 Related Commentary, page 2077

Natural peptides selected by diabetogenic DQ8 and murine I-A^{g7} molecules show common sequence specificity

Anish Suri,[1] James J. Walters,[2] Michael L. Gross,[2] and Emil R. Unanue[1]

[1]Departments of Pathology and Immunology and [2]Department of Chemistry, Washington University School of Medicine, St. Louis, Missouri, USA.

In this study, a large number of naturally processed peptides was isolated and identified from the human diabetes-susceptible class II MHC molecules HLA-DQ8 (DQA1*0301,DQB1*0302) and from murine I-A^{g7} species, both of which are expressed in genetically identical APC lines. The peptides presented during the processing of autologous proteins were highly selective in showing sequence specificity, mainly consisting of 1 or more acidic residues at their C terminus. Testing for binding to the MHC molecules revealed that the position 9 (P9) acidic residues of the peptides contributed decisively to binding. For HLA-DQ8, the P1 residue, which was also an acidic amino acid, influenced binding positively. Both HLA-DQ8 and I-A^{g7} selected for common peptides that bound in the same register. There was no evidence for selection of peptides having nonspecific or promiscuous binding. Thus, diabetogenic class II MHC molecules are highly selective in terms of the peptides presented by their APCs, and this is governed by the features of their P9 anchor pocket. These results are in striking contrast to those from studies examining synthetic peptide or phage display libraries, in which many peptides were shown to bind.

Introduction

The most important genes responsible for the onset of type 1 diabetes mellitus (T1DM) are those encoding the class II MHC alleles (1–3). In particular, the biochemical features of diabetogenic class II MHC molecules determine binding of autoantigenic peptides that ultimately trigger islet β cell–reactive T cells. In both humans and NOD mice (4), a notable feature of diabetes-related class II MHC alleles is the expression of a nonaspartic acid residue at position 57 of the β chain: an alanine in the case of the human HLA-DQ2 and HLA-DQ8 molecules (hereafter referred to as DQ2 and DQ8, respectively) and a serine in the case of the NOD class II MHC molecule I-A^{g7} (5–7). In contrast, most other class II MHC alleles express a conserved aspartic acid at β57 that pairs with an arginine at α76, defining the position 9 (P9) pocket of the peptide binding groove. Moreover, in humans, there are other alleles, such as HLA-DR3 and HLA-DR4, that in association with the DQ molecules increase the genetic risk for T1DM (8, 9).

The structure of both the I-A^{g7} (10, 11) and DQ8 (12) molecules was solved by x-ray crystallography. I-A^{g7} showed a P9 anchor pocket that was shallow, wide, and more open toward the C terminus as a result of the β57Ser and β56His. Subsequent studies established that the P9 pocket was most crucial in determining the selection of peptides during processing of natural proteins by APC (13, 14). Peptides selected by I-A^{g7} contained acidic-rich C termini that interacted with the P9 pocket and often contained multiple C-terminal acidic residues that increased their binding affinity (13). APCs expressing a modified I-A^{g7}, wherein the β57Ser was changed to the conserved aspartic acid, did not favor peptides with C-terminal acidic residues (13).

Wiley's group resolved the protein crystal structure of the DQ8 molecule bound to an insulin peptide, demonstrating the similarities of its P9 pocket to that of I-A^{g7} (12). However, the spatial and biochemical features of the other binding pockets of DQ8 are distinct. The DQ8 P4 pocket is very large and accommodates bulky residues such as tyrosines, while in I-A^{g7}, it is shallow and favors small to medium-sized hydrophobic residues, disfavoring large, bulky residues. Similarly, the P1 pocket of DQ8 contains an arginine at α52 (in contrast to an isoleucine in the case of I-A^{g7}), which forms an ion pair with an acidic amino acid from the peptide at P1. Our previous studies on peptide selection by I-A^{k}, which also contains α52Arg, revealed that most of the high-affinity peptides selected by this haplotype contained an acidic residue at P1 (15, 16). Thus, with respect to the P1 pocket, DQ8 resembles I-A^{k} more closely than I-A^{g7}. However, the precise contribution of each of these pockets in influencing the repertoire of peptides selected by DQ8 remains unclear.

Previous studies analyzing naturally processed peptides selected by human diabetogenic class II MHC molecules are limited and have given ambiguous results. In the first study by Chicz et al., few peptides were identified from DQ8 APCs ranging from 13–74 amino acids in length. These did not exhibit a consensus binding motif and were not analyzed extensively for their binding features (17). A second study, by Godkin et al., utilized pool sequencing to identify DQ8-bound peptides. This report suggested the presence of both acidic and basic amino acids at P4 (aspartic acid), P6 (arginine), and P9 (arginine and glutamic acid) (18). Concerning the peptide binding motif for DQ8 and DQ2, some reports suggested the preference for an acidic P9 anchor, while others did not (17–29). Moreover, recent reports examining binding with synthetic peptides indicated that the murine and human diabetogenic molecules bound epitopes of various specificity (24, 30, 31).

Information on the peptides naturally selected during processing will help in understanding the role of MHC in autoimmunity as well as in predicting potential diabetogenic peptides.

Nonstandard abbreviations used: MS, mass spectrometry; MS/MS, tandem mass spectrometry; P9, position 9; SCX, strong cation exchange; TFA, trifluoroacetic acid.

Conflict of interest: The authors have declared that no conflict of interest exists.

Citation for this article: *J. Clin. Invest.* **115**:2268–2276 (2005). doi:10.1172/JCI25350.

Table 1
Class II MHC molecules and background genes of cell lines

APC	Class II MHC	β57	Background
NOD.C3	I-A^{g7}	Ser	NOD
NOD.DQ8	DQ8	Ala	NOD

APC lines were generated by fusion of LPS-activated splenocytes from NOD or transgenic NOD.DQ8 mice to the M12.C3 B cell lymphoma partner line.

Naturally processed peptides represent the physiological substratum for CD4 T cell recognition: they represent what the APC prefers to select and display to the extracellular milieu. What are the chemical features of the natural peptides bound and selected by DQ8 molecules during the processing of self proteins? What are the binding motifs of such naturally selected peptides and the anchor residues that contribute to binding and selection? Are there peptide families that are naturally selected by both DQ8 and I-A^{g7} molecules, and if so, is the mode of binding identical between the two? To answer these basic questions in a definitive way, we isolated naturally processed peptides from APC lines expressing either DQ8 or I-A^{g7}. The APC lines were identical, except for the class II molecules, which allowed us to directly compare the repertoire of peptides selected by them. The large data sets of naturally processed peptides indicated a high degree of specificity of peptide selection exhibited by DQ8 — of note was the presence of multiple C-terminal acidic residues especially at P9 and to a lesser degree at P1. In contrast to results of studies using synthetic peptide libraries, there was little peptide promiscuity for DQ8 or for I-A^{g7} proteins. We provide the first *direct* evidence that identical peptide families were naturally selected by both DQ8 and I-A^{g7} and bound in the same register, which indicates a common biochemical outcome of the antigen processing and presentation pathways.

Results

Generation and characterization of APC lines

APC lines were generated on identical genetic backgrounds expressing either DQ8 or I-A^{g7} (Table 1). FACS analysis indicated that the I-A^{g7}–expressing NOD.C3 cell line had about 8-fold more cell surface class II MHC molecules than the NOD.DQ8 cell line (Figure 1). This difference in class II MHC levels also correlated with the amounts of peptides recovered from each cell line; i.e., more peptides were isolated from I-A^{g7} than from DQ8 (Supplemental Tables 1 and 2, and see following sections; supplemental material available online with this article; doi:10.1172/JCI25350DS1).

Features of naturally processed peptides

Naturally selected peptides, from I-A^{g7} and DQ8, were isolated and sequenced by tandem mass spectrometry (MS/MS) analyses. The class II MHC molecules select for peptides in "families," each sharing a common 9-amino-acid core, which interacts with the MHC binding groove from P1 to P9, along with varying lengths of flanking residues in the N and C termini (for example, see refs. 13, 32). A total of 206 peptides representing 108 families were isolated from DQ8 (Table 2 is a summary; the entire library of peptides is presented in Supplemental Table 1). We previously reported on I-A^{g7}–bound peptides but repeated the isolation in the current study in order to compare experiments done with the same cell line and in parallel with the examination of peptides from DQ8: 301 peptides representing 115 distinct families were isolated from I-A^{g7} (Table 3 is a summary; Supplemental Table 2 includes all the information). Peptides from both were derived from proteins present on the cell membrane, vesicular compartment, and cytosol, as well as the extracellular milieu.

Interestingly, 18 of the 108 peptide families (or 17%) from DQ8 were also selected by I-A^{g7} (Table 4 and marked by asterisks in Supplemental Table 1). The abundance of some of the common peptide families was comparable in the 2 molecules, while others were dominant in one compared with the other (see Supplemental Tables 1 and 2). For example, from DQ8, 7 members of the E2B 112–126 peptide family were identified, while from I-A^{g7}, 16 members were sequenced (Supplemental Tables 1 and 2). Nonetheless, the finding of common peptide families in I-A^{g7} and DQ8 was especially surprising, since both proteins exhibit only 60–70% similarity at the amino acid level between their α and β chains. In contrast, I-A^{g7} is approximately 94% similar to I-A^{d}, but the naturally processed peptides selected by each were entirely distinct (13, 14). Moreover, peptides selected by a mutant I-A^{g7} molecule called I-A^{g7PD}, in which residues 56 and 57 of the β chain were replaced by those from the I-A^{d} allele, did not overlap with those selected by I-A^{g7} (13). Note that most of the common peptides exhibited the presence of an acidic amino acid at P9 (Table 4). The relevance of this observation with regard to the binding motif is discussed in the following section.

Binding motif: DQ8 versus I-A^{g7}

The naturally processed peptides from DQ8 and I-A^{g7} were aligned on the basis of a favorable 9-mer core (as shown in Supplemental Tables 1 and 2). The compiled results of such analyses established the preferred binding motif for both diabetogenic class II MHC molecules, which we detail below. DQ8 and I-A^{g7} selected for peptides that contained acidic amino acids toward their C terminus (Figure 2). Many peptides from both contained runs of double or triple acidic residues toward the C terminus. It was shown previously that multiple C-terminal acidic residues cooperated to increase peptide binding affinity to I-A^{g7} (13). In addition, DQ8

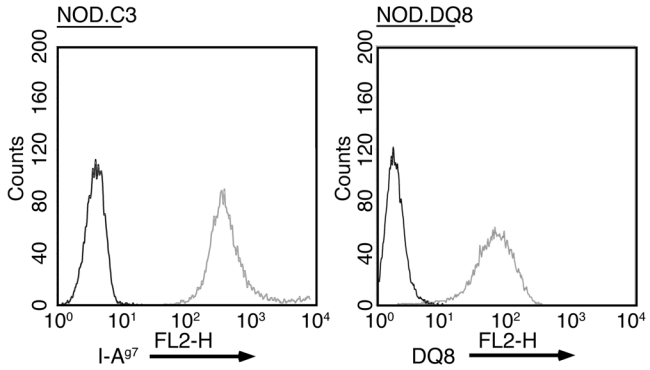

Figure 1
Class II MHC expression among APC lines. NOD.C3 and NOD.DQ8 cells were stained with anti–I-A^{g7} or anti-DQ8 mAb (gray lines), respectively, and then analyzed by FACS. Black lines indicate the isotype control for each cell line.

Table 2
Examples of naturally processed peptides selected by DQ8

Peptide					1			4		6			9				
E25B protein 112–126	Y	Q	T	I	E	E	N	I	K	I	F	E	E	D	A		
Superoxide dismutase 90–103		A	G	K	D	G	V	A	N	V	S	I	E	D	R		
TRAIL receptor 2 364–380	G	R	F	T	Y	Q	N	A	A	A	Q	P	E	T	G	P	G
Lamp I 129–144	I	Y	T	M	D	S	T	T	D	I	K	A	D	I	N	K	
Transferrin receptor 334–347		I	P	V	Q	T	I	S	R	A	A	A	E	K	L		
Cyclophilin C–associated protein 325–340	V	D	Q	W	S	T	E	T	I	A	S	H	E	D	I	E	
Serine proteinase inhibitor 90–103			L	R	D	F	S	N	M	A	S	A	E	E	N	Q	
14-3-3 ζ 28–41		S	V	T	E	Q	G	A	E	L	S	N	E	E	R		
Nicastrin 65–78		I	S	G	D	T	G	V	I	H	V	V	E	K	E		
IL-6 signal transducer 311–324		W	S	E	E	A	S	G	T	T	Y	E	D	R	P		

The favorable anchor residues at P1, P4, P6, and P9 are underlined. The entire list of naturally processed peptides selected by DQ8 is shown in Supplemental Table 1.

also showed a slight preference for acidic amino acids at the N terminus (Figure 2), a finding consistent with the biochemical features of the P1 pocket of DQ8 discussed above (12).

The relative abundance for naturally processed peptides selected by DQ8 varied greatly: the most abundant peptide family was 1,200-fold more abundant than the least; similarly, for I-A^{g7}, the difference was about 3,000-fold (see Supplemental Tables 1 and 2 for abundance). There was also no relationship between the abundance of peptides and their binding motifs (Supplemental Tables 1 and 2); i.e., peptide families with favorable binding cores were present in varying amounts. In addition, binding affinity also did not correlate with abundance, for example, even though the peptide from superoxide dismutase 1 protein 90–103 bound about 15-fold better to DQ8 than the epitope from DQ α chain 39–52, it was present in one-tenth the amount of the DQ α chain peptide (Table 5 and Supplemental Table 1; see also the following section). Similar results were obtained in our earlier analysis of peptides bound to I-A^{g7}, I-A^{d}, I-A^{g7PD}, and I-A^{k} (13, 14) and in the analysis of class I MHC peptides by Engelhard's laboratory (33) and ours with K^{d}-bound peptides. Other parameters such as the levels of protein expression, their turnover rates, and their accessibility to the MHC pathways of antigen processing and presentation must impact the final outcome of abundance and peptide selection. These features may become very relevant in situations of pathogen-specific immune responses or autoimmune responses.

DQ8 binding motif. The P1 pocket of DQ8 exhibited a preference for acidic amino acids (about 42% of residues at P1 were either glutamic acid or aspartic acid; Figure 3A), which was consistent with the structural aspects. Besides these, the P1 pocket also accommodated a variety of residues with the noticeable exception of basic amino acids, which would be electrostatically disfavored at this position. P4 was dominated by small hydrophobic or polar residues — alanine, serine, glycine, and threonine constituted 53% of amino acids at P4; while P6 favored alanine, valine, and serine, which made up 52% of the residues (Figure 3A). The P9 pocket of DQ8 was almost exclusively dominated by acidic amino acids — 84% of them (Figure 3A).

I-A^{g7} binding motif. P1 of I-A^{g7} accommodated a variety of residues, including polar, charged, and hydrophobic amino acids. P4 and P6 were dominated by the presence of small to medium-sized hydrophobic amino acids in addition to polar residues (leucine, isoleucine, and valine represented more than 60% of amino acids at P4; while alanine, serine, valine, and glycine represented more than 60% of residues at P6) (Figure 3B). Both P4 and P6 disfavored bulky or basic amino acids. P9 was highly biased for the presence of acidic amino acids (~90% of residues at P9 were glutamic acid or aspartic acid; Figure 3B), results that are akin to those of our prior studies (13, 14).

Binding analyses

All 14 peptides tested in the binding assay bound to soluble DQ8, which confirmed that these were bona fide epitopes selected by

Table 3
Examples of naturally processed peptides selected by I-A^{g7}

Peptide						1			4		6			9				
E25B protein 112–126		Y	Q	T	I	E	E	N	I	K	I	F	E	E	D	A		
Cyclophilin C–associated protein 325–340		V	D	Q	W	S	T	E	T	I	A	S	H	E	D	I	E	
Actg2 protein 70–84		Y	P	I	E	H	G	I	I	T	N	W	D	D	M	E		
HDL binding protein 20–35		V	P	Q	Q	I	K	V	A	T	L	N	S	E	E	E	N	
IL-21 receptor 350–366	I	P	L	A	A	G	Q	A	V	S	A	Y	S	E	E	R	D	
B lymphocyte antigen CD19 19–33				R	P	Q	K	S	L	L	V	E	V	E	E	G	G	N
Cathepsin D 65–77			E	P	V	S	E	L	L	K	N	Y	L	D	A			
Ndfip 1 protein 84–98		T	E	A	T	I	P	L	V	P	G	R	D	E	D	F		
Transferrin receptor 34–47			S	H	V	E	M	K	L	A	A	D	E	E	E	N		
Phosphoglycerate kinase 76–89		W	E	A	F	A	R	G	T	K	S	L	M	D	E			

The favorable anchor residues at P1, P4, P6, and P9 are underlined. The entire list of naturally processed peptides selected by I-A^{g7} is shown in Supplemental Table 2.

Table 4
List of naturally processed DQ8 peptides that were also selected by I-A^{g7}

Peptide									1			4		6			9			
Ran 25 kDa ras-related protein 19–34					H	D	L	E	V	A	Q	T	T	A	L	P	D	E	D	D
Lymphocyte cytosolic protein 1 508–519						E	D	I	G	G	G	Q	K	V	N	D	D			
14-3-3 ζ 228–245	W	T	S	D	T	Q	G	D	E	A	E	A	G	E	G	G	E	N		
Ubqln2 protein 81–94					M	R	Q	L	I	M	A	N	P	Q	M	Q	Q	L		
E25B protein 112–126[A]					Y	Q	T	I	E	E	N	I	K	I	F	E	E	D	A	
Hnrpr protein 592–601									Y	Q	D	T	Y	G	Q	Q	W	K		
SNRPF protein 128–143					Y	I	R	G	V	E	E	E	E	E	D	G	E	M	R	E
Envelope protein 497–512					V	S	A	L	E	K	S	L	T	S	L	S	E	V	V	L
Cytochrome b5 precursor 136–146							Y	R	H	F	W	A	D	S	K	S	S			
Wbscr5 146–162				K	P	S	T	P	E	S	G	V	E	D	F	E	D	Y	Q	N
Small nuclear ribonucleoprotein polypeptide G 48–59							M	A	T	S	G	Q	Q	N	N	I	G	M		
Eef1a1 protein 191–200									V	P	I	S	G	W	N	G	D	N		
Translocase of mitochondrial membrane 7 36–50					F	T	R	G	A	D	P	G	M	P	E	P	S	V	L	
Apolipoprotein A-II[A]					G	S	E	L	Q	T	Q	A	K	A	Y	F	E	K	T	Q
Elongation factor 1-β homolog 213–225							Y	V	Q	S	M	D	V	A	A	F	N	K	I	
G-3-P dehydrogenase 296–309					R	V	V	D	L	M	A	Y	M	A	S	K	E	E		
Cyclophilin C–associated protein 325–340[A]					V	D	Q	W	S	T	E	T	I	A	S	H	E	D	I	E
CD23 218–231[A]				D	L	K	G	R	L	V	S	I	H	S	Q	K	E	Q	D	

The peptides are aligned on the basis of the preferred anchor residues at P1, P4, P6, and P9. [A]Peptides that were shown to bind to both DQ8 and I-A^{g7}.

this MHC molecule (Table 5). Most peptides bound with a relative affinity in the range of 0.6–5.0 μM, and 2 were poor binders (IC_{50} > 33.0 μM; Table 5). Four peptides tested in this assay were previously isolated from I-A^{g7} and shown to bind to it (Table 5) (13). The relative binding affinity of these 4 common peptides was comparable between I-A^{g7} and DQ8 (Table 5).

The peptide-binding register of the natural epitopes to DQ8 was examined by testing peptides with single amino acid changes in order to confirm the usage of preferred anchor residues (Table 6). Placing lysine, a strong hindering residue for P1, P4, P6, and P9, at different positions in the peptide sequence helped indicate which position was normally being used.

E25B peptide. The E25B 112–126 peptide family was selected by both DQ8 and I-A^{g7}. To identify the P9 anchor, each of the 3 C-terminal acidic amino acids was changed to a lysine. Changing the Glu123 and Asp125 to a lysine had no effect on binding to either DQ8 or I-A^{g7} (Table 6). However, when Glu124 was mutated, the peptide bound very poorly to both, which demonstrated that this residue was being preferentially used as the P9 anchor. To confirm that this was indeed the binding register, i.e., with Glu124, the putative P4 (Ile119) or P6 (Ile121) was changed to a lysine. A basic amino acid at P4 or P6 also negatively affects binding: either change also resulted in a loss of binding (Table 6). (The lack of a binding register having Glu123 or Asp125 as the P9 anchor residues is explained by unfavorable residues at P4 and/or P6: if Glu123 was the P9 anchor, then Lys120 would be the P6 anchor; if Asp 125 was the P9 anchor, then Lys120 and Phe122 would be the P4 and P6 anchors, respectively.) Of note is that changing the multiple C-terminal acidic residues to alanine only resulted in a modest decrease in binding to both DQ8 and I-A^{g7} — an alanine, although not preferred at P9, was well tolerated (10, 11). In the case of DQ8, changing the P1 Glu116 to a lysine also resulted in about a 4- to 5-fold loss of binding (Table 6). Thus, although DQ8 prefers acidic amino acids at P1 and P9, the effect on binding is more evident for P9 than for P1 (Table 6). Moreover, these mutational analyses taken together established that the E25B 112–126 peptide binds to both I-A^{g7} and DQ8 in *exactly* the same register.

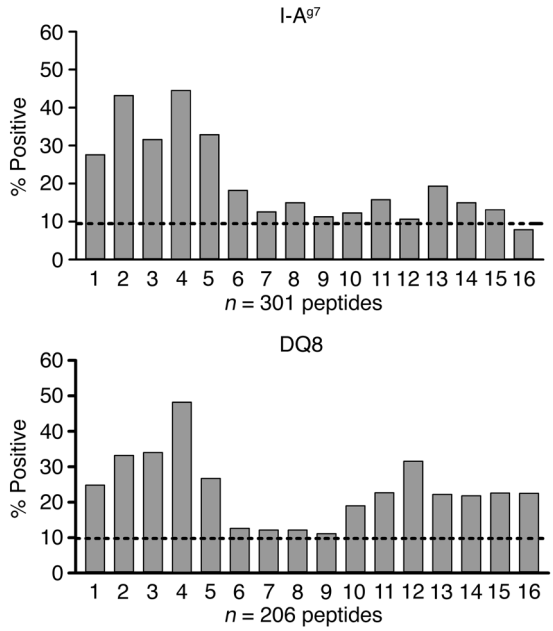

Figure 2
Acidic amino acid distribution among naturally processed peptides selected by I-A^{g7} and DQ8. Each peptide was scored for the presence of either an aspartic acid or glutamic acid residue starting at the most C-terminal position, which was assigned position 1 in the figure. Thus, by this scheme, the second most C-terminal residue would be position 2, the third would be position 3, and so on. Note that peptides selected by both I-A^{g7} and DQ8 contained acidic amino acids toward their C termini. In addition, peptides selected by DQ8 also exhibited the biased presence of acidic amino acid toward their N terminus. The dotted line represents the background frequency of the presence of acidic amino acids.

Table 5
Binding analysis of naturally processed peptides to DQ8

Peptide	AA sequence	Binding (μM)	
		DQ8	I-A^{g7} [A]
MHC II Eα 51–65	FDGDEIFHVDIEKSE	1.0	–
Superoxide dismutase 1 90–103	AGKDGVANVSIEDR	2.0	–
Cyclophilin C–associated protein 325–340	VDQWSTETIASHEDIE	0.6	3.3
E25B protein 112–126	YQTIEENIKIFEEDA	0.8	0.4
TRAIL receptor 2 364–380	GRFTYQNAAAQPETGPG	1.7	–
Transferrin receptor 334–347	IPVQTISRAAAEKL	3.8	–
Nicastrin 65–78	ISGDTGVIHVVEKE	1.0	–
Na,K-ATPase α-1 subunit 400–413	ADTTENQSGVSFDK	>33.0	–
Lamp I 129–144	IYTMDSTTDIKADINK	7.8	–
Apolipoprotein A-II 31–46	GSELQTQAKAYFEKTQ	4.6	1.8
CD23 219–231	DLKGRLVSIHSQKEQD	5.9	3.1
MHC class II antigen, DQ α chain 39–52	YQSYGPSGQYSHEF	>33.0	–
Polyubiquitin 6–21	KTLTGKTITLEVEPSD	1.0	–
Serine (or cysteine) proteinase inhibitor 90–103	LRDFSNMASAEENQ	2.4	–

[A]Peptides that were tested for binding in a previous study by Suri et al. (13).

MHC II Eα peptide. Three acidic amino acids are found at the C terminus of the Eα51–65 peptide: Asp60, Glu62, and Glu65. It seemed unlikely that Asp60 or Glu65 could be used as P9 anchors since they generated binding registers that contained unfavorable residues at P4 and P6 (as described above) (Table 6). Glu62 emerged as the likely candidate for the P9 anchor and when changed to a lysine generated a nonbinding peptide (Table 6). The binding register was further ascertained by mutating the putative P4 and P6 (Phe57 and Val59, respectively) to a lysine — both changes resulted in a complete loss of binding (Table 6). Note that changing the amino acids at P4 and/or P6 singly to an alanine did not have a noticeable affect on binding, since these residues are acceptable at either of these positions. Also, changing the acidic P1 anchor (Asp54) to a lysine or alanine only slightly affected binding; however, an alanine at P1 in conjunction with P9 reduced binding 10-fold (Table 6). This was consistent with the results with the E2B peptide; i.e., an acidic residue at P9 contributed more to binding than the same residue at P1.

Superoxide dismutase 1 peptide. In the superoxide dismutase 1 peptide, Glu101 was the P9 anchor; changing this to lysine generated a nonbinding peptide (Table 6). Mutating the P9 Glu101 to an alanine decreased binding about 3-fold, and this was enhanced when the P1 Asp93 was also changed to an alanine (>15-fold loss of binding) (Table 6). Substituting only the P1 Asp93 to a lysine had a very modest effect on binding. The register was further confirmed by changing the P6 Val98 to a lysine, which again resulted in a complete loss of binding (Table 6).

TRAIL receptor 2 peptide. In the case of the TNF-related apoptosis-inducing ligand (TRAIL) receptor 2 peptide, the only C-terminal acidic amino acid was a Glu376, which when mutated to a lysine resulted in about a 4-fold loss of binding (Table 6). Changing the P1 Tyr368 to lysine or an alanine had a small negative effect, but replacing both P1 and P9 by lysines decreased binding about 15-fold, which indicates a cooperative interaction between the P1 and P9 anchor residues (Table 6).

Nicastrin peptide. In the nicastrin peptide, changing the P9 Glu76 to a lysine lowered binding about 18-fold, but the same mutation at P1 (Asp68→Lys) resulted in a complete loss of binding (Table 6). Thus, the P1 acidic anchor contributed more to binding affinity than the P9 acidic residue (Table 6).

Discussion

The diabetogenic-propensity class II MHC molecules exhibited a very high degree of specificity with regard to natural peptides selected during processing of the proteins. Neither DQ8 nor I-A^{g7} molecules were promiscuous in terms of the peptides selected during processing, and for both, the P9 anchor played the dominant role in establishing the register of peptide binding. These characteristics explain the commonalities of peptide selection between the murine and human class II molecules, which are structurally similar at the P9 anchor site but are unique at the other positions. Also worthy of emphasis is the presence of multiple C-terminal acidic residues — 39% and 46% of peptide families from DQ8 and I-A^{g7}, respectively, contained more than 1 C-terminal acidic residue, some in pairs and others separated by 1 amino acid (Figure 2). In addition, for DQ8 molecules, the P1 anchor (mostly aspartic or glutamic acid) cooperated with the P9 acidic amino acid to increase binding affinity (Table 6).

We emphasize these results because other studies, including an early one from our laboratory (34), claimed the contrary, that there was a noticeable lack of specificity of I-A^{g7} or DQ8 molecules for peptides (24, 30, 31). These data were obtained by examining synthetic peptides or phage display libraries thought to represent the antigenic epitopes and equivalent to the natural peptides. The results indicated that the key primary anchor positions for DQ8 and I-A^{g7} were tolerant to many amino acids, an issue which is undisputed. An excellent example supporting this argument is the P9 pocket of DQ8 and I-A^{g7}, which accommodated acidic or small hydrophobic/polar residues. However, the results indicated here by the analyses of naturally selected peptides portray a very different outcome characterized by exquisite selectivity and an overwhelming bias for the presence of acidic amino acids at this position (Figure 3, A and B).

The explanations for this important discrepancy may be along 2 lines: either during the processing of the proteins to generate the peptide-MHC complex, the peptides with nonacidic C-terminal residues were not bound (in general the peptides with non–terminal acidic residues tend to bind less well, although this reduction is not a striking feature); or if they did bind, they were edited by HLA-DM (or H2-DM) and/or lost during the intracellular trafficking of the MHC molecules. Work that is in progress indicates that peptides with nonacidic residues dissociated at a higher rate from the MHC molecules at neutral pH. Regardless of the explanation, the repertoire of natural peptides represents those that are physiologically relevant for the T cell response, which is especially important in the case of aberrant immune responses mounted against self. (In addition, knowing the sequence of the natural peptides establishes other features important for T cell recognition,

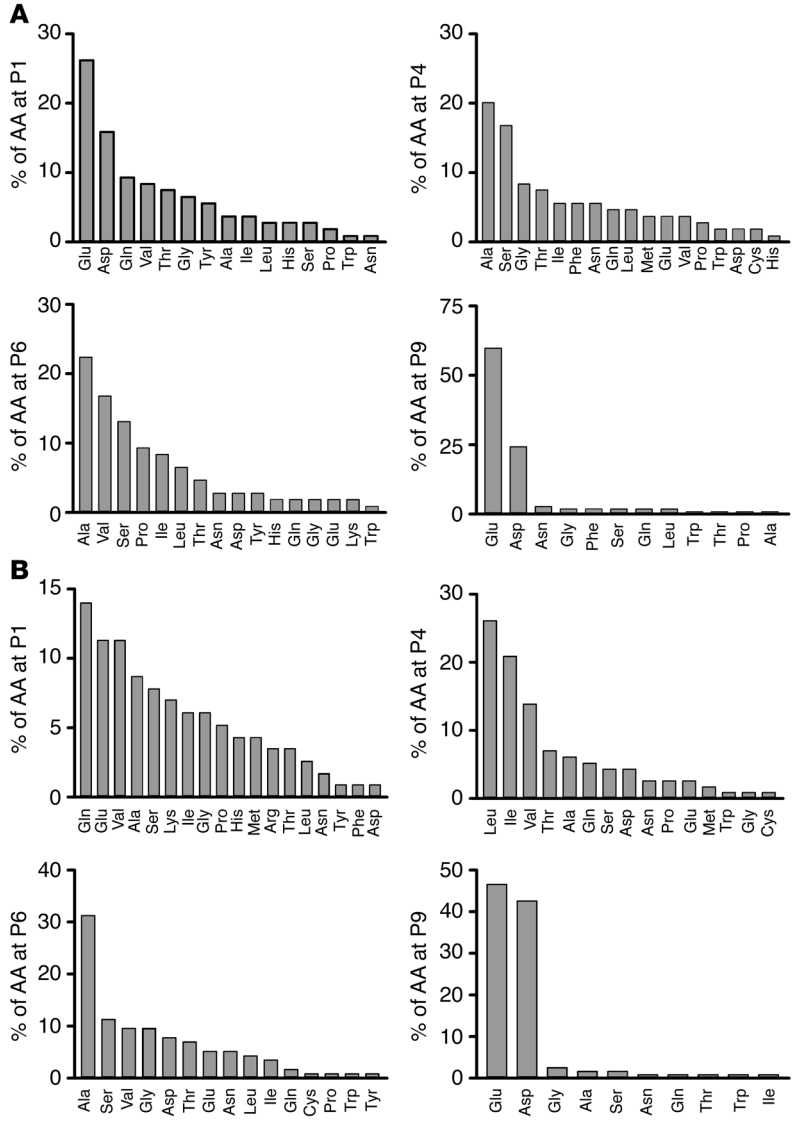

Figure 3

Distribution of amino acids at the P1, P4, P6, and P9 positions. (**A**) Peptide binding motif for DQ8. The P1 through P9 amino acids of the naturally processed peptides listed in Supplemental Table 1 were aligned to identify the preferred residues at P1, P4, P6, and P9. (**B**) Peptide binding motif for I-A^{g7}. The P1 through P9 amino acids of the naturally processed peptides listed in Supplemental Table 2 were aligned to identify the preferred residues at P1, P4, P6, and P9.

such as the relevance of flanking residues in the display of the T cell receptor contact residues [ref. 35].) The contrasting results indicate the problems in depending only on data obtained from peptides thought to represent the natural epitopes. Vaccination with microbial or autoimmune peptides without knowing their display may be a fruitless venture.

Finally, our results provide more evidence, this time direct, of the selection of identical epitopes between APCs bearing DQ8 or I-A^{g7} molecules. Many previous studies identified common proteins (such as glutamic acid decarboxylase, insulin, heat shock proteins, etc.) as targets of autoimmunity in NOD mice and humans (reviewed in refs. 1, 36). In addition, previous reports analyzing T cell responses have indirectly alluded to common epitopes that may be displayed by I-A^{g7} and DQ8 (37–39). Our data supports the hypothesis that the NOD mouse is highly valid for identification of putative pathogenic T cell epitopes and for testing such peptides for their ability to establish disease-preventive protocols.

Methods

Cell lines, antibodies, and isolation of peptides

The M12.C3 B cell lymphoma was fused to LPS-activated splenocytes from wild-type NOD (I-A^{g7}) or transgenic NOD.DQ8 (DQ8^{+}I-A^{g7-}) (40) mice to generate the NOD.C3 or NOD.DQ8 cell lines, respectively. Usage of mice was approved by the Division of Comparative Medicine at Washington University (St. Louis, Missouri, USA). This protocol generated APC lines on identical genetic backgrounds, although variation due to chromosomal differences between the 2 cell lines was not excluded. Both cell lines were maintained in DMEM supplemented with 10% calf serum. For isolation of peptides, NOD.C3 and NOD.DQ8 cells were expanded in roller bottles to obtain 5×10^9 to 10×10^9 cells. Following detergent lysis, the MHC molecules were affinity purified using the AG2.42.7 (for anti–I-A^{g7}) or 9.3.F10 (for anti-DQ8) antibodies conjugated to sepharose beads, as previously described (13, 14). Peptides bound to class II MHC molecules were acid eluted using 0.1% trifluoroacetic acid (TFA), separated from the class II MHC proteins, dried down, and then analyzed by mass spectrometry (MS) as described below.

Generation of soluble DQ8 molecules

Construction of DQ8 expression vector. The cell line NOD.DQ8 containing the DQ8 α and β ectodomains were cloned into the vector pCR2.1 (Invitrogen Corp.) and modified by PCR extension to include the Fos and Jun leucine zipper domains. The domains of both DQ8 chains were subcloned into the baculovirus expression vector pFastBac Dual (Invitrogen Corp.). The Fos and Jun leucine zippers were attached to the C terminus of each chain through a polylinker containing thrombin cleavage sites (ANLVPRGSTAPS). The α chain carries a 6-His-tag at the end of Fos. The β chain has a covalently linked human CLIP-containing sequence (SKMRMANPLLMQA) attached at its amino terminus by a glycine- and serine-rich linker containing a thrombin cleavage site (GGGGSLVPAGSGGGGSGS).

Expression of DQ8 in insect cells. High Five insect cells were infected with baculovirus expressing the DQ8 chains. Cells were grown to a density of 2×10^6 cells/ml in serum-free medium (JRH Biosciences) in baffled flasks before infection at an MOI of 10 and then were incubated for 72 hours at 27°C. The clarified medium was concentrated and exchanged into 50 mM NaH_2PO_4 (pH 8.0), 10 mM imidazole, and 0.5 M NaCl using 10 kDa Centramate tangential flow filtration system (Pall Corp.). The concentrated supernatant was passed by gravity over a 5-ml bed of Ni-NTA beads (QIAGEN).

Table 6
Mutational analyses of naturally processed peptides to identify the binding register

Peptide	Sequence	Binding (μM)	
		DQ8	I-A$^{g7\ A}$
E25B protein 112–126	Y Q T I E E N I K I F E E D A	0.8	0.4
	----------K------	1.0	0.3
	----------K----	>33.0	12.0
	----------K--	1.7	0.2
	----------A----	1.8	0.3
	----------A-A--	2.5	0.5
	----------A-A-A--	1.7	2.3
	----------K----------	>33.0	9.9
	----------K----------	>33.0	66.0
	--------K----------	3.8	ND
MHC II Eα 51–65	F D G D E I F H V D I E K S E	1.0	
	----------K------	NB	
	------K----------	2.8	
	------A----------	2.0	
	------A----------A------	10.7	
	------K----------K------	NB	
	----------K----------	10.7	
	----------A----------	3.3	
	----------K----------	NB	
	----------A----------	3.1	
Superoxide dismutase 1 90–103	A G K D G V A N V S I E D R	2.0	
	----------K------	NB	
	----------A------	7.8	
	----------A-A----	7.7	
	------K----------	5.1	
	------A----------	1.8	
	------K----------K------	NB	
	------A----------A------	>33.0	
	----------K----------	NB	
	----------A----------	2.8	
TRAIL receptor 2 364–380	G R F T Y Q N A A A Q P E T G P G	1.7	
	----------K--------	6.9	
	----------A--------	1.0	
	--------K----------	3.7	
	--------A----------	1.7	
	--------K----------K--------	26.4	
	--------A----------A--------	3.1	
Nicastrin 65–78	I S G D T G V I H V V E K E	1.0	
	----------K----	18.7	
	----------A----	4.3	
	------K----------	NB	
	------A----------	2.3	
	------K----------K----	NB	

The 9-mer binding core mapped for each peptide is underlined. [A]Binding was tested in a previous study by Suri et al. (13). ND, not determined; NB, non-binder.

The column was washed first with 50 mM NaH_2PO_4 (pH 8.0), 0.5 M NaCl, and 50 mM imidazole and then eluted with 50 mM NaH_2PO_4 (pH 8.0), 0.5 M NaCl, and 250 mM imidazole. The fractions containing the DQ8 were pooled and dialyzed to 10 mM HEPES, 150 mM NaCl (pH 7.2).

Binding analysis

Briefly, 1.0–1.5 μg of purified baculovirus DQ8/CLIP was treated with 0.1 units of thrombin to cleave both the zipper tails and peptide linker (Novagen) and simultaneously incubated with 0.125 pmol of ^{125}I-radiolabeled reference hemagglutinin peptide

(FESTGNLIAPEYGFKISY), labeled by the chloramine T method, as described previously (41), and increasing doses of unlabeled peptides in 40 mM 2-[N-Morpholino]ethanesulfonic acid, 150 mM sodium chloride, pH5.5. The hemagglutinin peptide has previously been used as an indicator peptide in DQ8 binding assays (26). Binding reactions were incubated for 48 hours at 37°C in 30- to 32-μl volumes. Complexes were purified from free peptide using gel filtration Bio-Spin columns (Bio-Rad Laboratories). The percentage of bound peptide was evaluated by gamma counting. The amounts of peptides bound in the reactions were about 15–20% that of the input radiolabeled standard peptide. Relative binding data refers to the amounts required to inhibit the binding of the reference peptide by 50%. Individual binding results with the reference peptide varied less than 10% from the averaged value.

Two-dimensional liquid chromatography–MS

MHC peptide extracts from NOD.DQ8 and NOD.C3 lines were treated identically. Each was analyzed by on-line reverse phase liquid chromatography–MS (RPLC-MS) and MS/MS and additionally by off-line strong cation exchange (SCX) LC followed by on-line RPLC-MS/MS. To avoid cross-contamination of samples, both reverse phase (RP) and SCX columns were new and dedicated to either NOD.DQ8 or NOD.C3 peptide analysis.

Initially, the dried extracts were resuspended in 60 μl 0.1% TFA, desalted, and further purified using 3 ZipTip (Millipore) pipette tips containing C18 RP media per the manufacturer's protocol. Three elutions were performed, each using 30 μl 50:50 acetonitrile (CH_3CN)/H_20, in 0.1% TFA, providing 90 μl elution for each extract. The eluent was completely dried and resuspended in 55 μl 3% CH_3CN, 97% H_20, 0.1% formic acid (solvent A for RPLC).

Five microliters of sample was loaded onto a silica capillary column with a PicoFrit tip (New Objective Inc.) packed in-house with C18 RP material (Delta-Pak, 0.075 × 100 mm, 5 μm, 300 Å; Waters Corp.). The gradient, pumped using a Waters CapLC (Waters Corp.), was from 0% solvent B (97% CH_3CN, 3% H_2O, 0.1% FA) to 5% solvent B over 3 minutes, then to 50% solvent B over 70 minutes. Eluent flow was approximately 5 μl/min and split before the column at an approximately 1:25 ratio to maintain a flow rate at the tip of 250 nl/min. Flow was directed into the entrance of the heated capillary of an LCQ-Deca Quadrupole ion-trap mass spectrometer (Thermo Electron Corp.) equipped with a custom-built nanospray source.

MS and dependent-scan mode MS/MS were performed using the mass spectrometer under the control of Xcalibur 1.3 software (Xcalibur Software Inc.). For MS, the scan range was 600–1,400 mass/charge ratio (*m/z*) in the profile mode. For MS/MS, a scan range of 600–1,400 *m/z* in centroid mode was used for the MS, and the MS/MS range was from 30% of the *mass/charge ratio* of the parent ion to 2,000 *m/z*. The isotope clusters of parent ions were dynamically selected and isolated with a 2.0-*m/z* window. The collision energy was set to 28% of the maximum, which is approximately 5 eV. For detection of peptides, the mass spectrometer selected peptides based on their signal intensity. In such experiments, the first and second most abundant ions were analyzed in the first run, and the third and fourth most abundant ions were analyzed in a subsequent run.

After RP-LCMS and MS/MS analysis, the remainder of the sample (~40 μl) was dried to completion and resuspended in 20 μl solvent C (30% acetonitrile; pH to 3.5 with FA). Ten microliters was analyzed by SCX using an Ultra-Plus II LC with flow split to 5 μl/min and a PolySulfoethyl column (15 cm × 320 μm; MicroTech Scientific Inc.). The gradient was from 0% solvent D (30% CH_3CN, 01% FA, 1M ammonium acetate) to 25% solvent B in 35 minutes, to 50% solvent B at 50 minutes, to 100% solvent B at 65 minutes. Ten 7-minute fractions were collected, dried down, and resuspended in 10 μl 3% CH_3CN, 0.1% formic acid. Five microliters was injected for on-line RPLC electrospray ionization–MS as described above using data-dependent scanning fragmentation of the 2 most abundant ions.

All product ion spectra were analyzed, and peptide sequences were determined automatically using Mascot software (Matrix Science Ltd.). All the automatically determined sequences were manually verified against the experimental product ion spectra. The relative abundance of each peptide ion was calculated as previously described (13, 14).

Acknowledgments

We thank Chella David for providing the NOD.DQ8 transgenic mice and T.J. Esparza and Shirley Petzold for technical assistance with production of soluble DQ8 molecules and binding analyses. This work was supported by grants from the NIH and the Kilo Diabetes and Vascular Research Foundation.

Received for publication April 14, 2005, and accepted in revised form May 24, 2005.

Address correspondence to: Emil Unanue, Washington University School of Medicine, 660 South Euclid Avenue, Box 8118, St. Louis, Missouri 63110, USA. Phone: (314) 362-7440; Fax: (314) 362-4096; E-mail: unanue@pathbox.wustl.edu.

1. Castano, L., and Eisenbarth, G.S. 1990. Type-I diabetes: a chronic autoimmune disease of human, mouse, and rat [review]. *Annu. Rev. Immunol.* **8**:647–679.
2. Wicker, L.S., Todd, J.A., and Peterson, L.B. 1995. Genetic control of autoimmune diabetes in the NOD mouse. *Annu. Rev. Immunol.* **13**:179–200.
3. McDevitt, H., Singer, S., and Tisch, R. 1996. The role of MHC class II genes in susceptibility and resistance to type I diabetes mellitus in the NOD mouse. *Horm. Metab. Res.* **28**:287–288.
4. Kikutani, H., and Makino, S. 1992. The murine autoimmune diabetes model: NOD and related strains [review]. *Adv. Immunol.* **51**:285–322.
5. Acha-Orbea, H., and McDevitt, H.O. 1987. The first external domain of the nonobese diabetic mouse class II I-A beta chain is unique. *Proc. Natl. Acad. Sci. U. S. A.* **84**:2435–2439.
6. Morel, P.A., Dorman, J.S., Todd, J.A., McDevitt, H.O., and Trucco, M. 1988. Aspartic acid at position 57 of the HLA-DQ beta chain protects against type I diabetes: a family study [erratum 1989, **86**:1317]. *Proc. Natl. Acad. Sci. U. S. A.* **85**:8111–8115.
7. Todd, J.A., Bell, J.I., and McDevitt, H.O. 1987. HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature.* **329**:599–604.
8. Abraham, R.S., Kudva, Y.C., Wilson, S.B., Strominger, J.L., and David, C.S. 2000. Co-expression of HLA DR3 and DQ8 results in the development of spontaneous insulitis and loss of tolerance to GAD65 in transgenic mice. *Diabetes.* **49**:548–554.
9. Deschamps, I., Beressi, J.P., Khalil, I., Robert, J.J., and Hors, J. 1991. The role of genetic predisposition to type I (insulin-dependent) diabetes mellitus. *Ann. Med.* **23**:427–435.
10. Latek, R.R., et al. 2000. Structural basis of peptide binding and presentation by the type I diabetes-associated MHC class II molecule of NOD mice. *Immunity.* **12**:699–710.
11. Corper, A.L., et al. 2000. A structural framework for deciphering the link between I-Ag7 and autoimmune diabetes. *Science.* **288**:505–511.
12. Lee, K.H., Wucherpfennig, K.W., and Wiley, D.C. 2001. Structure of a human insulin peptide-HLA-DQ8 complex and susceptibility to type 1 diabetes. *Nat. Immunol.* **2**:501–507.
13. Suri, A., et al. 2002. In APCs, the autologous peptides selected by the diabetogenic I-Ag7 molecule are unique and determined by the amino acid changes in the P9 pocket. *J. Immunol.* **168**:1235–1243.
14. Suri, A., Walters, J.J., Kanagawa, O., Gross, M.L., and Unanue, E.R. 2003. Specificity of peptide selection by antigen-presenting cells homozygous or

heterozygous for expression of class II MHC molecules: the lack of competition. *Proc. Natl. Acad. Sci. U. S. A.* **100**:5330–5335.
15. Nelson, C.A., Viner, N.J., Young, S.P., Petzold, S.J., and Unanue, E.R. 1996. A negatively charged anchor residue promotes high affinity binding to the MHC class II molecule I-Ak. *J. Immunol.* **157**:755–762.
16. Fremont, D.H., Monnaie, D., Nelson, C.A., Hendrickson, W.A., and Unanue, E.R. 1998. Crystal structure of I-Ak in complex with a dominant epitope of lysozyme. *Immunity.* **8**:305–317.
17. Chicz, R.M., et al. 1994. Self-peptides bound to the type I diabetes associated class II MHC molecules HLA-DQ1 and HLA-DQ8. *Int. Immunol.* **6**:1639–1649.
18. Godkin, A., et al. 1997. Use of eluted peptide sequence data to identify the binding characteristics of peptides to the insulin-dependent diabetes susceptibility allele HLA-DQ8 (DQ 3.2). *Int. Immunol.* **9**:905–911.
19. Johansen, B.H., Vartdal, F., Eriksen, J.A., Thorsby, E., and Sollid, L.M. 1996. Identification of a putative motif for binding of peptides to HLA-DQ2. *Int. Immunol.* **8**:177–182.
20. Kwok, W.W., Domeier, M.L., Raymond, F.C., Byers, P., and Nepom, G.T. 1996. Allele-specific motifs characterize HLA-DQ interactions with a diabetes-associated peptide derived from glutamic acid decarboxylase. *J. Immunol.* **156**:2171–2177.
21. Nepom, B.S., Nepom, G.T., Coleman, M., and Kwok, W.W. 1996. Critical contribution of beta chain residue 57 in peptide binding ability of both HLA-DR and -DQ molecules. *Proc. Natl. Acad. Sci. U. S. A.* **93**:7202–7206.
22. Quarsten, H., et al. 1998. The P9 pocket of HLA-DQ2 (non-Aspbeta57) has no particular preference for negatively charged anchor residues found in other type 1 diabetes-predisposing non-Aspbeta57 MHC class II molecules. *Int. Immunol.* **10**:1229–1236.
23. Reizis, B., Altmann, D.M., and Cohen, I.R. 1997. Biochemical characterization of the human diabetes-associated HLA-DQ8 allelic product: similarity to the major histocompatibility complex class II I-A(g)7 protein of non-obese diabetic mice. *Eur. J. Immunol.* **27**:2478–2483.
24. Sidney, J., Del Guercio, M.F., Southwood, S., and Sette, A. 2002. The HLA molecules DQA1*0501/B1*0201 and DQA1*0301/B1*0302 share an extensive overlap in peptide binding specificity. *J. Immunol.* **169**:5098–5108.
25. Sato, A.K., Sturniolo, T., Sinigaglia, F., and Stern, L.J. 1999. Substitution of aspartic acid at beta57 with alanine alters MHC class II peptide binding activity but not protein stability: HLA-DQ (alpha1*0201, beta1*0302) and (alpha1*0201, beta1*0303). *Hum. Immunol.* **60**:1227–1236.
26. Straumfors, A., et al. 1998. A peptide-binding assay for the disease-associated HLA-DQ8 molecule. *Scand. J. Immunol.* **47**:561–567.
27. Vartdal, F., et al. 1996. The peptide binding motif of the disease associated HLA-DQ (alpha 1* 0501, beta 1* 0201) molecule. *Eur. J. Immunol.* **26**:2764–2772.
28. Verreck, F.A., et al. 1994. Identification of an HLA-DQ2 peptide binding motif and HLA-DPw3-bound self-peptide by pool sequencing. *Eur. J. Immunol.* **24**:375–379.
29. Oiso, M., Nishi, T., Ishikawa, T., Nishimura, Y., and Matsushita, S. 1997. Differential binding of peptides substituted at putative C-terminal anchor residue to HLA-DQ8 and DQ9 differing only at beta 57. *Hum. Immunol.* **52**:47–53.
30. Astill, T.P., Ellis, R.J., Arif, S., Tree, T.I., and Peakman, M. 2003. Promiscuous binding of proinsulin peptides to type 1 diabetes-permissive and -protective HLA class II molecules. *Diabetologia.* **46**:496–503.
31. Stratmann, T., et al. 2000. The I-Ag7 MHC class II molecule linked to murine diabetes is a promiscuous peptide binder. *J. Immunol.* **165**:3214–3225.
32. Nelson, C.A., Roof, R.W., McCourt, D.W., and Unanue, E.R. 1992. Identification of the naturally processed form of hen egg white lysozyme bound to the murine major histocompatibility complex class II molecule I-Ak. *Proc. Natl. Acad. Sci. U. S. A.* **89**:7380–7383.
33. Engelhard, V.H., Brickner, A.G., and Zarling, A.L. 2002. Insights into antigen processing gained by direct analysis of the naturally processed class I MHC associated peptide repertoire [review]. *Mol. Immunol.* **39**:127–137.
34. Carrasco-Marin, E., Shimizu, J., Kanagawa, O., and Unanue, E.R. 1996. The class II MHC I-Ag7 molecules from non-obese diabetic mice are poor peptide binders. *J. Immunol.* **156**:450–458.
35. Latek, R.R., and Unanue, E.R. 1999. Mechanisms and consequences of peptide selection by the I-Ak class II molecule. *Immunol. Rev.* **172**:209–228.
36. Tisch, R., and McDevitt, H. 1996. Insulin-dependent diabetes mellitus. *Cell.* **85**:291–297.
37. Liu, J., Purdy, L.E., Rabinovitch, S., Jevnikar, A.M., and Elliott, J.F. 1999. Major DQ8-restricted T-cell epitopes for human GAD65 mapped using human CD4, DQA1*0301, DQB1*0302 transgenic IA(null) NOD mice. *Diabetes.* **48**:469–477.
38. Kelemen, K., et al. 2004. HLA-DQ8-associated T cell responses to the diabetes autoantigen phogrin (IA-2 beta) in human prediabetes. *J. Immunol.* **172**:3955–3962.
39. Wen, L., Wong, F.S., Sherwin, R., and Mora, C. 2002. Human DQ8 can substitute for murine I-Ag7 in the selection of diabetogenic T cells restricted to I-Ag7. *J. Immunol.* **168**:3635–3640.
40 Raju, R., Munn, S.R., Majoribanks, C., and David, C.S. 1998. Islet cell autoimmunity in NOD mice transgenic for HLA-DQ8 and lacking I-Ag7. *Transplant. Proc.* **30**:561.
41. Nelson, C.A., Petzold, S.J., and Unanue, E.R. 1993. Identification of two distinct properties of class II major histocompatibility complex-associated peptides. *Proc. Natl. Acad. Sci. U. S. A.* **90**:1227–1231.