Originals

Polymorphic structural features of modelled HLA-DQ molecules segregate according to susceptibility or resistance to IDDM

J. Routsias[1], G. K. Papadopoulos[2]

[1] Laboratory of Immunology, Department of Internal Medicine, University of Ioannina Medical School, Ioannina, Greece
[2] Laboratory of Biological Chemistry, University of Ioannina Medical School, Ioannina, Greece

Summary The structural features of HLA-DQ alleles which are susceptible and resistant to insulin-dependent diabetes mellitus (IDDM) have been examined using a model of their three-dimensional structure obtained by energy minimisation, based on the published structure of HLA-DR1. The model shows DQ molecules to have an overall shape nearly identical to that of DR molecules, but with significant differences in the fine structure: 1) the antigen-binding groove of DQ molecules has a polymorphic first pocket; this pocket can be either amphiphilic or hydrophilic, 2) The β49–56 dimerisation domain of DQ is polymorphic: hydrophobic, or amphiphilic, or hydrophilic and positively charged, leading to spontaneous or T-cell receptor-induced homodimer formation, or difficulty of the formation of such dimers, respectively; 3) a prominent Arg-Gly-Asp loop is formed by some DQ alleles (β167–169) and probably functions in cell adhesion. There are also small differences in the residues and sequences implicated in CD4 binding (mostly in DQβ134–148) but the significance of these differences cannot be evaluated at present. All seven DQ alleles which confer susceptibility to IDDM posses a hydrophilic first pocket in the antigen-binding groove, a hydrophobic or amphiphilic β49–56 dimerisation patch that allows for spontaneous or T-cell receptor-induced dimerisation, and the Arg-Gly-Asp loop. By contrast, in the protective alleles at least one of these three features is absent. This segregation of phenotypes according to susceptibility or resistance can well explain the model of tighter autoantigen binding by the protective alleles compared to the susceptible alleles, previously proposed for the pathogenesis of IDDM. [Diabetologia (1995) 38: 1251–1261]

Key words Antigen presentation, autoimmune diseases, diabetes mellitus, juvenile diabetes, histocompatibility antigens, HLA-DQ molecules.

Corresponding author: Dr. G. K. Papadopoulos, Laboratory of Immunology, Department of Internal Medicine, University of Ioannina Medical School, GR-451 10 Ioannina, Greece
Abbreviations: IDDM, Insulin-dependent diabetes mellitus; HLA, human leukocyte antigen; MHC, major histocompatibility complex; Arg-Gly-Asp, Arginine-Glycine-Aspartate.

About 15 years ago, it was established that several human autoimmune diseases were genetically linked to the HLA/D locus [1]. This locus maps to the short arm of chromosome 6 and contains several genes and pseudogenes. The best studied of the former are those of DP, DQ and DR. Each of these genes codes for one α chain and for at least one β chain. The mature protein, expressed in professional antigen-presenting cells (macrophages, B cells, dendritic cells) is composed of two integral membrane proteins linked non-covalently as an $\alpha\beta$ heterodimer. Such heterodimers participate in the trapping of antigenic peptide fragments in the endocytic compartment of antigen-presenting cells. Thereafter, the complex of antigen plus HLA Class II molecule translocates to the cell

Table 1. Combinations of DQA and DQB genes associated with susceptibility or resistance to IDDM

DQA1, DQB1	Relative Risk
Susceptibility	
0301, 0201	5–20
0301, 0302	8–12
0301, 0303[a]	2
0301, 0401[a]	4
0301, 0402	5–15
0501, 0201	3–5
0501, 0302	8–35
Resistance	
0102, 0602	0.2
0103, 0603[b]	0.2
0103, 0601[a]	0.2
0501, 0301	0.2–0.4

[a] Only significantly increased or decreased in Japanese subjects. These haplotypes are very rare in Black and Caucasian subjects

[b] Only significantly decreased in Black and Caucasian subjects. This haplotype is very rare in Japanese subjects

Modified from previously published data [10], with the permission of the publisher and the authors

membrane, in order to be presented to CD4[+] helper T lymphocytes. In insulin-dependent diabetes mellitus (IDDM), genetic linkage in Caucasian populations has been shown for the serologically-defined antigens DR3 and DR4, while DR2 has been shown to confer protection from the disease [2]. Later studies revealed that this was due to linkage disequilibrium of these alleles to various DQ alleles [3, 4]. Furthermore, genes telomeric and centromeric to HLA-DQ were shown to be less highly associated with the disease than DQ itself [5]. Of even greater significance is the fact that susceptibility to diabetes can be conferred by DQ A and B genes in trans position giving credence to the hypothesis that DQ molecules are true susceptibility genes for IDDM [6]. Meanwhile, a number of studies have linked susceptibility to the disease in Caucasian populations to the absence of aspartic acid in position β57 (DQβ57)[7], and the presence of anarginine residue inposition DQα52 [8]. By contrast, most susceptibility alleles in the Japanese population contain aspartate in position β57 [9]. A more recent worldwide collaborative study using 981 unrelated patients with IDDM and 2228 healthy control subjects, showed that in black, Caucasian, and Japanese populations there are seven susceptible DQ alleles (two of which are exclusive for the Japanese population). Furthermore, there are four dominantly protective DQ alleles in the same populations (with one allele exclusive for the Japanese population [10] (Table 1)). The protection conferred by the various DQ alleles appears to be against all DQ susceptible alleles [10]. A new genetic study scanning the entire human genome has found that the linkage of IDDM to major histocompatibility complex (MHC) genes is the strongest of all other linkages [11].

The molecular basis of the susceptibility to IDDM, or resistance to it, has remained a mystery, even though an attractive hypothesis has been proposed [12]. Specifically, it states that the "diabetogenic peptide(s)" binds tighter to the resistant than to the susceptible DQ molecules, leading to elimination of strongly autoreactive T-cell clones in the thymus during ontogenesis of the immune system. In the periphery, the protective DQ molecules would effectively compete with the susceptible molecules for binding to the peptide(s), thus not permitting the latter combination to be presented on the surface of antigen-presenting cells [12]. Other DQ alleles that are neutral with respect to diabetes susceptibility are regarded as not being capable of binding such diabetogenic peptide(s). While this hypothesis is not contrary to the current immunological paradigm, very little is known about the function of DQ molecules; they are only constitutively present on a subfraction of antigen-presenting cells, and at a much lower density than DR molecules [13]. However, their expression on antigen-presenting cells can be induced by interferon γ [14], and infectious agents such as the Epstein-Barr virus [15]. It has been postulated that DQ molecules exercise epistatic action over DR molecules [16] and that they are the mediators of immunosuppression [17].

The structure of nearly the entire extracellular part of the related human MHC molecule HLA-DR1 has been published in complex with one [18] or several [19] antigenic peptides. The bound antigen is held by hydrogen bonds from polar DR1 residues to its amide and carbonyl groups and by five pockets into which residues of limited or wide specificity can fit. Aspartate β57 participates in this hydrogen bonding and also forms a salt bridge to the apposed α76Arg [19]. The importance of this aspartate as a major determinant of peptide affinity has also been demonstrated in a far-reaching peptide-binding study to 11 different DR molecules [20]. The pocket motif has been independently corroborated for other DR molecules by the same study [20] and by pool sequencing of bound peptides to various DR molecules [21–23], and recently to one DQ and one DP allele [24]. Also of interest is the formation of the homodimer of heterodimers of DR1 maintained by interactions of residues in the β49–56 patch, as well as several interactions in the α2β2 domain. Dimerisation seems an obligatory step in the signal transduction of several ligated cell-surface receptors (e.g. growth hormone [25]), and in this case it is presumed to be induced by the binding of cognate T-cell receptors to identical DR molecules [19].

We have constructed a model of the structure of DQ molecules, based on the 60–70 % identity between them and the HLA-DR1 molecule, the similar peptide binding characteristics and the binding to the CD4 co-receptor [26]. Such modelling studies

have retrospectively been shown as quite successful when there is at least 40 % identity between the known and the modelled protein and similarity of biochemical functions [27–29]. While there is near identity in overall shape between DR and DQ molecules, there are several unique structural features of the latter: 1) a polymorphic first pocket in the antigen-binding groove that is either hydrophilic or amphiphilic (with slight variations on this theme), 2) a polymorphic β49–56 dimerisation patch that probably leads to either spontaneous or T-cell receptor-induced dimerisation, or difficulty of doing so, and 3) a prominent Arg-Gly-Asp loop in DQβ167–169 that probably functions in cell adhesion.

In this study we have focused on DQ alleles known to confer either susceptibility or protection to IDDM [10]. The seven diabetes-susceptible DQ alleles are all characterised by a hydrophilic first pocket in the antigen-binding groove, a hydrophobic or amphiphilic β49–56 sequence that leads to spontaneous or T-cell receptor-induced dimerisation, and the presence of an Arg-Gly-Asp loop. By contrast, three of the four dominantly resistant DQ alleles lack at least two out of three of the mentioned structural features, and the fourth one (A1*0501/B1*0301) lacks only the Arg-Gly-Asp loop. A structural explanation of the model of Nepom [12] can now be provided by the results of our modelling: the susceptible DQ molecules bind to diabetogenic peptide(s) and can initiate an immune response by virtue of spontaneous or induced homo-dimerisation in the α1β1 domain followed by homodimerisation in the α2β2 domain and CD4 binding. By contrast, resistant molecules probably bind the diabetogenic peptide(s) more avidly, due to a different first pocket and the presence of aspartate in β57, yet cannot activate cognate T-cell clones because the latter have been deleted in the thymus. In case any such clones have escaped deletion, the resistant DQ molecules cannot easily form dimers, a necessary prerequisite for T-cell activation, or lack the Arg-Gly-Asp putative cell-adhesion loop.

Materials and methods

The high resolution structure of HLA-DR1 (kindly provided by Professor D. Wiley) [18] was used as a basis for the modelling of DQ molecules. The aminoacid sequences of the DQ alleles were obtained from a previously published study [30] after substitution of the appropriate aminoacid for each nucleotide triplet in the DNA sequence. The alignment of DQ aminoacid sequences to those of DR molecules was based on that first presented in Brown et al. [31]. Only one insertion was necessary in position 9 of the α chain of DQ (an alanine residue) [31]. In keeping with earlier alignments [31] and in consideration of all conserved interactions between conserved side chains we have placed DQα9Ala in the position of DRα6Val, and placed the aminoacids 6–8 of DQα in the positions of DRα3–5 (no coordinates for DRα1–2 provided in [18]). The allele DQA1*0101/DQB1*0501 was generated

first, and all other DQ combinations were obtained by substitution of the residues at appropriate positions. The whole procedure was performed on a Silicon Graphics Iris 5.1 workstation using the program Insight II, version 2.3.1 (Biosym Corp., San Diego, Calif., USA). At each substitution, there was an automatic selection of the most suitable rotamer, from a library of rotamers provided by the program. After each aminoacid substitution there was an automatic energy minimisation to obtain the most suitable arrangement of all residues. Following all substitutions, two hundred iterative cycles of energy minimisation ensued by the conjugate gradient method. The changes in energy after the first 30–40 cycles were minimal. The pockets in the antigen-binding groove of the α1β1 domain of each modelled DQ molecule were identified as large depressions in the van der Waals surface representation of the model. The DQ molecules obtained thus are very similar in overall structure to the published structure of DR1 [18, 19]. The mean root mean square (rms) deviation for Cα atoms between DR1 and the DQ molecule A1*0101/B1*0501 is 0.51 Å and no value is above 1.5 Å. Other DQ molecules give just as good if not better structural agreement with DR1. The relevant DQ alleles were taken from the recent worldwide collaborative study, a part of the 11th International Histocompatibility Workshop [10]. The coordinates of the modelled DQ molecules will be deposited at the Brookhaven Data Bank (Chemistry Department, Brookhaven National Laboratory, Upton, N. Y. 11973, USA).

Results

All of the DQ molecules, including those conferring susceptibility or resistance to IDDM, have an overall structure exemplified by Figure 1. Specifically, they possess the mould of a class II histocompatibility molecule with an antigen-binding groove in the α1β1 domain that is bounded by a "floor" of eight β-sheets and two "walls" of antiparallel α-helices. There is an α2β2 domain very much like an immunoglobulin fold, that contains the homodimerisation region, the CD4-binding area and the Arg-Gly-Asp loop. This domain and the α1β1 domain dimerise in some alleles with their counterparts in an identical DQ molecule forming a homodimer of αβ heterodimers (Fig. 1).

The antigen-binding groove. The polar residues lining the antigen-binding groove of DR1 and participating in hydrogen bonding with the antigenic peptide, are located in exactly the same positions in modelled DQ molecules ([18, 19], Figs. 2, 3), with nearly identical orientations [26]. Within the antigen-binding groove of modelled DQ molecules we identified at most five pockets that can trap specific residues of the antigenic peptides. They appear in nearly the same positions and are formed by equivalently placed residues as in DR1 [26]. The first pocket of DQ molecules (formed by α10, 27, 34, 35, 46 and β85, 86, 89 and 90) is either amphiphilic or hydrophilic, as judged from the aminoacids that line-up this formation (Figs. 2, 3, Table 2). The hydrophilic var-
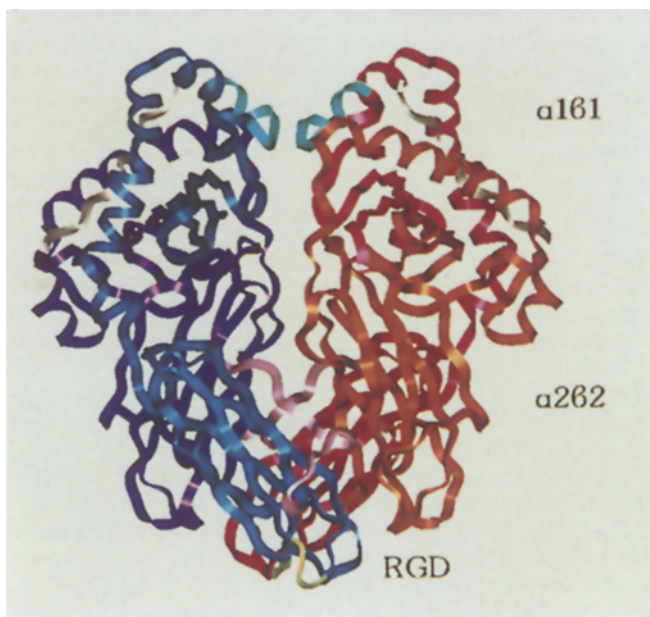
Fig. 1 Overall structure of the extracellular domains of the DQ molecule A1*0301/B1*0302 as a homodimer of two αβ heterodimers, obtained after aminoacid substitution on the high resolution coordinates of DR1 [19] and energy minimisation, showing the antigen-binding groove (in the α1β1 domain with a virtual trapped peptide in grey), the β49–56 dimerisation patches (turquoise), the CD4 binding stretch (β134–148, pink) and the Arg-Gly-Asp (RGD) loop (yellow-green). On the left the α chain is in purple (away from the viewer) and the β chain is in blue (towards the viewer). On the right, the α chain is in orange (towards the viewer) and the β chain in red (away from the viewer). There is an identical RGD loop and CD4 binding region formed by the α and β chains that are away from the viewer (shown in identical colours). The backbone of the trapped peptide fragment is actually that of the HA306–318 peptide in the coordinates provided for its complex with DR1 [18], and shown here only for purposes of orientation. The CD4 molecule is postulated to dock in the empty space shown between the two heterodimers. There are two identical docking positions, one towards the viewer and one away from the viewer (see also Fig. 4). All modelled DQ molecules are nearly identical in gross shape to the one depicted here. The cell membrane of the antigen-presenting cell that expresses DQ molecules is parallel to the horizontal and perhaps 10 residues below the level of the RGD loop. This figure is drawn on a Silicon Graphics Iris workstation using Insight II

iant of the first pocket appears shallower, owing to the presence of bulky residues in β85, 86 and 89 (Leu, Glu and Thr instead of Val, Ala and Gly, respectively for the amphiphilic variant). The character of the first pocket is also modulated by residue α34 (Glu or Gln). All of the IDDM-susceptible DQ alleles have a first pocket that is of the second variety, i.e. smaller and hydrophilic, while three of the four resistant alleles have an amphiphilic, and slightly more accommodating first pocket (Table 3). The fourth resistant molecule DQA1*0501/DQB1*0301, has the hydrophilic form of a first pocket. The second pocket appears as the most prominent or anchoring pocket.

This is most probably due to small residues (e. g. Gly, Ala) in positions α9 and β13 of DQ molecules in contrast to the bulky glutamine and phenylalanine, respectively, found in DR1. There is extensive polymorphism in the four residues from the β1 helix of this pocket (70, 71, 74 and 78) but no clear distinction between susceptible and resistant DQ molecules. Likewise, different residues exist in certain positions of the other pockets without a separation into the susceptible and resistant forms. We note however, the presence of aspartate in β57, which is part of the fifth pocket and also a major determinant of peptide affinity. All resistant alleles, and three of seven susceptible alleles, contain aspartate in this position; the residue is oriented in a manner identical to that in DR1 (forming a salt bridge with α79Arg and a hydrogen bond with an amide group of the antigenic peptide).

The β49–56 and α2β2 dimerisation patches. There are important differences in the β49–56 dimerisation patch [19] of all the susceptible DQ molecules when compared to the patch of the resistant molecules. DR molecules have a monomorphic β49–56 dimerisation patch, with dimerisation probably promoted after T-cell receptor binding [19], by symmetrical salt bridges between β52Glu of one heterodimer and β55Arg of the opposite heterodimer. By contrast, DQ alleles are polymorphic in this region leading to β49–56 sequences that may be very hydrophobic (hydrophobicity of 8.6 or 14.6 kcal/mol), or amphiphilic (hydrophobicity of –0.5 kcal/mol), or hydrophilic (–10.4 kcal/mol) and positively charged. Of the seven diabetes-susceptible DQ alleles five (those with DQB1*0201, 0302, 0303, Table 1) have a very hydrophobic patch, and two (DQB1*0401, 0402) an amphiphilic patch [26]. The positive hydropathy values are such that exposure to an aqueous environment is prohibitive, and probably leads to spontaneous homodimerisation via this patch. The two susceptible alleles (DQB1*0401 and DQB1*0402) with the amphiphilic dimerisation patch can probably form homodimers, albeit with greater difficulty than the other susceptible alleles. Their conformation in this region is probably more peculiar. They have an arginine at position 55, yet they contain two leucines, at positions 53 and 56, right opposite to each other at the first turn of the β1 helix (Figs. 1, 3). This will probably bring a distortion to the conformation of these DQ molecules, such that when close together because of cross-linking by cognate T-cell receptors, the arginines would turn away from this hydrophobic patch allowing the four leucines in the homodimer to come opposite to each other (like a mini leucine zipper [32]). On the other hand, three of the four resistant alleles (DQB1*0601, 0602, 0603) have the hydrophilic variant that is also positively charged because of β55Arg and one (DQA1*0501/DQB1*0301) has the hydrophobic variant (8.6 kcal/mol).

```
                1        10        20        30        40        50        60
DR1             -- KEE  -IIQAEF LN DQ  EFMFD       I H  MAK   V  LE  A  AS EA
DQA1*0101       EDIVADHVASCGVNLYQFYGPSGQYTHEFDGDEEFYVDLERKETAWRWPEFSKFGGFDPQ
DQA1*0301       _____Y_____S_____S_____V_QL_L_RK_RS_____
DQA1*0501       _____Y_____S_____Q_____G____V_CL_VLRQ_R___

DQA1*0102       _____Q_____
DQA1*0103       _____S_____F_____Q_____K_____
DQA1*0501       _____Y_____S_____Q_____G____V_CL_VLRQ_R___
                        ~   *               *    ~    **          *
```

```
                61       70        80        90        100       110       120
DR1             __A_I__D_A__E__T__S_Y_PI__VP____LTN___E_RE__V__FI_KFT_____
DQA1*0101       GALRNMAVAKHNLNIMIKRYNSTAATNEVPEVTVFSKSPVTLGQPNTLICLVDNIFPPVV
DQA1*0301       F__T_I__L_____L__S_____
DQA1*0501       F__T_I__L____SL__S_____

DQA1*0102       _____
DQA1*0103       _____..............................
DQA1*0501       F__T_I__L____SL__S_____
                   !       !       !          #              #    /
                           ~
```

```
                121      130       140       150       160       170       180
DR1             _V__R__KP_T_____V__PRE__L_R_FH__P____TEDV___R_____E_____
DQA1*0101       NITWLTNGQSVTEGVSETTFLSKSDHSFFKISYLTFLPSADEIYDCKVEHWGLDQPLLKH
DQA1*0301       _____H_____E_
DQA1*0501       _____H_____L____E_S_____K_____

DQA1*0102       _____D_____
DQA1*0103       ...........................................................
DQA1*0501       _____H_____L____E_S_____K_____
                 / /    / / /                       ##  # / /  /// # #
```

```
                181      190       200
DR1             __FDA_S_LP_T__N_____
DQA1*0101       WEPEIPTPMSELTETVVCAL
DQA1*0301       _____T_____
DQA1*0501       _____

DQA1*0102       _____
DQA1*0103       ...................
DQA1*0501       _____
                 # #
```

Fig. 2 Aminoacid sequences of the first two domains of DQα chains susceptible and resistant to IDDM. Aminoacids are in the one letter code. DRα and DQA1*0101 are placed for comparison. The top group in each sequence comprises the susceptible alleles and the bottom group the resistant ones. DQA1*0501 is in both groups as there is one allele conferring susceptibility and another conferring resistance that use this particular α chain. DR1α has three aminoacids less than DQα. The numbering used here refers to DQα. Code: _: identity; -: no aminoacid in corresponding position; .: aminoacid unknown; ~ : intramolecular hydrogen bonds or salt bridges; !: intermolecular hydrogen bonds (between class II MHC and trapped peptide); *: residues participating in the first pocket (by analogy to DR1); @: residues forming the β49–56 dimerisation patch; /: residues participating in the formation of the CD4 binding region; #: interacting residues that stabilise the homodimer of heterodimers in the α2β2 domain (α2 of one heterodimer with β2 of another heterodimer)

```
          1        10        20        30        40        50        60
DRB1      G_TRPR_LW_L_FE_H_F_____LLE_C___S_____E____EL__D___
DQB1*0501 RDSPEDFVYQFKGLCYFTNGTERVRGVTRHIYNREEYVRFDSDVGVYRAVTPQGRPVAEY
DQB1*0201 _____M_____L_S_S___I_____EF____LL_L_A___
DQB1*0302 _____M_____L___Y_____A_____L_P_A___
DQB1*0303 _____M_____L___Y_____A_____L_P_D___
DQB1*0401 _____F___M_____L_____Y_____A_____L__LD___
DQB1*0402 _____F___M_____Y_____A_____L__LD___

DQB1*0301 _____AM_____Y___Y_____A_____E____L_P_D___
DQB1*0601 _____F___M_____L___Y_____A_____D___
DQB1*0602 _____F___M_____L___Y_____A_____D___
DQB1*0603 _____M_____L_____A_____D___
                                 ~                   ~    @@@@@@@@!
                                                              ~
```

```
          61       70        80        90       100       110       120
DR1       _____DL__QR_A__TY_____G_GESFTV_____K__VY_K_QP_Q_____V__S
DQB1*0501 WNSQKEVLEGARASVDRVCRHNYEVAYRGILQRRVEPTVTISPSRTEALNHHNLLICSVT
DQB1*0201 _____DI__RK__A_____QLEL_TT_____V___
DQB1*0302 _____RT__EL_T_____QLEL_TT_____V___
DQB1*0303 _____RT__EL_T_____QLEL_TT_____V___
DQB1*0401 _____DI__ED_____T_____QLEL_TT_......................
DQB1*0402 _____DI__ED_____T_____QLEL_TT_____V___

DQB1*0301 _____RT__EL_T_____QLEL_TT_____V___
DQB1*0601 _____DI__RT__EL_T_____F_____V___
DQB1*0602 _____D_T__EL_T_____F_____V___
DQB1*0603 _____T__EL_T_____F_........................
            !              !!  **   **              #      ##
```

```
          121      130       140       150       160       170       180
DRB1      G___GS_E_____G___K_____G_Q_____T_____TV_RS_E____Q_____V
DQB1*0501 DFYPSQIKVRWFRNDQEETAGVVSTPLIRNGDWTFQILVMLEMTPQRGDVYTCHVEHPSL
DQB1*0201 _____A_____
DQB1*0302 _____A_____T_____
DQB1*0303 _____A_____T_____
DQB1*0401 ..............................................
DQB1*0402 _____A_____T_____

DQB1*0301 _____A_____T_____H_____
DQB1*0601 _____G_____H_____
DQB1*0602 _____G_____
DQB1*0603 ..............................................
                ///////////////            #
                       ##
```

Fig. 3 Aminoacid sequences of the first two domains of DQB chains susceptible and resistant to IDDM. The sequences of DR1 and DQB1*0501 are placed for comparison. The top group in each sequence contains the β chain sequences of sus-ceptible DQ alleles, while the bottom group contains those resistant to IDDM. Aminoacid notations and codes as in Figure 2. The β167–169 Arg-Gly-Asp (RGD) loop is in bold type

The α2β2 homodimerisation domain involves a large surface area of the DQ molecule (ca. 900 Å²) and the dimerisation is stabilised by multiple interactions involving charged and hydrophobic residues ([19, 26], and Figs. 1–3). There are very few substitutions (α114Lys → Asn and α161Glu → Asp, the latter only in some alleles, Fig. 2) in residues considered crucial for this extensive interaction. Therefore, the homodimerisation of DQαβ heterodimers in the α2β2 domain is expected to be promoted by the specific interactions of the same residues, with no differences between susceptible and resistant DQ molecules, just as in DR1.

The CD4 binding area and the Arg-Gly-Asp loop. The CD4 binding area is formed by the homodimerisation of DQ molecules (β2 of one DQ heterodimer and α2 of another DQ heterodimer), and has been shown in DR to be composed of the sequence β134–148 and several residues on the α chain apposed to this sequence (with a mark under each residue, in Figures 2 and 3; [19, 33, 34]). There are two symmetrical CD4-binding areas in each homodimer of αβ heterodimers ([19] and Fig. 1). Of the residues shown as crucial for CD4 binding to HLA-DRβ, by site-directed mutagenesis (β137Glu, 142Val, 143Val, [33]), all remain invariant in the known DQB alleles. The effect of the three-four substitutions in the DQβ134–148 stretch (Fig. 3) on CD4 binding cannot be evaluated. Likewise, nothing can be stated about the two substitutions in the apposing residues from the α chain of the other heterodimer (130Pro → Ser and 175Glu → Gln or Lys, in all but one allele). We cannot discern any obvious structural difference in this region between DQ alleles conferring susceptibility or resistance to IDDM.

The Arg-Gly-Asp loop is present in six of seven alleles susceptible to IDDM (but unknown in the seventh, DQB1*0401), as well as the protective allele DQB1*0602, but it is missing (DQB1*0301, DQB1*0601) or not known (DQB1*0603) for the other three protective alleles.

Discussion

The various structural features of modelled DQ molecules are well segregated between the resistant and the susceptible alleles to IDDM, with the exception of DQB1*0301 (Table 3). The antigen-binding groove has a distinctly different first pocket in susceptible DQ molecules (hydrophilic and shallow) compared to that of resistant molecules. The latter group also has aspartate in position β57, a residue that our modelling studies show as crucial in maintaining a salt bridge with DQα79Arg. This Asp-Arg pair is crucially situated to also form hydrogen bonds with the proximal amide nitrogen and carbonyl oxygen of the

bound antigenic peptide in the manner shown for DR1 [18]. A recent very detailed study of the binding affinities of several DR molecules to given antigenic peptides has shown that a β57Asp → Ser substitution (DRB1*0401 → 0405) leads to a decrease in the affinity of the peptide by three orders of magnitude [20]. The pocket motif for DQ antigen A1*0501/B*0301 (protective with respect to IDDM) has been determined by pool sequencing. It was shown that anchoring exists in relative positions 1, 5, and 7 of the peptide, with at least five residues preceeding the aminoacid in relative position 1 [24]. We interpret this relative position 1 as the anchoring residue for pocket 2. The clustering of aromatic and bulky aliphatic aminoacids in this position [24] is consistent with the character of this pocket in this DQ allele, as shown by our modelling. Furthermore, in relative position –2, the authors found a nest of polar or charged residues (Asp, Lys, Glu, Asn and Gln), which is also consistent with our modelling for the first pocket for this antigen.

We are fortunate to have a protective (DQA1 *0501/DQB1*0301) and a susceptible (DQA1*0501/ DQB1*0302) molecule sharing the same α chain and differing in the β1 domain in four residues only (Table 1 and Fig. 3). Three of the four substitutions are in the antigen-binding groove (β13Ala → Gly, 26Tyr → Leu, 57Asp → Ala). The residues in β13 and 57 participate in the formation of the second and fifth pockets, respectively. The substitution of alanine for aspartate in position β57 removes a source of stability in peptide-MHC II and α-β interchain interactions [20]. It is also expected to alter the affinity of the DQ molecule for the bound antigenic peptide, as discussed above. Thus, the fine structure of the peptide-binding groove in this region is also changed, leading to further changes in the binding affinity of the putative diabetogenic peptide(s). The fourth substitution β45Glu → Gly is in a position not directly in contact with the antigenic peptide. However, this glutamate residue is considered essential for the maintenance of a negative electrostatic potential on the surface of this specific DQ molecule [35]. Its conversion to glycine in DQB1*0302, changes this potential towards substantially more positive values [35]. This very large change in the potential will affect the binding affinities of a given antigenic peptide for these two DQ alleles. Therefore, the four aminoacid substitutions in toto signify a change in the affinity of the DQ molecule to the diabetogenic peptide(s). A calculation of the electrostatic surface potential of various DQ alleles has shown that DQ molecules susceptible to IDDM have a significantly more positive potential than the resistant molecules [35, 36]. This finding is complementary to our results of distinct differences in the first pocket and the presence of aspartate β57 in resistant molecules. It shows that the susceptible DQ molecules will have a

Table 2. Structural features of DQ molecules

Domain	Features	Character	DQ alleles	
			Susceptible	Resistant
$\alpha 1 \beta 1$	Antigen-binding groove	Polymorphic		
	First residue binding pocket	Dimorphic, hydrophilic, or amphiphilic	Hydrophilic	Amphiphilic[a]
	Residue at $\beta 57$	Polymorphic	4/7 Ala, 3/7 Asp	4/4 Asp
	$\beta 49$–55 dimerisation patch	Polymorphic, very hydrophobic, or amphiphilic[b], or hydrophilic[b]	very hydrophobic (5/7) or amphiphilic (2/7) thus dimer formers	Hydrophilic[a] thus dimer non-formers
$\alpha 2 \beta 2$	CD4 binding region $\beta 134$–148 and several residues from the apposed α chain		No discernible differences	
	$\beta 167$–169 RGD loop	probably involved in cell adhesion	6/7 posses loop 1/7 unknown sequence	2/4 do not have loop 1/4 unknown sequence

[a] DQB1*0301 allele though in the resistant category has the characteristics of the susceptible alleles in this respect
[b] Both are positively charged

Table 3. Structural features of DQ molecules resistant to IDDM

DQA, B	First pocket	$\beta 57$ Asp	$\beta 49$–56 dimerisation patch	$\beta 167$–169 RGD loop
0102, 0602	amphiphilic	present	hydrophilic and charged	present
0103, 0603	amphiphilic	present	hydrophilic and charged	unknown sequence
0103, 0601	amphiphilic	present	hydrophilic and charged	absent
0501, 0301	hydrophilic	present	hydrophobic	absent

lower affinity, according to our reasoning, for the diabetogenic peptide(s) than the resistant molecules.

The difficulty of homodimerisation by resistant DQ molecules (hydrophilic and positively-charged $\beta 49$–56 dimerisation patch) means that in case any cognate T-cell clones exist in the periphery, their activation upon recognition of resistant DQ molecule complexed with a diabetogenic peptide would be difficult. By contrast, susceptible DQ molecules, can form homodimers with great ease, so when complexed with diabetogenic peptide(s), they could lead to the activation of cognate CD4+ T-cell clones. DQ molecules that show difficulty of homodimerisation would not however signify a generalised immune deficiency of the host in the homozygous state. Of the three human class II MHC loci, DR is constitutively expressed in antigen-presenting cells. Since the $\beta 49$–56 dimerisation patch is monomorphic in all DR alleles, the latter can form homodimers after cognate T-cell receptor recognition and binding.

The Arg-Gly-Asp loop, present in all the susceptible DQ molecules with known sequences in the region, probably functions in cell adhesion as in other integral membrane proteins and proteins of the extracellular matrix involved in such function [37–41]. This is a totally novel function for DQ molecules that requires further investigation.

A scheme where the three structural features of the DQ molecules are shown to participate in the fashionning of the immune response by these molecules is shown in Figure 4. These features segregate in the two phenotypes of the susceptible and resistant DQ alleles for IDDM (Table 3). All susceptible alleles have: 1) a hydrophilic and rather shallow first pocket, 2) a hydrophobic or amphiphilic $\beta 49$–56 dimerisation patch, presumed to lead to spontaneous or T-cell receptor-induced homodimerisation and CD4 binding, and 3) the Arg-Gly-Asp loop, supposedly participating in cell adhesion. The difference in the physicochemical properties of the antigen-binding groove of susceptible and resistant DQ alleles (as exemplified by the amphiphilic first pocket and presence of aspartate in $\beta 57$ in resistant alleles) would translate into different affinities for given diabetogenic peptide(s). Such differences would certainly play a role both in the ontogenesis of the immune system and in the mounting of a specific DQ-restricted immune response in the mature organism. Indeed, the human embryonic thymus is very rich in DQ molecules that probably function as restriction elements [42]. Therefore, the CD4+ T-cell clones recognising the thymus-resistant DQ molecules that have very tightly bound diabetogenic peptides, would be eliminated [43]. By contrast, CD4+ T cells recognising the combination of susceptible DQ molecules and diabetogenic peptides, (that *a priori* are not tightly bound), would mature in the thymus [44]. In the heterozygous state the higher affinity of the resis-
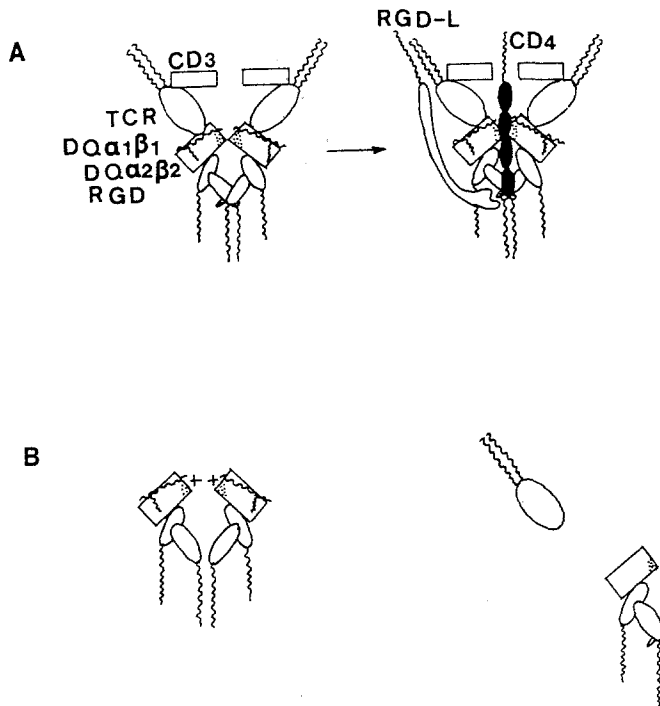
Fig. 4 (A, B) Putative mechanism of susceptibility to and protection from IDDM as it emerges from the study of the structural features of the respective alleles. **A.** Susceptibility arises because under proper environmental stimulation susceptible DQ molecules can present diabetogenic peptide(s) to cognate T-cell clones. Such clones must have not been deleted in the thymus during ontogeny. The peptides are shown as wavy lines through the α1β1 domain of DQ with a perpendicular projection signifying the anchor residue in the anchor pocket. The β49-56 dimerisation patch is depicted as the dotted corner in the α1β1 domain. In one category, (DQA1*0301/DQB1*0401 and DQA1*0301/DQB1*0402) the amphiphilicity of the β49-56 dimerisation patch and the recognition by cognate T-cell receptor molecules is postulated to overcome the repulsion of arginines. CD3 stands for the several invariant chains of the T-cell receptor. RGD-L is the unknown RGD ligand postulated as necessary for DQ-restricted T-cell clone activation. **B.** Protection arises from the tight binding of diabetogenic peptides by DQ molecules in the thymus that leads to elimination of the respective T-cell clones. Any diabetogenic clones that escape from the thymus will be in permanent suppression because of the difficulty of dimerisation or absence of the RGD loop (left). In heterozygous individuals the trapping of the diabetogenic peptides by the resistant DQ molecules does not leave sufficient quantity of antigen for trapping and presentation by the susceptible DQ molecules. Consequently, existing cognate diabetogenic T-cell clones cannot be activated (right). Most of the elements in the mode of depiction of the DQ molecules and the T-cell receptor are taken from published data [19] with the permission of the publisher and the authors

tant DQ molecules for diabetogenic peptide(s) would not leave a sufficient quantity of antigen for presentation in combination with susceptible DQ molecules in the thymus. Thus, few if any diabetogenic CD4+ T cells would be selected for export.

In the periphery, the susceptible DQ molecules expressed under proper stimulation on antigen-present-ing cells could bind to diabetogenic peptide(s) and present such complexes to cognate CD4+ T cells. The ensuing immune reaction could be one of the first steps leading to IDDM. The dominant effect of protective DQ molecules could be exercised in the periphery in the same manner as in the thymus, in accordance with Nepom's hypothesis [12] (Fig. 4). The difference in the physicochemical character of the antigen-binding groove, assures a preferential affinity for the diabetogenic peptide(s) by the resistant DQ molecules. The difficulty of dimerisation by the resistant molecules ensures that even though the antigen has been trapped, the activation of the cognate T-cell clones that might have escaped elimination in the thymus, would be very difficult. The allele DQA1*0501/DQB1*0301 is resistant yet possesses the first pocket and β49-56 dimerisation domain of the susceptible molecules. However, as discussed above, its antigen-binding groove also has characteristics of resistant DQ molecules (presence of β57Asp overall negative electrostatic potential). Also, it lacks the Arg-Gly-Asp loop and probably cannot fully activate any cognate diabetogenic T-cell clones in the periphery. Furthermore, in one recent study [45] it has been classified as neutral with respect to susceptibility to IDDM.

The identification of diabetogenic peptide(s) would allow a study of their interaction with the susceptible and resistant DQ molecules. Such interactions can be studied in molecular dynamics simulations, as have already been performed with certain mouse MHC class I molecules and well-binding peptides [46], or in binding experiments to isolated DQ molecules in a manner similar to that performed for various DR alleles [20], or by isolating biochemically the spectrum of peptides that bind to the antigen-binding groove of DQ molecules (already reported for DQ7 [24]). We are currently pursuing the first approach with the peptides isolated and sequenced from DQ7 [24]. The recent identification of several diabetogenic peptides in the nonobese diabetic mouse [47, 48] from the autoantigen of glutamic acid decarboxylase [49] can also be used as a test case.

It should be noted that in offering an explanation for the involvement of DQ molecules in the pathogenesis or protection from IDDM, based on modelled structural features, this study takes no account of the polymorphisms of DQ molecules in the intracellular aminoacid sequences that participate in signal transduction [50], or the possible differences in the level of expression of various DQ alleles. It has long been known that the first intervening sequence in DQB from diabetic patients is different from that of normal HLA-DR matched control subjects [3, 4, 51]. Also, a wealth of regulatory sequences has been uncovered in the DQB and DQA genes [52, 53], with no inkling as to their possible significance in the pathogenesis of this disease. This study points to several

important structural features of the extracellular domains of DQ molecules that have quite different characters between susceptible and resistant alleles; these differences are suggestive of mechanisms of susceptibility to and protection from IDDM. While the intracellular domain of DQ molecules and regulatory regions of their genes may also contribute to the susceptibility and resistance to IDDM, the scheme outlined here already segregates these two groups of alleles.

Recently the DQ locus has also been linked to several other autoimmune diseases (e. g. multiple sclerosis [13, 54] and coeliac disease [13, 55]). One allele conferring protection from IDDM (DQA1*0102/DQB1*0602) confers susceptibility to multiple sclerosis [54]. While no explanation is readily available for this paradox, we should note that the mode of pathogenesis of each autoimmune disease is different from the others, as exemplified by probable participation of different environmental agents, the different autoantigen(s) involved, and the different target organ(s), and course of pathogenesis in each case.

References

1. Todd JA, Acha-Orbea H, Bell JI et al. (1988) A molecular basis for MHC Class II-associated autoimmunity. Science 240: 1003–1009
2. Platz P, Jakobsen BK, Morling N et al. (1981) HLA-D and -DR antigens in genetic analysis of insulin-dependent diabetes mellitus. Diabetologia 21: 108–115
3. Owerbach D, Lernmark Å, Platz P et al. (1983) HLA-DR beta-chain DNA endonuclease fragments differ between healthy and insulin dependent diabetic individuals. Nature 303: 815–817
4. Cohen-Haguenauer O, Robbins E, Busson M et al. (1985) A systematic study of HLA class II-B DNA restriction fragments in insulin-dependent diabetes mellitus. Proc Nat Acad Sci USA 82: 3335–3339
5. Erlich HA, Bugawan T, Scharf S, Nepom GT, Tait B, Griffith R (1990) HLA-DQB sequence polymorphism and genetic susceptibility to IDDM. Diabetes 39: 96–103
6. Khalil I, Deschamps I, Lepage V, Al-Daccak R, Degos L, Hors J (1992) Dose effect of cis and trans encoded HLA-DQαβ heterodimers in type 1 diabetes susceptibility. Diabetes 41: 378–384
7. Todd JA, Bell JI, McDevitt HO (1987) HLA-DQB gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. Nature 329: 559–604
8. Khalil I, d'Auriol, L, Gobet M et al. (1990) A combination of HLA-DQB Asp57-negative and HLA-DQα Arg52 con-

fers susceptibility to insulin-dependent diabetes mellitus. J Clin Invest 85: 1315–1319
9. Rønningen KS, Iwe T, Halstensen TS, Spurkland A, Thorsby E (1989) The aminoacid position 57 of HLA-DQB chain and susceptibility to develop insulin-dependent diabetes mellitus (type 1 diabetes). Human Immunol 26: 215–225
10. Thorsby E, Rønningen KS (1993) Particular HLA-DQ molecules play a dominant role in determining susceptibility or resistance to type 1 (insulin-dependent) diabetes mellitus. Diabetologia 36: 371–377
11. Davies JL, Kawaguchi Y, Bennett ST et al. (1994) A genome-wide search for human type 1 diabetes susceptibility genes. Nature 371: 130–136
12. Nepom GT (1990) A unified hypothesis for the complex genetics of HLA associations with type 1 diabetes. Diabetes 39: 1153–1157
13. Altman DM, Sansom D, Marsh SGE (1991) What is the basis of HLA-DQ associations with autoimmune disease? Immunol Today 12: 267–270
14. Gillis S (1989) T cell-derived lymphokines. In: Paul WE (ed) Fundamental Immunology, 2nd edn. Raven Press, New York, pp 621–638
15. Fossum B, Gedde-Dahl III T, Hansen T, Eriksen JA, Thorsby E, Gaudernack G (1993) Overlapping epitopes encompassing a point mutation (12 Gly → Arg) in p21 ras can be recognized by HLA-DR, -DP and -DQ restricted T cells. Eur J Immunol 23: 2687–2691
16. Hirayama K, Matsushita S, Kikuchi I, Iuchi M, Ohta N, Sasazuki T (1987) HLA-DQ is epistatic to HLA-DR in controlling the immune response to schistosomal antigen in humans. Nature 327: 426–430
17. Salgame P, Convit J, Bloom BR (1991) Immunological suppression of human CD8+ T cells is receptor dependent and HLA-DQ restricted. Proc Natl Acad Sci USA 88: 2598–2602
18. Stern LJ, Brown JH, Jardetzky TS et al. (1994) Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature 368: 215–221
19. Brown JH, Jardetzky TS, Gorga JC et al. (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 364: 33–39
20. Marshall KW, Liu AF, Canales J et al. (1994) Role of polymorphic residues in HLA-DR molecules in allele-specific binding of peptide antigens. J Immunol 152: 4946–4957
21. Hammer J, Valsasnini P, Tolba K et al. (1993) Promiscuous and allele specific anchors in HLA-DR-binding peptides. Cell 74: 197–203
22. Kropshofer H, Max H, Halder T, Kalbus M, Muller C, Kalbacher H (1993) Self peptides from four HLA-DR alleles share hydrophobic anchor residues near the NH$_2$-terminal including proline as a stop signal for trimming. J Immunol 151: 4732–4742
23. Sette A, Sidney J, Oseroff C et al. (1993) HLA DR4w4-binding motifs illustrate the biochemical basis of degeneracy and specificity in peptide-DR interactions. J Immunol 151: 3163–3170
24. Falk K, Rötzsche O, Stefanovic S, Jung G, Rammensee H-G (1994) Pool sequencing of natural HLA-DR, DQ and DP ligands reveals detailed peptide motifs, constraints of processing and general rules. Immunogenetics 39: 230–242
25. De Vos AM, Ultsch M, Kossiakoff AA (1992) Human growth hormone and the extracellular domain of its receptor: crystal structure of the complex. Science 255: 306–312
26. Paliakasis K, Routsias J, Petratos K, Ouzounis C, Kokkinidis M, Papadopoulos GK (1994) Modelling of the struc-

ture of the human histocompatibility molecules HLA-DQ based on the published structure of HLA-DR1. 13th International Immunology and Diabetes Workshop, Montvillagerne, France, May 25–28, 1994 (Abstract)

27. Srinivasan N, Blundell TL (1993) An evaluation of the performance of an automated procedure for comparative modelling of protein tertiary structure. Prot Engineering 6: 501–510

28. Johnson MS, Srinivasan N, Sowdhamini R, Blundell T (1994) Knowledge-based protein modeling. Crit Rev Biochem Mol Biol 29: 1–68

29. Weber IT (1991) Modeling of structure of Human Immunodeficiency Virus-1 protease with substrate based on crystal structure of Rous Sarcoma Virus protease. Methods Enzymol 202: 727–741

30. Marsh SGE, Bodmer J (1991) HLA Class II nucleotide sequences. Tissue Antigens 37: 181–189

31. Brown JH, Jardetzky T, Saper MA, Samraoui B, Bjorkman PJ, Wiley DC (1988) A hypothetical model of the foreign antigen binding site of Class II histocompatibility molecules. Nature 332: 845–850

32. O'Shea EK, Klemm JD, Kim PS, Alber T (1991) X-ray structure of the GCN4 leucine zipper, a two-stranded parallel coiled coil. Science 254: 539–544

33. Konig R, Huang Y-L, Germain RN (1992) MHC class II interaction with CD4 mediated by a region analogous to the MHC class I binding site for CD8. Nature 356: 796–798

34. Cammarota G, Scheirle A, Tacaks B et al. (1992) Identification of a CD4 binding site on the β2 domain of HLA-DR molecules. Nature 356: 799–801

35. Sanjeevi CB, Lybrand TP, DeWesee C et al. (1995) Polymorphic amino acid variations in HLA-DQ are associated with systematic physical property changes and occurrence of IDDM. Diabetes 44: 125–131

36. Lernmark Å (1994) Molecular biology of IDDM. Diabetologia 37 [Suppl 2]: S73–S81

37. Ruoslahti E, Pierschbacher MD (1987) New perspectives in cell adhesion: RGD and integrins. Science 238: 491–497

38. Huang T-F, Holt JC, Kirby EP, Niewiarowski S (1989) Trigramin: primary structure and its inhibition of von Willebrand factor binding to glycoprotein IIb/IIIa complex in human platelets. Biochemistry 28: 661–666

39. Leahey DJ, Hendrickson WA, Aukhil I, Erickson HP (1992) Structure of fibronectin type II domain from tenascin phased by MAD analysis of the selenomethionyl protein. Science 258: 987–991

40. Hynes RO (1992) Integrins: versality, modulation and signaling in cell adhesion. Cell 69: 11–25

41. Springer TA (1994) Traffic signals for leukocyte recirculation and leukocyte emigration: the multistep paradigm. Cell 76: 301–312

42. Janossy G, Thomas JA, Bollum FJ et al. (1980). The human thymic microenvironment: an immunohistochemical study. J Immunol 125: 202–212

43. Hogquist KA, Jameson SC, Heath WR, Howard JL, Bevan MJ, Carbone FR (1994) T cell receptor antagonist peptides induce positive selection. Cell 76: 17–27

44. Swat W, Ignatowicz L, von Boehmer H, Kisielow P (1991) Clonal deletion of immature CD4$^+$8$^+$ thymocytes in suspension culture by extrathymic antigen-presenting cells. Nature 351: 150–153

45. Todd JA (1994) The emperor's new genes: 1993 RD Lawrence lecture. Diabet Med 11: 6–20

46. Corr M, Boyd LF, Padlan EA, Margulies DH (1993) H-2D^d exploits a four residue peptide binding motif. J Exp Med 178: 1877–1892

47. Kaufman DL, Clare-Salzer M, Tian J et al. (1993) Spontaneous loss of T cell tolerance to glutamic acid decarboxylase in murine insulin dependent diabetes. Nature 366: 69–72

48. Tisch R, Yang X-D, Singer SM, Liblau RS, Fugger L, McDevitt HO (1993) Immune response to glutamic acid decarboxylase correlates with insulitis in non-obese diabetic mice. Nature 366: 72–75

49. Baekkeskov S, Aanstoot H-J, Christgau S et al. (1990) Identification of the 64 K autoantigen in insulin-dependent diabetes as the GABA-synthesizing enzyme glutamic acid decarboxylase. Nature 345: 151–156

50. Wade WF, Davoust J, Salamero J, Pascale A, Watts TH, Cambier JC (1993) Structural compartmentalization of MHC class II signaling function. Immunol Today 14: 539–546

51. Michelsen B, Lernmark Å (1987) Molecular cloning of a polymorphic DNA endonuclease fragment associates insulin dependent diabetes mellitus with HLA-DQ. J Clin Invest 79: 1144–1152

52. Andersen LC, Beaty JS, Nettles JW, Seyfried CE, Nepom GT, Nepom BS (1991) Allelic polymorphism in transcriptional regulatory regions of HLA-DQB genes. J Exp Med 215: 181–192

53. Morzycka-Wroblewska E, Harwood JI, Smith JR, Kagnoff MS (1993) Structure and evolution of the promoter regions of the DQA genes. Immunogenetics 37: 364–372

54. Ollerup J, Hillert J, Frederickson J et al. (1989) Primarily chronic-progressive and relapsing-remitting multiple sclerosis: two immunogenetically distinct disease entities. Proc Natl Acad Sci USA 86: 7113–7117

55. Sollid LM, Marcussen G, Ek G et al. (1989) Evidence for a primary association of celiac disease to a particular HLA-DQ α/β heterodimer. J Exp Med 169: 345–350