JMB

Sequence Variability Analysis of Human Class I and Class II MHC Molecules: Functional and Structural Correlates of Amino Acid Polymorphisms

Pedro A. Reche[1,2] and Ellis L. Reinherz[1,2]*

[1]*Laboratory of Immunobiology Dana-Farber Cancer Institute Harvard Medical School, 44 Binney Street, Boston, MA 02115, USA*

[2]*Department of Medicine Harvard Medical School, 44 Binney Street, Boston, MA 02115, USA*

Major histocompatibility complex class I (MHCI) and class II (MHCII) molecules display peptides on antigen-presenting cell surfaces for subsequent T-cell recognition. Within the human population, allelic variation among the classical MHCI and II gene products is the basis for differential peptide binding, thymic repertoire bias and allograft rejection. While available 3D structural analysis suggests that polymorphisms are found primarily within the peptide-binding site, a broader informatic approach pinpointing functional polymorphisms relevant for immune recognition is currently lacking. To this end, we have now analyzed known human class I (774) and class II (485) alleles at each amino acid position using a variability metric (V). Polymorphisms ($V > 1$) have been identified in residues that contact the peptide and/or T-cell receptor (TCR). Using sequence logos to investigate TCR contact sites on HLA molecules, we have identified conserved MHCI residues distinct from those of conserved MHCII residues. In addition, specific class II (HLA-DP, -DQ, -DR) and class I (HLA-A, -B, -C) contacts for TCR binding are revealed. We discuss these findings in the context of TCR restriction and alloreactivity.

© 2003 Elsevier Ltd. All rights reserved

Keywords: HLA polymorphism; T-cell receptor; sequence variability; alloreactivity; major histocompatibility complex

*Corresponding author

Introduction

Classical major histocompatibility complex (MHC) molecules are cell-surface glycoproteins that are central to the process of adaptive immunity, functioning to capture and display peptides on the surface of antigen-presenting cells (APCs).[1] These plasma membrane-bound peptide–MHC complexes (pMHC) are scrutinized by T-lymphocytes *via* their T-cell receptors (TCRs) during immunosurveillance.[2] Since T-cells recognizing self-peptides are eliminated during the process of thymic selection, those pMHC incorporating foreign peptides are the primary focus of T-cell-mediated immune responses.[3]

There are two major classes of MHC molecules,

termed class I and class II.[4] MHC class I (MHCI) molecules are expressed on most cells, bind endogenously derived peptides with sizes ranging from eight to ten amino acid residues and are recognized by CD8 cytotoxic T-lymphocytes (CTL). On the other hand, MHC class II (MHCII) are present only on specialized APCs, bind exogenously derived peptides with sizes varying from 9 to 22 residues, and are recognized by CD4 helper T-cells. These differences indicate that MHCI and MHCII molecules engage two distinct arms of the T-cell-mediated immune response, the former targeting invasive pathogens such as viruses for destruction by CD8 CTLs, and the latter inducing cytokine-based inflammatory mediators to stimulate CD4 helper T-cell activities including B-cell activation, maturation and antibody production.[5]

Sequence identity between classical MHCI and MHCII is low (<20%), yet their 3D structures are strikingly similar. MHCI molecules are heterodimers composed of a single-membrane-spanning α chain, paired with the soluble β2 microglobulin (β2m) protein. The α chain has been divided into three distinct segments termed α1, α2 and α3. The α3 region has an immunoglobulin (Ig)-like fold,

---

Abbreviations used: APC, antigen-presenting cell; CTL, cytotoxic T-cells; HLA, human leukocyte antigen; MHC, major histocompatibility complex; MHCI, and MHCII, MHC class I and class II; pMHC, plasma membrane-bound peptide–MHC complex; TCR, T-cell receptor; *V*, variability metric.

E-mail address of the corresponding author: ellis_reinherz@dfci.harvard.edu

whereas the membrane distal α1 and α2 segments form a peptide-binding cleft consisting of two α-helices overlying a floor comprised of eight anti-parallel β-stranded sheets (α1α2 domain). MHCII molecules are heterodimers also but comprised of two subunits, each with a single membrane-spanning anchor. The extracellular segment of the α chain includes α1 and α2 domains, and likewise, the β chain is composed of β1 and β2 domains. In the mature MHCII molecule, the α2 and β2 domains fold as independent Ig-like domains, whereas α1 and β1 fold together, creating a single antigen-presenting platform with architecture very similar to that of the α1α2 domain in MHCI molecules.[6,7]

In the human, MHC molecules are referred to as HLA, an acronym for human leukocyte antigens, and are encoded by the chromosome 6p21.3-located HLA region.[8,9] The HLA segment is divided into three regions (from centromere to telomere), class II, class III and class I. Classical class I and class II HLA genes are contained in the class I and class II regions, respectively, whereas the class III locus bears genes encoding proteins involved in the immune system but not structurally related to MHC molecules. The classical HLA class I molecules are of three types, HLA-A, HLA-B and HLA-C. Only the α chains of these mature HLA class I molecules are encoded within the class I HLA locus by the respective *HLA-A*, *HLA-B* and *HLA-C* genes. In contrast, the β2m chain encoded by the *β2m* gene is located on chromosome 15. The classical HLA class II molecules are also of three types (HLA-DP, HLA-DQ and HLA-DR), with both the α and β chains of each encoded by a pair of adjacent loci. In addition to these classical HLA class I and HLA class II genes, the human MHC locus includes a long array of HLA pseudogenes as well as genes encoding non-classical MHCI and MHCII molecules.[9] HLA-pseudogenes are an indication that gene duplication is the main driving force for HLA evolution,[10] whereas non-classical MHCI and MHCII molecules often subserve a restricted function within the immune system quite distinct from that of antigen presentation to αβ TCRs.[11,12] Furthermore, in contrast to the classical HLA molecules (with the exception of the somewhat polymorphic *MICA* and *MICB* genes), non-classical HLA molecules are composed of a very limited number of allelic variants.

Genes encoding classical HLA molecules are extremely polymorphic, such that most classical HLA genes include a large number of allelic variants. Thus, the HLA IMGT/HLA database†[13] currently includes 1524 HLA allelic sequences (904 HLA I alleles and 620 HLA II alleles) (release 1.16, 14/10/2002). The basis for this extreme level of allelic diversity has been linked to evolving optimization of immune protection against pathogens.

Consistent with this notion, most polymorphisms are associated with the peptide-binding residues of the α1α2 and α1β1 domains of HLA class I and HLA class II molecules, respectively.[14] Moreover, the rate of non-synonymous substitutions (those that result in amino acid changes) is greater than the rate of synonymous substitutions (those that do not change amino acid sequence) in codons specifying amino acids involved in peptide binding.[15,16]

Table 1. HLA protein sequences included in this variability analysis

HLA	Molecule	Sequence
Class I		
	HLA-A	230
	HLA-B	447
	HLA-C	97
Class II		
	HLA-DPA1	12
	HLA-DPB1	90
	HLA-DQA1	17
	HLA-DQB1	42
	HLA-DRA	2
	HLA-DRB1	271
	HLA-DRB3	30
	HLA-DRB4	7
	HLA-DRB5	14

In the case of HLA-DR, the α chains (HLA-DRA) derive from the *HLA-DRA* gene loci, whereas the β chain counterpart (HLA-DRB) derives from four different *HLA-DRB* genes, namely *HLA-DRB1*, *HLA-DRB3*, *HLA-DRB4* and *HLA-DRB5*. Sequences included in this analysis differ in at least one amino acid.

encoding for proteins that differ at the amino acid level. The number of amino acid differences between HLA alleles varies among HLA genes. Thus, the two different *HLA-DRA* gene products differ at only one amino acid position. In contrast, *HLA-A* and *HLA-B* gene products differ at ∼19 residue positions (i.e. ∼5% of the protein sequence) (Table 3). Moreover, as shown in Table 3, most of the differences localize to the antigen-presenting α1α2 domain of HLA class I molecules, and the equivalent α1β1 domain of HLA class II molecules. For example, among HLA-A alleles,

there are 5.42% overall differences throughout the entire molecule but 9.05% within the α1α2 domain.

The functional relevance of discrete protein variable segments was first elucidated by E. A. Kabat, who predicted that the highly variable sites in immunoglobulin molecules were contacts for antigen,[21] antedating crystallographic resolution of antibody–antigen complexes. By analogy, the extreme number of HLA allelic variants would suggest that HLA polymorphic sites *per se* are functionally significant, even if structural detail is lacking for all HLA gene products.

Detection of HLA polymorphisms

HLA polymorphisms (variable sites) can be detected from multiple sequence alignments of allelic variants using the traditional Wu-Kabat variability coefficient[21] where the variability index is defined as the number of different symbols (amino acids) found at a site divided by the relative frequency of the most common amino acid at the site. In the current study, we have investigated HLA sequence variability and polymorphisms using a second variability metric (V) (equation (1) in Materials and Methods) that has proven to be a robust tool for the analysis of sequence variability in related systems.[22] This variability metric is formally identical with that of the Shannon entropy equation.[20] However, it is important to stress that we have not performed a *bona fide* Shannon entropy analysis of HLA diversity in the human population, which would require consideration of HLA sampling issues, taking into consideration the frequency of all the alleles in the human population. Instead, we have adopted this equation as a variability metric to study sequence variability

Table 2. PDB files with the 3D structures of MHC molecules from which HLA contacts with peptide, TCR, CD4 and CD8 were obtained

	PDB							
	Class I				Class II			
Contact	HLA-A	HLA-B	HLA-C	K$^{\text{b a}}$	HLA-DP	HLA-DQ	HLA-DR	IA$^{\text{k a}}$
Peptide	1HHJ;1AKJ;1HHK; 1AO7;1BD2;1B0G; 1HHG;1HHI;1B0R; 2CLR;1HHH;1I1Y; 1I1F;1DUZ;1DUY; 1IM3;1QSF;1QSE; 1QRN;1JF1;1JHT; 1I7T;1I7R;1I7U; 1I4F;1TMC	1AGB;1AG;C1AGD; 1AGE;1AGF;1HSA; 1A1N;1A9E;1A9B; 1E27;1E28;1A1M; 1A1O	1IM9;1EFX		Model	1JK8	1AQD;1DLH;1SEB; 1FYT;1HXY;1KG0; 1BX2;2SEB;1J8H; 1A6A	
TCR	1AO7;1QRN;1QSE1QSF; 1BD2	NA	NA	1G6R;1JTR 2CKB; 1FO0;1KJ2	NA	NA	1FYT;1J8H	1D9K
CD8	1AKJ	NA	NA	1BQH	Do not bind CD8			
CD4	Do not bind CD4				NA	NA	NA	1JL4

NA, not available.
[a] These are mouse MHC molecules.

Table 3. Amino acid changes between different HLA alleles: full length and peptide binding region

MHC	Full length			α1α2 (α1β1) peptide-binding region		
	Difference	% Difference	Residues	Difference	% Difference	Residues
HLA-A	19.79 ± 9.56	5.42	365	16.20 ± 6.69	9.05	179
HLA-B	18.15 ± 8.01	5.00	363	15.89 ± 5.66	8.88	179
HLA-C	13.33 ± 6.74	3.64	366	10.72 ± 3.71	5.99	179
HLA-DPA	2.90 ± 2.33	1.11	260	1.64 ± 1.01	2.22	74
HLA-DPB	9.01 ± 5.56	3.4	258	7.63 ± 3.54	8.7	87
HLA-DQA	16.72 ± 9.87	6.56	255	12.67 ± 7.49	16.90	75
HLA-DQB	13.75 ± 7.53	5.27	261	12.53 ± 6.04	14.40	87
HLA-DRA	1.0 ± 0.0	0.4	254	0	0	74
HLA-DRB	11.04 ± 4.68	4.15	266	10.40 ± 4.22	11.96	87

Values correspond to mean ± standard deviation. Amino acid (aa) changes between HLA alleles depends on the individual identity of alleles considered, and therefore mean differences with standard deviation were calculated from all-against-all pairwise comparisons.

and polymorphisms among the available HLA sequences.

Given a multiple sequence alignment, V per site ranges from 0 (totally conserved site) to 4.3 (all 20 amino acids are present, and at the same frequency). In an absolute sense, any site with $V > 0$ is polymorphic. However, several V ranges can be designated to categorize polymorphic sites of HLA molecules based on the frequency of the different amino acids present at a given site. Considering only two different amino acids per site, a value of $V = 0.5$ will be obtained with a ratio between the two amino acids of 1.1:8.9 (given 100 sequences with 11 containing one amino acid and 89 containing a second amino acid at the same position). Moreover, a value of $V = 1$ can be reached only when both amino acids are represented equally (1:1 ratio). On the other hand, in a site where there are only three different amino acid types, a ratio of 1.0:1.1:7.9 (given 100 sequences: 10, 11, and 79 with amino acids of each type), yields a value of $V > 1$. From these considerations, we regard sites with $V < 0.5$ as low significance polymorphisms (90% or more of the amino acids are of the same type), the sites with V between 0.5 and 1 (0.5–1) as dimorphic, and the sites with $V > 1$ as polymorphic. Note, that these ranges of V defined here represent a simplification that will, however, be useful to study HLA polymorphisms and their correlation with key structural and functional features. However, these V ranges are meaningful only within the HLA system.

Sequence variability analysis of HLA molecules

In Figure 1, we have analyzed the variability of HLA class I molecules considering all sequences together (774 sequences), or separated into the HLA-A, HLA-B or HLA-C subtypes (230, 447 and 97 sequences, respectively; see also Table 1). Likewise, we have analyzed the variability of both of the α and β chains of HLA class II molecules (Figure 2), when pooled or as products of separate genes. Variability plots of the combined HLA class I molecules (Figure 1(A)) are very similar to those

of the individual HLA-A, HLA-B, and HLA-C alleles, with the only significant differences showing in the transmembrane and (TM) and cytoplasmic (CYT) regions. However, some of this variability might be a consequence of the CLUSTALW analysis, since the CYT region of HLA class I sequences can vary in length, fostering sequence misalignment. Variability plots from combined HLA class II α chain sequences, and combined HLA class II β chain sequences are quite different from those obtained when allelic sequences are separated by genes (Figure 2). This difference indicates that the three HLA class II genes (HLA-DP, HLA-DQ, and HLA-DR) are more dissimilar to each other than HLA class I genes (HLA-A, HLA-B, and HLA-C). Consistent with this notion, classical HLA class II genes seem to have diverged earlier than classical HLA class I.[10] Thus, we have studied the variability of the individual HLA molecules in a gene-specific manner to elucidate functional features related to sequence variability, thereby eliminating variation due to sequence divergence between the distinct genes. Moreover, this approach prevents the detection of sequence variability resulting from CLUSTALW artifacts. The one exception involves the HLA-DR β chains (HLA-DRB) encoded by four different genes (*HLA-DRB1*, *HLA-DRB3*, *HLA-DRB4*, *HLA-DRB5*). We analyzed the products of all these genes together, since these molecules are highly similar and form heterodimers with the same HLA-DR α chains (HLA-DRA).

In this study we have distinguished between the leader, α1, α2, α3, TM and CYT segments of HLA class I molecules. Likewise, we have distinguished the α1, α2, TM, and CYT of the α chains (HLA-DPA, HLA-DQA and HLA-DRA), and the corresponding segments (β1, β2, TM, and CYT) of the β chains (HLA-DPB, HLA-DQB and HLA-DRB) for HLA class II molecules. The number of variable sites within different V ranges and their segmented location is summarized in Table 4. The number of positions with more than one amino acid ($V > 0$) is truly remarkable for class I. Thus, for HLA-A, HLA-B, and HLA-C, 168 of 365, 176 of 363, and
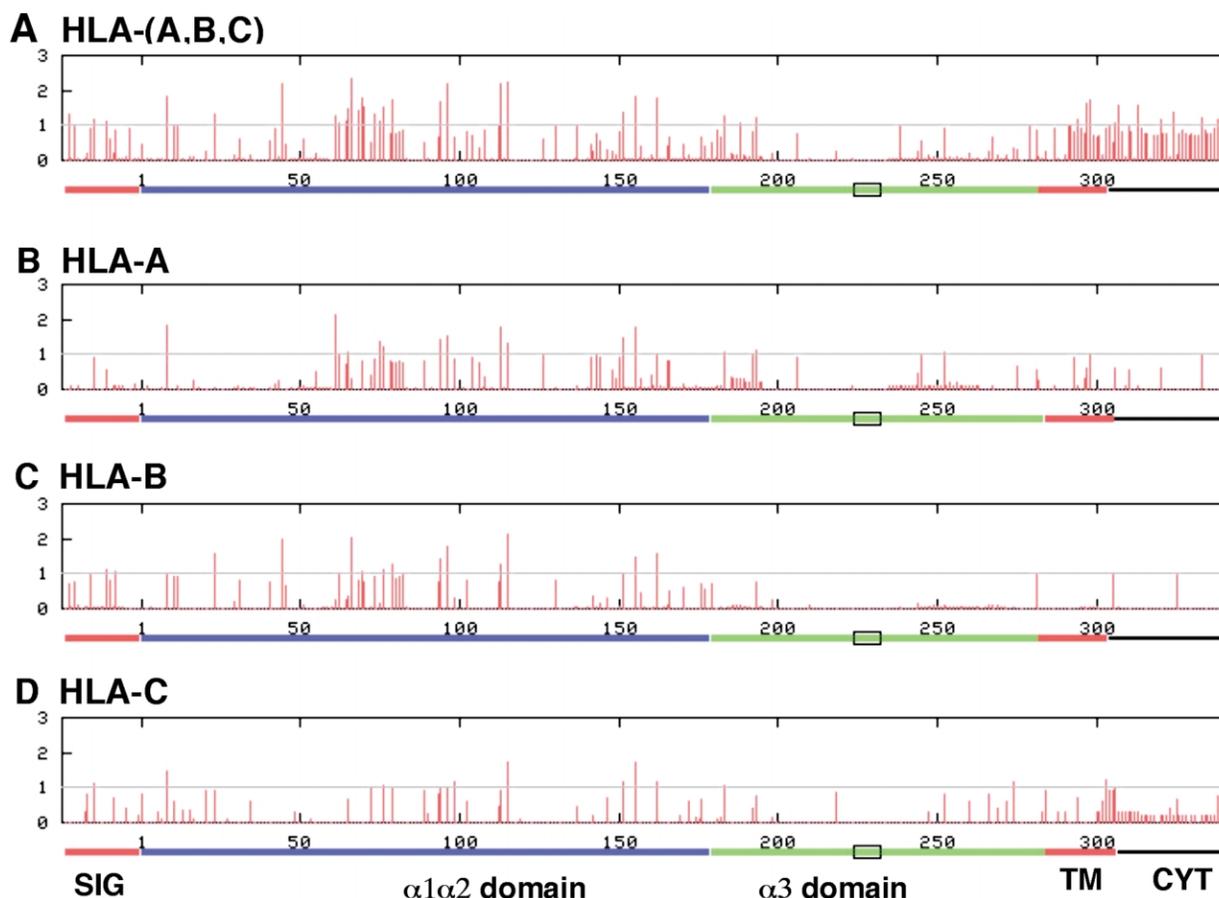
Figure 1. The Figure shows sequence variability corresponding to all HLA class I molecules (A), and to each of the HLA-A, HLA-B, and HLA-C types (B, C and D, respectively). Sequence variability was calculated from the relevant multiple sequence alignments using the variability metric (V), and plotted for each site/position in the alignment. V can range from 0 (totally conserved site) to 4.3 (most variable, with all 20 amino acids represented equally at the site), and the value of $V = 1$ is highlighted. Positions with $V > 1$ are polymorphic. The signal sequence (SIG), antigen-presenting domain (α1α2 domain), α3 domain, transmembrane (TM) and cytosolic (CYT) region of HLA class I molecules are shown under the variability plots. The box in the α3 domain indicates loop residues 225–232 interacting with CD8. The CD8 contacts were obtained from the analysis of the X-ray structure of HLA-A∗0201 in complex with human CD8 (PDB: 1AKJ). Sequence numbering is from the X-ray structures of pMHC complexes and starts after the leader sequence.

103 of 366 positions respectively, are subject to amino acid change. In contrast, in class II, with the exception of the HLA-DRB possessing 106 sites with $V > 0$, the number of variable positions is much smaller. Also, in HLA class II, aside from HLA-DQ, the α chains have many fewer variable positions than the β chains. The extreme case occurs with HLA-DR. Thus, there are only two known HLA-DRA variants differing by just one amino acid within the TM domain (Figure 2(D); Table 4). Despite the large number of variable positions in HLA molecules, Table 4 shows that most have $V < 0.5$. Many fewer positions qualify as dimorphic or polymorphic.

Variable sites ($V > 0$) are distributed throughout the sequence of HLA class I molecules, although with some preference for the α1 and α2 regions. However, dimorphic sites ($0.5 \leq V \leq 1$) and most notably polymorphic sites ($V > 1$) are concentrated in the α1 and α2 regions (Table 4). Moreover, the density of polymorphic sites (defined as sites with $V > 1$ divided by the total domain length) is larger in the α1 and α2 regions than in the remaining domains. In HLA class II molecules, the variable, dimorphic and polymorphic sites are most concentrated in the α1 and β1 segments. In summary, polymorphic residues are preferably localized to the α1α2 domain of HLA class I and the α1β1 domain of HLA class II molecules where peptide binding and TCR-based immune recognition occurs.[23,24] The relationship between the polymorphic sites and the function of these domains are discussed below.

There are also some polymorphic sites ($V > 1$) that are not restricted to the antigen-presenting platform of HLA class I and HLA class II (Table 4). The selective pressure driving these polymorphisms is unknown. Polymorphisms in the α3 domain of MHC class I molecules are of particular interest, since this domain binds the
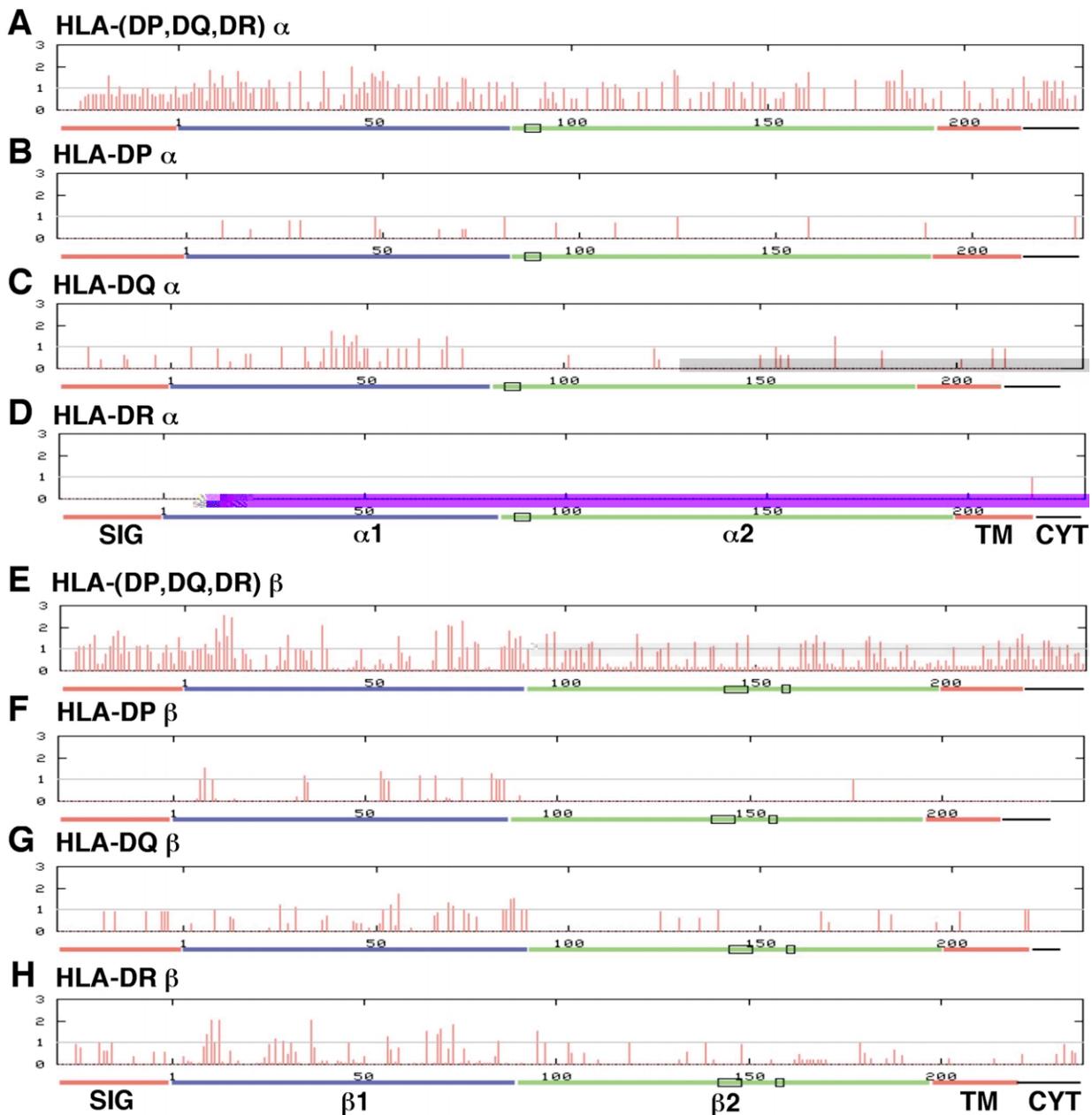
Figure 2. Sequence variability within all HLA class II α chains (A), and from each of the HLA class II gene HLA-DP, HLA-DQ, and HLA-DR α chains (B, C and D, respectively) was calculated from the relevant multiple sequence alignments using V and plotted for each site/position in the alignment. Likewise, E, F, G and H show variability plots within the HLA class II β chains. The value of $V = 1$, above which a site is always polymorphic, is indicated. Signal sequence (SIG), α1 antigen-presenting region, α2 domain, transmembrane (TM) and cytosolic (CYT) of the HLA class I α chains are indicated under their variability plots. Also indicated under the variability plots of HLA class II β chains are the equivalent regions SIG, β1, and β2, TM and CYT. The region in the α domain that appears boxed encompasses positions 88–92 that contact the CD4 co-receptor. In the β domains, the position 142–148 and 156–158 also interact with the CD4 co-receptor, and are indicated with boxes. CD4 contacts were extrapolated from those obtained from the X-ray structure of mouse IA^k in complex with human CD4 (PDB: 1JL4). Sequence numbering is that of the X-ray structure of pMHC complexes and starts after the leader sequence.

CD8 co-receptor and associates with β2m. The CD8 co-receptor is an important component of the pMHCI recognition complex, binding in a bidentate manner to the same pMHC molecule ligated by the TCR, and thus helping to form a stable interaction.[25,26] No polymorphic site was located in a position contacting CD8. Indeed, the primary CD8 contact involving a flexible loop (residues 223–229) of the MHC α3 domain[25] is virtually free of variability in the three distinct HLA class I molecules (Figure 1). Moreover, the sequence of this loop is conserved among the three HLA class I molecules (Figure 1(A)). HLA-A and HLA-B allotypes (HLA-A68, HLA-B48, and

Table 4. Variable sites in classical HLA molecules and their location for different V ranges

HLA Class I

	V	Leader	α1	α2	α3	TM	CYT	Total
HLA-A	>0	9 (36)	44 (49.4)	51 (57.3)	50 (48.5)	8 (36.4)	6 (16.7)	168
	0.5–1	2 (8.0)	8 (9.0)	12 (13.5)	4 (3.9)	4 (18.2)	4 (11.1)	34
	>1	0 (0.0)	6 (6.7)	7 (7.9)	3 (2.9)	0 (0.0)	0 (0.0)	16
HLA-B	>0	17(70.8)	42 (46.6)	63 (7.0)	45 (44.1)	6 (27.3)	3 (8.6)	176
	0.5–1	4 (16.6)	11 (12.2)	8 (8.8)	2 (2.0)	1 (4.6)	2 (5.7)	28
	>1	2 (8.3)	8 (8.9)	7 (7.8)	0 (0.0)	0 (0.0)	0 (0.0)	17
HLA-C	>0	6 (25.0)	18 (20.0)	22 (24.4)	15 (14.2)	10 (45.4)	32 (91.4)	103
	0.5–1	2 (8.3)	8 (8.0)	6 (5.7)	5 (22.7)	2 (5.7)	2 (5.7)	31
	>1	1 (4.2)	3 (3.3)	6 (6.6)	2 (1.9)	1 (4.55)	1 (2.9)	14

HLA Class II

	V	Leader	α1	β1	α2(β2)	TM	CYT	Total
HLA-DPA	>0	0 (0.0)	10 (11.9)	–	5 (4.7)	0 (0.0)	1 (7.1)	16
	0.5–1	0 (0.0)	5 (6.0)	–	5 (4.7)	0 (0.0)	1 (7.1)	11
	>1	0 (0.0)	0 (0.)	–	0 (0.0)	0 (0.0)	0 (0.0)	0
HLA-DPB	>0	0 (0.0)	–	22(25.0)	2 (1.8)	0 (0.0)	0 (0.0)	24
	0.5–1	0 (0.0)	–	8 (9.1)	1 (0.9)	0 (0.0)	0 (0.0)	9
	>1	0 (0.0)	–	8 (9.1))	0 (0.0)	0 (0.0)	0 (0.0)	8
HLA-DQA	>0	5 (17.9)	27 (32.9)	–	9 (8.3)	2 (9.1)	1 (6.7)	44
	0.5–1	3 (10.7)	16 (19.5)	–	7 (6.5)	1 (4.5)	1 (6.7)	28
	>1	0 (0.0)	6 (7.3)	–	1 (0.9)	0 (0.0)	0 (0.0)	7
HLA-DQB	>0	6 (18.8)	–	32 (35.6)	10 (9.3)	2 (8.7)	1 (12.5)	51
	0.5–1	6 (18.8)	–	13 (14.4)	8 (7.4)	2 (8.7)	1 (12.5)	30
	>1	0 (0.0)	–	9 (10.0)	0 (0.0)	0 (0.0)	0 (0.0)	9
HLA-DRA	>0	0 (0.0)	0 (0.0)	–	0 (0.0)	1 (5.2)	0 (0.0)	1
	0.5–1	0 (0.0)	0 (0.0)	–	0 (0.0)	1 (5.2)	0 (0.0)	1
	>1	0 (0.0)	0 (0.0)	–	0 (0.0)	0 (0.0)	0 (0.0)	0
HLA-DRB	>0	9 (31.0)	–	61 (67.8)	29 (26.9)	2 (8.7)	5 (31.2)	106
	0.5–1	8 (27.6)	–	10 (11.1)	10 (9.3)	0 (0.0)	3 (18.8)	31
	>1	0 (0.0)	–	14 (15.6)	2 (1.9)	0 (0.0)	0 (0.0)	16

Values indicate the number of sites with the specified variability in the different HLA domains, followed the density (in parentheses) of variable sites defined as the percentage of variable position divided by the domain length. Domain boundaries used to derive domain sizes needed for these calculations were obtained from the relevant consensus sequences using SMART[69] and were the same as those of reported X-ray structures. Consensus sequences were derived from the relevant amino acid sequence alignments. Boxes labeled with "–" indicate that the corresponding HLA molecule (leftmost in row) does not contain a particular domain (top in column).

HLA-B81) have been described with impaired CD8 binding due to substitution at position 245 of the α3 domain, however.[27,28] Position 245 is an Ala in almost all HLA class I alleles. This residue is buried in the structure of the α3 domain and not directly accessible for contact by CD8. However, A245 is adjacent to the flexible α3 loop interacting with CD8 such that an A245 mutation can result in a local distortion, thereby disrupting the binding of CD8.[25] Position 245 has a V value of 0.451, 0.145 and 0, for the HLA-A, HLA-B and HLA-C, respectively. Thus, this position will be substituted more frequently in HLA-A than it is in HLA-B, and is constant in HLA-C. Not surprisingly, no HLA-C allotype has been described with impaired CD8 binding due to a change at residue 245. Note that position 245 has some degree of variability but this is under the threshold (0.5) that we set to define dimorphic sites. This finding illustrates that any allelic polymorphism can have an impact on the structure and function of a given HLA molecule.

HLA class II molecules engage CD4 as their co-receptor.[29] CD4 interacts with a hydrophobic pocket created by residues located in the α2 and β2 domains of HLA class II molecules. The α2 domain residues contacting CD4 show virtually no variability (Figure 2). Moreover, this region is identical between the HLA-DPA, HLA-DQA, and HLA-DRA chains (Figure 2(A)). CD4 contacts with individual HLA class II β chains (HLA-DPB, HLA-DQB, and HLA-DRB) show little variability (Figure 2(F)–(H)). However, when all HLA class II β chains were analyzed together (Figure 2(E)), position 145 had a significant variability ($V > 1$), indicating that this residue varies between the HLA-DPB, HLA-DQB, and HLA-DRB chains. This difference was somewhat unexpected, given that CD4 interaction is a feature common to the three classical HLA class II molecules. Nevertheless,

examination of the reported X-ray crystal structure of CD4-MHCII[29] indicates that the CD4 domain 1 C″ strand forms an extended β sheet with the MHCII β2 domain D strand, which encompasses Ser143-Thr145 and, as such, the side-chain of residue 145 is not constrained by any specific amino acid side-chain at that position.

Sequence variability and peptide/TCR contacts in the peptide-recognition domain of HLA molecules

To investigate the functional significance of HLA polymorphisms in the antigen-presenting plat-form, we have combined HLA sequence variability analysis with HLA sequence contact maps enumerating those interactions with peptide and TCR. The combination of sequence variability and contact maps is shown in Figure 3 for HLA class I and in Figure 4 for HLA class II. Most HLA residues identified as peptide contacts (see Materials and Methods) were identical among the specific HLA molecule subtypes belonging to the same class. Several contacts are distinct, implying a different peptide binding mode and/or differences arising from the limited set of peptide–MHC structures analyzed. Although HLA TCR contact residues were obtained from a limited number of
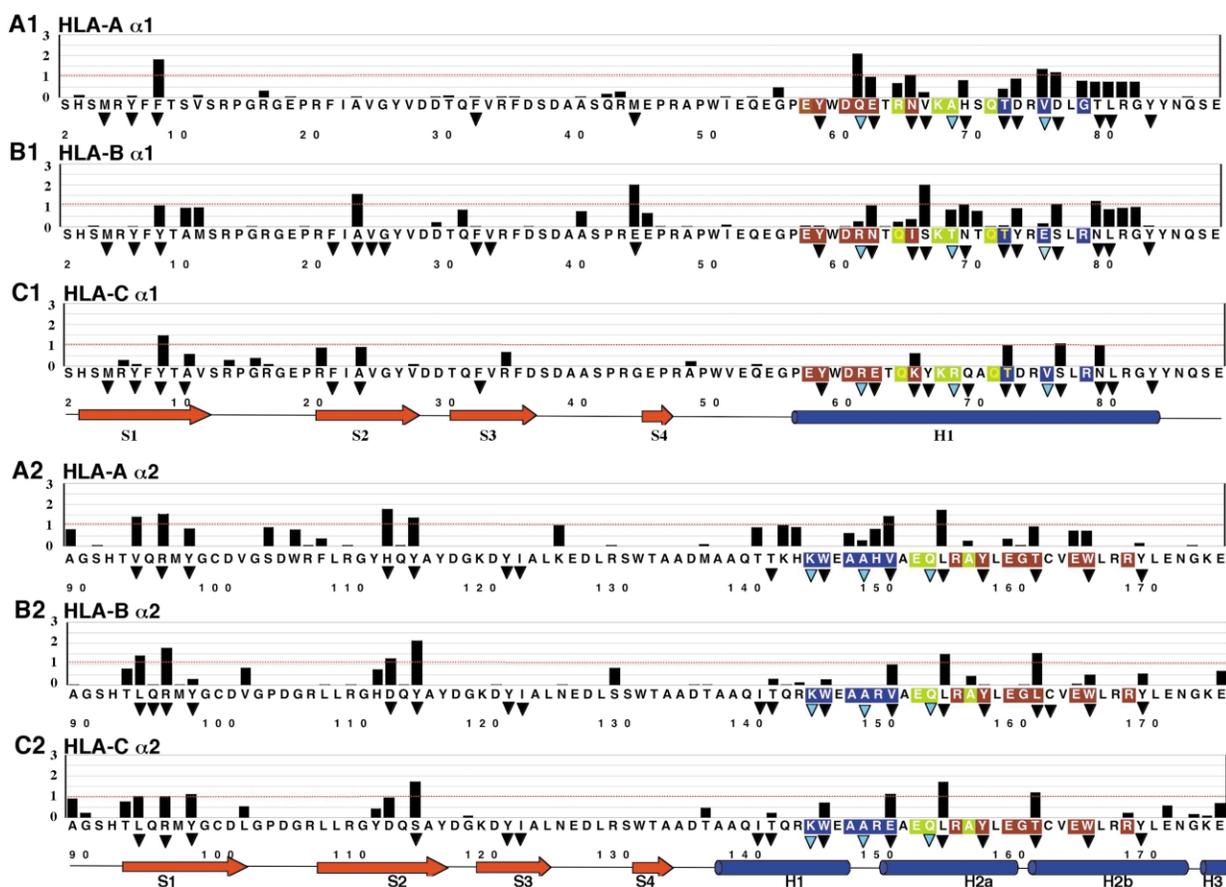
Figure 3. Sequence variability and peptide and/or TCR contacts in the antigen-presenting α1α2 domain of HLA class I molecules. Sequence variability within the α1 and α2 domain of HLA-A (A1 and A2), HLA-B (B1 and B2), and HLA-C (C1 and C2), are plotted against their corresponding consensus sequences. Variability was calculated using equation (1) from the relevant multiple sequence alignment. Consensus sequences were derived from the relevant alignment as the most abundant amino acid at every position. The polymorphism threshold value of $V = 1$ is indicated with a red line. Secondary structure elements are indicated under the consensus sequences (red arrows for β-strands and blue cylinders for α-helices), and are labeled following the nomenclature given by the PDBsum database.[68] Positions that are in contact with the peptide are indicated with an inverted triangle under the consensus sequences. Black triangles indicate residues that form part of the binding pockets of the relevant HLA molecule contributing to the binding properties and specificity of the HLA molecule. In contrast, blue triangles point out HLA residues that are in proximity to the peptide but likely do not contribute to the specificity and binding properties of the molecule. MHC residues that are potential contacts for the TCR Vα chain are shadowed in red, and shadowed in blue are potential contacts for the TCR Vβ chains. Shadowed in green are those residues that can be contacted by both the TCR Vα and TCR Vβ chains. TCR and peptide contacts were obtained from the analysis of the structures indicated in Table 2, and represent a summation of all possible contacts obtained from the analysis of all the structures as indicated in Materials and Methods. Therefore, due to the flexibility of interaction of both the peptide and TCR with the MHC, a given peptide and/or TCR will not necessarily engage in all the contacts indicated in the Figure. Moreover, as the numbers of X-ray structures increase, it is possible that new contacts will be found.
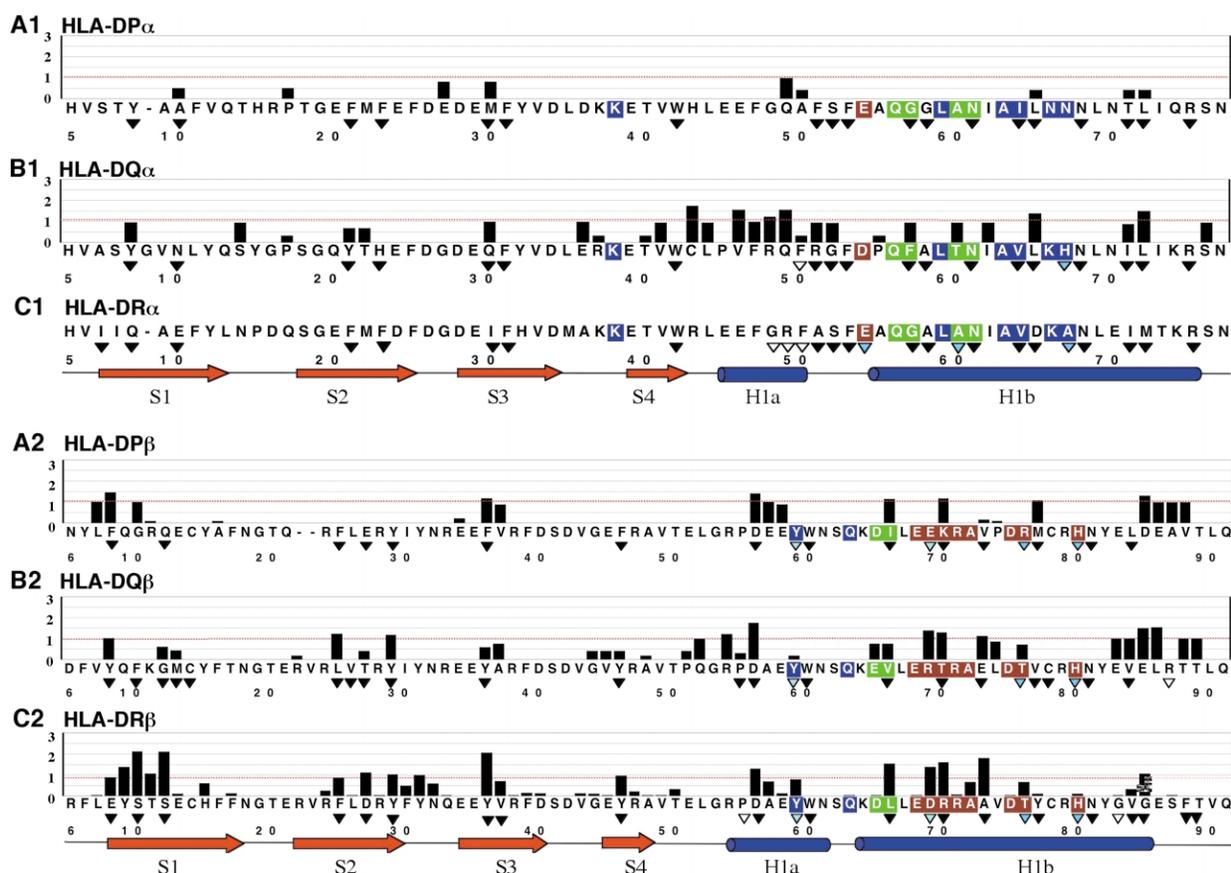
Figure 4. Sequence variability and peptide and/or TCR contacts in the antigen-presenting α1β1 domain of HLA class II molecules. Sequence variability within the α1β1 antigen-presenting domain of HLA-DP (A1 and A2), HLA-DQ (B1 and B2), and HLA-DR (C1 and C2) was obtained using the *V* metric from the relevant multiple sequence alignments, and plotted against their corresponding consensus sequences. In the case of HLA-DRA, only the consensus sequence is shown (C1), since the only two known allelic variants (DRA∗0101 and DRA∗0102) have identical sequences in the α1 region. Secondary structure elements are indicated under the consensus sequences (using symbols defined in the legend to Figure 3). Sequence variability and consensus sequences were obtained from the relevant amino acid sequence alignments as indicated in Figure 3. The polymorphism threshold value of $V = 1$ is indicated with a red line on the variability plots. Positions that are in contact with the peptide are indicated by inverted triangles under the consensus sequences. Black triangles indicate residues that contribute to the binding properties and specificity of the relevant HLA molecule, whereas blue triangles point out HLA residues that are in proximity to the peptide but likely do not contribute to the specificity and binding properties of the molecule. White triangles point to HLA class II residues interacting with flanking regions of the peptide core extending beyond the binding groove. As for Figure 3, MHC residues that are potential contacts for the TCR Vα chain are shadowed in red, shadowed in blue are potential contacts for the TCR Vβ chains, and shadowed in green are the residues that can be contacted by both the TCR Vα and TCR Vβ chains. Also as for HLA class I molecules, these represent a summations of all possible contacts obtained from the analysis of the set of X-ray crystal structures (Table 2; Materials and Methods), and therefore a given peptide and/or TCR will not necessarily engage in all the contacts indicated in the Figure.

pMHC–TCR 3D structures (Table 2), these are considered to be comparable for all the MHC molecules belonging to the same class (MHCI or MHCII), given that their 3D structure and geometry are almost identical. The number of different peptide and TCR contact residues identified in HLA class I molecules was 40 and 30, respectively (Figure 3). Likewise, in HLA class II, the number of residues identified as making peptide and TCR contacts was 45 and 23, respectively. Note that only primary HLA peptide residue contacts (those indicated with black triangles in Figures 3 and 4) have been considered in this enumeration of peptide contacts. Peptides binding

to MHCII molecules often extend beyond the boundaries of the binding groove, with only a core of nine residues fitting into the binding groove.[6,7] Thus, in HLA class II molecules we have considered as primary peptide contacts those residues interacting with the peptide core. Nevertheless, in Figure 4 we have indicated the HLA class II residues that interact with flanking regions of the peptide (shown as inverted white triangles). There are additional HLA class I and class II residues in close proximity to the peptide, that have not been tallied as primary peptide contacts (shown as inverted blue triangles in Figures 3 and 4) as visual inspection of the 3D structures indi-

cated that they are not likely to contribute to peptide binding. Interestingly, for both class I and class II molecules, many of the residues identified as TCR contacts are also forming peptide contacts (Figures 3 and 4). The number of residues identified as TCR contacts is somewhat greater in HLA class I than in HLA class II molecules. This difference is primarily due to the fact that there are currently more available 3D structures of TCR–pMHCI complexes than TCR–pMHCII and the former apparently manifests greater docking variability.[30] Note that the listed HLA-TCR residue contacts represent a summation of all HLA residues found to be within 5 Å of the TCR in existing TCR–pMHC crystallographic structure: these are not tallied as atomic contacts but rather as residue contacts.

A summary of the $V > 0$, $0.5 \le V \le 1$ and $V > 1$ sites contacting peptide or TCR is shown Table 5. Also listed are the total number of variable sites within the indicated V ranges present in the HLA class I and HLA class II ($\alpha + \beta$ chains) full-length molecules as well as the corresponding $\alpha1\alpha2$ and $\alpha1\beta1$ domains. Analysis of these results indicates that the HLA peptide contact residues account for most of the $V > 1$ sites observed in the antigen-presentation domains of HLA class I and HLA class II molecules (nine of 13 in HLA-A, 14 of 15 in HLA-B, nine of nine in HLA-C, eight of eight in HLA-DP, nine of 15 in HLA-DQ, and ten of 14 in HLA-DR). Thus, the numbers of HLA TCR contact residues that are $V > 1$ (polymorphic) is much smaller than that of peptide contact residues. Furthermore, virtually all $V > 1$ residues that are TCR contacts are also in close proximity to the peptide (≤ 5 Å). Overall, these results indicate a strong correlation between $V > 1$ residues and peptide–antigen binding. Such a correlation provides a good predictive tool to anticipate that any $V > 1$ sites in the MHCI $\alpha1\alpha2$ domain or MHCII $\alpha1\beta1$ domain will likely be a peptide contact residue.

It is interesting to note that each HLA molecule has a variability profile that is distinct from that of the other HLA molecules (Figures 3 and 4). This is best seen by examining the residues that are peptide and/or TCR contact positions, as they display the greatest variability. As an example, residue 62 is the most variable position in HLA-A but it has low variability in HLA-B and is virtually constant in HLA-C (Figure 3). These differences in localization of the variability in HLA class I and HLA class II molecules are shown dramatically when variability is displayed on their 3D structures. In Figure 5, the variability of HLA-A, HLA-B and HLA-C molecules has been mapped onto their molecular surface (Figure 5(A1), (B1) and (C1), respectively) and corresponding worm representations (Figure 5(A2), (B2) and (C2), respectively). Likewise, Figure 6 represents the variability of HLA-DP, HLA-DQ and HLA-DR molecules displayed on their respective molecular surfaces (Figure 6(A1), (B1) and (C1), respectively), with corresponding worm representations (Figure 6(A2), (B2) and (C2)). In these Figures, variability has been scaled into a two-color representation, where blue indicates variable sites and red is for invariant sites with blue, white and red shadings representing positions with intermediate variability.

A theme emerges from the analysis of these Figures, which is common to class I and class II molecules. First, variable positions (in blue) map onto different sites in each of the HLA class I and HLA class II molecules. As an example, in class I the three residues with the highest variability are residue 62, 9 and 114 in HLA-A (Figure 5(A2)), residues 45, 67, and 116 in HLA-B (Figure 5(B2)), and residues 9, 116, and 156 in HLA-C (Figure 5(C2)). Second, as most variable positions correspond to residues contacting peptide (Figures 3 and 4; Table 4), variable residues map onto the β-strands that form the base of the antigen-presenting domain (for example residues 9 and 114 in HLA-A, and residues β11, β13, β37 in HLA-DRB), as well as in residues lining the walls of the peptide-binding groove (for example, residue 67 in HLA-B, residue 156 in HLA-C, and residue β57 in HLA-DP and HLA-DQ) (Figures 5 and 6). As anticipated, in HLA-DP and HLA-DR HLA class II molecules, virtually no variability maps onto their α1 domain 3D structures. Interestingly, however, in HLA-DQ residues α44 and α50, two of the three most variable positions were not identified as peptide contacts (Figure 4). However, they are in the neighborhood of the peptide, indicating they could contribute to peptide binding in some other HLA-DQ allelic

Table 5. Number of variable HLA sites contacting peptide and TCR

MHC	V	Peptide	TCR	$\alpha1\alpha2$ ($\beta1$)	Total
HLA-A	>0	24 (60.0)	17 (56.7)	95	168
	0.5–1	9 (22.5)	8 (26.7)	20	34
	>1	9 (22.5)	4 (13.3)	13	16
HLA-B	>0	31 (77.5)	19 (63.3)	103	176
	0.5–1	5 (12.5)	2 (6.7)	19	28
	>1	14 (35.0)	2 (6.7)	15	17
HLA-C	>0	18 (45.0)	6 (20.0)	40	103
	0.5–1	7 (17.5)	3 (10.0)	14	31
	>1	9 (22.5)	2 (6.7)	9	14
HLA-DP	>0	30 (66.7)	10 (43.5)	32	40
	0.5–1	4 (8.9)	0 (0.0)	13	20
	>1	8 (17.8)	2 (8.7)	8	8
HLA-DQ	>0	27 (60.0)	8 (34.8)	59	95
	0.5–1	14 (31.1)	5 (21.7)	29	58
	>1	9 (20.0)	2 (8.7)	15	15
HLA-DR	$V > 0$	18 (40.0)	12 (52.2)	61	107
	0.5–1	4 (8.9)	3 (13.0)	10	32
	>1	10 (22.2)	3 (13.0)	14	16

In HLA class II, 45 and 23 positions were identified as potential peptide and TCR contacts, respectively. In class I, 40 and 30 sites were identified as potential peptide and TCR contacts, respectively. Percentages shown in parentheses are relative to the total number of residues contacting peptide or TCR.

variants. Third, although more limited, some variability is seen in the molecular surface of the MHC molecules in the regions contacted by the TCR. The extent of this variability differs among the various HLA molecules. Thus, among class I molecules, the HLA-A alleles appear to manifest the greatest variation among surface exposed residues in TCR contact areas (Figure 5(A1)). Moreover, residue 62 (a TCR contact site) in the α1 helix is the most variable residue in HLA-A. In contrast, in HLA-B and HLA-C, none of the TCR contact sites was among the most variable and, furthermore, the more variable residue in this region (residue 163) is located in the α2 helical region and is, in addition to a TCR contact site, a peptide contact site. In class II, the variability in the TCR contact regions also differs for each subtype. Thus, this region is less prone to variation in HLA-DP (Figure 6(A1)), intermediate in HLA-DR, and most variable in HLA-DQ. Among the TCR contact sites, residues 70 and 71, which are also peptide contacts, have the largest variability in both HLA-DQ and HLA-DR. Note how the β1 variability in HLA-DQ and HLA-DR *vis-a-vis* TCR contacts involves the Vα docking site and is a region without variability in the corresponding α2 surface of HLA-A, -B and -C molecules. Whether this difference relates to the more restricted docking topology of class II MHC binding T-cells[31] remains to be determined.

Sequence logos of HLA contacts sites with peptide and TCR

Variability values within the HLA molecules never exceed 2.5, indicating that there are only a few prevalent amino acids permitted at any site, presumably a consequence of structural and/or functional constraints. Given that the majority of polymorphic sites are peptide and/or TCR contacts, a better understanding of the interaction of HLA molecules with the peptide and TCR will follow from knowing what specific amino acids types are present. Because V offers identification of the polymorphic sites without information on the amino acids types present at a given position, we compiled the amino acid residues at positions identified as primary peptide contacts from primary HLA sequence alignments (positions indicated with inverted black triangles in Figures 3 and 4) and generated a consensus sequence logo (Figure 7). Likewise, a second set of sequence logos was generated for each of the HLA molecules
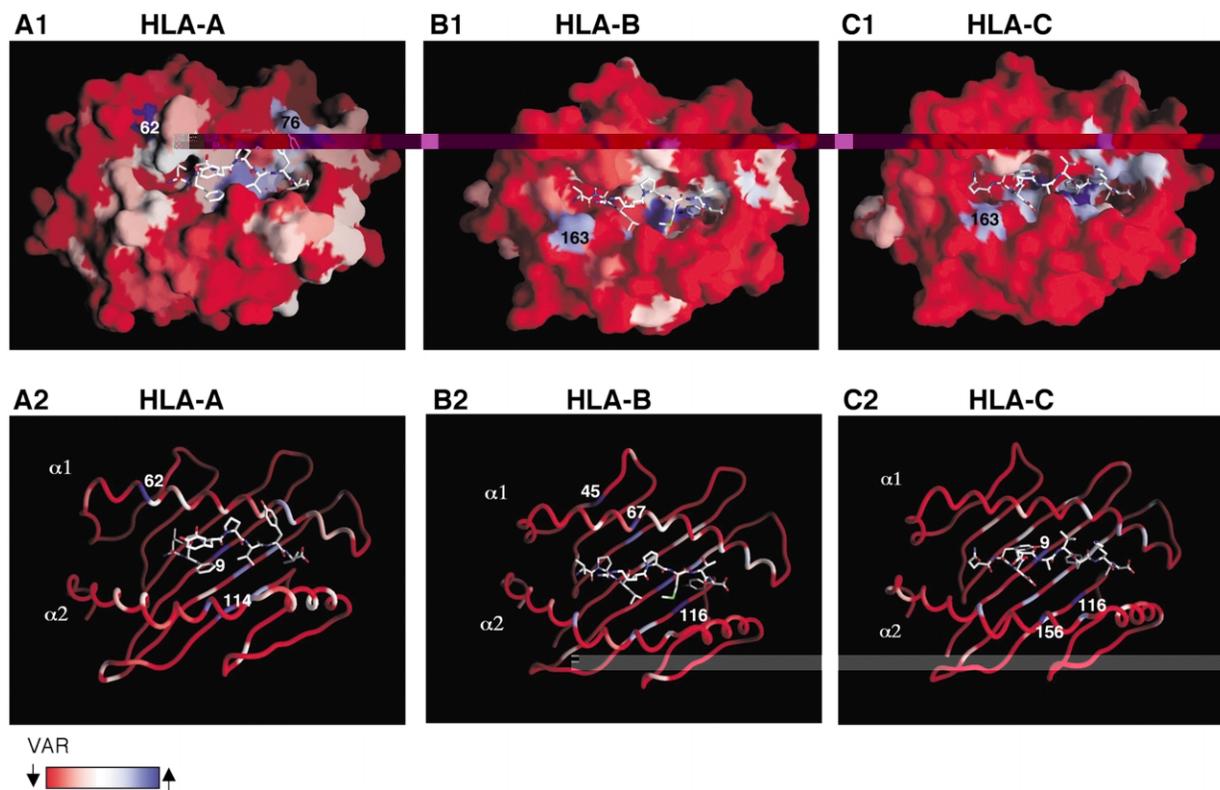

Figure 5. The 3D-structure mapping of the sequence variability in α1α2 domains of HLA class I molecules. Sequence variability in the α1α2 domains of HLA-A, HLA-B, HLA-C molecules was mapped onto their molecular surfaces (A1, B1 and C1), as well as onto their worm representations (A2, B2 and C2, respectively). Consensus amino acids in positions with the greatest variability are indicated in the worm representations. The consensus amino acids in the polymorphic positions that are TCR contacts are indicated in the molecular surface diagrams. Variability is represented in a relative two-color scale of blue and red, where red indicates non-variable regions and blue indicates variable regions. White in this two-color variability scale indicates regions of intermediate variability. The Figures were rendered from the following PDB files: HLA-A, 1DUY; HLA-B, 1A1N; HLA-C, 1IM9.
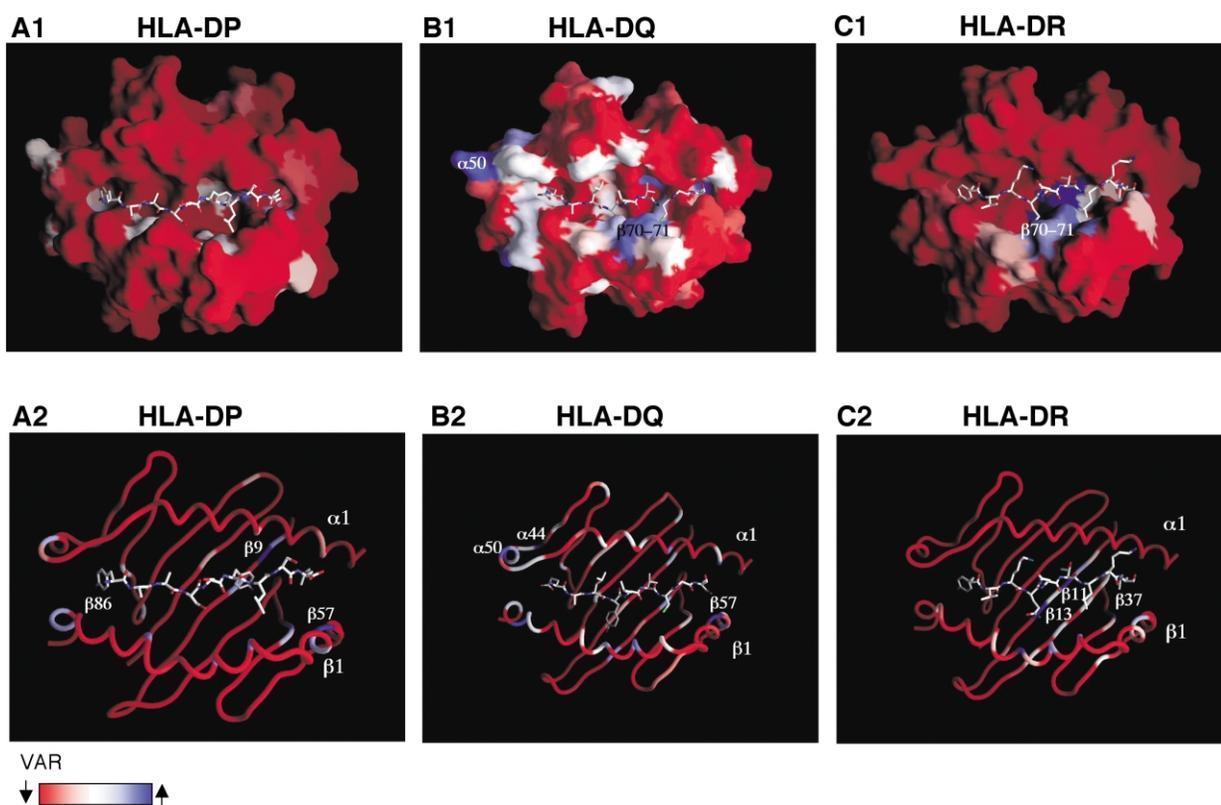
Figure 6. The 3D-structure mapping of the sequence variability in α1β1 domains of HLA class II molecules. Sequence variability in the α1β1 domain of the HLA-DP, HLA-DQ, HLA-DR molecules appears mapped onto their molecular surfaces (A1, B1 and C1), as well as onto their worm representations (A2, B2 and C2, respectively). Consensus amino acids in positions with the greatest variability are indicated in the worm representation. The consensus amino acids in the polymorphic positions that are TCR contacts are indicated in the molecular surface diagrams. Variability is represented as in Figure 5. HLA-DQ and HLA-DQ Figures were rendered from PDB files 1JK8 and 1HXY, respectively. HLA-DP was rendered from a model obtained as indicated in Materials and Methods. For simplicity, only the core of the peptide (nine residues) fitting into the binding groove of the HLA class II molecules is shown.

(Figure 8) from the residues identified as TCR contacts sites (color shaded residues in Figures 3 and 4). In such a sequence logo, each amino acid type at a given position is represented by its corresponding symbols with a height proportional to both the sequence information in that position and its frequency (see Materials and Methods). In this representation, the greater the number of amino acids present at a given site, the lower the sequence information at that site. Correspondingly, the height of the stack of symbols in that position is small, with each amino acid represented by a size that is proportional to its frequency. In any position, the most frequent amino acid is on the top of the stack. On the other hand, if only one residue is present at a given position, then the height is equal to the maximum sequence information at that site, which for a system of 20 symbols is 4.3 (Materials and Methods, equation (2)).

Inspection of HLA peptide-binding sequence logos identifies two types of residues: those that are constant (invariant) and those that are variable (polymorphic or dimorphic). Constant residues are identical between all HLA class I molecules, providing a common conserved framework for the binding of a peptide in a gene and allele-indepen-

dent manner. In HLA class I molecules, these are exemplified by residues Y59, Y84, Y123 and Y159 (Figure 7(A)–(C)). Those residues engage in a conserved network of H bonds with the N and C terminii of the peptide.[6] Similarly, in class II, residues conserved in HLA-DP, HLA-DQ and HLA-DR include αR76, βW61 and βN82 (Figure 7(D)–(F)), form hydrogen bonds with the backbone of the peptide and thereby provide a binding affinity that is also gene and allotype-independent. Note that no conservation is anticipated between HLA class I and HLA class II molecules, since binding of peptides to these molecules occurs in a very different mode. Additional invariant residues to those conserved between all HLA molecules of the same class, can be found among the alleles belonging to the same gene. These sets of additional gene-specific invariant residues can be quite distinct among different HLA molecules (Figure 7), providing binding features shared by all alleles derived from the same gene. Finally, peptide-binding specificity is modulated at the allelic level by the polymorphic residues. The combination of variable and constant residues is also gene-specific. For example, residue 45 is an invariant Met in HLA-A (Figure 7(A)), conserved Gly in HLA-C
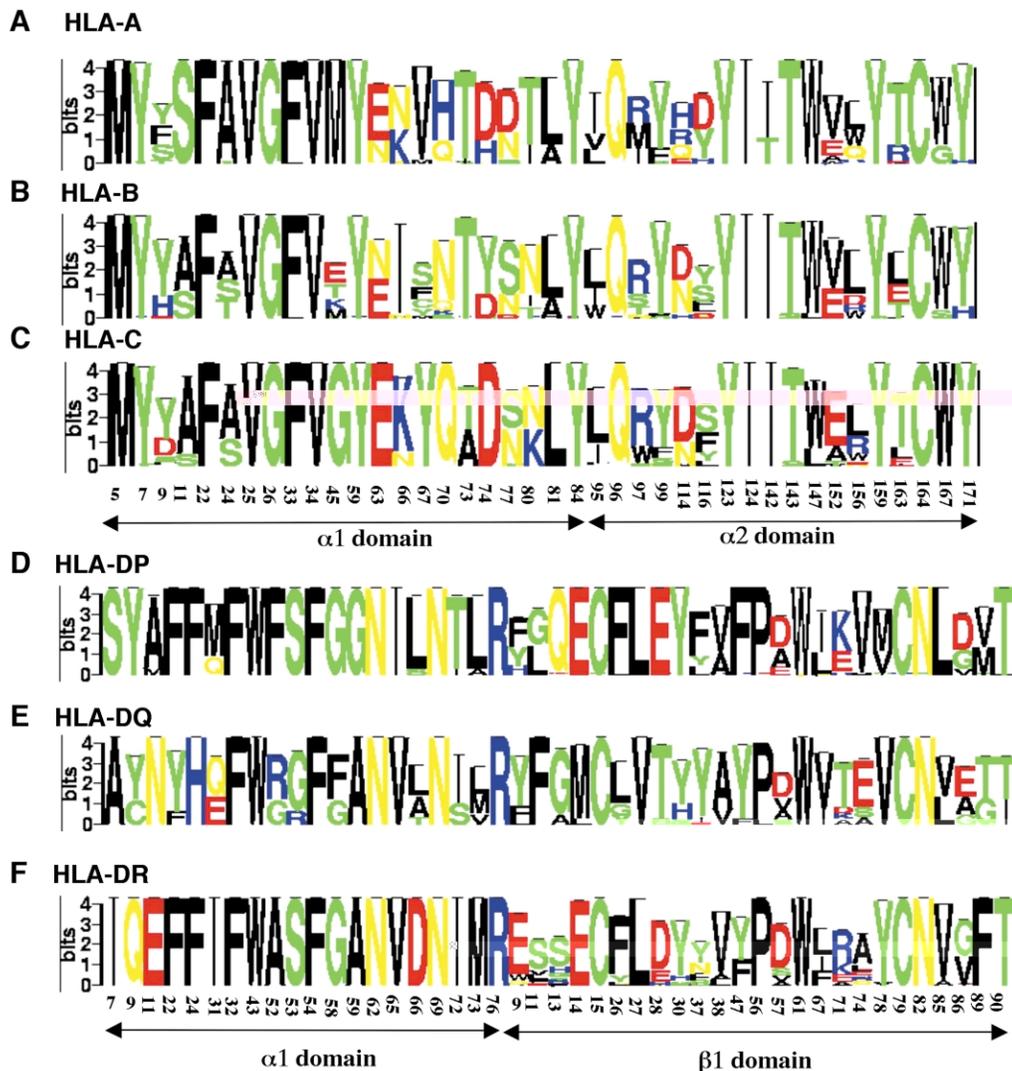
Figure 7. Consensus sequence logos of HLA class I and HLA class II peptide contact sites. HLA peptide binding consensus sequence logos are given for MHC class I HLA-A (A), HLA-B (B), and HLA-C (C), and for MHC class II HLA-DP (D), HLA-DQ (E), and HLA-DR (F). Only primary peptide contacts were considered for these sequence logos (residues indicated with black triangles in Figures 3 and 4). HLA sites from which the sequence logos were derived are indicated under the amino acid symbols. Also indicated under the amino acid symbols is the domain location of the residues (α1 or α2 domain of class I MHC molecule, and likewise the α1 or β1 domain of class II MHC molecules). These sequence logos show all the amino acid types present at a given position with a size that is proportional to its frequency and total sequence information in that position (see Materials and Methods for more details). Thus, the more amino acid residues present in a given site, the lower will be the sequence information at that site (bits) and, correspondingly, the smaller will be the height of the stack of symbols. Yet each amino acid residue in the stack will be proportional to its frequency in that position. On the other hand, if only one amino acid residue is present, the symbol representing that residue will reach heights of 4.3 bytes, the maximum information content for a system of 20 symbols.

(Figure 7(C)) but in HLA-B is either Glu, Thr, Lys or Met. Likewise, residue α52 is a conserved Phe in HLA-DP (Figure 7(D)), a conserved Ala in HLA-DR (Figure 7(F)), and either an Arg or a Gly in HLA-DQ (Figure 7(E)). This gene-specific combination of polymorphic and constant residues suggest that, in general, peptides binding to HLA molecules encoded by the same gene are likely to be more similar to each other than to other peptides binding to HLA allelic variants derived from different genes.

Comparison of sequence logos on HLA-TCR contact sites is also informative. As shown in Figure 8, many of these TCR contacts on HLA are conserved. In the case of MHC class I, 18/30 HLA-A, 26/30 HLA-B and 25/30 HLA-C contacts are essentially invariant. Likewise, for HLA class II, 21/23 DP, 16/23 DQ and 17/23 DR residues are conserved. These conservations define HLA gene-specific TCR docking sites. In addition, note that some of the HLA-TCR contact sites are polymorphic. The extent of this variability component differs for the individual HLA genes. With respect to TCR contacts, polymorphisms within HLA-A,
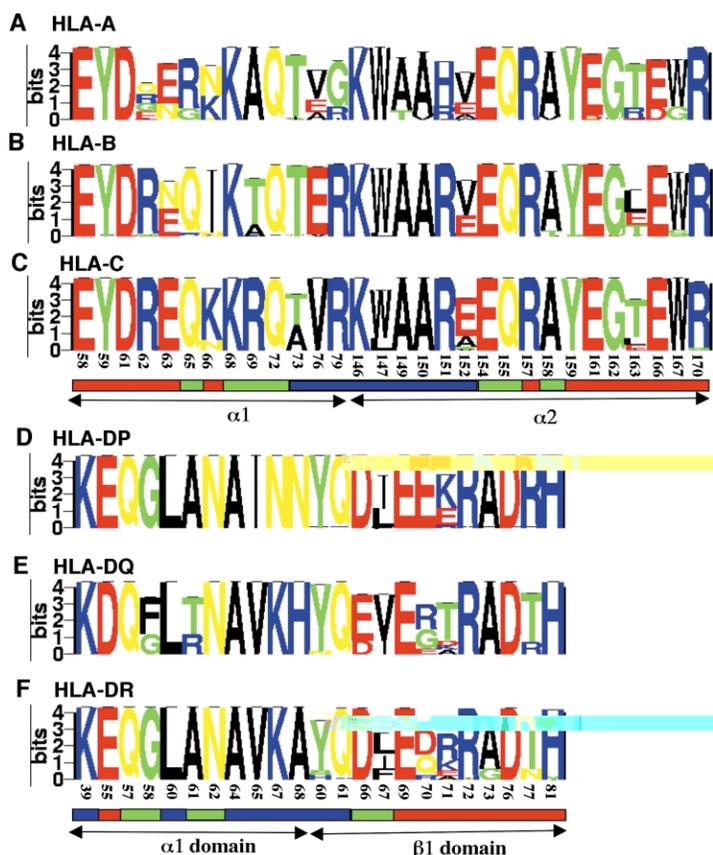
Figure 8. Consensus sequence logos for HLA class I and HLA class II TCR contact sites. HLA TCR consensus sequence logos are presented for MHC class I HLA-A (A), HLA-B (B), and HLA-C (C), and for MHC class II HLA-DP (D), HLA-DQ (E), and HLA-DR(F). HLA sites from which the sequence logos were derived are indicated under the amino acid symbols. The domain location of the residues (α1 or α2 domain in class I MHC molecule, and likewise the α1 or β1 domain of class II MHC molecules) is indicated under the amino acid symbols. TCR contact sites were located in the helical region of the α1 and α2 domain of MHC class I, and likewise in the helical region of the α1 and β1 domains, with the exception of residue 39, which is located in the S3–S4 loop of the α1 domain of HLA class II. MHC contacts to Vα, Vβ and both Vα plus Vβ domains are indicated by red, blue and green rectangles, respectively.

DQ and DR have the greatest contribution to allelic variability. In Figure 8, we have also distinguished between residues that contact the Vα (indicated with a red rectangle under the sequence), Vβ (blue rectangle) or both domains (green rectangle). Note how the TCR makes more potential contacts with the residues in the MHCI α2 helix than with residues in the α1 helix. This is significant, taking into account the fact that TCR contacts are representative of a summation of all TCR–pMHC structures analyzed and especially given that docking variability is greatest along the MHCI α1 helix, thereby exaggerating the number of such potential contacts in favor of TCR Vβ interaction. That more Vα residue contacts with MHCI are observed further underscores the key role of Vα in pMHC ligand recognition.

From the HLA TCR contact logos, it is clear that many residues are conserved among the molecules encoded by the three HLA class I genes. For example, contact residues such as E58, Y59, D61, K68, Q72, K146, E154, Q155, R157, Y159, G162 and R170 are constant and conserved in all three subgroups of HLA molecules (Figure 8(A)–(C)). Similarly, a majority of the HLA class II TCR contact residues are constant among HLA-DP, HLA-DQ and HLA-DR (Figure 8(D)–(F)) despite the fact that HLA class II molecules are more divergent than HLA I molecules (Figure 2). These constant contact residues in all HLA class II molecules include αK39, αQ57, αL60, αN62, αA64; βQ61,

βE69, βR72, βD76 and βH81. Nevertheless, there are differences between the HLA-DP, HLA-DQ and HLA-DR sequence logos affecting invariant residues (for example, residue α68 is Asn in HLA-DP, His in HLA-DQ and Ala in HLA-DR). Note also that the amino acid types at the polymorphic sites can be quite distinct (for example, at residue β70 and β71).

In the case of the HLA class II TCR contacts, the TCR Vβ domain interacts with MHC residues that are more hydrophobic (five of eight) than those contacted by the TCR Vα domain (one of nine). In contrast, eight of nine TCR Vα contacts are with hydrophilic residues on MHCII. This observation might indicate that the interaction between the TCR β chain and the MHCII α1 helix provides a major contribution to total binding energy within the TCR–pMHC complex, whereas the interaction of the TCR Vα domain and the MHCII β1 helix may provide greater specificity through this docking.

Implications for TCR–pMHC recognition

TCRs recognize peptide antigens only when presented in the context of a given MHC molecule, a phenomenon referred to as MHC restriction.[32–34] Such MHC-dependent recognition of the peptide by the TCR is so precise that, in general, a TCR will recognize pathogenic antigens presented by self-MHC molecules but ignore the same peptides

when presented by foreign MHC molecules. The importance of the MHC context for TCR-based recognition of peptide antigens is evident in the 3D structures of TCR–pMHCI(II) complexes: these structures reveal a topologically conserved TCR–pMHC docking mode that generally maximizes the contact area between the TCR and the MHC molecule.[30,35–37] Up to two-thirds of atomic contacts between the TCR and the pMHC complex are with the MHC itself.[36]

The TCR is composed of an α and β chain, each encoded by variable, randomly rearranged V and (D)J segments.[38] Variability is concentrated in the complementarity-determining regions (CDRs): CDR1 (α,β) and CDR2 (α,β) are encoded by the germline V segments, whereas the CDR3 (α,β) loops originate from the junction of the V(D)J elements and display greater variability due to junctional diversification mechanisms. The docking of the TCR with the pMHC ligand is such that the Vα domain of the TCR is closest to the N-terminal residues of the antigenic peptide, whereas the Vβ domain of the TCR is closest to the C-terminal segment.[39] Moreover, the CDR1 (α,β) and CDR2 (α,β) loops interact with conserved residues over the long α1 and α2 (β1) helices of the MHC molecule, while the more variable CDR3 (α,β) loops are positioned to primarily contact the peptide. The central location of the CDR3s in the pMHC complex provides an explanation of their role in determining specificity. Despite the conserved diagonal docking mode of the TCR onto the pMHCI, there is variation among the different TCR–pMHCI complexes.[30,36] In addition, the docking of the D10 TCR onto the mouse class II MHC molecule IA^k has been described as close-to-orthogonal, likely reflective of more limited TCR–pMHCII orientations resulting from the p-1 to p-3 extension of peptides bound to MHCII molecules.[31,36] The TCR–pMHC interaction further deviates with respect to twist, tilt and shift.[30,40]

The reactivity of the pre-selected TCR repertoire towards the pMHC argues strongly that MHC-reactivity is inherent in the TCR germline elements, hence independent of positive and negative selection.[41,42] Moreover, such intrinsic reactivity would suggest a co-evolution of the TCR genetic components towards the recognition of the MHC molecule, implying a bias of the TCR germline to optimize recognition of MHC molecules through specified TCR contact residues. The TCR contact sequence logos of HLA class I *versus* HLA class II molecules (Figure 8) are reflective of this co-evolution and, furthermore, support the existence of TCR restriction that is class dependent (MHCI *versus* MHCII TCR restriction). Vα-dependent T-cell repertoire selection through preferential recognition of MHC class I or class II molecules has been shown in several experimental systems.[43,44] In this regard, Sim *et al.*[43] found that in the mouse individual residues in either the CDR1 or CDR2 in the TCR Vα chain were sufficient to shift selection from the CD4$^+$ to the CD8$^+$ subset. Since compari-

son of TCR contact sequences logos of class I and class II HLA molecules (Figure 8) identifies distinct class I *versus* class II conserved residues as being contacted by the TCR Vβ domain, some bias in TCR Vβ-encoded germline usage can be anticipated. Note that TCR V-germline encoded selection of the repertoire through preferential recognition of MHC class I or class II molecules is not absolute, as the same Vα elements can be utilized by both CD4$^+$ and CD8$^+$ compartments. Given that the interaction of the TCR with the pMHC is extremely flexible, and that most of the TCR contact residues in the MHC contribute equally to the total energy of the binding,[45] favorable contacts between the variable CDR3 loops and the pMHC could compensate for strict differential Vα segment usage in the CD8$^+$ and CD4$^+$ subsets. Candeais *et al.*[46] reported that the repertoire of transgenic mice expression only the Vβ^{17+} segment showed dissimilar CDR3s in CD4$^+$ *versus* CD8$^+$ cells. As the CDR3 of the TCR primarily senses the peptide in the MHC groove, these observations highlight the impact of the peptide in the restriction of the TCR.

Inspection of HLA TCR contact sequence logos (Figure 8) makes it apparent that class I-HLA-restricted TCRs will have little potential to interact with HLA class II molecules and *vice versa*, given the differences in contact residues of the two HLA classes. With respect to TCR crossreactivity with alleles from different genes, this appears unlikely because of the existing differences in the TCR contacts sites, and because HLA-A *versus* HLA-B molecules, for example, bind to very different sets of peptides for presentation.[47] On the other hand, TCR crossreactivity between distinct HLA allelic variants derived from the same MHCI (or MHCII) gene is more probable, being likely a major source of alloreactivity. In this case, a similar constellation of bound peptides contributes further to TCR reactivity. For example, a crossreactive H1.7 TCR specific for an influenza virus peptide (HA) restricted by HLA-DR1(DRA∗0101xDRB∗0101) can bind to the same peptide presented by HLA-DR4(DRA∗0101xDRB∗0401).[48]

Implications for alloreactivity

Approximately 1–10% of T-cells are alloreactive, reacting strongly with foreign HLA allotypes to which they were never exposed.[49] This contrasts with a multiple order of magnitude lower frequency of foreign peptide/self-MHC specific T-cells. Alloreactive T-cells are the major cause of allograft rejection.[1] T-cell allorecognition has been divided into two types, referred to as direct and indirect, that differ in their molecular mechanism. The indirect pathway occurs when T-cells recognize peptide antigens derived from foreign (donor) MHC molecules presented by self-MHC (recipient).[50,51] This mechanism is applicable to any polymorphic gene products (i.e. so-called minor histocompatibility antigens) of the donor presented by the self-MHC.[52] On the other hand,

the direct pathway involves T-cell recognition of intact allogeneic pMHC molecules.[53] It is believed that the indirect pathway may play an important role in chronic allograft rejection,[54] whereas the direct pathway is the main contributor to acute graft rejection.[51] The high expression level of HLA molecules per cell and their allelic variability are the obvious molecular bases of the indirect pathway of alloreactivity.

While the precise molecular basis of direct alloreactivity is not fully resolved, recent structural studies of alloreactive TCRs in complex with allogeneic pMHCI molecules[55–57] indicate that these molecules engage in a binding mode that is like that of TCRs in complex with self-MHCI molecules. In these complexes, there is a shift in the extent of TCR Vα and TCR Vβ domain interactions with the pMHC ligand towards the TCR Vβ chain; Vβ makes more contacts than Vα in the 2C/H-2K^{bm3}-dEV8[55] and BM3.3/H-2K^b-pBM1 complexes,[57] while Vβ contributes equally in the Kb5-C20-pKB1/K^b complex.[56] Overall, these studies do not define a unique structural feature of alloreactive TCRs, but instead show that allorecognition is similar to foreign antigen/self-MHC recognition, implying that subtle differences in the MHC and/or conformation of the peptide can trigger an alloresponse.

Two main models have been proposed to explain direct alloreactivity. On the one hand, it has been proposed that alloreactive T-cells recognize allelic polymorphisms at TCR contact sites of the MHC molecule.[58] Indeed, polymorphisms in the TCR contact sites of the MHC have been reported to lead to alloreacitivity for both class I[59] and class II allogeneic MHC.[60] However, alloresponse is often directed against allogeneic MHC molecules displaying amino acid sequencing differences only in the buried portion of the MHC binding groove, which is inaccessible to the TCR.[57,61] To account for these observations, a second model for alloreactivity has been proposed, driven by the specific peptide presented by the allogeneic MHC molecules.[62] In this second model, the high frequency of alloreactive T-cells derives from the fact that allogeneic MHC molecules present "new" peptides to which T-cells from the recipient were never exposed. Although these two models are not necessarily exclusive, current 3D structural studies on alloreactive TCR complexed with allogeneic MHC molecules have highlighted the role played by peptides in alloreactivity. For example, comparison of the 2C TCR in complex with an allogeneic MHC (H-2K^{bm3}-dEV8) and the 2C TCR in complex with self-MHC (H-2K^b-dEV8) has offered insights about how allogeneic MHC molecules can be sensed by the TCR *via* the bound peptide.[57] There are two amino acid differences between the H-2K^{bm3} and H-2K^b molecules. The Asp77Ser mutation that imbues the H-2K^{bm3}-dEV8 with alloreactive properties is located beneath the peptide, making no contact with the TCR. However, this allogeneic mutation is enough to cause a confor-mational change in the peptide, so that the same peptide is presented differently by the allogeneic and syngeneic MHC. This subtle difference is the basis of the alloreactive response by the 2C TCR.

The present study of HLA polymorphism, without dismissing the role of TCR contacts as potential effectors for alloreactivity, endorses the notion that peptide-driven recognition of allogeneic HLA molecules will likely be the most frequent cause of alloreactivity. Indeed, sequence variability analyses of HLA molecules indicate that most polymorphic sites occur in peptide-binding positions (Table 5; Figures 3, 4 and 7) not readily accessible to the TCR (Figures 5 and 6). These polymorphic peptide-binding sites will result in allelic HLA variants having different peptide specificity, likely presenting a new pool of peptides that TCRs will recognize as foreign, since these peptides were not encountered during thymic maturation. In contrast, TCR contacts in the HLA molecules in both class I and class II are the subject of much more limited polymorphic variability (Figures 3, 4 and 8; Table 5), indicating that at least during graft-rejection, alloreactivity caused by this variability is likely to be less frequent.

Materials and Methods

MHC sequences and multiple sequence alignments

Amino acid sequences of the human MHCI and MHCII alleles were collected from the IMGT/HLA database.[13] Amino acid sequences were collected from functional genes with sequences from pseudogenes being discarded. Thus, amino acid sequences of the α and β chains of the HLA-DP molecule (HLA-DPA, and HLA-DPB, respectively) were obtained from the *HLA-DPA1* and *HLA-DPB1* genes. Likewise, the *HLA-DQA1* and *HLA-DQB1* genes were the only sources of the amino acid sequences for the α and β chains of the HLA-DQ molecule (HLA-DQA, and HLA-DQB, respectively). The amino acid sequences for the α chain of HLA-DR (HLA-DRA) came from the *HLA-DRA* gene, whereas the amino acid sequences for the β chain (HLA-DRB) came from four distinct genes: *HLA-DRB1*, *-DRB3*, *-DRB4*, and *-DRB5*. The amino acid sequences of the class I HLA molecules HLA-A, HLA-B, and HLA-C were encoded by the *HLA-A*, *HLA-B*, and *HLA-C* genes, respectively. The number of sequences included in this analysis are shown in Table 1. Only allelic sequences differing in at least one amino acid residue were considered. Multiple sequence alignment of allelic variants of the HLA-A, HLA-B, and HLA-C class I molecules, as well as of HLA class II molecules HLA-DPA, HLA-DPB, HLA-DQA, HLA-DQB, HLA-DRA, and HLA-DRB were carried out from the selected sequences (Table 1) using CLUSTALW.[63] Sequence numbering is based on that for the 3D structures of the MHC molecules.

Sequence variability analysis and consensus sequences

To quantify the diversity of human HLA molecules, we have used a variability metric (V) formally identical with the Shannon entropy equation.[20] For a multiple

sequence alignment, V per site will be given by:

$$V = -\sum_{i=1}^{M} P_i \log_2 P_i \qquad (1)$$

where P_i is the fraction of residues of amino acid type i, and M is equal to 20, the number of amino acid types. Given that sequence alignments were mostly ungapped but contained many sequence fragments, gap symbols were considered as no data. V ranges from 0 (only one amino acid type is present at that position) to 4.322 (all 20 amino acids are represented equally in that position). It is important to note that at least 20 sequences are required in order to achieve the maximum value $V = 4.3$.

Consensus sequences of HLA positions contacting the peptide and TCR were obtained and displayed using sequence logos.[64] In this approach, each of the residues present in a position is represented with a height that is proportional to its frequency, and the height of the entire stack is proportional to the total information content (R) in that position. R per site is given by the following equation:

$$R = 4.3 - V \text{ (bits per position)} \qquad (2)$$

where 4.3 is the maximum variability for 20 symbols, and V is the variability in that position, given by equation (1). Sequence information is given in bits.

Sequence contact maps

MHC residues contacting peptide, TCR, CD4 and CD8 were obtained by analyzing the relevant 3D structures shown in Table 2. Briefly, MHC residues were considered to contact the peptide or the TCR if they were at a distance of 5.0 Å or less. The chosen contact distance was larger than the standard van der Waals distance (4.0 Å) because MHCs, peptides and TCRs are all highly variable molecules and, hence, all possible contacts cannot be represented by the limited number of available 3D structures. Contacts between MHC and peptides and MHC and TCRs were compiled after summing all contacts resulting from the analyses of individual 3D structures. Nevertheless, MHC–peptide contacts were obtained separately for the HLA-A, HLA-B, HLA-C, HLA-DP, HLA-DQ, HLA-DR molecules from the representative X-ray crystal structures (Table 2). In the case of HLA-DP, for which there is no available 3D structure, peptide contacts were obtained from the analysis of a model (DPA1∗0103xDPB1∗0201) in complex with the nonamer peptide ligand YAASSYLSL derived from human Ig lambda chain. Potential HLA class I contacts with the TCR were derived from the analysis of all available X-ray crystal structures of class I MHC–TCR complexes, including those from mouse MHCI molecules (Table 2). Likewise, HLA class II contacts with the TCR were obtained from all available class II MHC–TCR complexes. HLA class II residues contacting CD4 and HLA class I residues contacting CD8 were obtained from the relevant X-ray 3D structures (Table 2) using a distance of 4.5 Å. All contacts were subject to visual inspection using the program RASMOL.[65]

Molecular graphics and modeling

A molecular model of the HLA-DP α1β1 domain (DPA1∗0103xDPB1∗0201) with the bound nonamer peptide YAASSYLSL from human Ig λ chain was derived by homology modeling from the 3D coordinates of HLA-DR1(DRB1∗0101) (PDB: 1HXY) using the package MODELLER.[66] Sequence variability of MHC molecules was mapped onto the molecular surface or worm representations of the MHC molecules using the program GRASP[67] *via* a B-factor. A program for projecting the variability for a multiple sequence alignment to a 3D structure is available on-line at the H2PDB web server hosted by the Molecular Immunology Foundation†.

References

1. Paul, W. E. (1998). Editor of *Fundamental Immunology*, Lippincott, Williams & Wilkins, Philadelphia, PA.
2. Margulies, D. H. (1997). Interactions of TCRs with MHC–peptide complexes: a quantitative basis for mechanistic models. *Curr. Opin. Immunol.* **9**, 390–395.
3. von Boehmer, H. (1991). Positive and negative selection of the αβ T cell repertoire *in vivo*. *Curr. Opin. Immunol.* **3**, 210–215.
4. Maenaka, K. & Jones, E. Y. (1999). MHC superfamily structure and the immune system. *Curr. Opin. Struct. Biol.* **9**, 745–753.
5. Reinherz, E. L. & Schlossman, S. F. (1980). The differentiation and function of human T lymphocytes: a review. *Cell*, **19**, 821–827.
6. Madden, D. R. (1995). The three-dimensional structure of peptide–MHC complexes. *Annu. Rev. Immunol.* **13**, 587–622.
7. Stern, L. J. & Wiley, D. C. (1994). Antigen peptide binding by class I and class II histocompatibility proteins. *Structure*, **2**, 245–251.
8. MHC Sequencing Consortium (1999). Complete sequence and gene map of a human major histocompatibility complex. *Nature*, **401**, 921–923.
9. Shiina, T., Tamiya, G., Oka, A., Takishima, N., Yamagata, T., Kikkawa, E. *et al.* (1999). Molecular dynamics of MHC genesis unraveled by sequence analysis of the 1,796,938-bp HLA class I region. *Proc. Natl Acad. Sci. USA*, **96**, 13282–13287.
10. Yeager, M. & Hughes, A. L. (1999). Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. *Immunol. Rev.* **167**, 45–58.
11. Braud, V. M., Allan, D. S. & McMichael, A. J. (1999). Functions of non-classical MHC and non-MHC-encoded class I molecules. *Curr. Opin. Immunol.* **11**, 100–108.
12. Alfonso, C. & Karlsson, L. (2000). Nonclassical MHC class II molecules. *Annu. Rev. Immunol.* **18**, 113–142.
13. Robinson, J., Malik, A., Parham, P., Bodmer, J. G. & Marsh, S. G. E. (2000). IMGT/HLA database—a sequence database for the human major histocompatibility complex. *Tissue Antigens*, **55**, 280–287.
14. Parham, P. (1988). Function and polymorphism of

† http://www.mifoundation.org/Tools/entropy2pdb.html

human leukocyte antigen-A,B,C molecules. *Am. J. Med.* **85**, 2–5.

15. Hughes, A. L. & Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, **335**, 167–170.

16. Hughes, A. L. & Nei, M. (1989). Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc. Natl Acad. Sci. USA*, **86**, 958–962.

17. Hedrick, P. W., Whittam, T. S. & Parham, P. (1991). Heterozygosity at individual amino acid sites: extremely high levels for HLA-A and -B genes. *Proc. Natl Acad. Sci. USA*, **88**, 5897–5901.

18. Meyer, D. & Thomson, G. (2001). How selection shapes variation of the human major histocompatibility complex: a review. *Ann. Hum. Genet.* **65**, 1–26.

19. Marsh, S. G. E. & Bodmer, J. G. (1989). HLA-DR and -DQ epitopes and monoclonal antibody specificity. *Immunol. Today*, **10**, 305–312.

20. Weaver, W. & Shannon, C. E. (1949). *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL.

21. Kabat, E. A. (1970). Antigenic determinants and antibody complementarity. *Folia Allergol.* **17**, 425.

22. Stewart, J. J., Lee, C. Y., Ibrahim, S., Watts, P., Shlomchik, M., Weigert, M. & Litwin, S. (1997). A Shannon entropy analysis of immunoglobulin and T cell receptor. *Mol. Immunol.* **34**, 1067–1082.

23. Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987). The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature*, **329**, 512–518.

24. Brown, J. H., Jardetzky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L. & Wiley, D. C. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature*, **364**, 33–39.

25. Gao, G. F., Tormo, J., Gerth, U. C., Wyer, J. R., McMichael, A. J., Stuart, D. I. *et al.* (1997). Crystal structure of the complex between human CD8alpha(alpha) and HLA-A2. *Nature*, **387**, 630–634.

26. Kern, P. S., Teng, M.-K., Smolyar, A., Liu, J.-H., Liu, J., Hussey, R. E. *et al.* (1998). Structural basis of CD8 co-receptor function revealed by crystallographic analysis of a murine CD8αα ectodomain fragment in complex with H-2Kb. *Immunity*, **9**, 519–530.

27. Salter, R., Norment, A. M., Chen, B. P., Clayberger, C., Krensky, A. M., Littman, D. R. & Parham, P. (1989). Polymorphism in the α3 domain of HLA-A molecules affects binding to CD8. *Nature*, **338**, 345–347.

28. Martinez-Naves, E., Barber, L. D., Madrigal, J. A., Vullo, C. M., Clayberger, C., Lyu, S. C. *et al.* (1997). Interactions of HLA-B*4801 with peptide and CD8. *Tissue Antigens*, **50**, 258–264.

29. Wang, J.-H., Meijers, R., Xiong, Y., Liu, J.-H., Sakihama, T., Zhang, R. *et al.* (2001). Crystal structure of the human CD4 N-terminal two domain fragment complexed to a class II MHC molecule. *Proc. Natl Acad. Sci. USA*, **98**, 10799–10804.

30. Teng, M. K., Smolyar, A., Tse, A. G., Liu, J. H., Liu, J., Hussey, R. E. *et al.* (1998). Identification of a common docking topology with substantial variation among different TCR–peptide–MHC complexes. *Curr. Biol.* **8**, 409–412.

31. Reinherz, E. L., Tan, K., Tang, L., Kern, P., Liu, J., Xiong, Y. *et al.* (1999). The crystal structure of a T

32. Katz, D. H., Hamaoka, T. & Benacerraf, B. (1973). Cell interactions between histoincompatible T and B lymphocytes. II. Failure of physiologic cooperative interactions between T and B lymphocytes from allogeneic donor strains in humoral response to hapten-protein conjugates. *J. Expt. Med.* **137**, 1405–1418.

33. Rosenthal, A. S. & Shevach, E. M. (1973). Function of macrophages in antigen recognition by guinea pig T lymphocytes. I. Requirement for histocompatible macrophages and lymphocytes. *J. Expt. Med.* **138**, 1194–1212.

34. Zinkernagel, R. M. & Doherty, P. C. (1974). Restriction of *in vitro* T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature*, **248**, 701–702.

35. Garcia, K. C., Teyton, L. & Wilson, I. A. (1999). Structural basis of T cell recognition. *Annu. Rev. Immunol.* **17**, 369–397.

36. Wang, J.-H. & Reinherz, E. (2001). Structural basis of T cell recognition of peptides bound to MHC molecules. *Mol. Immunol.* **38**, 1039–1049.

37. Rudolph, M. G. & Wilson, I. A. (2002). The specificity of TCR/pMHC interaction. *Curr. Opin. Immunol.* **14**, 52–65.

38. Davis, M. M. & Bjorkman, P. J. (1988). T cell antigen receptor genes and T-cell recognition. *Nature*, **334**, 395–402.

39. Garcia, K. C., Degano, M., Pease, L. R., Huang, M., Peterson, P. A., Teyton, L. & Wilson, I. A. (1998). Structural basis of plasticity in T cell receptor recognition of a self peptide–MHC antigen. *Science*, **279**, 1166–1172.

40. Ding, Y. H., Smith, K. J., Garboczi, D. N., Utz, U., Biddison, W. E. & Wiley, D. C. (1998). Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids. *Immunity*, **8**, 403–411.

41. Zerrahn, J., Held, W. & Raulet, D. H. (1997). The MHC reactivity of the T cell repertoire prior to positive and negative selection. *Cell*, **88**, 627–636.

42. Merkenschlager, M., Graf, D., Lovatt, M., Bommhardt, U., Zamoyska, R. & Fisher, A. G. (1997). How many thymocytes audition for selection? *J. Expt. Med.* **186**, 1149–1158.

43. Sim, B. C., Zerva, L., Greene, M. & Gascoigne, N. R. (1996). Control of MHC restriction by TCR Valpha CDR1 and CDR2. *Science*, **273**, 963–966.

44. Sim, B. C. & Gascoigne, N. R. (1999). Reciprocal expression in CD4 or CD8 subsets of different members of the Vα 11 gene family correlates with sequence polymorphism. *J. Immunol.* **162**, 3153–3159.

45. Manning, T. C., Schlueter, C. J., Brodnicki, T. C., Parke, E. A., Speir, J. A., Garcia, K. C. *et al.* (1998). Alanine scanning mutagenesis of an alphabeta T cell receptor: mapping the energy of antigen recognition. *Immunity*, **8**, 413–425.

46. Candeias, S., Waltzinger, C., Benoist, C. & Mathis, D. (1991). The Vβ 17+ T cell repertoire: skewed J beta usage after thymic selection; dissimilar CDR3s in CD4+ *versus* CD8+ cells. *J. Expt. Med.* **174**, 989–1000.

47. Vasmatzis, G., Zhang, C., Cornette, J. L. & DeLisi, C. (1996). Computational determination of side chain specificity for pockets in class I MHC molecules. *Mol. Immunol.* **33**, 1231–1239.

48. Hennecke, J. & Wiley, D. C. (2002). Structure of a complex of the human alpha/beta T cell receptor (TCR) HA1.7, influenza hemagglutinin peptide, and

major histocompatibility complex class II molecule, HLA-DR4 (DRA∗0101 and DRB1∗0401): insight into TCR cross-restriction and alloreactivity. *J. Expt. Med.* **195**, 571–581.

49. Sherman, L. A. & Chattopadhyay, S. (1993). The molecular basis of allorecognition. *Annu. Rev. Immunol.* **11**, 385–402.
50. Benichou, G., Takizawa, P. A., Olson, C. A., McMillan, M. & Sercarz, E. E. (1992). Donor major histocompatibility complex (MHC) peptides are presented by recipient MHC molecules during graft rejection. *J. Expt. Med.* **175**, 305–308.
51. Liu, Z., Braunstein, N. S. & Suciu-Foca, N. (1992). T cell recognition of allopeptides in context of syngeneic MHC. *J. Immunol.* **148**, 35–40.
52. den Haan, J. M., Meadows, L. M., Wang, W., Pool, J., Blokland, E., Bishop, T. L. *et al.* (1998). The minor histocompatibility antigen HA-1: a diallelic gene with a single amino acid polymorphism. *Science*, **279**, 1054–1057.
53. Horowitz, J. B., Kaye, J., Katz, M. E. & Janeway, C. A. J. (1987). Ability of fixed B-lymphoma cells to present foreign protein antigen fragments and allogenic MHC molecules to a cloned helper-T-cell line. *Cell Immunol.* **109**, 429–436.
54. Shoskes, D. A. & Wood, K. J. (1994). Indirect presentation of MHC antigens in transplantation. *Immunol. Today*, **15**, 32–38.
55. Reiser, J. B., Darnault, C., Guimezanes, A., Gregoire, C., Mosser, T., Schmitt-Verhulst, A. M. *et al.* (2000). Crystal structure of a T cell receptor bound to an allogeneic MHC molecule. *Nature Immunol.* **1**, 291–297.
56. Reiser, J. B., Gregoire, C., Darnault, C., Mosser, T., Guimezanes, A., Schmitt-Verhulst, A. M. *et al.* (2002). A T cell receptor CDR3beta loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex. *Immunity*, **16**, 345–354.
57. Luz, J. G., Huang, M., Garcia, K. C., Rudolph, M. G., Apostolopoulos, V., Teyton, L. & Wilson, I. A. (2002). Structural comparison of allogeneic and syngeneic T cell receptor-peptide-major histocompatibility complex complexes: a buried alloreactive mutation subtly alters peptide presentation substantially increasing Vβ interactions. *J. Expt. Med.* **195**, 1175–1186.
58. Bevan, M. J. (1984). High determinant density may explain the phenomenon of alloreactivity. *Immunol. Today*, **5**, 128–130.
59. Grandea, A. G. & Bevan, M. J. (1992). Single-residue changes in class I major histocompatibility complex molecules stimulate responses to self peptides. *Proc. Natl Acad. Sci. USA*, **89**, 2794–2798.
60. Coppin, H. L., Carmichael, P., Lombardi, G., L'Faqihi, F. E., Salter, R., Parham, P. *et al.* (1993). Position 71 in the alpha helix of the DR beta domain is predicted to influence peptide binding and plays a central role in allorecognition. *Eur. J. Immunol.* **23**, 343–349.
61. Pullen, J. K., Hunt, H. D., Horton, R. M. & Pease, L. R. (1989). The functional significance of two amino acid polymorphisms in the antigen-presenting domain of class I MHC molecules. Molecular dissection of Kbm3. *J. Immunol.* **143**, 1674–1679.
62. Matzinger, P. & Bevan, M. J. (1977). Hypothesis: why do so many lymphocytes respond to major histocompatibility antigens? *Annu. Rev. Immunol.* **13**, 93–126.
63. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **2**, 4673–4680.
64. Schneider, T. D. & Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. *Nucl. Acids Res.* **18**, 6097–6100.
65. Sayle, R. A. & Milner-White, E. J. (1995). RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.* **20**, 374.
66. Sali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815.
67. Nicholls, A., Sharp, K. & Honig, B. (1991). Protein folding and association insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.
68. Laskowski, R. A. (2001). PDBsum: summaries and analyses of PDB structures. *Nucl. Acids Res.* **29**, 221–222.
69. Letunic, I., Goodstadt, L., Dickens, N. J., Doerks, T., Schultz, J., Mott, R. *et al.* (2002). Recent improvements to the SMART domain-based sequence annotation resourcen. *Nucl. Acids Res.* **30**, 242–244.