Prediction of HLA-DQ8 β cell peptidome using a computational program and its relationship to autoreactive T cells

Kuan Y. Chang[1,2] and **Emil R. Unanue**[1]

[1]Division of Immunology, Department of Pathology and Immunology and [2]Computational Biology Program, Washington University School of Medicine, Campus Box 8118, 660 South Euclid Avenue, St Louis, MO 63110, USA

Abstract

The goal was to identify HLA-DQ8-bound β cell epitopes important in the T cell response in autoimmune diabetes. We first identified HLA-DQ8 (DQA1*0301/DQB1*0302) β cell epitopes using a computational approach and then related their identification to CD4 T cell responses. The computational program (TEA-DQ8) was adapted from one previously developed for identifying peptides bound to the I-A^g7 molecule and based on a library of naturally processed peptides bound to HLA-DQ8 molecules of antigen-presenting cells. We then examined experimentally the response of NOD.DQ8 mice immunized with peptides derived from the Zinc transporter 8 protein. Log-of-odds scores on peptides were experimentally validated as an indicator of peptide binding to HLA-DQ8 molecules. We also examined previously published data on diabetic autoantigens, including glutamic acid decarboxylase-65, insulin and insulinoma-associated antigen-2, all tested in NOD.DQ8 transgenic mice. In all examples, many peptides identified with a favorable binding motif generated an autoimmune T cell response, but importantly many did not. Moreover, some peptides with weak-binding motifs were immunogenic. These results indicate the benefits and limitations in predicting autoimmune T cell responses strictly from MHC-binding data. TEA-DQ8 performed significantly better than other prediction programs

Introduction

The expression of the class II MHC molecule HLA-DQ8 (DQA1*0301/DQB1*0302) is linked to susceptibility to autoimmune diabetes mellitus (1, 2). Similar linkage to the murine I-A^g7 molecules is found in the autoimmune strain NOD that develops spontaneous diabetes (3). Both HLA-DQ8 (referred to as DQ8) and I-A^g7 molecules by lacking an aspartic acid at β57 fail to form an ion pair with an arginine at α76. This absence of β57Asp is correlated with susceptibility to type 1 diabetes mellitus (T1DM) (4, 5). The absence of β57Asp defines the P9 anchoring sites of DQ8 and I-A^g7 molecules during protein processing by antigen-presenting cells (APC) (6–11).

Peptide bound to class II MHC molecules can be identified by biochemical analysis of eluates. These analyses have been invaluable in identifying sequence motifs responsible for peptide selection during processing. Indeed, an extensive analysis of the peptides selected by cell lines expressing either I-Ag7 or HLA-DQ8 molecules indicated a high level of peptides having one or several acidic residues at the carboxyl terminus (10, 12–14). However, the identification of the β cell-MHC peptidome is difficult since antigens responsible for the autoimmune reaction are not directly presented, but only after transferred to an APC. Moreover, the number of islet class II bearing cells is highly limited. In this situation, computational programs that identify the peptides could be highly useful. There have been few attempts to build a computational model to predict DQ8-restricted T cell epitopes (15–17). These programs identified a general preference for acidic residues at P9 but varied concerning the other MHC anchor residues, perhaps reflecting the limited databases on which they were based. Currently, two computational programs are publicly available for predicting peptides bound to DQ8 molecules: RankPep and MHC2Pred (18–20). RankPep derived a weight matrix profile using DQ8-bound peptides (18) from the MHCPEP database (21). MHC2Pred (20) applied a matrix optimization technique,

which is motif less on the DQ8 peptides from the MHCBN and JenPep databases (22, 23), and employed a support vector machine approach to distinguish the DQ8 peptide binders from the non-binders (19).

An effective computational program was previously developed to predict the I-A^{g7} peptidome based on a large data of natural peptides isolated from APC lines (24). Using the previously developed program as a template, this study reports on T cell Epitope Analyzer-DQ8 (TEA-DQ8), a program also based on naturally sequenced peptides but modified to the features of the DQ8 peptidome. TEA-DQ8 offers an effective approach to predict the DQ8 β cell peptidome and its relationship to the CD4 T cell response.

Research design and methods

TEA-DQ8

The TEA-DQ8 system was built on a program, described in Chang *et al.* (24) and detailed in the Supplementary data (available at *International Immunology* Online). In brief, the program applied a model of 100 parameters representing the 20 amino acids at the five MHC-binding pockets, P1, P4, P6, P7 and P9. The features of the model were derived from the DQ8 peptidome by iterating through an expectation–maximization (EM) algorithm. Based on this model, the structural and statistical hindering residues were defined and by treating the proteome as the baseline, log-of-odds (LOD) scores were used to indicate potential peptides bound to DQ8 molecules. The data of the DQ8 peptidome came from a previous study by Suri *et al.* (14): the naturally processed peptides bound to DQ8 molecules were extracted from purified molecules isolated from cell lines expressing DQ8 and examined by tandem mass spectrometry. There were 108 peptide families, each consisting of a nine amino acid core sequence but flanked with various amino- and carboxyl-terminal extensions (14). The 108 peptides examined had an average length of 15 residues. Out of each family, one peptide was selected to be representative, having the core segment and the longest amino- and carboxyl flanks.

For developing the computational program, the LOD scores with the Laplace plus-one pseudo-counts were applied to scale each residue in terms of DQ8 preferences. The LOD scores were generated based on the EM alignment results: positive LOD scores indicate bias toward DQ8 and negative LOD scores indicate bias toward the baseline proteome. For example, the acidic residues favored at the P9 pocket of DQ8 were weighed positively and the hindering residues or unfavorable residues, such as lysine and arginine at P4, P6 and P9, were weighed negatively. Furthermore, bulky residues such as phenylalanine and tryptophan, which can be accommodated by the deep P4 pocket of DQ8, were positively weighed (Fig. 1). The acidic residues were weakly favored at the carboxyl terminus of the peptide at P10 and P11 (Fig. 1).

Important was the use of hindering residues introduced in our prior computational study (24) and which refined the prediction by adding a penalty for each unfavorable residue. The computational hindering residues were defined by either a statistical or a structural approach. A statistical hindering residue is an amino acid that occurs well below what is expected in random assignations. For DQ8, the statistical hindering residues are glycine at P6, alanine, leucine and serine at P9 (Fig. 1). A structural hindering residue is an amino acid whose size is larger than the pocket's size, which is estimated by the size of the largest favored residues at that position. For DQ8, the structural hindering residues are tryptophan at P1, P7 and P9 as well as phenylalanine, lysine, arginine and tyrosine at P9 (Fig. 1).
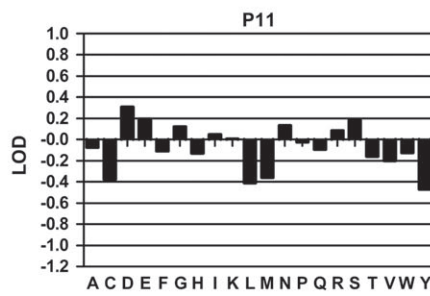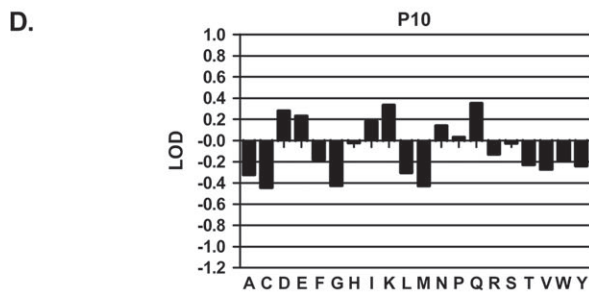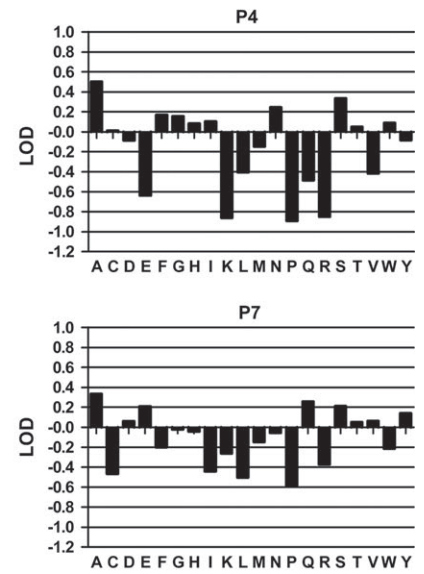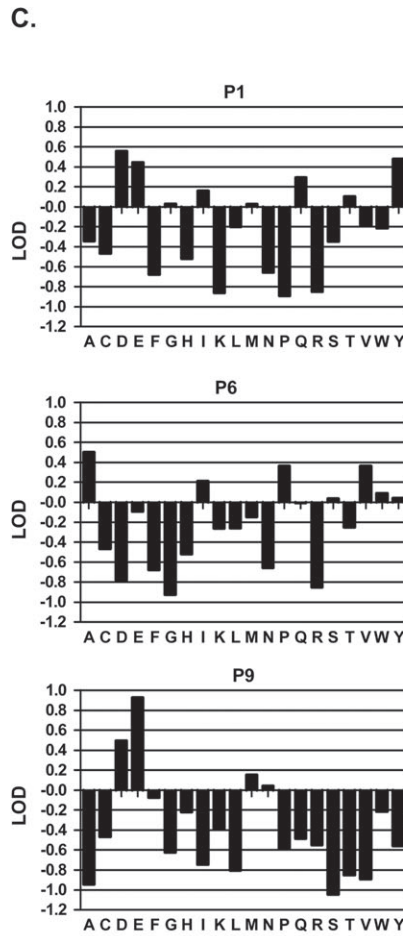
Some of the data are displayed by WebLogo (25) (http://weblogo.berkeley.edu/), a web interface of sequence logos program (26). It displays the level of conservation at specific positions by the height of the residues. Each residue is sorted according to their frequency at each position.

Peptide-binding assays

The binding assays were done on soluble DQ8 proteins produced by the recombinant baculovirus system. The assays were performed under acidic (pH 5.5) conditions as in Suri *et al.* (14). The I^{125}-radiolabeled hemagglutinin reference peptide (FESTGNLIAPEYGFKISY) was incubated at 37°C for 48 h with increasing doses of unlabeled peptides in 20 mM 2-(*N*-morpholino)ethanesulfonic acid and 150 mM sodium chloride. Complexes were purified from free peptide by gel filtration Bio-spin columns (Bio-Rad). The percentage of bound peptide was evaluated by gamma counting. The concentration of unlabeled peptide required for 50% inhibition of labeled peptide binding is reported (IC50). The peptides were synthesized using fluorenylmethoxycarbonyl chemistry on a Symphony Multiplex peptide synthesizer (Protein Technologies, Tucson, AZ, USA), purified to homogeneity by C$_{18}$ reverse-phase high-performance liquid chromatography and verified by matrix-assisted laser desorption ionization time-of-flight mass spectrometry.

Enzyme-linked immunosorbent spot assays

IL-2 enzyme-linked immunosorbent spot (ELISPOT) assays were performed on the popliteal lymph nodes of the NOD.DQ8 transgenic mice. The mice were a generous gift from Dr C. S. David (Mayo Clinic, Rochester, MN, USA) and all the mice were used according to the protocols defined by the Division of Comparative Medicine, Washington University School of Medicine. The mice were immunized in the hind footpads with 10 nmol of Zinc transporter 8 (ZnT8) peptides (27) in a total volume of 100 µl (50 µl per footpad) of complete Freund's adjuvant. IL-2 capture antibody was coated onto the 96-well polyvinylidene fluoride membrane plates (Millipore, Bedford, MA, USA) 6 days after immunization. The IL-2 ELISPOT antibody was diluted and set up according to the manufacturer's direction (BD Biosciences, San Jose, CA, USA). One week after immunization, the mice were sacrificed and the draining lymph nodes were harvested. Single-cell suspensions were added to wells in triplicate (5 × 10^5 cells per well) and incubated at 37°C in 5% CO$_2$ for 24 h with 10 µM ZNT8 peptide in a volume of 200 µl. The IL-2-secreting cells were detected by a secondary anti-IL-2 mAb and the spots of the IL-2-secreting cells were determined by ImmunoSpot 3.2 software.
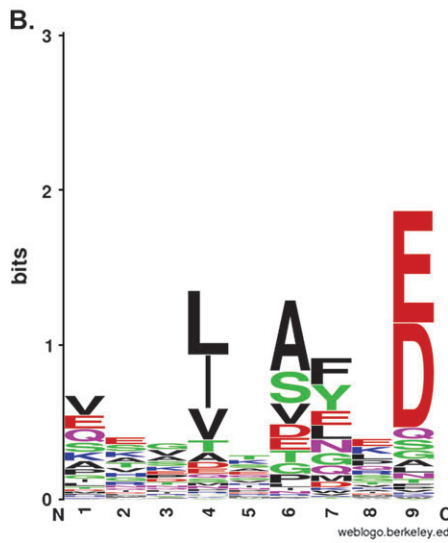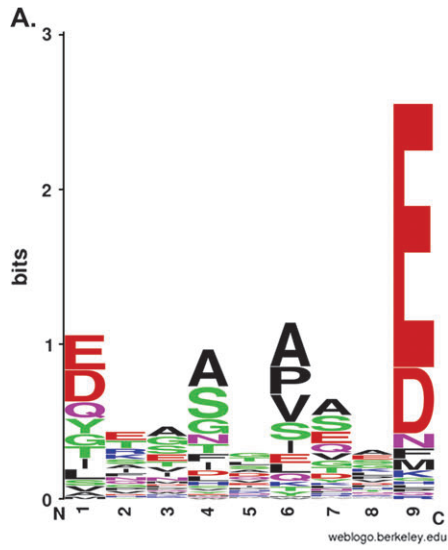
Results

Alignment of DQ8 peptides

The TEA-DQ8 program, which is explained in the Supplementary data (available at *International Immunology* Online), was based on a previous study on peptides bound to I-A^{g7} proteins (24). The information for the EM alignment of the DQ8-bound peptides came from the study of Suri *et al.* (14). The alignment confirmed the binding motif deduced in that study of 108 peptides and in the crystallographic reports (6, 7, 11) (Fig. 1A). It indicates a strong preference for acidic residues at P9, a preference for acidic residues at P1, for small residues at P4, P6 and P7 (alanine and serine) and for proline and valine at P6.

The EM alignment of the I-A^{g7}-bound peptides was used for comparison. (One hundred and ninety-two peptides were examined with an average length of 17 residues (10, 24).) Similar to the DQ8-bound peptides, the I-A^{g7}-bound peptides strongly favor acidic residues at P9 and small residues at P6 (alanine and serine). However, instead of small-sized residues, the I-A^{g7}-bound peptides prefer medium-sized hydrophobic residues at P4 (leucine, isoleucine and valine). The P7 pocket of I-A^{g7} displays a special preference for bulky residues (phenylalanine and tyrosine), which was not found in DQ8. In addition, unlike the DQ8-bound peptides, the I-A^{g7}-bound peptides show no preference for acidic residues at P1. Another difference between DQ8 and I-A^{g7} is the presence of proline found in DQ8 at P6 but rare in the MHC-binding pockets of I-A^{g7} (Fig. 1B).

Testing the program on the ZnT8 response

To predict the DQ8-restricted T cell epitopes of a protein, TEA-DQ8 followed the procedure in Chang *et al.* (24). A sliding window whose size was nine amino acids long traversed the peptide segments to determine the LOD scores based on the MHC-binding pockets; that is the cumulative LOD scores at P1, P4, P6, P7 and P9. The predicted value was treated as an indicator for the potential MHC-binding and T cell epitopes. The LOD values were refined with analysis of the hindering residues and then were sorted to rank the peptide segments as the potential MHC-binding epitopes.

To demonstrate that the LOD scores can distinguish the binding properties of a peptide, an experiment was designed with the ZnT8 protein, recently identified as a diabetic autoantigen (27). Each peptide segment was assigned an LOD score and sorted into two groups: one having the top five values and the other group having the lowest of the top 30 sorted values (Supplementary Table 1, available at *International Immunology* Online). Those with high LOD scores were expected to be good binders, while those with low LOD scores were expected to be weak binders. Four peptides were chosen from each group to measure the binding affinity. The binding experiment shows that the group

with high LOD scores binds DQ8 molecules better than those with low LOD scores (Table 1). (As in the previous study with I-A^{g7}, a strict relationship between LOD score and IC50 value was not found.) Each of the peptides was then tested by immunizing NOD.DQ8 mice with them. Only peptides 166–179 induced a robust response when examined in ELISPOT assays (Fig. 2). Table 1 also shows that the two public programs performed considerably less well.

Epitope prediction and T cell responses

Several epitopes from β cell proteins have been identified in DQ8 transgenic mice and also in human patients. In none of these studies was the relationship established between the T cell responses and the MHC-binding interactions with the DQ8 molecule; that is whether the core segments of the antigenic peptides contained a favorable or unfavorable MHC-binding sequence. We examined the published information applying the computational program for predicting the DQ8-binding segment (28–44) (Supplementary Table 2, available at *International Immunology* Online). We selected to evaluate three proteins, glutamic acid decarboxylase-65 (GAD65), protein tyrosine phosphatase insulinoma-associated antigen-2 (IA-2) and pre-proinsulin, in which NOD.DQ8 mice had been immunized with the protein or with peptides derived from it and then the T cell response was tested.

Human GAD65.   The TEA-DQ8 scan of the entire GAD65 protein with the ranking of the 9-mer peptide segments are shown in Table 2. This scan was then compared with the results published by Herman *et al.* (28) and Liu *et al.* (29). Both reports tested by immunizing with the GAD65 protein and then checking the T cell response to overlapping peptides of 15 (28) or 20 (29) amino acids.

The study of Herman *et al.* (28) from McDevitt's laboratory is the most useful for examining and evaluating the prediction program because it included testing truncated peptides to identify the binding segments (Tables 2). In their study, only six epitopes were identified. In the study by Liu *et al.* (29) from Elliot's laboratory, only four epitopes elicited T cell responses, two of which were also found in the Herman *et al.* report.

Three results emerged from these evaluations: (i) out of the eight peptide segments known to induce a response, six were predicted by our program in the top 25 (ranked 4, 13, 14, 19, 22 and 25); (ii) some segments predicted to bind well and which showed an excellent sequence motif did not induce T cell responses—note the first three in the list—thus there is no direct relationship between rank order and T cell responses, in accordance with the results shown with ZnT8 and (iii) epitopes predicted to bind weakly and which had a poor sequence motif, induced T cell responses, such as the epitope included in peptides 126–140, ranked in 243 in Herman *et al.* (28), and peptides 121–140 and 231–250,

Fig. 1. The figures by WebLogo indicate the alignment of the naturally processed peptides of (A) HLA-DQ8 molecules and those of (B) I-A^{g7} molecules. (C) The panels show the various residues associated with the HLA-DQ8 MHC-binding pockets; that is P1, P4, P6, P7 and P9, including in panel (D), the LOD scores at P10 and P11. Those residues with positive scores are favorable and those with negative scores are unfavorable. (E) The computational hindering residues of the HLA-DQ8 molecules. The statistical hindering residues are in gray and the structural hindering residues are in vertical strip.

Table 1. The binding registers of ZnT8 of DQ8 class II MHC molecules

| Sequence segment		LOD scores	TEA-DQ8 rank	IC$_{50}$ (μM)	T cell response	RankPep rank	MHC2Pred rank
345–358	LTIQIESAADQDPS	1.63	1	2.5	–	30	51
330–343	RTGIAQALSSFDLH	1.27	2	2.5	–	20	26
166–179	LYPDYQIQAGIMIT	1.12	3	0.7	+++	218	251
313–326	ILSVHVATAASQDS	1.03	5	4.9	–	4	62
128–141	SKRLTFGWYRAEIL	0.29	19	20.2	–	50	76
98–111	AILTDAAHLLIDLT	0.24	23	7.2	–	62	39
57–70	KATGNRSSKQAHAK	0.19	26	NB	–	19	168
292–305	AVDGVISVHSLHIW	0.13	30	18.0	–	119	20

NB, non-binder.
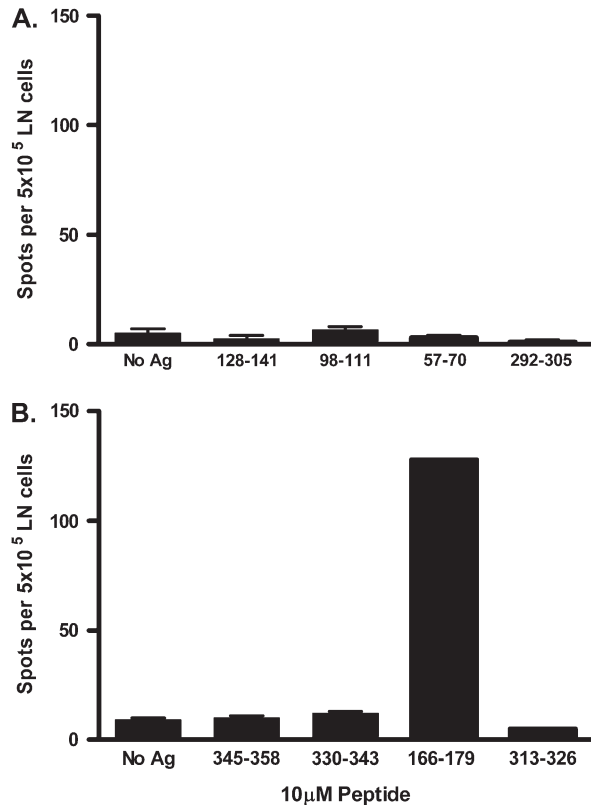
Fig. 2. The T cell responses to the ZnT8 peptides in NOD.DQ8 mice measured in the IL-2 ELISPOT assay. (A) The T cell responses to the ZnT8 peptides which bind weakly to the DQ8 molecules. (B) The panel includes the T cell responses to the ZnT8 peptides which bind stronger better than those included in (A).

ranked 74 and 50, respectively, in Liu *et al.* (29). (Table 2 includes the 9-residue core identified by TEA-DQ8 in the three segments.)

IA-2. Peptides from the cytoplasmic region of IA-2 were examined in a study by Kudva *et al.* (33). Each transgenic DQ8 mouse was immunized with one of 25 peptides, each of 18 residues in length: the T cell response was tested 10 days later. These 18-mer peptides spanned across the cytoplasmic region of IA-2 and each peptide overlapped adjacent peptides by three residues. Of the nine peptides that induced a T cell response, seven were among the top 20 peptides ranked by TEA-DQ8 scanning the entire IA-2 pro-

tein. However, in agreement with the GAD65 results, two of them ranked 29 and 70, also induced T cell responses, despite having very poor binding motifs. Likewise, of the 16 peptides that did not induce a response, four were among the top 20, including three that were ranked 1, 4 and 9 and which contained an excellent binding motif. The remaining 12 were ranked <20 (Table 3).

Insulin. In the human pre-proinsulin, the transgenic DQ8 mice were injected with peptide and tested a week later (34). These peptides which spanned across pre-proinsulin were 20 amino acids in length overlapping adjacent peptides by 10 residues. This experiment was designed to test all possible binding registers although the precise 9-mer segments were not identified. Out of 10 peptides, six stimulated a T cell response and were associated with higher LOD scores (Table 4).

Other proteins. Examined in NOD.DQ8 mice, the oral response to gliadin resulted in response to only two peptides out of 26 that were tested (36). These two peptides were ranked 1 and 11 (Supplementary Table 2, available at *International Immunology* Online). However, along the lines of the results discussed above, a number of other peptides ranked high, which showed excellent core sequences, were negative. For one example, peptides 147–155: QSTYQLLQE, ranked as 13. In another study, immunization with a number of peptides from the house dust mites allergen P2 elicited T cell responses including one which had a poor sequence motif and was ranked 33 (peptides 102–110: ENVVVKVYTV, LOD score of −0.67) (42). Again a number of peptides with excellent core sequences, failed to stimulate T cell responses. For example, peptides 105–113: VVTVKVMGD, ranked as 10 having a LOD score of 0.33. Two other studies also from David's laboratory echoed these same results (43, 44) (data not discussed).

There were no obvious differences in sequences between peptides that generated a T cell response and those that failed, when considering all peptides in the database. (The database has 78 peptides with P9 glutamic acid or aspartic acid among those that did not stimulate T cells, compared with 43 among those that did stimulate; other peptides with non-acidic residues at P9 represented a minority but still did not show any differences.) Whether in the two groups there was a preferential association with other residues was also examined. The group that stimulated T cells showed a slight

Table 2. The DQ8 T cell epitope prediction on human GAD65[a]

Rank	Segments		LOD scores	T cell response
1	403–411	QCSALLVRE	1.41	–
2	484–492	IIKNREGYE	1.19	–
3	308–316	GKMIPSDLE	1.13	–
4	464–472	TTGFEAHVD	1.09	+
5	385–393	ERANSVTWN	0.99	–
6	297–305	DSVILIKCD	0.97	–
7	140–148	ELLQEYNWE	0.93	–
8	49–57	YGDAEKPAE	0.93	–
9	571–579	IDFLIEEIE	0.88	–
10	560–568	VISNPAATH	0.85	–
11	430–438	DKHYDLSYD	0.83	–
12	253–261	IARFKMFPE	0.80	–
13	540–548	YGTTMVSYQ	0.79	+
14	433–441	YDLSYDTGD	0.72	+
15	316–324	ERRILEAKQ	0.66	–
16	68–76	ARKAACACD	0.61	–
17	512–520	SLRTLEDNE	0.60	–
18	509–517	IPPSLRTLE	0.60	–
19	481–489	LYNIIKNRE	0.55	+
20	466–474	GFEAHVDKC	0.54	–
21	218–226	YVTLKKMRE	0.53	–
22	209–217	IAPVFVLLE	0.52	+
23	94–102	ATDLLPACD	0.48	–
24	256–264	FKMFPEVKE	0.47	–
25	104–112	ERPTLAFLQ	0.46	+
26	570–578	DIDFLIEEI	0.46	–
27	144–152	EYNWELADQ	0.45	–
28	15–23	DGSGDSENP	0.41	–
29	243–251	GAISNMYAM	0.40	–
30	14–22	EDGSGDSEN	0.40	–
50	240–248	SPGGAISNM	0.15	+
74	126–134	FDRSTKVID	−0.07	+
243	128–136	RSTKVIDFH	−1.51	+

[a]In the study of Herman *et al.* (28), the examined peptide segments were 15 amino acids long overlapped adjacent peptides by 10 amino acids. Six epitopes were identified. In addition, their truncation analysis revealed the minimal binding registers required to stimulate T cell responses. These six epitopes were 101–115 (104–112), 126–140 (127–136), 206–220 (208–217), 431–445, 461–475 and 536–550 (539–548), where the minimal epitopes are indicated inside the parentheses. In the study of Liu *et al.* (29), the examined segments were 20 amino acids long also overlapped by 10 amino acids. Four epitopes were identified: 121–140, 201–220, 231–250 and 471–490.
[b]The predicted binding registers matched the minimal epitopes identified by the truncation analysis.

preference for proline in the P1 to P8 segment, ~8.8%, compared with 3.7% in the group that did not stimulate T cells. Particularly at P6, the former showed a preference for alanine, proline and valine: 48.8% versus 25.6%. Clearly, the numbers are too small to point to a significant relationship, but these are recorded in Supplementary Figure 1 (available at *International Immunology* Online).

Finally, TEA-DQ8 outperformed two of the public programs (Table 1 and data not shown). Neither RankPep nor MHC2Pep identified any of the peptides that induced a T cell response in the ZnT8 or GAD65.

Discussion

We presented a computational model for predicting the DQ8 peptidome and applied it to the identification of the β cell-

Table 3. The HLA-DQ8 T cell epitope prediction on IA-2 by TEA-DQ8[a]

Peptide examined		TEA-DQ8			
Segment	T cell response	Segment		LOD	Rank
931–948	+	935–943	EIDIAATLE	1.96	2
721–738	+	724–732	NTCATAQGE	1.58	5
		722–730	EPNTCATAQ	0.72	14
601–618	+	610–618	ERLAALGPE	1.53	8
		602–610	QHARQQDKE	0.48	20
826–843	+	835–843	YEVNLVSEH	1.11	10
		828–836	GASLYHVYE	0.34	28
961–979	+	961–969	FALTAVAEE	0.94	11
616–633	+	619–627	GAHGDTTFE	0.91	12
		622–630	GDTTFEYQD	0.60	17
661–678	+	664–672	AAQASPSSH	0.57	18
646–663	+	652–660	SRVSSVSSQ	0.34	29
766–783	+	868–876	TQFHFLSWP	−0.31	70
676–693	–	680–688	EEPAQANMD	1.98	1
691–708	–	692–700	GHMILAYME	1.66	4
		694–702	MILAYMEDH	0.48	22
706–723	–	714–722	QALCAYQAE	1.19	9
		712–720	EWQALCAYQ	0.65	16
946–963	–	952–960	LVRSKDQFE	0.76	13
811–828	–	813–821	VEDGVKQCD	0.48	21
916–933	–	919–927	YILIDMVLN	0.43	23
796–813	–	800–808	ESGCTVIVM	0.39	26
856–873	–	716–724	LCAYQAEPN	0.38	27
		863–871	ETRTLTQFH	0.41	24
751–768	–	755–763	VESSPSRSD	0.32	30
781–798	–	785–793	GPLSHTIAD	0.31	32
631–648	–	640–648	SLFNRAEGP	0.22	35
736–753	–	740–748	HPDFLPYDH	0.18	38
871–888	–	871–879	HFLSWPAEG	0.18	39
886–903	–	889–897	DFRRKVNKC	−0.24	64
901–918	–	908–916	HCSDGAGRT	−0.52	89
841–858	–	850–858	LVRSFYLKN	−0.61	99

[a]Twenty-five peptides of IA-2 were tested on the DQ8 transgenic mice in the study of Kudva *et al.* (33): nine out of these 25 peptides stimulated T cell responses. The table includes the top 30 predictions by TEA-DQ8. The peptides ranked 3, 6, 7, 15, 19 and 25 were excluded because those were not experimentally tested.

DQ8 peptidome. A computational program that identifies the β cell-DQ8 peptidome will facilitate the discovery of novel T cell epitopes in T1DM: it will be practically impossible to identify the β cell-DQ8 peptidome (or the β cell-I-A[g7] peptidome) by sampling natural peptides from the islets. Presentation of β cell class II epitopes is by way of cell transfer since the β cells do not express class II molecules. Moreover, the number of APC within islets is very limited. In the mouse system, we are testing insulinomas that are made to express class II molecules but with a modest degree of success (45).

The main characteristics of the program included applying a specialized EM algorithm that aligned the DQ8 naturally processed peptides based on their MHC-binding pockets, determined computationally the hindering residues of DQ8 molecules and utilized the LOD scores with a baseline proteome. We had reported a program to identify I-A[g7] peptidome (24), which as indicated previously has some differences from DQ8 peptidome, particularly at P1 and P4. One advantage is that the program was based on peptides

Table 4. The HLA-DQ8 T cell epitope prediction on human pre-proinsulin by TEA-DQ8[a]

Peptide examined		TEA-DQ8				
Segment	T cell response	Segment			LOD	Rank
34–53	+	37–45	B13–B21	EALYLVCGE	1.15	1
94–110	+	98–106	A9–A17	SICSLYQLE	1.02	2
1–24	+	12–20	p1–p20	ALLALWGPD	0.71	3
74–93	+	75–83	C19–C27	GSLQPLALE	0.66	4
44–63	+	51–59	B27–C3	TPKTRREAE	0.50	6
14–33	+	21–29	p21–B5	PAAAFVNQH	0.28	8
		20–28	p20–B4	DPAAAFVNQ	0.26	9
		19–27	p19–B3	PDPAAAFVN	0.18	11
84–103	−	87–95	C31–A6	QKRGIVEQC	0.50	5
64–83	−	67–75	C11–19	ELGGGPGAG	0.30	7
54–73	−	57–65	C1–C9	EAEDLQVGQ	0.19	10
		61–69	C5–C14	LQVGGQVELG	−0.01	12
24–43	−	29–37	B5–B13	HLCGSHLVE	−0.12	13

[a]Ten peptides of human pre-proinsulin were tested on the DQ8 transgenic mice in the study of Raju *et al.* (34): six out of these 10 peptides stimulated T cell responses; 34–53 of pre-proinsulin is the same as B10–B29 of insulin. The study by Alleva *et al.* (35) showed that DQ8 restricted T cells against B9–B23 existed in human.

found in naturally processed proteins rather than on peptide libraries or phage display libraries in which the peptides are not edited by the processing process. The nature of peptides in the former differs from those in the latter (45).

Several conclusions became apparent as the program was applied to the β cell proteins and for re-evaluating the published results of others. TEA-DQ8 distinguished the peptide binders of DQ8 molecules as indicated in the examination of ZnT8: the peptides with high LOD scores bound better than those with low LOD scores. Similar results were previously found with the peptides bound to I-A^{g7}. Using mathematical proof based on the hypergeometric distribution, TEA-DQ8 performed significantly better than random guessing and it outperformed other programs that predict DQ8 epitopes.

Although TEA-DQ8 identifies correctly peptides that bind to DQ8, predicting those that are the targets of T cell responses is fraught with uncertainty. As a generalization, the program identified immunogenic peptides that induced T cell responses, but these were not ranked best in all examples. However, the lack of absolute relationship between the rank order and the LOD scores and the presence or extent of the T cell responses is very evident with the four examined β cell antigens. Many peptide segments predicted to bind well and which, indeed, showed the favorable binding motifs, did not elicit T cell responses. In contrast, some poor binding peptides were immunogenic.

Three explanations can be offered for interpreting these results: first, an epitope may have an excellent binding sequence but the segment of the protein is not appropriately selected for processing; that is, the peptide is poorly displayed or not at all. When examining the natural peptides selected by class II molecules, we commented previously that the display and selection of peptides did not equate necessarily to their binding affinity (10, 13). Peptide selection from antigen processing depends on other factors that

are beyond having a class II MHC peptide-binding motif. Clearly, several factors can come into play. Conditions that should influence selection of peptides during antigen processing include the extent of expression of the protein, the localization in the cell, the susceptibility to proteolysis and the presence or absence of interaction with auxiliary molecules like HLA-DM. Thus, this is one major factor for the lack of relationship between binding and degree of the T cell response.

Second, the selection of T cells may depend on the availability of primary T cells to the peptide-MHC and/or to the regulatory process that control the reactivity of the primary T cell repertoire. In the experiments with ZnT8, IA-2 or insulin discussed here, immunization was made with a library of peptides so the issue of processing need not to be considered to explain the lack of T cell responses. The simplest conclusion is that the functional T cell receptor repertoire is limited to only a fraction of the MHC-presented peptides. In the context of autoimmune proteins, some epitopes may be very effective in inducing negative selection, while some may not, depending on the factors described above including the extent of expression in the thymus and the binding affinity to class II MHC molecules and on the interactions with regulatory T cells.

Finally, another important feature with self-proteins is that very weak class II-binding epitopes are effective in inducing T cells (46). In the perspective of autoimmune diabetes, we called attention to insulin peptides in NOD mice which induce strong T cell responses but which bind weakly to the I-A^{g7} molecule (47, 48). Further examples were noted here in the response to GAD65 and IA-2. Such weak-binding peptides will be difficult to identify by any computational program. Weak-binding peptides may evade negative selection and become the predominant effector reaction.

The ideal computational program is the one that first identifies peptide segments that bind to the MHC, such as this one described here, but second segregates the peptides into those that favor or do not favor T cell responses; that is assuming that differences among the two groups may exist. The present information is limited for the second, but this will be the goal in the future as more data become available.

Supplementary data

Supplementary data, Tables 1 and 2 and Figure 1 are available at *International Immunology* Online.

Abbreviations

APC	antigen-presenting cells
ELISPOT	enzyme-linked immunosorbent spot
EM	expectation-maximization
GAD65	glutamic acid decarboxylase-65
IA-2	protein tyrosine phosphatase insulinoma-associated antigen-2
LOD	log of odds
T1DM	type I diabetes mellitus
ZnT8	Zinc transporter 8

References

1 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.

2 Sheehy, M. J., Scharf, S. J., Rowe, J. R. *et al.* 1989. A diabetes-susceptible HLA haplotype is best defined by a combination of HLA-DR and -DQ alleles. *J. Clin. Invest.* 83:830.

3 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. USA* 84:2435.

4 Nepom, G. T. 1993. Immunogenetics and IDDM. *Diabetes Rev.* 1:93.

5 Miyazaki, T., Uno, M., Uehira, M. *et al.* 1990. Direct evidence for the contribution of the unique I-ANOD to the development of insulitis in non-obese diabetic mice. *Nature* 345:722.

6 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.

7 Henderson, K. N., Tye-Din, J. A., Reid, H. H. *et al.* 2007. A structural and immunological basis for the role of human leukocyte antigen DQ8 in celiac disease. *Immunity* 27:23.

8 Corper, A. L., Stratmann, T., Apostolopoulos, V. *et al.* 2000. A structural framework for deciphering the link between I-Ag7 and autoimmune diabetes. *Science* 288:505.

9 Latek, R. R., Suri, A., Petzold, S. J. *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.

10 Suri, A., Vidavsky, I., van der Drift, K. *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.

11 Jones, E. Y., Fugger, L., Strominger, J. L. and Siebold, C. 2006. MHC class II proteins and disease: a structural perspective. *Nat. Rev. Immunol.* 6:271.

12 Munz, C., Hofmann, M., Yoshida, K. *et al.* 2002. Peptide analysis, stability studies, and structural modeling explain contradictory peptide motifs and unique properties of the NOD mouse MHC class II molecule H2-A(g7). *Eur. J. Immunol.* 32:2105.

13 Suri, A., Walters, J. J., Kanagawa, O. *et al.* 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. USA* 100:5330.

14 Suri, A., Walters, J. J., Gross, M. L. *et al.* 2005. Natural peptides selected by diabetogenic DQ8 and murine I-A(g7) molecules show common sequence specificity. *J. Clin. Invest.* 115:2268.

15 Kwok, W. W., Domeier, M. L., Raymond, F. C. *et al.* 1996. Allele-specific motifs characterize HLA-DQ interactions with a diabetes-associated peptide derived from glutamic acid decarboxylase. *J. Immunol.* 156:2171.

16 Godkin, A. J., Davenport, M. P., Willis, A. *et al.* 1998. Use of complete eluted peptide sequence data from HLA-DR and -DQ molecules to predict T cell epitopes, and the influence of the nonbinding terminal regions of ligands in epitope selection. *J. Immunol.* 161:850.

17 Tong, J. C., Zhang, G. L., Tan, T. W. *et al.* 2006. Prediction of HLA-DQ3.2beta ligands: evidence of multiple registers in class II binding peptides. *Bioinformatics* 22:1232.

18 Reche, P. A., Glutting, J. P., Zhang, H. *et al.* 2004. Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. *Immunogenetics* 56:405.

19 Bhasin, M. and Raghava, G. P. 2004. SVM based method for predicting HLA-DRB1*0401 binding peptides in an antigen sequence. *Bioinformatics* 20:421.

20 Bhasin, M. and Raghava, G. P. 2004. http://www.imtech.res.in/raghava/mhc2pred/info.html

21 Brusic, V., Rudy, G. and Harrison, L. C. 1998. MHCPEP, a database of MHC-binding peptides: update 1997. *Nucleic Acids Res.* 26:368.

22 Bhasin, M., Singh, H. and Raghava, G. P. 2003. MHCBN: a comprehensive database of MHC binding and non-binding peptides. *Bioinformatics* 19:665.

23 Blythe, M. J., Doytchinova, I. A. and Flower, D. R. 2002. JenPep: a database of quantitative functional peptide data for immunology. *Bioinformatics* 18:434.

24 Chang, K. Y., Suri, A. and Unanue, E. R. 2007. Predicting peptides bound to I-Ag7 class II histocompatibility molecules using a novel expectation-maximization alignment algorithm. *Proteomics* 7:367.

25 Crooks, G. E., Hon, G., Chandonia, J. M. *et al.* 2004. WebLogo: a sequence logo generator. *Genome Res.* 14:1188.

26 Schneider, T. D. and Stephens, R. M. 1990. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 18:6097.

27 Wenzlau, J. M., Juhl, K., Yu, L. *et al.* 2007. The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. *Proc. Natl Acad. Sci. USA* 104:17040.

28 Herman, A. E., Tisch, R. M., Patel, S. D. *et al.* 1999. Determination of glutamic acid decarboxylase 65 peptides presented by the type I diabetes-associated HLA-DQ8 class II molecule identifies an immunogenic peptide motif. *J. Immunol.* 163:6275.

29 Liu, J., Purdy, L. E., Rabinovitch, S. *et al.* 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.

30 Boyton, R. J., Lohmann, T., Londei, M. *et al.* 1998. Glutamic acid decarboxylase T lymphocyte responses associated with susceptibility or resistance to type I diabetes: analysis in disease discordant human twins, non-obese diabetic mice and HLA-DQ transgenic mice. *Int. Immunol.* 10:1765.

31 Wen, L., Wong, F. S., Burkly, L. *et al.* 1998. Induction of insulitis by glutamic acid decarboxylase peptide-specific and HLA-DQ8-restricted CD4(+) T cells from human DQ transgenic mice. *J. Clin. Invest.* 102:947.

32 Abraham, R. S., Wilson, S. B., de Souza, N. F. Jr. *et al.* 1999. NOD background genes influence T cell responses to GAD 65 in HLA-DQ8 transgenic mice. *Hum. Immunol.* 60:583.

33 Kudva, Y. C., Deng, Y. J., Govindarajan, R. *et al.* 2001. HLA-DQ8 transgenic and NOD mice recognize different epitopes within the cytoplasmic region of the tyrosine phosphatase-like molecule, IA-2. *Hum. Immunol.* 62:1099.

34 Raju, R., Munn, S. R. and David, C. S. 1997. T cell recognition of human pre-proinsulin peptides depends on the polymorphism at HLA DQ locus: a study using HLA DQ8 and DQ6 transgenic mice. *Hum. Immunol.* 58:21.

35 Alleva, D. G., Crowe, P. D., Jin, L. *et al.* 2001. A disease-associated cellular immune response in type 1 diabetics to an immunodominant epitope of insulin. *J. Clin. Invest.* 107:173.

36 Senger, S., Maurano, F., Mazzeo, M. F. *et al.* 2005. Identification of immunodominant epitopes of alpha-gliadin in HLA-DQ8 transgenic mice following oral immunization. *J. Immunol.* 175:8087.

37 Paisansinsup, T., Deshmukh, U. S., Chowdhary, V. R. *et al.* 2002. HLA class II influences the immune response and antibody diversification to Ro60/Sjogren's syndrome-A: heightened antibody responses and epitope spreading in mice expressing HLA-DR molecules. *J. Immunol.* 168:5876.

38 Yang, H., Goluszko, E., David, C. *et al.* 2002. Mapping myasthenia gravis-associated T cell epitopes on human acetylcholine receptors in HLA transgenic mice. *J. Clin. Invest.* 109:1111.

39 Raju, R., Marietta, E., Vinasco, J. *et al.* 2002. Cryptic determinants and promiscuous sequences on human acetylcholine receptor: HLA-dependent dichotomy in T-cell function. *Hum. Immunol.* 63:237.

40 Poussin, M. A., Goluszko, E., David, C. S. *et al.* 2001. HLA-DQ6 transgenic mice resistance to experimental autoimmune

myasthenia gravis is linked to reduced acetylcholine receptor-specific IFN-gamma, IL-2 and IL-10 production. *J. Autoimmun.* 17:175.

41 Neeno, T., Krco, C. J., Harders, J. *et al.* 1996. HLA-DQ8 transgenic mice lacking endogenous class II molecules respond to house dust allergens: identification of antigenic epitopes. *J. Immunol.* 156:3191.

42 Krco, C. J., Harders, J., Chapoval, S. *et al.* 2000. Immune response of HLA-DQ transgenic mice to house dust mite allergen p2: identification of HLA-DQ restricted minimal epitopes and critical residues. *Clin. Immunol.* 97:154.

43 Papouchado, B. G., Chapoval, S. P., Marietta, E. V. *et al.* 2000. HLA-DQ/human CD4-restricted immune response to cockroach allergens in transgenic mice. *Tissue Antigens* 55:303.

44 Chapoval, S. P., Neeno, T., Krco, C. J. *et al.* 1998. HLA-DQ6 and HLA-DQ8 transgenic mice respond to ragweed allergens and recognize a distinct set of epitopes on short and giant ragweed group 5 antigens. *J. Immunol.* 161:2032.

45 Suri, A., Walters, J. J., Rohrs, H. W. *et al.* 2008. First signature of islet beta-cell-derived naturally processed peptides selected by diabetogenic class II MHC molecules. *J. Immunol.* 180:3849.

46 Fairchild, P. J., Wildgoose, R., Atherton, E. *et al.* 1993. An autoantigenic T cell epitope forms unstable complexes with class II MHC: a novel route for escape from tolerance induction. *Int. Immunol.* 5:1151.

47 Levisetti, M. G., Suri, A., Petzold, S. J. *et al.* 2007. The insulin-specific T cells of nonobese diabetic mice recognize a weak MHC-binding segment in more than one form. *J. Immunol.* 178:6051.

48 Levisetti, M. G., Lewis, D. M., Suri, A. *et al.* 2008. Weak proinsulin peptide-major histocompatibility complexes are targeted in autoimmune diabetes in mice. *Diabetes* 57:1852.