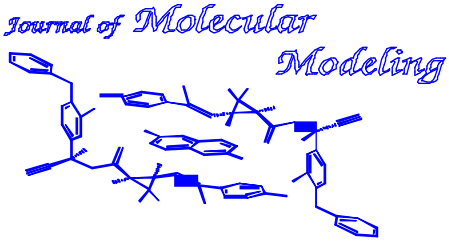
J. Mol. Model. **1996**, *2*, 205 - 216



Characterization of a Novel Reverse-orientation Model for a Peptide/MHC Complex Putatively Associated with Type I Diabetes Mellitus

Carol DeWeese, William W. Kwok[§], Gerald T. Nepom[§], and Terry P. Lybrand*

Center for Bioengineering, University of Washington, Box 351750, Seattle, WA 98195-1750, USA, Tel: 206-685-1515, Fax: 206-616-4387 (lybrand@proteus.bioeng.washington.edu)

§ Virginia Mason Research Center, 1000 Seneca Street, Seattle, WA 98101, USA

Received: 14 June 1996 / Accepted: 18 July 1996 / Published: 2 September 1996

Abstract

Molecular modeling techniques were used to generate structures of several HLA-DQ proteins associated with insulin-dependent diabetes mellitus (IDDM). A peptide fragment from glutamic acid decarboxylase (GAD), a known IDDM autoantigen, binds to certain HLA-DQ molecules positively associated with IDDM. Modeling studies were used to explore possible binding interactions between this GAD peptide and several HLA-DQ molecules. Based on the characterization of anchor pockets in the HLA-DQ binding groove and of peptide side chains, a novel binding mode was proposed. This binding mode predicts the GAD peptide is positioned in the binding groove in the direction opposite the orientation observed for class I proteins and the class II DR1, DR3, and I-E^k proteins. Peptide docking exercises were performed to construct models of the HLA-DQ/peptide complexes, and the resulting models have been used to design peptide binding experiments to test this "reverse-orientation" binding mode. A variety of experimental results are consistent with the proposed model and suggest that some peptide ligands of class II molecules may bind in a reversed orientation within the binding groove.

Keywords: Major Histocompatibility Complex proteins, insulin-dependent diabetes mellitus, peptide docking, molecular modeling.

Introduction

The primary genetic factor associated with insulin-dependent diabetes mellitus (IDDM) susceptibility in humans is the DQ gene region of the human leukocyte antigen (HLA) complex. [1-5] Although genetic susceptibility is widely accepted as the primary factor required for IDDM onset, individuals who are genetically programmed to be susceptible do not develop the disorder until exposed to necessary environmental triggers. [6,7] Studies of human IDDM suggest that viruses, chemicals and toxins, and dietary factors are all potential triggers. [8] However, none have been identified definitively.

The DQ genes, part of the major histocompatibility complex (MHC), encode heterodimeric transmembrane proteins that bind antigenic peptide ligands for presentation to T cell receptors (TCRs) on $CD4^+$ T lymphocytes. A class II MHC protein such as HLA-DQ consists of an α chain and a β chain, encoded by separate genes. The peptide binding region is a

* *To whom correspondence should be addressed*

cleft formed by the membrane-distal portions of both the α and β chains.

Peptide binding in the groove of a class II MHC molecule occurs via noncovalent interactions involving both side chain and backbone atoms of the peptide. The side chains of "anchor residues" in the peptide interact with corresponding "anchor pockets" in the binding groove. The peptide lies flat in the groove and the complex is stabilized by a hydrogen bond network involving primarily MHC side chains and the peptide backbone. Both ends of a class II MHC binding groove are open, allowing the termini of the bound peptide to extend beyond the ends of the groove. [9-14]

Unlike antibodies and T cell receptors, the repertoire of MHC molecules in an individual is limited. Necessity dictates that in order to initiate an immune response to a large number of antigens, each MHC protein must be able to bind multiple peptide ligands. This characteristic has been demonstrated in studies of peptide binding motifs. [15-17] Peptide binding to MHC proteins appears to depend partially on the absence of unfavorable side chains in anchor residue positions, rather than strictly on the presence of specific, favorable anchor residues. The identified peptide binding motifs indicate that anchor residues may be categorized as favorable, impartial, or unfavorable. Peptides containing either favorable or impartial anchor residue side chains bind MHC proteins [15-17], presumably with varying affinities, and the complex is stabilized by a network of hydrogen bonds. Peptide binding is prevented when unfavorable residues are present at anchor positions.

The HLA gene loci are among the most polymorphic within the population. [18-20] Only specific amino acid positions are polymorphic sites (e.g., position 86 of the DQ β chain can be Ala, Gly, or Glu), and high homology exists among the various alleles. In the binding groove region, most DQ α and β chains share at least 90-95% sequence identity. [20] Therefore, the differences observed in binding characteristics among various HLA-DQ haplotypes are due to the effects of amino acid substitutions at the polymorphic sites within the binding groove.

The haplotype HLA-DQ3.2 (allelic designation DQA1*0301-DQB1*0302), is positively associated with IDDM susceptibility in Caucasian populations, while DQ3.1 (DQA1*0301-DQB1*0301) is negatively associated with susceptibility. [3-5,21,22] A third haplotype, DQ3.3 (DQA1*0301-DQB1*0303), is associated with susceptibility in Japanese populations [23], and in the Swedish population, when in heterozygous combination with DQ3.2 or selected other DQ alleles. [24] These three MHC molecules all share a common α chain and have highly homologous β chains (~97-99% sequence identity). In the peptide binding region, DQ3.1 and DQ3.2 differ by four residues at positions 13, 26, 45, and 57. DQ3.3 is identical to DQ3.2 except for a single substitution at position 57. The polymorphic residues and their locations in the antigen binding groove are shown in Figure 1A.

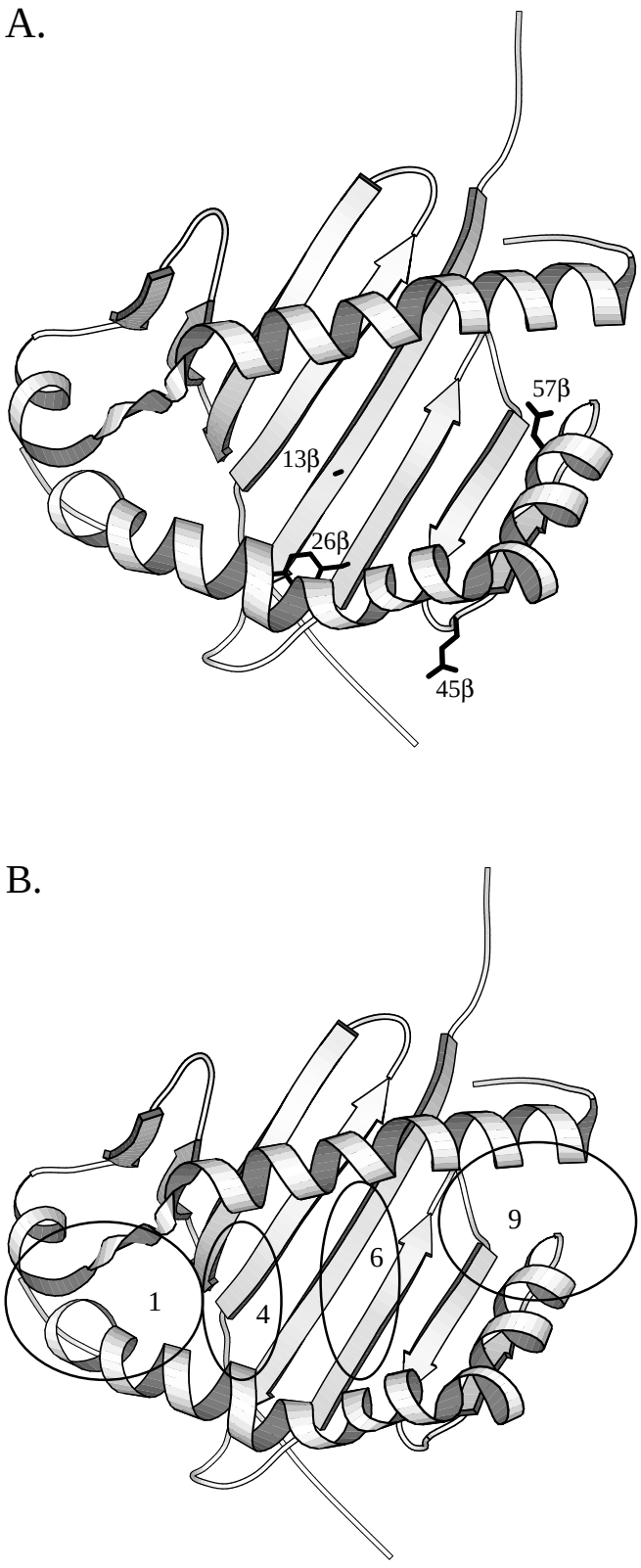


Figure 1. *A) Ribbon structure of DQ3.1 binding groove with the four polymorphic side chains shown. B) Anchor pockets 1, 4, 6, and 9 of the DR1 structure are indicated. Figures 1, 3, and 4 were generated with the program Molscript.[47]*

In spite of the high sequence homology of these three MHC molecules, they exhibit some dramatic differences in peptide binding. In particular, a 13-residue fragment from the 65-kD isomer of glutamic acid decarboxylase (GAD65) binds well to DQ3.2, but binds poorly to DQ3.3 and does not bind to DQ3.1 at all. GAD65 is a known IDDM autoantigen [25], and the 13-residue peptide fragment (designated 34p) that binds well to DQ3.2 exhibits high sequence homology to the most immunogenic peptide derived from the P2-C protein of coxsackievirus B4 [26], a putative environmental trigger of IDDM onset.

We used molecular modeling techniques to construct three-dimensional models for the DQ3.1, DQ3.2, and DQ3.3 MHC molecules, and to generate docked complexes for peptide 34p with each MHC molecule. Based on the peptide docking exercises, we propose a hypothesis to explain the differential binding of peptide 34p to these three MHC molecules, and we predict a "reverse-orientation" binding motif for DQ3.2 with peptide 34p.

Methods

Model construction

A model structure of DQ3.2 was generated using homology modeling techniques based on the structure of DR1 (DRA-DRB1*0101), a similar class II MHC protein encoded by different gene loci. The DR1 structure, which contains a bound peptide from the influenza protein hemagglutinin (HA), was determined by x-ray crystallography. [9] DR1 and DQ3.2 share ~62% sequence identity overall, with ~52% sequence identity in the peptide binding region, and many of the substitutions at polymorphic sites are conservative changes. [20] Chothia and Lesk demonstrated that proteins with greater than 50% sequence identity generally have very similar tertiary structures. [27] Thus, the DR1 structure should be a good template for DQ3.2 model construction. We do expect that there are some structural differences between DR and DQ class II MHC molecules, particularly in two regions. A region in the DQα chain between residues 48-56 (residues 45-53 in DR1 α chain) contains a number of nonconservative substitutions, including a cluster of arginine residues in the DQ molecules. Another interesting region of probable structural variation between DR and DQ proteins is located in the β chain at position 55, where many DQ proteins have a proline substituted in place of the arginine observed in DR1. This substitution yields two adjacent prolines in these DQ molecules, which we predict will likely disrupt the beginning of the helical region in the β1 domain. However, neither of these regions impact the peptide binding groove profoundly in our DQ models. More significantly, the three DQ heterodimers we have modeled are highly homologous, and thus, likely have nearly indentical three-dimensional structures. We have focused our modeling efforts on the differences between these three DQ molecules.

The DQ3.2 model was generated using standard homology modeling techniques. Backbone atoms of the DR1 template were fixed at crystallographic positions and necessary amino acid side chain substitutions were made to generate the DQ3.2 sequence. Since side-chain conformations in proteins with high sequence identity are highly conserved [28,29], DR1 side chain conformations were retained for all homologous sites and conservative substitutions (e.g., Thr for Ser), to the extent possible. For nonconservative substitutions, side-chain atoms were placed initially in the most probable conformation. [30] The models were constructed using MidasPlus [31] and PSSHOW [32] interactive molecular graphics programs.

In addition to the amino acid substitutions, DQ α chains contain three insertions relative to the DR α chain. These insertions are located at the amino terminus of the α chain (positions 1, 2, and 9), far removed from the binding groove region. The amino terminus has an extended chain conformation, and insertions at positions 1 and 2 were made by attaching single residues in an extended conformation to the amino terminus of the protein. The insertion at position 9 was made by breaking the protein backbone at the site of insertion, attaching the appropriate residue to the amino terminus of residue 10, and rotating the fragment comprising residues 1-8 to form a trans peptide bond connecting residues 8 and 9. Side-chain atoms were placed in the most probable conformations.

Once all amino acid substitutions and insertions were made, substituted side chains were adjusted manually to relieve steric clashes. Limited energy minimization was then performed using AMBER 4.0[33] with an all-atom potential function [34] to refine the models. Only the membrane-distal binding groove portion of the molecule was relaxed by minimization; the membrane-proximal region was fixed in the crystallographic conformation during minimization, as this region in DQ3.2 is highly homologous to DR1 (~70%), with few nonconservative substitutions. Conjugate gradient energy minimization was performed in vacuo with a distance-dependent dielectric constant. The model structure was evaluated by calculating side chain packing densities using the program QPACK [35] and by verifying reasonable side-chain conformations. [30] Visual inspection was performed to insure that all polar and charged residues not exposed to solvent had suitable interaction partners to permit formation of hydrogen bonds and salt bridges. The final DQ3.2 model was used as a template for construction of DQ3.1 and DQ3.3 models, using the protocol outlined above.

The DQ3.2 structure generated in this homology modeling exercise has several acidic and basic residues in regions predicted to form key anchor pockets in the peptide binding groove. To assess the probable charge state of these residues, pKa values were calculated for all ionizable residues using Poisson-Boltzmann electrostatics calculations and a pKa calculation protocol developed by Antosiewicz. [36,37] Briefly, this method entails 1) calculation of the self-ionization energy of each titratable group when free in aqueous

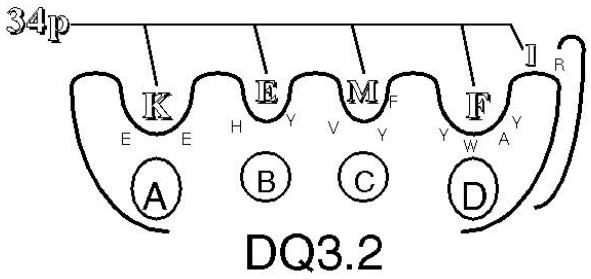


Figure 2. *Diagram of DQ3.2 anchor pockets. The four anchor residues from 34p are shown in the appropriate pockets predicted by the reverse-orientation model.*

solution, 2) calculation of the ionization energy of each titratable group in the neutral protein (i.e., all other titratable groups are held neutral, but partial charges for all atoms in the protein are included in Poisson-Boltzmann calculations), 3) calculation of the interaction energy between all ionizable groups, and 4) a Monte Carlo simulation to determine the lowest energy state(s) from among the 2^M possible ionization states in the protein, where M = number of titratable groups. The electrostatic potentials and electrostatic interaction energies were computed using a finite difference algorithm to solve the linearized Poisson-Boltzmann equation with the UHBD program. [38] A coarse grid lattice (2.5 Å spacing) was used to calculate long-range electrostatic contributions, followed by a focusing technique with successively higher resolution grids (1.20 Å, 0.75 Å, 0.25 Å lattice point spacing) to obtain converged results for short-range electrostatic interactions. All calculations were performed at T = 293 K, pH 7.0, and 150 mM ionic strength with a solvent dielectric of 80.0, a protein dielectric of 20.0, and a 2.0 Å Stern layer. It has been observed in previous studies that a protein dielectric of 20.0 yields good agreement with experimentally measured pK_a values. [36,39] Partial charges and van der Waals radii needed for the calculations were taken from the latest AMBER potential functions. [40]

Peptide docking

Peptide docking involved the manual placement of putative peptide anchor residues in appropriate anchor pockets of the binding groove. To identify possible peptide anchor residues, each anchor pocket was assessed for size, hydrophobicity, and presence of charged and/or polar residues. Anchor residues with complementary properties and appropriate spacing in the peptide were then chosen.

A model of peptide 34p was constructed in extended-chain conformation, with all side chains placed in the most probable conformations. The peptide was docked manually into the DQ3.2 binding groove with the selected anchor residues positioned in the corresponding anchor pockets. Backbone torsion angles of the peptide were adjusted to accommodate the fit of the peptide in the binding groove, and anchor residue side chains were then rotated to fit well in their respective anchor pockets. Finally, limited energy minimization was performed using the all-atom potential functions to relieve any residual unfavorable contacts. Initially, only the peptide was permitted to relax, while subsequent minimization included the peptide and the binding groove region of the MHC molecule. The membrane-proximal portion of the DQ3.2 molecule was fixed throughout the minimization process.

Results

Construction of a DQ3.2/34p reverse-orientation model.

The nomenclature often used to distinguish each anchor pocket within the binding groove of a class II protein is a 1-4-6-9 scheme, based on the crystal structure of DR1 and its bound ligand, the HA peptide. [9] The amino acid sequence of the HA peptide is

Pro-Lys-**Tyr**-Val-Lys-**Gln**-Asn-**Thr**-Leu-Lys-**Leu**-Ala-Thr

1 2 3 **4** 5 **6** 7 8 **9**

and the anchor residues are shown in bold type. Each anchor pocket is numbered according to the corresponding peptide anchor residue that binds in the pocket, with the first anchor residue designated as position 1. The locations of these pockets within the groove are shown in Figure 1B. For purposes of clarification in this discussion, pockets 1, 4, 6, and 9 will be designated A, B, C, and D, respectively (see Figure 2).

In the DQ3.2 model, pockets A and D are more pronounced than pockets B and C. Pocket A, the largest in the DQ3.2 model, contains primarily polar and charged residues, including two exposed glutamic acids (34α and 86β). The pocket is stabilized by a network of hydrogen bonds and salt bridges formed by the side chains of the residues lining the pocket. Arg55α forms a salt bridge with Glu34α, and Glu86β forms hydrogen bonds with Ser10α and His27α. This is in contrast to the hydrophobic character of the comparable pocket in HLA-DR1. [9] Our electrostatics calculations suggest that both glutamate residues are significantly ionized, even at pH 4.5. We predicted that a positively charged residue from a bound peptide would be a preferred anchor residue for this pocket, forming a charge interaction with one of the glutamates. Figure 3A shows the charged and polar side chains that stabilize pocket A.

Pockets B and C are shallow and less distinct than pockets A and D. Pocket B of DQ3.2 contains primarily hydrophobic residues, plus an exposed histidine side chain and an exposed tyrosine hydroxyl group. We predicted a polar side chain able to form a hydrogen bond with the His or Tyr would be a likely anchor residue for this pocket. Pocket C is pre-

dominantly hydrophobic, and we predicted a hydrophobic anchor residue would be needed for this pocket.

Pocket D contains numerous hydrophobic and aromatic residues in DQ3.2. In the DR1 structure, this pocket contains primarily small hydrophobic side chains, plus one salt bridge formed by Arg76α and Asp57β. DQ3.2 contains an homologous Arg at position 79α, but has an alanine at position 57β. This alanine substitution in DQ3.2 disrupts the salt bridge observed in the DR1 crystal structure. Arg79α is located at the end of the binding groove, and its side chain can be easily oriented either into pocket D or out of the binding groove. Since pocket D contains no other polar or charged side chains, we chose to position the side chain in an alternate high probability conformation [30] with the guanidino group pointing away from the pocket, where it has greater solvent accessibility. We predicted a hydrophobic anchor residue for this pocket, with preference perhaps for an aromatic residue in order to form π-stacking interactions in the pocket. Figure 3B shows the aromatic, hydrophobic side chains that stack together to form pocket D of DQ3.2.

After anchor pocket characterization was completed, the peptide was inspected for the presence of complementary anchor residues. The amino acid sequence of peptide 34p is

Ile-Ala-Arg-Phe-Lys-Met-Phe-Pro-Glu-Val-Lys-Glu-Lys

1 2 3 4 5 6 7 8 9 10 11 12 13

The anchor residues of a peptide are usually found in a nine amino acid span approximately in the center of the peptide sequence. [41] In order to achieve an extended backbone conformation, the spacing between anchor residues for

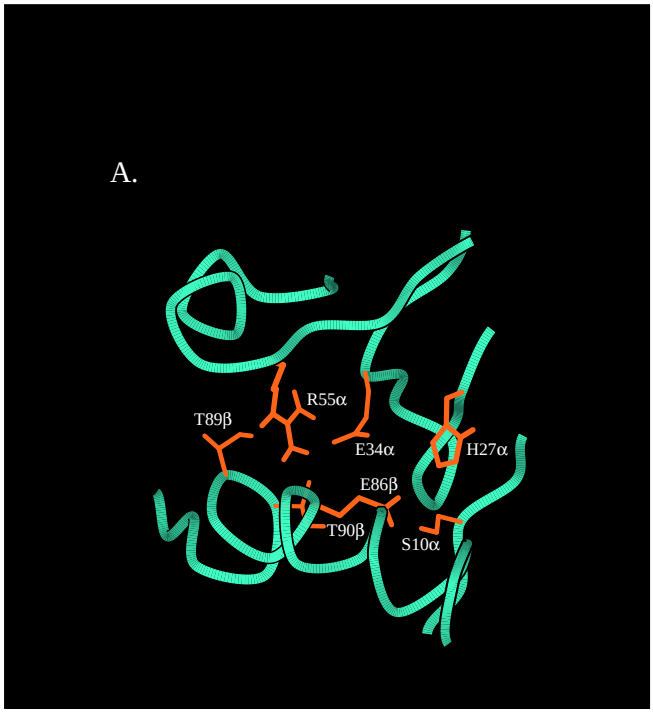
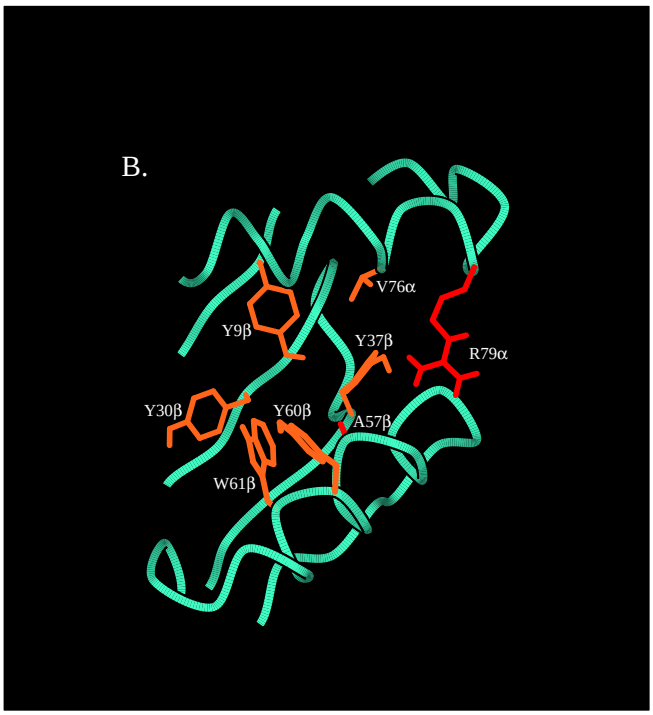


Figure 3. *Top view of A) pocket A in DQ3.2, B) pocket D in DQ3.2, and C) pocket D in DQ3.3 models. The protein backbone is shown in teal. Side chains shown (in orange) in pocket A are: Ser 10α, His 27α, Glu 34α, Arg 55α, Glu 86β, Thr 89β, and Thr 90β. In pocket D, Arg 79α and Ala/Asp 57β are shown in red, and all other side chains (Val 76α, Tyr 9β, Tyr 30β, Tyr 37β, Tyr 60β, and Trp 61β) are shown in orange.*



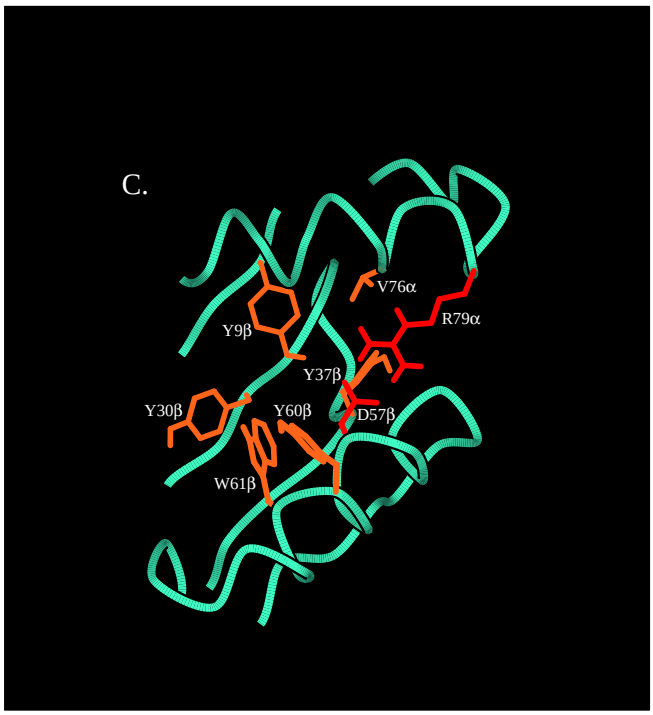


Table 1. *Properties of anchor pockets and the selected anchor residues from peptide 34p in a "reverse-orientation" binding mode.*

Pocket	Pocket Characteristics	Anchor Residue Properties Needed	Anchor Residue from peptide 34p
A	Polar and charged residues, with exposed Glu residues	Positively charged side chain	Lys (res. 11)
B	Primarily hydrophobic, plus exposed His and Tyr residues	Negatively charged or polar side chain	Glu (res. 9)
C	Hydrophobic	Hydrophobic side chain	Met (res. 6)
D	Hydrophobic; numerous aromatic side chains	Hydrophobic, preferably aromatic side chain	Phe (res. 4)

pockets A and D was restricted to seven or eight amino acids (i, i+7 or i, i+8). Anchor residues A and D, which received highest priority, were selected first, followed by anchor residues that complement pockets B and C. A suitable motif in the center of the peptide sequence containing the desired anchor residue properties and appropriate spacing was easily identified. The properties of each pocket, the corresponding properties sought in the anchor residue, and the selected side chains from the peptide are summarized in Table 1.

The anchor residues chosen for pockets A and D (positions 11 and 4, respectively) fit the characteristic profiles for these pockets ideally. The Lys side chain at position 11 is perfectly positioned in pocket A to form hydrogen bonds and charge interactions with the glutamate residues, and the Phe side chain at position 4 packs well with the aromatic residues in pocket D. This motif is comparable to the peptide binding motif identified in pool sequencing experiments for HLA-DQ2 [42], where Lys and Phe are the principle anchor residues in pockets A and D, respectively. Recent studies have also identified three peptides from a dust mite allergen protein that produce an immune response via DQ3.2-restricted TCR activation in transgenic mice. [43] All three peptides identified in this study fit our proposed binding motif well, with a lysine residue and various hydrophobic residues in either an i, i+7 or i, i+8 pattern. All three peptides can be docked easily in our DQ3.2 model, with lysine in pocket A and a hydrophobic anchor in pocket D. One of the peptides also has a glutamate at position i+2 which fits nicely in pocket B, exactly as seen in our DQ3.2-34p complex. Finally, the peptide binding motif for DQ3.2 in pool sequencing studies indicates that a Lys or Arg residue is the preferred primary anchor. [44]

With lysine and phenylalanine residues anchored in pockets A and D, the Glu side chain at position 9 was easily docked in pocket B to form a hydrogen-bonding partner for both the His and Tyr residues. The Met side chain at position 6 is well accommodated by the hydrophobic anchor pocket C. All anchor residues were positioned in the pockets while maintaining reasonable backbone torsion angles for the peptide.

In the DR1 crystal structure, the HA peptide backbone is in a polyproline II peptide conformation [9] as is the CLIP peptide in the recent DR3 complex structure [13] and two peptides of single-chain constructs with the mouse class II MHC molecule I-Ek. [14] The peptide backbone conformation in our DQ3.2/34p model complex is similar to these crystal structure complexes, although with a somewhat less pronounced twist that is not a classic polyproline II structure. Figure 4 shows the anchor residues from peptide 34p and key side chains in anchor pockets A, B, C, and D.

Peptide binding to class II proteins depends on interactions between peptide anchor residues and MHC anchor pockets, and also on the formation of a large number of hydrogen bonds between the MHC protein and backbone atoms of the peptide. In the DR1/HA crystal structure, 15 hydrogen bonds between DR1 side chains and the HA peptide backbone are reported [9], and the pattern is similar for the DR3/CLIP and I-E^k complexes. [13,14] In our model, DQ3.2 side chains were oriented to form 12 hydrogen bonds with peptide backbone atoms, most of which are analogous to those observed in the DR1 and DR3 structures. Six nonpolymorphic residues (N62α, N69α, R76α, W61β, N82β, H81β) that participate in nine hydrogen bonds in DR1/HA and DR3/CLIP are conserved in DQ3.2. Seven of these hydrogen bonds are maintained in our model, one additional hydrogen bond is formed with R79α, and one is lost due to the presence of a proline in peptide 34p. In addition to hydrogen bonds involving the peptide backbone, our model exhibits hydrogen bonding and charge interactions between peptide anchor residues and anchor pocket side chains, as discussed above. Figure 5 shows the hydrogen bond interactions between DQ3.2 and the peptide backbone.

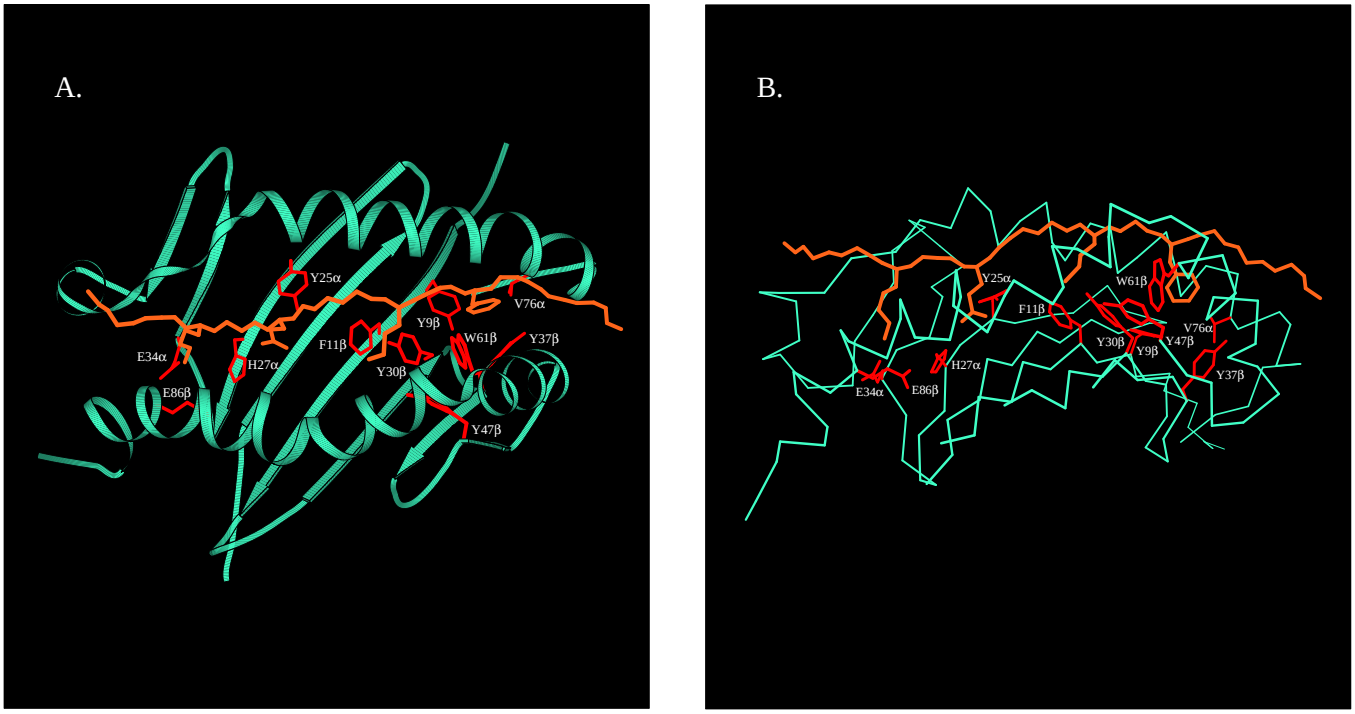


Figure 4. *Top-view and side-view images of the DQ3.2/34p complex showing anchor residues of 34p and key side chains from DQ3.2. The DQ3.2 backbone is shown in teal and DQ3.2 side chains are shown in red. The DQ3.2 side chains shown are: Glu 34α and Glu 86β in pocket A; Tyr 25α and His 27α in pocket B; Phe 11β, Tyr 30β, and Tyr 47β in pocket C; Val 76α, Tyr 9β, Tyr 30β, Tyr 37β, and Trp 61β in pocket D. The peptide is orange, with anchor residues Phe (position 4), Met (position 6), Glu (position 9), and Lys (position 11) shown.*

Construction and comparison of DQ3.1, DQ3.2 and DQ3.3 complexes with 34p

Position 57β, the single amino acid polymorphism that distinguishes DQ3.2 from DQ3.3, is located in anchor pocket D of the binding groove. Thus, any observed differences in peptide binding between DQ3.2 and DQ3.3 are due to this substitution. The primary difference predicted to arise from this Ala->Asp substitution is the probable formation of a salt bridge between Asp 57β and Arg 79α in DQ3.3. Formation of this salt bridge requires that Arg 79α adopt a side chain conformation similar to that observed for Arg 76α in the DR1 crystal structure. In our DQ3.2 model, an alternate conformation was chosen for Arg79α that orients it away from the hydrophobic environment of pocket D, and dramatically increases its solvent accessibility, as described above. This alternate conformation for Arg 79α is easily accommodated in our model structures with no backbone adjustment, and allows formation of an additional hydrogen bond with the peptide backbone, as discussed above. The main impact of this Asp-Arg salt bridge is a significant change in pocket size. Pocket D is significantly smaller in DQ3.3 and DQ3.1 than in DQ3.2, because the Asp-Arg contact pair fills a portion of pocket D. Figures 3B and 3C show a detailed view of pocket D in the DQ3.2 and DQ3.3 models.

DQ3.1 contains four polymorphic substitutions relative to DQ3.2. As in DQ3.3, there is an Ala→Asp substitution at position 57β. There are additional polymorphisms at positions 13β (Gly→Ala), 26β (Leu→Tyr), and 45β (Gly→Glu). Position 13β is situated along the edge of pocket B between pockets B and C, and position 26β is located along the edge of pocket C, near pocket B. Both of these substitutions in DQ3.1 reduce the size and depth of the (already) shallow anchor pockets B and C. Position 45β is located on the edge of the β-sheet that forms the floor of the binding groove, and is not expected to directly impact peptide interactions within the groove (see Figure 1).

Thus, our models suggest a clear structural basis for the observed binding properties of 34p to DQ3.1, DQ3.2, and DQ3.3. Anchor pocket D is much smaller in DQ3.1 and DQ3.3, and our docking exercises suggest that peptides with a phenylalanine anchor at this position, such as 34p, should have greatly reduced binding affinities, as is observed experimentally. [17] The two additional polymorphic substitutions at positions 13β and 26β reduce the size of anchor pockets B and C in DQ3.1 relative to DQ3.2, and these changes introduce greater steric restrictions in the DQ3.1 binding groove. Again, these polymorphic substitutions would be expected to diminish binding for peptides with large anchor residues in these positions, such as the glutamate and methionine in 34p.

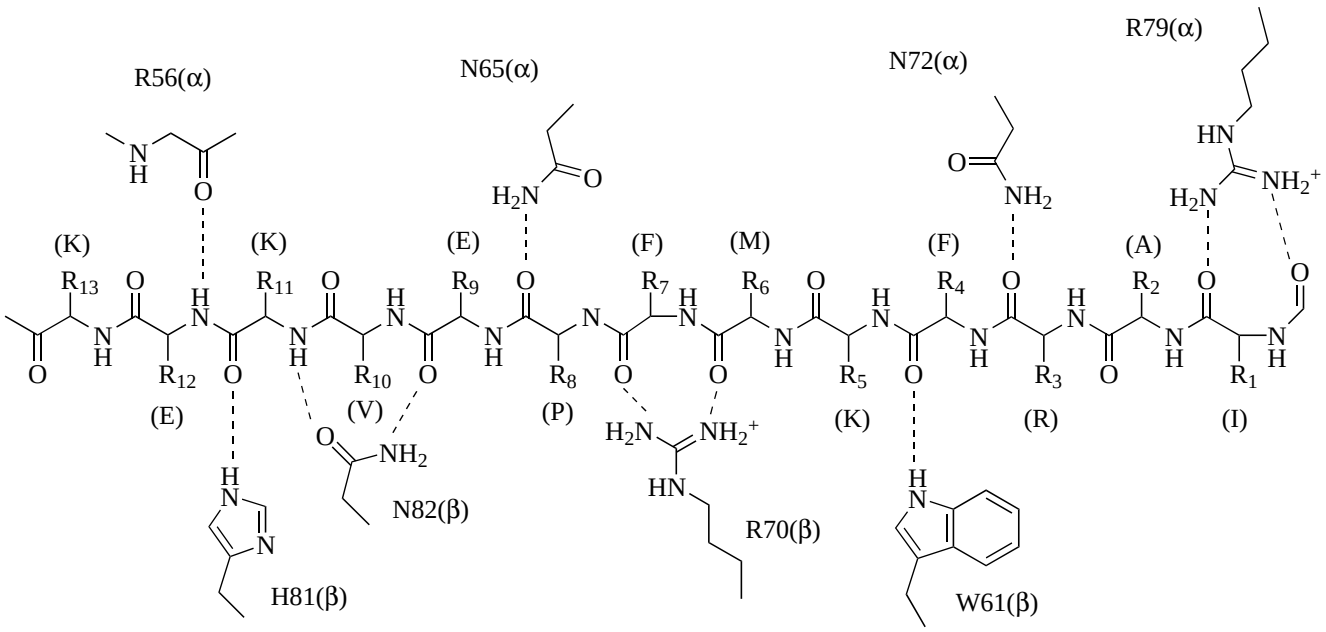


Figure 5. *Hydrogen-bonding pattern in the DQ3.2/34p complex. Side chains from DQ3.2 that form hydrogen bonds with the peptide backbone are shown and the hydrogen bonds are indicated by dashed lines. Peptide side chains are designated as R1-R13 and the residue at each position is given in parentheses.*

Discussion

The selected anchor residues from 34p provide an excellent model for the bound DQ3.2/34p complex, as shown in detail in Figure 4. This model exhibits good interactions between anchor residues and anchor pockets, appropriate spacing between anchor residues, a slightly twisted peptide backbone in extended conformation, and numerous hydrogen bonds between DQ3.2 and the peptide. This model of DQ3.2 with bound 34p suggests a novel binding mode in which the peptide orientation in the binding groove is opposite that typically observed. However, most peptide orientation data comes from crystal structures of class I complexes, in which both termini of the peptide are contained within the binding groove and contribute to specific interactions that stabilize the complex. Because the peptide termini extend beyond each end of a class II MHC binding groove, peptide binding is not restricted to the standard orientation by interactions between the peptide termini and the binding groove. Limited data on the characteristics of class II peptide anchor residues have revealed relative symmetry in the positioning and properties of anchor residues, and binding in either orientation has been suggested. [15] Src homology 3 domain (SH3) molecules also bind peptides that adopt a polyproline II conformation, and peptide binding has been observed in both orientations in SH3 protein-ligand complexes. [45, 46] It appears that peptide side chain interactions with the SH3 binding groove determine the binding orientation in these complexes. [45] Our DQ3.2/34p model is the first proposed structure of an MHC complex with a reverse-orientation motif.

A variety of experimental data are available which lend support to our reverse-orientation binding model for peptide 34p to DQ3.2. For example, the choice of a lysine anchor residue for pocket A and phenylalanine for pocket D seems well justified, based on peptide motifs for DQ2 and DQ3.2 molecules identified in sequencing studies. [42] Since DQ3.2 and DQ2 molecules are highly homologous (91-94% sequence identity), it is not surprising that our predicted DQ3.2 peptide binding motif is quite similar to the motif determined experimentally for DQ2 molecules. The presence of a highly similar possible binding motif in the dust mite peptide antigens is also intriguing. It is quite interesting to note the sequence binding motif identified for DQ2 antigens suggest a traditional orientation in our models, while the possible motifs observed in the dust mite peptide antigens suggest a traditional binding orientation for some peptides, and a reverse-orientation binding mode for others.

Experimental studies have been performed to map the anchor residue positions in peptide 34p. [17] To assess the impact on binding to DQ3.2, single Arg substitutions were introduced at each position in 34p that is not normally a lysine or arginine residue. Our reverse-orientation model predicts that peptides containing Arg substitutions at position 4 (Phe→Arg), position 6 (Met→Arg), or position 9 (Glu→Arg) would not bind DQ3.2, since Arg would be an unfavorable anchor in each of these pockets. The experimental results indeed show that peptide binding by DQ3.2 is blocked only when an arginine substitution is introduced at positions 4, 6, or 9, as predicted by our model, and at one additional site, position 1.[17] Analysis of our model reveals the formation

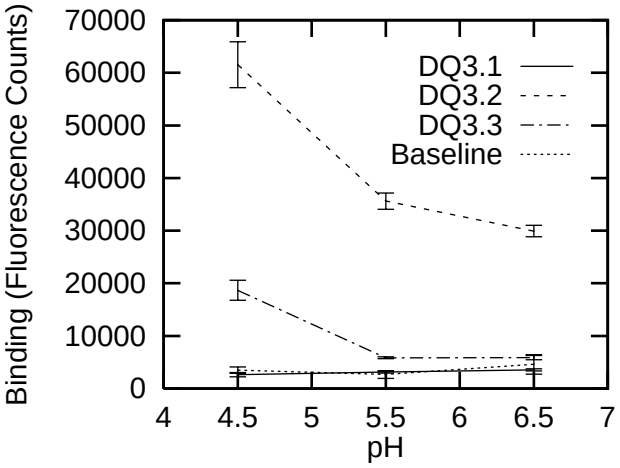


Figure 6. *Binding data for peptide 34p with DQ3.1, DQ3.2, and DQ3.3 as a function of pH. Baseline values are averages obtained for peptide 34p to the BLS1 cell line, which lacks class II MHC molecules.*

of two hydrogen bonds between Arg 79α in DQ3.2 and the backbone of the peptide at position 1. An Ile→Arg substitution at position 1 would place two Arg side chains in very close proximity, resulting in potential serious ionic and steric repulsion (see Figure 2). Replacement of Ile at position 1 with Thr, Gln, Phe, Asp, or Ala did not block peptide binding, however. [17] Since position 1 is tolerant of essentially any amino acid substitution except Arg, we propose that this position is not likely to function as an anchor residue.

A more systematic study has also been performed for positions 4, 6, and 9 of peptide 34p, using a number of amino acid substitutions. [17] The results indicate that binding is maintained only when each of these residues is replaced with a conservative substitution (e.g., Phe→Leu at position 4), or alanine, which appears to be an impartial, or permissive, anchor. Based on the data reported, position 9 seems to be least tolerant of substitution. Substitution of arginine or phenylalanine at this position abolishes binding, while threonine, leucine, and alanine mutations reduce binding substantially. These results are also consistent with the reverse-orientation model, as they imply the importance of the interaction between His27α and Glu-9 in pocket B.

We have used our model complexes to design additional experiments to further test this proposed binding mode for 34p. One of the most interesting implications of our reverse orientation model is the placement of Glu-9 from 34p in anchor pocket B, which contains a histidine residue in all three DQ molecules we have studied. At pH 4.5, both the His and Glu side chains are predicted to be predominantly charged. This suggests that peptide binding at low pH might involve formation of an ion-pair-reinforced hydrogen bond between these side chains. Our calculations also suggest that peptide binding may diminish somewhat as the pH increases to 6.5, since the histidine will begin to deprotonate noticeably at the higher end of this range. Similar behavior should also be observed for DQ3.3, since 34p exhibits measurable, though much reduced, binding affinity for DQ3.3.

To test these model predictions, peptide 34p binding to DQ3.1, DQ3.2, and DQ3.3 was evaluated at pH 4.5, 5.5, and 6.5 (Nepom, Kwok, DeWeese, and Lybrand, unpublished results). Binding to DQ3.2 was observed to be pH-dependent, with optimal binding observed at pH 4.5, and a smooth decrease in binding as the pH is raised to 6.5, where the histidine is expected to be only ~50% protonated. These results are shown in Figure 6, and agree well with our predictions for the reverse-orientation model. As expected, 34p shows no appreciable binding to DQ3.1. The peptide does exhibit limited binding to DQ3.3, and the pH profile follows the trend observed for DQ3.2, as predicted by the models. These data are also given in Figure 6.

We also explored several alternate binding modes for 34p, to determine if any other plausible models could be generated. One alternative assumed a traditional orientation, with 34p residues 1, 4, 6, and 9 chosen as anchors (i.e., the residues identified in the Arg substitution experiments by Kwok, at al. [17]). This binding motif, however, requires a different set of interactions in the primary anchor pockets. Specifically, this binding motif positions Ile-1 in pocket A, Phe-4 in pocket B, Met-6 in pocket C, and Glu-9 in pocket D. As described above, pocket A contains charged and polar side chains; a hydrophobic side chain such as Ile would be at best impartial in this pocket. Pocket D, the other key anchor pocket of DQ3.2, contains hydrophobic, aromatic side chains, and a charged Glu side chain does not appear to be a particularly good anchor residue for this pocket. It has been proposed that Glu-9 of the peptide could form a salt bridge with Arg79α in anchor pocket D. However, this proposal is difficult to reconcile with the observed pH-dependent binding profile. If this Glu-Arg interaction were formed, we predict it would be strongest at pH 6.5, where the glutamate is fully ionized. As the pH is decreased to 4.5, the glutamate would begin to protonate, weakening the salt bridge and reducing binding affinity at least slightly. The experimentally observed pH binding profile is distinctive and shows the opposite trend (i.e., peptide binding becomes significantly stronger as the pH is lowered), consistent with our reverse-orientation model that suggests a Glu-9/His27α interaction in pocket B as discussed above. Finally, this proposed binding motif is drastically different from the motif identified by Verreck et al. for DQ2 molecules. [42] As discussed above, we expect that DQ3.2 may have a similar binding motif, due to its very high sequence similarity to DQ2 molecules. As discussed above, we expect that DQ3.2 may have a similar binding motif, due to its very high sequence similarity to DQ2 molecules, and the motif data from pool sequencing for DQ3.2. [44]

We considered other combinations of anchor residues that would generate a traditional binding orientation. Based on anchor pocket characterization, most potential combinations of anchor residues for the traditional orientation are predicted to be unfavorable. The only combination of anchor residues

that fits a traditional orientation binding motif with impartial or favorable characteristics is Arg-3 for the polar and negatively charged pocket A and Val-10 for the hydrophobic pocket D. We constructed a model of the complex in this orientation. This combination of anchor residues contains i, i+7 spacing, as does our reverse-orientation model, and a similar hydrogen bonding pattern for peptide backbone with DQ side chains is observed. However, this model is inconsistent with the arginine substitution data obtained by Kwok and coworkers. [17] They found that replacement of Val-10 with Arg does not disrupt binding, as would be expected if Val were an anchor residue. Furthermore, this model requires that Phe-4, Met-6, and Glu-9 side chains project directly out of the binding groove. However, Kwok at al. observed that arginine substitution at any of these positions eliminates peptide binding to DQ3.2 completely. It is difficult to envision how arginine substitution for a residue that projects out of the binding groove could totally block peptide binding, or how arginine substitutions could be completely tolerated at all the other positions predicted by this model to project into the binding groove.

For direct comparison to the known human class II MHC-peptide crystal structures, we attempted to dock peptide 34p using both the HA and CLIP backbone alignments as templates. First, the DQ3.2 backbone atoms were aligned with the DR1 backbone atoms, and the HA coordinates were copied into the file containing the DQ3.2 model. Amino acid side chains were replaced to generate the 34p sequence. This procedure was repeated for the DR3/CLIP structure.

Since the HA and CLIP peptides have i, i+8 spacing between anchor residues, the 34p sequence was assessed for potential anchor combinations. However, 34p contains no side chains that fit the characteristic profiles needed to match anchor pockets A and D with i, i+8 spacing. The combination of Ala-2 and Val-10 appear to be the only possible combination of residues with i, i+8 spacing that would not be disruptive in pocket A or D. However, while Val is an acceptable choice for pocket D, Ala seems much less desirable for the large, polar pocket A. This motif is completely inconsistent with the arginine substitution data [17], since an Arg is not disruptive to binding at either position. In addition, this motif orients Phe-4, Met-6, and Glu-9 out of the binding groove, in contrast to the implications of the arginine substitution data, as discussed above. This motif also places the side chain of Lys-4 into pocket B, which contains His27α, another interaction predicted to be unfavorable.

One result observed by Kwok et al. is more difficult to interpret simply. Binding is enhanced when the Lys anchor at position 11 is replaced with Phe (34p11F). [17] Since our model predicts that a hydrophobic side chain such as Phe would be much less favorable than lysine in pocket A, the reverse-orientation model cannot account for the binding of peptide 34p11F. An analysis of the peptide sequence reveals two potential alternative combinations of anchor residues to explain the binding observed for the 34p11F peptide: one that predicts a traditional orientation for binding and one that predicts an alternate reverse-orientation binding mode.

The traditional orientation model has Arg-3 and Phe-11 anchors with i, i+8 spacing. Using 1-4-6-9 spacing for anchor residues, Met-6 and Pro-8 would be the anchor residues for pockets B and C, respectively. In this binding model, favorable interactions would occur in pockets A, C and D, with an impartial or slightly unfavorable (due to the steric bulk of Met) interaction in pocket B. A model of this complex was constructed using the HA peptide backbone as a template. Eleven hydrogen bonds can be formed between the peptide backbone and DQ3.2 in this model, but no interactions are identified that might explain the pH profile observed for 34p. However, there are no experimental data to demonstrate that this modified 34p11F peptide exhibits the same pH-dependent binding profile as 34p.

The alternate reverse-orientation binding model for peptide 34p11F would involve a register shift of two residues from the reverse-orientation model for wild-type 34p. This would place Lys-13 and Met-6 anchor residues in pockets A and D, respectively. Pro-8 and Phe-11 would be anchors for pockets B and C, respectively. This would also produce favorable interactions in pockets A, C, and D, with a neutral interaction in pocket B. A model of this complex was constructed for comparison using the wild-type 34p reverse-orientation peptide backbone as a template. Ten peptide backbone hydrogen bonds are formed with DQ3.2 side chains in this complex. As for the traditional orientation model of this complex, there is no suggestion that a pH-dependent binding profile would be observed for this peptide complex either.

The proposed shift in binding register, or flip to a traditional binding mode with different anchor residues, appear to be the two most rational explanations for the extremely strong binding observed for peptide 34p11F. Because each of these models entail fundamentally different binding modes, a new series of residue substitution and pH binding profile experiments will be needed to evaluate them properly.

Conclusions

Our reverse orientation binding model provides a reasonable explanation for the interactions of a potentially diabetogenic peptide with the DQ3.2 MHC molecule, and the rather dramatic impact selected polymorphic substitutions have on peptide binding to several highly homologous DQ molecules. It includes both of the key components for peptide binding to class II MHC proteins: (1) complementary interactions between peptide anchor residues and protein anchor pockets, and (2) an extensive hydrogen bonding network between protein side chains and the peptide backbone. In addition, our model provides a reasonable explanation for the pH-dependent binding profile observed experimentally.

For peptide 34p, no peptide binding motifs are available that incorporate the typical orientation, fit the profile needed for anchor residues and observed for homologous DQ2 mol-

ecules, and are consistent with results from anchor residue mapping experiments. Thus, it seems that peptide 34p most likely binds to DQ3.2 in a reverse-orientation mode. Our modeling results also suggest that DQ molecules may bind peptides in either direction, depending on the precise nature of anchor residue interactions with the binding groove, as is observed experimentally for SH3 protein complexes. Our modeling studies also suggest that individual amino acid substitutions at certain positions in the peptide may in some cases alter the binding mode dramatically.

This is the first proposed structure of an MHC protein binding a peptide ligand in the reverse orientation. Detailed biophysical studies will be needed to confirm the exact nature and orientation of peptide binding in DQ3.2. At present, we are performing photoaffinity labeling studies to obtain definitive information for 34p orientation in the DQ3.2 binding groove.

Acknowledgments This work was supported in part by grants from the National Institutes of Health (DK41801) and the Whitaker Foundation. We also wish to thank Mary Ellen Domeier and Eric Swanson for expert technical assistance.

Supplementary Material

Full coordinates for the following models are available in Brookhaven PDB format:

DQ3.1 (0205dg31.pdb)
DQ3.2 (0205dg32.pdb)
DQ3.3 (0205dg33.pdb)

References

1. Kim, S. J.; Holbeck, S. L.; Nisperos, B.; Hansen, J. A.; Maeda, H.; Nepom, G. T. *Proc Natl Acad Sci U S A* **1985**, *82*, 8139.
2. Michelsen, B.; Kastern, W.; Lernmark, Å.; Owerbach, D. *Biomed Biochim Acta* **1985**, *44*, 33.
3. Nepom, B. S.; Palmer, J.; Kim, S. J.; Hansen, J. A.; Holbeck, S. L.; Nepom, G. T. *J Exp Med* **1986**, *164*, 345.
4. Nepom, G. T.; Seyfried, C. A.; Nepom, B. S. *Pathol Immunopathol Res* **1986**, *5*, 37.
5. Nepom, G. T. *Diabetes Reviews* **1993**, *1*, 93.
6. Martin, J. M.; Trink, B.; Daneman, D.; Dosch, H.-M.; Robinson, B. *Annals of Medicine* **1991**, *23*, 447.
7. Lernmark, •.; BŠrmeier, H.; Dube, S.; Hagopian, W.; Karlsen, A.; Wassmuth, R. *Endocrinology and Metabolism Clinics of North America* **1991**, *20*, 589.
8. Yoon, J. W. *Curr. Top. Microbiol. Immunol.* **1990**, *164*, 95.
9. Stern, L. J.; Brown, J. H.; Jardetzky, T. S.; Gorga, J. C.; Urban, R. G.; Strominger, J. L.; Wiley, D. C. *Nature* **1994**, *368*, 215.
10. Sinigaglia, F.; Hammer, J. *Apmis* **1994**, *102*, 241.
11. Rotzschke, O.; Falk, K. *Curr Opin Immunol* **1994**, *6*, 45.
12. Madden, D. R. *Annu Rev Immunol* **1995**, *13*, 587-622.
13. Ghosh, P.; Amaya, M.; Mellins, E.; Wiley, D. C. *Nature* **1995**, *378*, 457.
14. Fremont, D. H.; Hendrickson, W. A.; Marrack, P.; Kappler, J. *Science* **1996**, *272*, 1001.
15. Falk, K.; Rotzschke, O.; Stevanovi'c, S.; Jung, G.; Rammensee, H. G. *Immunogenetics* **1994**, *39*, 230.
16. Johansen, B. H.; Buus, S.; Vartdal, F.; Viken, F.; Eriksen, J. A.; Thorsby, E.; Sollid, L. M. *Int. Immunol.* **1994**, *6*, 453.
17. Kwok, W. W.; Domeier, M. E.; Raymond, F. C.; Byers, P.; Nepom, G. T. *J Immunol* **1996**, *156*, 2171.
18. Goodman, J. W. In *Basic and Clinical Immunology*; 8 ed.; Stites-D-P, Terr-A-I, Parslow-T-G, Eds.; Appleton & Lange: Norwalk, CT, 1994.
19. Duquesnoy, R. J. *Clin Lab Med* **1991**, *11*, 509.
20. Marsh, S. G.; Bodmer, J. G. *Immunogenetics* **1993**, *37*, 79.
21. Kockum, I.; Wassmuth, R.; Holmberg, E.; Michelsen, B.; Lernmark, A. *Am J Hum Genet* **1993**, *53*, 150.
22. Sanjeevi, C. B.; Lybrand, T. P.; DeWeese, C.; Landin-Olsson, M.; Kockum, I.; Dahlquist, G.; Sundkvist, G.; Stenger, D.; Lernmark, A. *Diabetes* **1995**, *44*, 125.
23. Kobayashi, T.; Tamemoto, K.; Nakanishi, K.; Kato, N.; Okubo, M.; Kajio, H.; Sugimoto, T.; Murase, T.; Kosaka, K. *Diabetes Care* **1993**, *16*, 780.
24. Sanjeevi, C. B.; Falorni, A.; Kockum, I.; Hagopian, W. A.; Lernmark, Å. *Diabetic Medicine* **1996**, *13*, 209.
25. Hagopian, W. A.; Michelsen, B.; Karlsen, A. E.; Larsen, F.; Moody, A.; Grubin, C. E.; Rowe, R.; Petersen, J.; McEvoy, R.; Lernmark, A. *Diabetes* **1993**, *42*, 631.
26. Atkinson, M. A.; Bowman, M. A.; Campbell, L.; Darrow, B. L.; Kaufman, D. L.; Maclaren, N. K. *J Clin Invest* **1994**, *94*, 2125.
27. Chothia, C.; Lesk, A. M. *Embo J* **1986**, *5*, 823-6.
28. Ring, C. S.; Cohen, F. E. *Faseb J* **1993**, *7*, 783-90.
29. Sutcliffe, M. J.; Hayes, F. R.; Blundell, T. L. *Protein Eng* **1987**, *1*, 385.
30. Ponder, J. W.; Richards, F. M. *J Mol Biol* **1987**, *193*, 775.
31. Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. *J Mol Graph* **1988**, *6*, 13.
32. Swanson, E. PSSHOW, Seattle, 1995.
33. Pearlman, D. A.; Case, D. A.; Caldwell, J. C.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. ; University of California: San Francisco, 1991.
34. Weiner, S. J.; Kollman, P. A.; Nguyen, D.; Case, D. A. *J Comput Chem* **1986**, *7*, 230.
35. Gregoret, L. M.; Cohen, F. E. *J Mol Biol* **1990**, *211*, 959.

35. Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *J Mol Biol* **1994**, *238*, 415.
37. Antosiewicz, J.; Porschke, D. *Biophys J* **1995**, *68*, 655.
38. Davis, M. E.; Madura, J. D.; Sines, J.; Luty, B. A.; Allison, S. A.; McCammon, J. A. *Methods Enzymol* **1991**, *202*, 473.
39. Annand, R. R.; Kontoyianni, M.; Penzotti, J. E.; Dudler, T.; Lybrand, T. P.; Gelb, M. H. *Biochemistry* **1996**, *35*, 4591.
40. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. C.; Kollman, P. A. *J Am Chem Soc* **1995**, *117*, 5179.
41. Rammensee, H. G. *Curr Opin Immunol* **1995**, *7*, 85.
42. Verreck, F. A. W.; van de Poel, A.; Termijtelen, A.; Amons, R.; Drijfhout, J.-W.; Koning, F. *Eur. J. Immunol.* **1994**, *24*, 375.
43. Neeno, T.; Krco, C. J.; Harders, J.; Baisch, J.; Cheng, S.; David, C. S. *J. Immunol.* **1996**, *156*, 3191.
44. Chicz, R.M.; Lane, W.S.; Robinson, R.A.; Trucco, M.; Strominger, J.L.; Gorga, J.C. *Int. Immunol.* **1994**, *6*, 1639.
45. Feng, S.; Chen, J. K.; Yu, H.; Simon, J. A.; Schreiber, S. L. *Science* **1994**, *266*, 1241.
46. Wilson, I. A. *Science* **1996**, *272*, 973.
47. Kraulis, P. J. *Journal of Applied Crystallography* **1991**, *24*, 946.