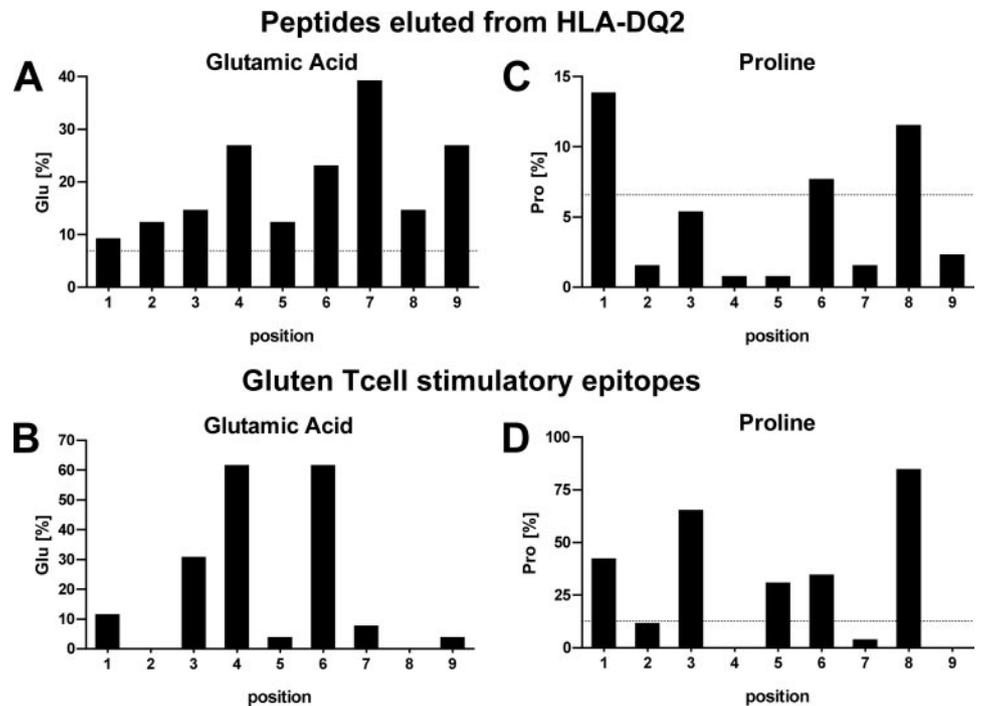
explained by the restricted amino acid composition of gluten and gluten-like molecules, which are strongly dominated by glutamine, proline, and hydrophobic amino acids. Moreover, while gluten lacks negatively charged residues, these are prominent in HLA-DQ2-restricted, T cell stimulatory gluten peptides due to the activity of tissue transglutaminase, which converts glutamine residues into glutamic acid, a prerequisite for high-affinity HLA-DQ2 binding.

Strikingly, 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 nonanchor position p3. Also, the p7-binding pocket, which has the highest preference for acidic residues, was found to mostly accommodate glutamine and aromatic amino acids in the case of gluten epitopes (Fig. 4). Similarly, glutamate was hardly found at the p2, p5, and p8 nonanchor 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 nonanchor 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 biased at both anchor and nonanchor 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. 3, A and 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 nonanchor 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.

**FIGURE 7.** Patterns of 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. *B* and *D*, The relative abundance of the indicated amino acids in the binding cores of cereal-derived T cell stimulatory sequences. The dotted line indicates the abundance of the amino acids in the sources, i.e., human proteome for the eluted peptides and the gluteome for the T cell epitopes. There is no dotted line in *B* because most of the glutamate residues in the T cell epitopes stem from the deamidation of glutamine.



It is well-established that all known HLA-DQ2-restricted gluten-derived T cell stimulatory peptides contain multiple proline residues (Table II). As noted previously, proline residues in these gluten peptides are usually found at positions p1, p3, p5, p6, and/or p8 (Fig. 7D), positions at which HLA-DQ2 favors or tolerates the presence of proline (Fig. 7C). Remarkably, the alignment of known T cell stimulatory peptides from gluten and gluten homologs in barley, rye, and oats indicates that proline is present at position p8 in 22 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 Ag 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 (19), and where roughly 95% of the patients carry HLA-DQ2, the remainder is HLA-DQ8 positive. There is strong evidence that the capacity of these HLA molecules to bind and present gluten peptides to CD4<sup>+</sup> T cells in the small intestine is a prerequisite for the development of celiac disease (20).

Peptide motifs specific for a number of *HLA* alleles have been defined and can be used to identify potential T cell epitopes in proteins. A peptide-binding motif for HLA-DQ2 was first described in 1996 (7, 8). 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 while gluten lacks such negatively charged residues. The observation that gluten-derived peptides undergo an enzymatic modification in the intestine, in which glutamine is converted into the negatively charged glutamic acid due to the activity of tissue transglutaminase, explained this apparent discrepancy (2, 3).

In the present study, we have investigated whether the preference of HLA-DQ2 for negative charges could fully explain the ability of HLA-DQ2 to bind and present a wide range of gluten-

derived peptides. For this purpose, we have characterized a large number of peptides that are naturally bound to HLA-DQ2. Our study revealed that HLA-DQ2 accepts negatively charged residues at the entire p1-p9 register. This can be mainly explained by the strategically positioned polar or positively charged residues in all the five binding pockets (Fig. 6). Additionally, 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 and data not shown). 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 positions in these peptides (Fig. 7B). Also, peptides containing acidic residues can be presented by other HLA haplotypes whereas only HLA-DQ2, and to much lower extent HLA-DQ8, are associated with celiac disease. A more plausible explanation for the exceptional gluten peptide-binding properties is offered by the unique ability of HLA-DQ2 to accommodate proline-rich peptides while gluten is known to be exceptionally proline rich. In a recent study, the structure of the complex between HLA-DQ2 and a gluten peptide was elucidated, 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 Ag-binding groove (12). In agreement with this, our results indicate that the presence of proline in T cell stimulatory gluten peptides largely matches the preference for or acceptance of proline by HLA-DQ2. Several unique features of HLA-DQ2 explain this property. 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  $\alpha 53$  deletion, which eliminates the requirement for a hydrogen bond from the p1 amide. There

exists, however, the possibility of hydrogen bonding to the  $\alpha 50$ –52 HLA-DQ2 backbone by the amide groups of p-1 and p-2 peptide residues (Refs. 11 and 12 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 despite the loss of one hydrogen bond to the side chain carbonyl of  $\alpha 62$ Asn, provided that there is no proline at p4 or p9 (21). 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  $\beta 30$ Ser. HLA-DQ2 is the only HLA allele to have such a residue at this position ([www.anthonynolan.uk.org](http://www.anthonynolan.uk.org)). Importantly, our results now indicate that proline is preferred at the p8 nonanchor position and 85% of T cell stimulatory gluten peptides carry proline at p8. Although selection for a particular amino acid would not be expected at nonanchor 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. We have previously demonstrated that the spacing between the proline and glutamine residues determines which glutamines will be deamidated by tissue transglutaminase (31). Notably, in the very common gluten sequence QXP, 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. Although 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. Together, these exceptional binding properties explain why this MHC molecule is so well-suited to bind and present a large set of various gluten-derived peptides.

Despite the large repertoire of gluten T cell epitopes implicated in the pathogenesis of celiac disease, it is clear that in adults T cells to certain gluten sequences (e.g.,  $\alpha$ -I, -II, and -III) are almost invariably present, whereas responses to others (e.g.,  $\gamma$ -IV, -VIIa, and VIIb) are less frequently observed (22, 23). This suggests immunodominance of certain peptides. In children with celiac disease, however, we were often not able to find responses to these immunodominant peptides. Instead, we observed a diverse response to several other gluten sequences, some of which are not found in adults (24). To explain this apparent discrepancy, we have hypothesized that a T cell response to gluten can be initiated against a relatively large number of peptides and that when the response evolves it focuses on the most immunodominant sequences (24). This hypothesis still awaits confirmation in an independent study with children with celiac disease.

Two recent studies have further investigated the specificity of the HLA-DQ8-restricted gluten-specific T cell response (6, 25). These studies have confirmed our original observation of the immunodominant character of T cell responses toward one particular  $\alpha$ -gliadin peptide (5). Thus, immunodominant T cell responses to certain gluten peptides are found both in adult HLA-DQ2- and HLA-DQ8-positive patients. In addition, the above-mentioned studies identified several HLA-DQ8-restricted subdominant peptides. In agreement with our previous studies (3), deamidation of many of these peptides at the p1 and p9 positions enhanced their T cell recognition, presumably increasing their binding affinity for HLA-DQ8, known to prefer negatively charged amino acids at

those positions. Remarkably, both studies reported that identical gluten peptides can bind to either HLA-DQ2 or HLA-DQ8, depending on the deamidation pattern (6, 25). Of interest, these promiscuous sequences are nonimmunodominant.

The presence of gluten-specific T cells in CD patients implies that such T cells are positively selected in the thymus, most likely on the basis of the expression of HLA-DQ2-restricted self peptides in the thymus. Therefore, it is conceivable that the peptides identified in the current study may be involved in the selection of such T cells and we intend to address this issue in our future studies.

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 peptides resistant to proteolytic degradation in the gastrointestinal tract. Given that the average daily gluten consumption ranges between 10 and 15 grams, it is clear that in HLA-predisposed individuals, HLA-DQ-gluten peptide complexes may almost continuously be formed in the small intestine. Although under normal circumstances, tolerance will be maintained, a perturbation of the homeostasis in the intestine may allow the development of an inflammatory T cell response. This is in line with the observation that HLA-DQ2 homozygous individuals, expressing higher numbers of HLA-DQ2 dimers on their APCs compared with HLA-DQ2 heterozygous individuals, have a much higher risk of developing celiac disease (18, 26). 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 toward generating such blocking compounds have already been undertaken (27), but it is clear that to obtain a therapeutically useful agent a substantial improvement will be required. The natural ligands for HLA-DQ2 identified in the present study provide new insight into peptide-binding properties of this molecule and novel templates, both of which are likely to be of importance for the development of an effective HLA-DQ2 blocker.

## Acknowledgments

We thank Bart Roep, Jeroen van Bergen, and John Sidney for critical reading of the manuscript and Willemien Benckhuijsen for peptide synthesis.

## Disclosures

The authors have no financial conflict of interest.

## References

- Sollid, L. M., and E. Thorsby. 1993. HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis. *Gastroenterology* 105: 910–922.
- Molberg, O., S. N. McAdam, R. Korner, H. Quarsten, C. Kristiansen, L. Madsen, L. Fugger, H. Scott, O. Noren, P. Roepstorff, et al. 1998. Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nat. Med.* 4: 713–717.
- van de Wal, Y., Y. M. C. 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.
- Sollid, L. M., G. Markussen, J. Ek, H. Gjerde, F. Vartdal, and E. Thorsby. 1989. Evidence for a primary association of celiac disease to a particular HLA-DQ  $\alpha\beta$  heterodimer. *J. Exp. Med.* 169: 345–350.
- van de Wal, Y., Y. M. C. Kooy, P. A. van Veelen, S. A. Pena, L. M. Mearin, O. Molberg, K. E. Lundin, L. M. Sollid, T. Mutis, W. E. Benckhuijsen, et al. 1998. Small intestinal T cells of celiac disease patients recognize a natural pepsin fragment of gliadin. *Proc. Natl. Acad. Sci. USA* 95: 10050–10054.
- Henderson, K. N., J. A. Tye-Din, H. H. Reid, Z. Chen, N. A. Borg, T. Beissbarth, A. Tatham, S. I. Mannering, A. W. Purcell, N. L. Dudek, et al. 2007. A structural and immunological basis for the role of human leukocyte antigen DQ8 in celiac disease. *Immunity* 27: 23–34.
- 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 ( $\alpha 1^*0501$ ,  $\beta 1^*0201$ ) molecule. *Immunogenetics* 44: 246–253.
- 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.

9. Spits, H., J. Borst, M. Giphart, J. Coligan, C. Terhorst, and J. E. De Vries. 1984. HLA-DC antigens can serve as recognition elements for human cytotoxic T lymphocytes. *Eur. J. Immunol.* 14: 299–304.
10. Shevchenko, A., M. Wilm, O. Vorm, and M. Mann. 1996. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* 68: 850–858.
11. 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.
12. Kim, C. Y., H. Quarsten, E. Bergsgen, C. Khosla, and L. M. Sollid. 2004. Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. *Proc. Natl. Acad. Sci. USA* 101: 4175–4179.
13. 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.
14. 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.
15. 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, et al. 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.
16. 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.
17. 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.
18. 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. USA* 100: 12390–12395.
19. Sollid, L. M., and B. Jabri. 2005. Is celiac disease an autoimmune disorder? *Curr. Opin. Immunol.* 17: 595–600.
20. Koning, F., D. Schuppan, N. Cerf-Bensussan, and L. M. Sollid. 2005. Pathomechanisms in celiac disease. *Best Pract. Res. Clin. Gastroenterol.* 19: 373–387.
21. 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.
22. Shan, L., O. Molberg, I. Parrot, F. Hausch, F. Filiz, G. M. Gray, L. M. Sollid, and C. Khosla. 2002. Structural basis for gluten intolerance in celiac sprue. *Science* 297: 2275–2279.
23. Qiao, S. W., E. Bergsgen, 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.
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. Tollefsen, S., H. Arentz-Hansen, B. Fleckenstein, O. Molberg, M. Raki, W. W. Kwok, G. Jung, K. E. Lundin, and L. M. Sollid. 2006. HLA-DQ2 and -DQ8 signatures of gluten T cell epitopes in celiac disease. *J. Clin. Invest.* 116: 2226–2236.
26. 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.
27. Xia, J., M. Siegel, E. Bergsgen, L. M. Sollid, and C. Khosla. 2006. Inhibition of HLA-DQ2-mediated antigen presentation by analogues of a high affinity 33-residue peptide from  $\alpha$ 2-gliadin. *J. Am. Chem. Soc.* 128: 1859–1867.
28. 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.
29. 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.
30. Sjostrom, H., K. E. Lundin, O. Molberg, R. Korner, S. N. McAdam, D. Anthonen, 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.
31. Vader, L. W., A. De Ru, Y. van der Wal, 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.
32. 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.
33. 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.