The impact of human leukocyte antigen (HLA) micropolymorphism on ligand specificity within the HLA-B*41 allotypic family

Christina Bade-Döding,[1,2]* Alex Theodossis,[3]* Stephanie Gras,[3] Lars Kjer-Nielsen,[2] Britta Eiz-Vesper,[1] Axel Seltsam,[4] Trevor Huyton,[1] Jamie Rossjohn,[3] James McCluskey,[2#] and Rainer Blasczyk[1#]
**joint first authors; #joint senior authors*

[1]Institute for Transfusion Medicine, Hannover Medical School, Hannover, Germany; [2]Department of Microbiology and Immunology, University of Melbourne, Parkville, Australia; [3]Protein Crystallography Unit, Department of Biochemistry and Molecular Biology, Monash University, Clayton, Australia, and [4]German Red Cross Blood Donor Service NSTOB, Institute Springe, Springe, Germany

ABSTRACT

Background
Polymorphic differences between human leukocyte antigen (HLA) molecules affect the specificity and conformation of their bound peptides and lead to differential selection of the T-cell repertoire. Mismatching during allogeneic transplantation can, therefore, lead to immunological reactions.

Design and Methods
We investigated the structure-function relationships of six members of the HLA-B*41 allelic group that differ by six polymorphic amino acids, including positions 80, 95, 97 and 114 within the antigen-binding cleft. Peptide-binding motifs for B*41:01, *41:02, *41:03, *41:04, *41:05 and *41:06 were determined by sequencing self-peptides from recombinant B*41 molecules by electrospray ionization tandem mass spectrometry. The crystal structures of HLA-B*41:03 bound to a natural 16-mer self-ligand (AEMYGSVTEHPSPSPL) and HLA-B*41:04 bound to a natural 11-mer self-ligand (HEEAVSVDRVL) were solved.

Results
Peptide analysis revealed that all B*41 alleles have an identical anchor motif at peptide position 2 (glutamic acid), but differ in their choice of C-terminal pΩ anchor (proline, valine, leucine). Additionally, B*41:04 displayed a greater preference for long peptides (>10 residues) when compared to the other B*41 allomorphs, while the longest peptide to be eluted from the allelic group (a 16mer) was obtained from B*41:03. The crystal structures of HLA-B*41:03 and HLA-B*41:04 revealed that both alleles interact in a highly conserved manner with the terminal regions of their respective ligands, while micropolymorphism-induced changes in the steric and electrostatic properties of the antigen-binding cleft account for differences in peptide repertoire and auxiliary anchoring.

Conclusions
Differences in peptide repertoire, and peptide length specificity reflect the significant functional evolution of these closely related allotypes and signal their importance in allogeneic transplantation, especially B*41:03 and B*41:04, which accommodate longer peptides, creating structurally distinct peptide-HLA complexes.

Key words: HLA-B*41, micropolymorphism, peptide motif, X-ray crystallography.

*Citation: Bade-Döding C, Theodossis A, Gras S, Kjer-Nielsen L, Eiz-Vesper B, Seltsam A, Huyton T, Rossjohn J, McCluskey J, and Blasczyk R. The impact of human leukocyte antigen (HLA) micropolymorphism on ligand specificity within the HLA-B*41 allotypic family. Haematologica 2011;96(1):110-118. doi:10.3324/haematol.2010.030924*

Funding: this work was supported by the German José Carreras Leukemia Foundation (DJCLS R05/27f) and the German Federal Ministry of Education and Research (reference number: 01EO0802).

Acknowledgments: the authors would like to thank Susanne Aufderbeck and Nicole Neumann for their excellent technical assistance and the beamline staff at the Standford Synchrotron Radiation Lightsource (SSRL) and the Argonne Advanced Proton Source (APS) for valuable assistance. The atomic coordinates and structure factors (codes 3LN4 and 3LN5) have been deposited in the Protein Data Bank Japan (http://www.pdbj.org/).

Manuscript received on July 20, 2010. Revised version arrived on September 16, 2010. Manuscript accepted on October 1, 2010.

Correspondence:
James McCluskey, Department of Microbiology & Immunology, University of Melbourne, Parkville 3010, Australia.
Phone: + 61.383445709.
Fax: +61.61 393471540.
E-mail : jamesm1@unimelb.edu.au
Rainer Blasczyk, Institute for Transfusion Medicine, Hanover Medical, School, Carl-Neuberg-Str. 1, D-30625 Hanover, Germany.
Phone: +495115326700.
Fax: +495115322079.
E-mail: blasczyk.rainer@mh-hannover.de

The online version of this article has a Supplementary Appendix.

Introduction

Polymorphism in genes that encode antigen-presenting molecules are crucial determinants of the risk of immunological reactions associated with both allogeneic hematopoietic stem cell transplantation and most solid organ transplantations. Distinct peptide motifs and peptide-binding features have been described for a variety of alleles across major allotypic groups which differ by 20-30 amino acids in the region of their antigen-binding cleft.[1] Hence differences in peptide selection can be reconciled with the nature of the individual human leukocyte antigen (HLA) allotypic polymorphism. However, only a few studies have addressed the extent to which variants of the same allelic group differ in their peptide motifs and are truly functionally distinct. Understanding the factors that make some mismatches more important than others, and appreciating the optimal match for a transplant candidate, requires a deeper knowledge of how HLA polymorphism affect peptide selection and T-cell responsiveness.

Mismatches within the HLA heavy chain which are predicted to contact a bound peptide appear to be more important in dictating cellular responses than those that are positioned outside of the peptide-binding region because they can affect the peptide binding motif.[2,3] However, the magnitude of a given polymorphism is not dependent only on its position but also on the nature of the exchanged amino acids, as well as their neighboring amino acids. Even small differences between HLA allotypes can have dramatic effects on their function and the selection criteria for identifying acceptable mismatches when no matched donor is available still remain poorly defined.

Scoring matrices, hidden Markov models, and artificial neural networks are examples of algorithms that have been successful in predicting major histocompatibility complex (MHC) peptide-binding.[3-7] However, because these algorithms are based on a limited amount of experimental peptide-binding data, prediction is only possible for a small fraction of known MHC proteins. Many groups have conducted sequencing analyses of naturally presented MHC peptides from both native membrane-bound molecules[8] and recombinant membrane-bound or secreted molecules.[9] These binding data are needed in order to identify the specific binding motifs and preferred peptide anchor residues for given alleles. They allow potential binding peptide sequences to be predicted by T-cell epitope prediction algorithms such as SYFPETHI,[10] NetMHC,[11] RANKPEP,[12] PeptideCheck[13] and BIMAS.[14] One failing of these programs is that they cannot address the presentation of peptides longer than the normal 8 to 10 mers.[15] Unusually long peptides with 12 or more residues which are naturally processed by the cellular machinery reportedly have quite high binding affinities (Bade-Doeding *et al.*, 2010, *unpublished data*)[16] and thus elicit strong T-cell responses,[17] making them clinically important in both transplantation and adoptive T-cell therapies.

The aim of the present study was to understand the structural and functional factors underpinning differences in ligand selection within a selected group of HLA-B allotypes. Peptide-binding motifs and peptide features of HLA-B*41:01, *41:02, *41:03, *41:04, *41:05 and *41:06 were investigated because all but two of their polymorphic differences are located within the peptide-binding region and were, therefore, predicted to influence the sequence and conformation of the presented peptides.

Design and Methods

Production of eukaryotic soluble recombinant HLA-B*41 molecules

mRNA of HLA-B*41:01 (exon 1 through exon 4, encoding the signal peptide and the α1-α3 domains) was amplified by reverse transcriptase polymerase chain reaction using the primers HLA-B3-TAS (5'GAGATGCGGGTCACGGCAC-3') and HLA-E4-WAS (5'CCATCTCAGGGTGAGGGGCT-3'). The polymerase chain reaction product was cloned in the eukaryotic expression vector pcDNA3.1V5/His (Invitrogen, Karlsruhe, Germany) as previously described.[18] Further HLA-B*41 expression constructs were produced by site-direct mutagenesis using the QuikChange Multi Site-Directed Mutagenesis Kit (Stratagene, Amsterdam, The Netherlands).

Transfection of expression constructs (pcDNA-B4101, pcDNA-B4102, pcDNA-B4103, pcDNA-B4104, pcDNA-B4105 and pcDNA-B4106) in the human embryonic kidney cell line HEK293 was performed by lipofection (Lipofectamine, Invitrogen).

Expression and purification of soluble human leukocyte antigens

High-expressing G418-resistant clones were individually selected 10 to 16 days after the addition of selection medium to obtain stable cell lines. Quantification of soluble HLA-B41 proteins was performed using a sandwich enzyme-linked immunosorbent assay (ELISA) in which the monoclonal antibodies anti-HLA-A-B-C W6/32 (Serotec, Düsseldorf, Germany)[19,20] or anti-V5 (Invitrogen), respectively, were employed as capture antibodies. Horseradish peroxidase-conjugated anti-β2 microglobulin monoclonal antibody (DAKO, Hamburg, Germany) served as the detection antibody. HEK293 clones with the highest expression levels were used for large-scale production in roller bottles. Soluble HLA molecules were purified from the supernatants using N-hydroxysuccimide-activated HiTrap columns (GE Healthcare, Munich, Germany) coupled with monoclonal antibody W6/32. Affinity chromatography was performed using the Äctaprime system (Amersham Pharmacia Biosciences).

Isolation and characterization of self-peptides

Peptides were eluted from soluble HLA molecules by acidic treatment and separated from the heavy chain and β2 microglobulin molecules by size exclusion. For desalting and purification, reverse-phase high performance liquid chromatography was performed using the HP 1100 (Agilent Technologies, Böblingen, Germany). Subsequently, peptides were sequenced using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF, MALDI MS/MS) (Proteomic Analyzer 4700, Applied Biosystems, Foster City, CA, USA). A database search using MASCOT[21] and BLAST[22] was performed to determine peptide sources.

Production of prokaryotic recombinant HLA-B*41- and β2 microglobulin molecules

Truncated HLA-B*41:03 and B*41:04 heavy chains (residues 1-276) and full-length β2-microglobulin molecules (residues 1-99) were expressed using an *E. coli* expression system refolded in the presence of peptide AEMYGSVTEHPSPSPL (B*41:03) or HEEAVSVDRVL (B*41:04) and purified as described previously.[23]

Crystallographic analysis

HLA-B*41:03/AEMYGSVTEHPSPSPL and HLA-B*41:04/

HEEAVSVDRVL crystals were grown at 21 °C using protein concentrations of 4.1 mg/mL and 1.1 mg/mL, respectively; the crystals were obtained by seeding from HLA-B*35:08/FPT (sharing greater than 93% sequence identity in the alpha chain)[24] in 0.1 M citrate buffer pH 5.6, PEG 4000 (14-20%), and 0.2M NH_4OAc. The crystals were cryoprotected by equilibration against mother liquor containing 30% PEG 4000 and flash-frozen in liquid nitrogen. Data sets were collected at 100 K at the Stanford Synchrotron Radiation Lightsource (SSRL) and processed with XDS and XSCALE.[25] The crystal structures were solved by molecular replacement in PHASER[26] against a HLA-B*4405 complex (sharing 94% identity in the alpha chain) (PDB entry 1SYV[27]). The resultant models were subjected to iterative cycles of refinement in REFMAC5[28] and then PHENIX,[29] followed by model building/correction in O[30] and COOT.[31] The solvent structures were built using ARP/wARP[32] and COOT. The high resolution of the B*4103/16-mer structure permitted the refinement of anisotropic displacement parameters during the final rounds of refinement in PHENIX, resulting in a distribution of anisotropy with a mean of 0.46 and standard deviation of 0.14, as determined using the PARVATI server.[33]

Structure/model validation was carried out using COOT, MOLPROBITY and the CCP4i implementation of SFCHECK.[34] The data processing and refinement statistics for the two structures are summarized in Table 1.

Table 1. Data collection and refinement statistics.

Data collection statistics	HLA-B*41: 04/11-mer	HLA-B*41: 03/16-mer
Temperature	100K	100K
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions (a,b,c) (Å)	50.95 82.12 110.21	50.77 81.99 109.81
Resolution (Å)	50-1.90 (2.00-1.90)	100-1.30 (1.35-1.30)
Total number of observations	144371 (18289)	814697 (49960)
Number of unique observations	35909 (4620)	110817 (9992)
Multiplicity	4.0 (3.9)	7.4 (5.0)
Data completeness (%)	96.5 (88.9)	97.2 (88.6)
I/σI	16.79 (4.26)	16.8 (4.2)
R_{merge}[a] (%)	7.1 (45.5)	6.6 (36.4)
Refinement statistics		
Non-hydrogen atoms		
Protein	3204	3280
Solvent	226	626
Resolution (Å)	38.48 - 1.90	32.85 - 1.30
R_{factor}[b] (%)	18.9	15.3
R_{free}[b] (%)	24.6	17.2
Rms deviations from ideality		
Bond lengths (Å)	0.005	0.005
Bond angles (°)	0.925	1.094
Ramachandran plot (%)		
Most favored region	97.88	98.65
Allowed region	2.12	1.35
Generously allowed region	0.0	0.0
B-factors (Å^2)		
Average main chain	19.5	15.1
Average side chain	25.8	18.8
Average for solvent	26.9	33.8

[a]$R_{merge} = \Sigma \mid I_{hkl} - < I_{hkl} > \mid / \Sigma I_{hkl}$; [b]$R_{factor} = \Sigma_{hkl} \mid\mid F_o \mid - \mid F_c \mid\mid / \Sigma_{hkl} \mid F_o \mid$ *for all data except ≈ 1-10%, which were used for the* R_{free} *calculation. Values in parentheses are for the bin of highest resolution (approximate interval = 0.5 Å).*

Computational analysis

The disordered central region of the 16-mer peptide in the B*41:03 structure (residues -G5-SVTEHP-S12-) was modeled using COOT and the 'model_anneal' protocol in CNSsolve.[35,36] Briefly, the disordered residues were built into the model and their stereochemistry regularized using COOT, resulting in two Ramachandran outliers as determined by the MOLPROBITY server.[37] The model was then subjected to torsion angle dynamics combined with simulated annealing from 5000 K, in which all residues were fixed except for Y_4-S_{14}. Of five trials carried out at different initial velocities, one model was obtained in which the Ramachandran outliers were reduced to one (His_{10}).

Protonation equilibrium (pK) constants were calculated using the H++ server (*http://biophysics.cs.vt.edu/H++*).[38] The input models were derived by removing solvent molecules and alternative conformers from the coordinates of the B*41:03/16mer and B*41:04/11mer structures. Separate runs were carried out in the presence and absence of the peptide coordinates. In the case of B*41:03, the modeled coordinates for the full length peptide were used (see above). The following physical conditions ε_{solute} = 10, $\varepsilon_{solvent}$ = 80, *salt* = 0.15 M, and the Poison-Boltzmann method were used. An intermediate value for the internal dielectric was chosen based on the calculated desolvation penalties for the residues of interest, which generally had an absolute value of 1-3p*K* units.

Results

Subtle differences in peptide-binding motifs for HLA-B*41 subtypes

Individual peptide sequences derived from the six B*41 allotypes (B*41:01 to B*41:06) determined by mass spectrometry are shown in *Online Supplementary Table S1*. Sixteen ligands were eluted from B*41:01, the majority of which were nonameric peptides (Table 2). Peptide anchors were determined to be Glu at position p2 and Val/Pro at position pΩ based on the frequency of residues in this position (Table 3). Twenty-one ligands of B*41:02 were isolated, among which octa-, nona- and decameric peptides were frequent (Table 2). Among the B*41:02-derived peptides, Glu was found to be the p2 anchor while Leu was the pΩ anchor motif (Table 3). The same anchor motifs were identified from the 85 eluted B*41:03-bound ligands (Table 3). The majority of these peptides were nonamers and decamers (Table 2). Interestingly, B*41:03 yielded no octamers and had the longest peptide eluted from any of these related allotypes (a 16-mer). In addition, 52 ligands were identified from B*41:04, about half of which were 11 to 15 amino acids in length (Table 2). Their anchor motifs were the same as for B*41:02 and B*41:03 (Table 3); an auxiliary anchor (Glu) was detected at position p3 of the B*41:04-derived peptides (Table 3). In the case of B*41:05, we detected 41 ligands, which were anchored predominantly by Glu at p2 and by Leu, Val and Pro at pΩ (Table 3). Eighteen of these ligands (nearly 44%) were non-canonical in length (11 to 13 amino acids) (Table 2). The tendency of B*41:05 to tolerate more promiscuous peptide anchoring at pΩ compared with B*41:01 appears to be associated with polymorphic position 80. Modeling studies previously suggested that Lys_{80} (unique in B*41:05) has the effect of relaxing the binding properties of pocket F, by extending the cavity that accommodates the C-terminus of the peptide.[39] The 22 ligands eluted from B*41:06

were up to 14 amino acids in length, although more than 40% were octamers (Table 2). Their anchor motifs were Glu at p2 and Val and Pro at pΩ (Table 3).

Shared peptide specificities within the B*41 allotypic family

Peptide sequencing showed that the six B*41 subtypes investigated in this study bind overlapping sets of peptides (*Online Supplementary Table S2*). As the B*41 alleles were all expressed in the same cell type (HEK293 cells), their source of potential ligands was identical and the subtypes acquired a shared set of intracellularly available peptides. Peptides detected in more than one B*41 subtype are listed in *Online Supplementary Table S2*. For example, the nonamer KEGKPPISV (unnamed protein product, human) and the decamer YEEGPGKNLP (cytochrome C oxidase VIIc) were eluted from both B*41:01 and B*41:02, while the decamer IEVDGKQVEL (Rho-related GTP-binding protein) was found in three B*41 variants (B*41:02, 41:03 and 41:04).

Notably, the shared peptides do not conform rigorously to the general anchor motifs of the B*41 specificities.

Differential selection of peptides

Some of the B*41-derived ligands were derived from the same protein and had overlapping sequences (*Online Supplementary Table S3*). These peptides differed both in their length (8 to 14 amino acids) and in their primary anchor residues. The shorter peptide versions were N- and/or C-terminally truncated, and related peptides were detected in distinct B*41 alleles.

A nonameric peptide derived from protein C20orf24, ^{7}KEEPPQPQL^{15}, was eluted from B*41:02 whereas B*41:04 bound the related 14-mer peptide ^{7}KEEPPQPQLANGAL^{20}. Both peptides feature the binding characteristics of B*41:02 and B*41:04 (p2/Glu, pΩ/Leu). Elution of the 14-mer from B*41:04 supports the finding that this allotype has a greater binding efficiency for unusually long ligands than have the other B41 subtypes investigated in this study.

Apolipoprotein A-II yielded FEKTQEEL, an octamer presented by B*41:02, and FEKTQEELTPFF, a 12-mer presented by B*41:03 suggesting differential processing of this polypeptide. Interestingly, the 12-mer has a Phe residue at position pΩ not generally found in B*41:02 and B*41:03 peptides which have a pΩ/Leu peptide-binding motif.

One 13-mer peptide detected bound to B*41:03 (HERFPFEIVKMEF, KIAA0820 protein) has a pΩ/Phe motif and another 13-mer (AEAGAGSATEFQF) eluted from B*41:04 (40S ribosomal protein S10) exhibited pΩ/Phe. Further, differential processing of peptidyl-prolyl cis-trans isomerase A resulted in the presentation of an octamer and an 11-mer. The 11-mer, ^{79}GEKFEDENFIL^{89} (peptidylprolyl cis-trans isomerase, PPCTI), was eluted from HLA-B*41:02, B*41:03 and B*41:04, whereas a truncated octameric product of this ligand, ^{82}FEDENFIL^{89} (PPCTI), was also eluted from the B*41:04 subtype.

The HLA-B*41:03-AEMYGSVTEHPSPSPL 16-mer peptide bulges from the antigen-binding cleft

Of particular interest was the observation that within the B*41 group the longest peptide (a 16-mer) was eluted from B*41:03, despite the fact that B*41:04 displays a statistically greater preference for long peptides (>10 amino acids; Table 2). We sought to elucidate the structures of the B*41:03 and B*41:04 allotypes.

The structure of the B*41:03-AEMYGSVTEHPSPSPL (16-mer) complex was solved to 1.3 Å in the $P2_12_12_1$ space

Table 2. Statistical analysis of the length of HLA-B*41-derived ligands.

Allele (N) Total number of peptides		B*41:01 N= 16	B*41:02 N= 21	B*41:03 N= 85	B*41:04 N= 52	B*41:05 N= 41	B*41:06 N= 22
Length of ligands (amino acids)	16			1			
	15				2		
	14				5		1
	13			1	5	4	1
	12			2	2	4	2
	11		1	2	10	10	3
	10	4	4	39	4	5	3
	9	11	10	40	13	10	3
	8	1	6		11	8	9
Average		9.1875	9	9.623529	10.46154	10.09756	9.681818
SD		0.543906	0.83666	0.755558	2.164432	1.609423	1.86155
CI		0.266509	0.357839	0.160623	0.588289	0.492636	0.777878

*The length (8 to 16 residues) and number of eluted peptides specific for each of the individual B*41 alleles are represented. The overall average length (amino acids) of the derived peptides for the whole group is 9.810127 residues with a standard deviation (SD) of 1.499231 and confidence interval (CI) of 0.190872. The B*41:04-derived ligands had a significantly higher total SD and average peptide length (10.46), which was outside the confidence interval of the group as a whole.*

Table 3. Polymorphic positions and corresponding peptide anchors within in the B*41 group.

Subtype	Peptide anchor position*					Heavy chain position					
	1	2	3	.....	Ω	80	95	97	(103)	114	(180)
B*41:01		E		.....	V, P	Asn	Trp	Arg	(Val)	Asn	(Glu)
B*41:02		E		.....	L	-	Leu	Ser	-	-	-
B*41:03		E		.....	L	-	Leu	-	-	-	-
B*41:04		E	E	.....	L	-	Leu	Ser	(Leu)	Asp	-
B*41:05		E		.....	L, V, P	Lys	-	-	-	-	-
B*41:06		E		.....	V, P	-	-	-	-	-	(Gln)

*The left column shows the subtype-specific peptide motifs determined in this study. The right column features the polymorphic residues distinguishing the B*41 variants, dashes indicate identity with the reference B*41:01. Residues located in outer loops and thus not responsible for peptide binding are given in brackets. * Primary anchors were defined by a 100% increase in pool sequencing signals and by their exclusive occurrence in natural ligands. Auxiliary anchors were defined by a greater than 50% increase in pool sequencing signals and by virtue of their occurrence in a significant number of ligands.*

group, with a single heterodimer in the asymmetric unit. The N- and C-terminal residues of the 16-mer were located within the antigen-binding cleft, while the central region of the peptide was found to be disordered (*Online Supplementary Figure S1A*). Although the structural database (PDB)[40] includes several entries of peptides as long as 14 residues in complex with MHC class I molecules (PDB entry 1XH3), these structures are characterized in some cases by significant disorder [(1XH3,[41] 2FYY,[42] 2NW3[43]] in the central bulged region of the peptide, consistent with a high degree of mobility. This was true of the B*41:03-AEMYGSVTEHPSPSPL complex where eight residues of the peptide, namely pAla$_1$-pMet$_3$ and pSer$_{14}$-pLeu$_{16}$ appear well ordered, while pTyr$_4$ - pPro$_{13}$ are partially, or completely, disordered. In their observed conformations, the side chains of pAla$_1$, pTyr$_4$, pPro$_{13}$ and pPro$_{15}$ point out of the antigen-binding cleft, whereas those of pGlu$_2$ and pLeu$_{16}$ are buried deep within the cleft, and pMet$_3$ and pSer$_{14}$ are oriented towards the α2 helix (Figure 1A). pAla$_1$, pGlu$_2$, pMet$_3$ and pLeu$_{16}$ form the majority of direct contacts with the HLA heavy chain (*Online Supplementary Table S4*). More specifically, p1 of the peptide forms hydrogen bond interactions with the HLA Tyr$_7$, Tyr$_{159}$ and Tyr$_{171}$ (Figure 2A). Position p2 of the peptide forms a hydrogen bond with Arg$_{62}$ and Glu$_{63}$ via its main chain and with Tyr$_{99}$ via its side chain (Figure 2C). The residue also forms potential ionic interactions with His$_9$ and Lys$_{45}$ and significant van der Waals (VDW) contacts with Tyr$_7$. p3 forms a hydrogen bond with Tyr$_{99}$ via its main chain and with Asp$_{156}$ via its side chain while also sharing an extensive VDW surface with Tyr$_{159}$ (Figure 2E). The side chain of p3 is observed in two distinct conformations in the structure and in one of these, the residue also forms a water-mediated hydrogen bond with Arg$_{97}$. p15 is in contact with Thr$_{73}$, Glu$_{76}$, and Ser$_{77}$ via VDW interactions and forms a hydrogen bond with Trp$_{147}$ (Figure 2G). The main chain of p16 forms hydrogen bond interactions with Ser$_{77}$, Asn$_{80}$, Tyr$_{84}$ and Thr$_{143}$ as well a salt bridge with Lys$_{146}$. In addition, its side chain is buried within a hydrophobic pocket, where it makes VDW contacts with a number of residues, including Leu$_{95}$, Tyr$_{116}$ and Trp$_{147}$ (Figure 2I).

*Crystal structure of HLA-B*41:04/HEEAVSVDRVL*

The structure of the B*41:04/HEEAVSVDRVL (11-mer) complex was solved to a resolution of 1.9 Å in the P2$_1$2$_1$2$_1$ space group, with a single heterodimer in the asymmetric unit. The 11-mer peptide was clearly bound in the antigen-binding cleft of the HLA heavy chain in a partially extended conformation (*Online Supplementary Figure 1B*), and a nearly 3$_{10}$-helical turn formed between positions Val$_7$ and Arg$_9$ (DSSP assignment[44]). All residues of the peptide appear structurally ordered and conformationally unambiguous. Briefly, the side chains of pHis1, pAla$_4$, pSer$_6$-pAsp$_8$ and pVal$_{10}$ point out of the antigen-binding cleft. pGlu$_2$, pVal$_5$ and pLeu$_{11}$ point down towards the floor of the cleft, while the side chains of pGlu$_3$ and pArg$_9$ are oriented towards the heavy chain's α$_2$ helix (Figure 1B). Furthermore, positions pHis$_1$-pGlu$_3$ and pArg$_9$-pLeu$_{11}$ interact predominantly with the HLA heavy chain, while pAla$_4$ and pSer$_6$-pAsp$_8$ are prominently surface exposed and expected to be important in T-cell receptor recognition.

Online Supplementary Table S4 contains an extensive list of contacts for each peptide position. The most important peptide-HLA interactions are described below. p1 forms hydrogen bond interactions with Tyr$_7$, Tyr$_{159}$ and Tyr$_{171}$ via its main chain, while its side chain shares a significant VDW surface with Trp$_{167}$ (Figure 2B). p2 forms hydrogen bond interactions via its backbone and side chain with Glu$_{63}$ and Tyr$_{99}$, respectively, while also forming potential ionic interactions with His$_9$ and Lys$_{45}$, and VDW contacts with Tyr$_7$ (Figure 2D). The main chain of p3 forms a hydrogen bond with Tyr$_{99}$ and interacts with Asp$_{156}$ via its side chain. P3 also shares an extensive VDW surface with Tyr$_{159}$ (Figure 2F). The side chain of p8 is involved in a hydrogen bond with Gln$_{155}$, while the side chain of p9 interacts with two additional residues, Asp$_{114}$ and Asp$_{156}$. p10 forms a hydrogen bond with Trp$_{147}$ via its main chain and VDW contacts with Glu$_{76}$ and Ser$_{77}$ via its side chain (Figure 2H). Finally, the backbone of p11 forms hydrogen bond interactions with Ser$_{77}$, Asn$_{80}$, Tyr$_{84}$ and Thr$_{143}$ as well as a salt bridge with Lys$_{146}$ (Figure 2J). The side chain of p11 is buried within a hydrophobic pocket and forms VDW contacts with a number of residues, including Leu$_{95}$, Tyr$_{116}$,

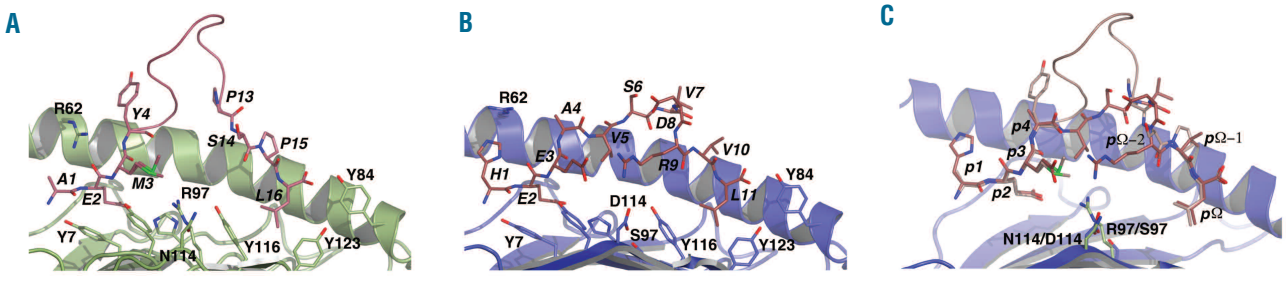

Figure 1. The crystal structures of B*41:03/AEMYGSVTEHPSPSPL and B*41:04/HEEAVSVDRVL. View of the antigen-binding cleft in (A) the B*41:03/16mer and (B) the B*41:04/11mer complexes, as well as a superposition of the B*41:03/16mer and B*41:04/11mer structures using the Cα atoms of heavy chain residues 2-181 (C). The α-1/α-2 domains of the two alleles superimpose very well with an rms deviation of 0.33 Å. Therefore, only that of B*41:04 is presented in 1C. In each case the MHC is presented in cartoon format (B*41:03, green; B*41:04, blue) and the α-2 helix of the heavy chain has been removed for clarity. The peptide residues are drawn as sticks (16mer, light red; 11mer dark red), as are selected side chains of the MHC. Peptide residues are labeled in italics. The polymorphic residues distinguishing the two alleles are also shown, highlighting the size/charge differences associated with the R97S/N114D substitutions. In the case of the B*41:03/16mer complex (A, C), the central disordered region of the peptide was modeled and is presented here in ribbon format. The superposition (C) demonstrates that the N- and C-terminal regions of the two peptides adopt equivalent conformations, while the central region of the 16mer is expected to diverge significantly in conformation with respect to the 11mer. Moreover, the superposition also helps to highlight the significant difference in surface area that each peptide might present to T-cell receptor. Based on the modeled central region, the 16mer peptide is expected to have an exposed surface that is 650 Å2 greater than that of the 11mer when bound to one of the B*41 alleles (determined using the CCP4i implementation of AREAIMOL.[55,56]

Tyr123 and Trp147.

In addition to these interactions, the B*41:04/HEEAVSVDRVL (11-mer) complex is characterized by a large number of contacts between residues of the peptide, which constrain its conformation within the antigen-binding cleft (type III constraints[45]). The side chain of p1 forms a hydrogen bond with the main chain carbonyl of p2; in addition, the side chain of p5 shares an extensive VDW surface with p3 and p9. Moreover, positions p6 to p10 display a characteristic 3_{10}-helix main chain hydrogen-bonding pattern ($i \rightarrow i$-3), while additional hydrogen bonds and VDW contacts are observed between the main chain and side chain groups of p6-p8. Of particular interest, however, are p3 and p9, the side chains of which form a network of contacts with each other and with the side chains of Asp114 and Asp156. More specifically, the carboxylate of Asp156 is within interacting distance of the equivalent groups on Asp114 and pGlu3, and Asp156 together with pGlu3 are in a position to form a salt bridge with the guanidinium group of pArg9. The latter is, therefore, particularly important in stabilizing this network and thus the pHLA complex by providing a countercharge to the two pairs of proximal acidic residues.

Impact of human leukocyte antigen polymorphism on the B*41:03/16-mer and B*41:04/11-mer complexes

B*41:03 and B*41:04 share a highly conserved residue configuration within their α1/α2 domains. Residues 2-181 of the two HLA heavy chains display main chain and all atom rms deviations of 0.32 Å and 0.57 Å, respectively, while the largest conformational deviations occur outside the antigen-binding cleft. Despite their overall similarity, however, the antigen-binding clefts of B*41:03 and B*41:04 display two significant differences, both of which arise due to the polymorphism in these alleles. Firstly, the arginine to serine substitution at position 97 results in an antigen-binding cleft that is larger by 140 $Å^3$ in the case of B*41:04 (molecular volume calculated using the CASTp server,[46] Figure 1C). Furthermore, the combination of Ser97Arg and Asn114Asp substitutions give rises to a sig-

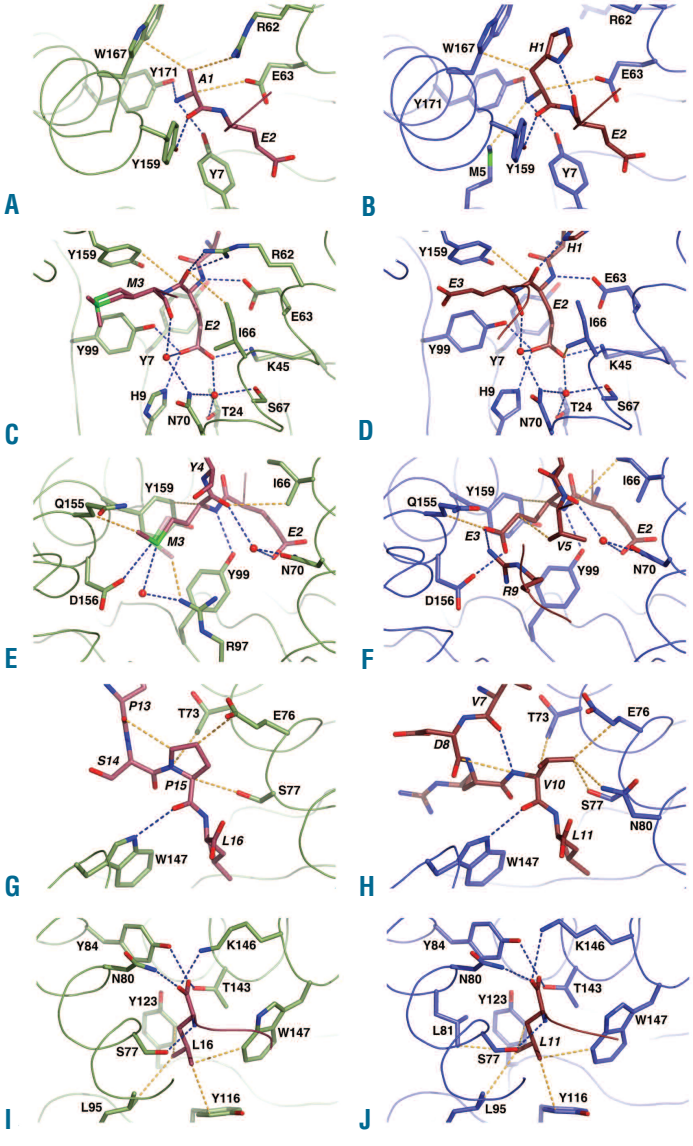

Figure 2. Peptides bound to B*41:03 and B*41:04. Peptides bound to B*41:03 and B*41:04 display similarities at positions p1, p3 and pΩ-1 by occupying equivalent positions and interacting in a similar manner with the heavy chain. Views of the individual antigen-binding cleft pockets in the B*41:03/16mer and B*41:04/11mer structures. In all figures the MHC are presented in ribbon format, with key residue side chains shown as sticks. The peptides are presented in a combination of ribbon and stick format. B*41:03 is illustrated in green and B*41:04 in blue, while the 16mer and 11mer peptides are colored light and dark red, respectively. Hydrogen bond and salt bridge interactions are shown as blue dashes, while selected VDW contacts are shown as orange dashes. Peptide residues are labeled in italics. (A, B) Pocket A composition of B*41 variants. p1 interactions in the B*41:03/16-mer (A) and B*41:04/11-mer structures (B). In both B*41 molecules, the residues contacting p1 of the peptide are conserved at positions 7, 63, 159, 167 and 171 (pocket A). (C, D) Pocket B composition of B*41 variants. p2 interactions in the B*41:03/16-mer (C) and B*41:04/11-mer structures (D). Positions 7, 9, 24, 45, 63, 66, 67, 99 and 159 are in contact with p2 of the peptide. p2 of B*41:03 is also anchored by residue 62. Arg62 is not part of pocket B in B*41:04 (D) and it is guanidinium group appears disordered, highlighted in (D) using a partially transparent stick representation. (E, F) p3 of the peptides is in contact with R97 of B*41:03 but not with S97of B*41:04. p3 interactions in the B*41:03/16-mer (E) and B*41:04/11-mer structures (F). Positions 66, 99, 155, 156 and 159 are in contact with p3 of the peptide. The polymorphic residue 97 distinguishing B*41:03 and B*41:04 does not anchor p3 of the peptide bound to B*41:04 (F), but is part of pocket D for B*41:03 (E). The side chain of p3 is observed in two distinct conformations in the B*41:03/11mer structure (E), each of which was refined at 50% occupancy. The two conformers interact alternately with either position 97 or 156 of th MHC, and are distinguished in (E) using a partially transparent stick representation for one, but not the other. (G, H): Contacts between the heavy chain and position pΩ-1 of the bound peptides. pΩ-1 interactions in the B*41:03/16-mer (G) and B*41:04/11-mer structures (H). p15 of the ligand bound to B*41:03 and p10 of the ligand bound to B*41:04 are in contact with conserved HLA heavy chain residues 73, 76, 77 and 147, which anchor peptide positions pΩ-1 in both subtypes. (I, J) Position 97 and 114 are not part of pocket F in B41 variants. pΩ interactions in the B*41:03/16-mer (I) and B*41:04/11-mer structures (J). This figure shows the composition of residues contacting the C-terminal position of the peptide, namely L16 for B*41:03 and L11 for B*41:04. Positions 77, 80, 84, 95, 116, 123 143, 146 and 147 are in contact with the pΩ position of the peptide. B*41:03 and B*41:04 share the same F pocket composition. Residues 97 and 114 are not part of the F pocket of B*41 subtypes. For this reason, mismatches at these positions will not alter the C-terminal binding motif in B*41 subtypes.

nificantly different electrostatic surface within the antigen-binding clefts of each allele (*Online Supplementary Figure S2*). Therefore, while B*41:04 displays an overall negative charge within the central region of the cleft, that of B*41:03 is clearly more electropositive in character.

The 11-mer and 16-mer peptides display significant differences in mobility of their crystal structures, indicating that the central regions of these two peptides interact very differently with their respective HLA. By contrast, their N- and C-terminal regions adopt equivalent conformations and interact with similar sets of heavy chain residues (Figures 1 and 2; *Online Supplementary Table S4*). This is particularly evident at the primary anchor positions (p2 and pΩ), which are conserved in the two peptides (Figure 2C, D, I, J). The only difference at either of these positions relates to Arg$_{62}$, which is unable to interact with the main chain of p2 in the B*41:04/11-mer complex due to the presence of the histidine side chain at p1 of the peptide. Instead, the side chain of Arg$_{62}$ appears partially disordered in this structure. The two structures also display striking similarities at positions p1, p3 and pΩ-1 of the peptides (Figure 2A, B, E, F, G, H), and although not conserved, these residues occupy equivalent positions and interact in a similar manner with the heavy chain.

Of particular interest is a network of interactions involving position p3 of the peptides and the two polymorphic heavy chain positions (97 and 114), as it appears that B*41:03 and B*41:04 compensate for their polymorphic differences in order to maintain this network. Thus, as part of this network, the side chain of Asp156 maintains an interaction with that of residue 114 in both structures, despite the unfavorable substitution of an asparagine for an aspartate at position 114 in B*41:04 (Figure 3).

Conversely, the side chain of residue 97 does not participate in the network of interactions in B*41:04, but is incorporated into this network in B*41:03, in which allele it is substituted from a serine to an arginine, thereby providing a positive charge within pocket D to balance the negative charge of Asp$_{156}$.

Surprisingly, B*41:04 does not possess a countercharge for the aspartic acids at 114 and 156 within pocket D (corresponds with p3 of a given peptide). In the case of the B*41:04/11-mer complex it appears that this charge imbalance is compensated for by the peptide, in the form of pArg$_9$. Therefore, in addition to stabilizing the peptide conformation through intra-peptide contacts with pGlu$_3$ and pVal$_5$, pArg$_9$ may also contribute to the greater stability of the 11-mer complex by participating in the network of interactions between Asp$_{156}$, Asp$_{114}$ and the peptide's auxiliary anchor (pGlu$_3$). However, based on the peptide elution data (*Online Supplementary Table S1*) there is no evidence to suggest that such peptide contributions constitute a requirement for binding to B*41:04.

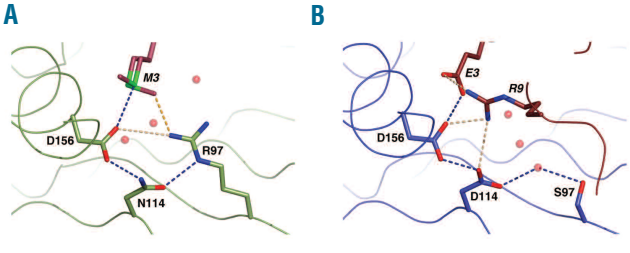

Figure 3. A conserved network of interactions in the B*41:03/16mer and B*41:04/11mer complexes. Detail of the antigen-binding cleft in (A) the B*41:03/16mer and (B) the B*41:04/11mer structures, showing a conserved network of interactions involving position p3, and the heavy chain positions 114 and 156. In each case the MHC and peptide are shown in ribbon format, with key residue side chains drawn as sticks. Key interactions are represented by dashes. Hydrogen bonds are shown in blue, VDW contacts in orange, and potential salt bridge interactions in 'wheat'. The network is maintained in both alleles, despite the unfavorable Asn114Asp in B*41:04. Ordered water molecules within interacting distance of the charged groups are shown as semi-transparent spheres.

The protonation equilibrium (pK) constants of titratable residues were calculated computationally using the atomic coordinates of the two B*41 high resolution structures. The results demonstrate that Arg$_{97}$ in B*41:03 not only provides a countercharge to the acidic Asp$_{156}$, but actively promotes its negatively charged state (*Online Supplementary Table S5*). By contrast, the R97S/N114D double substitution in B*41:04 creates a local environment for Asp$_{156}$ which strongly favors the protonated state at physiological pH. Given the close proximity between the carboxylate groups of Asp$_{156}$ and pGlu$_3$ observed in the B*41:04/11mer complex, the absence of a negative charge on Asp$_{156}$ is expected to significantly reduce the repulsive forces between the two groups.

Binding of pGlu$_3$ in a similar manner to B*41:03, on the other hand, is expected to be less favorable, particularly given that the crystal structure suggests only a weak interaction between Asp$_{156}$ and Arg$_{97}$.[47] Furthermore, the N114D substitution (a feature unique to the B*41:04 allele) was found to be essential to the observed perturbation in the protonation equilibrium constant of Asp$_{156}$. Thus, by using the high resolution crystallographic data in conjunction with powerful pK calculation algorithms we were able to obtain compelling evidence that explains how the polymorphism in B*41 gives rise to an environment within the antigen-binding cleft of the B*41:04 allele which uniquely favors the pGlu$_3$ auxiliary anchor.

Interestingly, the pK predictions also suggest an effect of polymorphism on the nature of the interactions involving the pGlu$_2$ primary anchor. While in B*41:03 the local environment of His$_9$ strongly favors the deprotonated state, the R97S substitution in B*41:04 gives rise to a pK consistent with His9 being more positively charged at physiological pH. Consequently, in B*41:04 the pGlu$_2$-His$_9$ interaction is likely to be more ionic in character. Given that Lys$_{45}$ (the only other residue able to form a salt bridge with pGlu2) is expected to form an ion pair with Glu$_{63}$, the potential for pGlu$_2$ to ion pair with pHis9 may have a significant effect on the nature of the peptide anchoring.

Discussion

In this study we investigated endogenous peptides of HLA-B*41:01, B*41:02, B*41:03, B*41:04, B*41:05 and B*41:06 molecules. Systematic characterization of the ligands derived from the B*41 allotypes demonstrated that they are more conserved at their p2 than at their C-terminal anchors. The peptide anchor motif for the B*41 group exhibits a striking preference for Glu at the p2 position of the ligands. Three distinct pΩ motifs were identified by mass spectrometry: Val/Pro (B*41:01 and B*41:06), Leu (B*41:02, B*41:03 and B*41:04), and Leu/Val/Pro (B*41:05) (Table 3).

The polymorphic residues between the B*41 variants

are 80, 95, 97, 103, 114 and 180 (Table 3). Amino acid residues 103 and 180 are located in outer loop positions, and are not, therefore, directly involved in peptide binding. By definition, residues 80 and 95 are part of the F pocket and thus critical for binding at the pΩ position of the peptide. According to the pocket definition of Chelvanayagam,[1] residues 97 and 114 within the center of the peptide-binding region are designated as part of four pockets (C, D, E and F) and are thus predicted to contact bound peptides at positions 3, 6, 7 and pΩ when considering canonical peptides 8 to 10 amino acids in length.

Based on the peptide-binding analysis, we were able to divide the B*41 molecules into two specificity subsets, B*41:01/B*41:05-06 (Trp95) and B*41:02-04 (Leu95), in terms of pΩ specificity. The presence of Leu95 in the peptide-binding region was associated with Leu at position pΩ of bound peptides, whereas Trp95 was associated with Val/Pro in the pΩ motifs.[39] While position 97 is presumed to also affect pΩ preference, our peptide binding analysis did not identify a link between pΩ specificity and mismatches at amino acid 97 in the six B*41 variants. This is highlighted by B*41:02 and B*41:03, which differ by a single amino acid residue, Ser97Arg, yet share the same peptide-binding anchor motif at pΩ.

Our peptide elution data derived from B*41:04 also highlighted a unique preference for a Glu auxiliary anchor at peptide position p3, which correlates with the presence of an Asp at position 114 within the peptide-binding region of B*41:04 (in contrast to B*41:01/41:02/41:03/41:05/41:06-Asn114). An effect of amino acid position 114 on pocket D (which corresponds with position 3 of a given peptide) was predicted and has also been observed for the B7 group.[48]

While a total of 237 peptides, varying in length between 8 and 16 amino acids, were eluted from the B*41 group, their distribution varies significantly between the six alleles. Predominantly peptides of canonical length (8-10 amino acids) were eluted from B*41:01 and 02, while no octamers were eluted from B*41:03. Also, although multiple peptides of non-canonical length (≥11 amino acids) were eluted from B*41:03-06, B*41:04 exhibited a statistically significant preference for ligands of non-canonical length over the other alleles (Table 2). The B*41:04-derived ligands also showed more divergent peptide-binding characteristics, namely, an average peptide length of 10.5 amino acids. B*41:03 yielded the only 16-mer self-peptide.

Several long T-cell epitopes for HLA class I molecules have been predicted by extending known shorter epitopes or by screening peptide libraries, including an overlapping 16-mer,[49,50] and several 15-mer peptides.[51,52] MHC class I molecules bind peptides with a length of up to 25 amino acids.[53]

B*41:03 yielded 85 ligands. To explain these findings, we solved the X-ray crystal structures of the B*4103/AEMYGSVTEHPSPSPL (16-mer self-peptide) and B*4104/HEEAVSVDRVL (11-mer self-peptide) complexes. These structures show that B*41:03 and B*41:04, which share conserved primary anchor motifs, interact in a highly conserved manner with the terminal regions of two distinct peptides. However, the two structures also reveal certain roles in peptide binding for the polymorphic residues 97 and 114 that could not be predicted based on the knowledge of peptide feature information alone.

Contrary to previous pocket definitions,[1,54] the side chain of Arg_{97} extends out of the floor of the antigen-binding cleft of B*41:03 and contacts the p3 residue of the peptide, thereby forming part of pocket D. Moreover, residue 114 in the B*41:04/11-mer structure contacts the p9 residue of the peptide directly, while only interacting with the p3 residue indirectly via p9 and Asp_{156}.

Micropolymorphism at positions 97 and 114 affects not only the size of the antigen-binding cleft in the region of pocket D, but also the pocket's electrostatic properties. Furthermore, combined with computational analysis the structural data provide compelling evidence that the auxiliary anchor motif at p3, unique to B*41:04, is brought about by the modulation of the neighboring Asp_{156}'s protonation state by the polymorphic positions 97 and 114. This auxiliary anchor is anticipated to confer greater stability to peptide complexes formed with B*41:04, while also permitting more diverse binding modes. We are thus able to correlate micropolymorphism with the significant differences in peptide repertoire observed between B*41:03 and B*41:04, as well as the apparent ability of B*41:04 to accommodate a broader range of peptide lengths compared to the other B*41 alleles.

Overall, our analysis has resulted in a detailed characterization of the binding characteristics of the B*41 variants.

Authorship and Disclosures

The information provided by the authors about contributions from persons listed as authors and in acknowledgments is available with the full text of this paper at www.haematologica.org.

Financial and other disclosures provided by the authors using the ICMJE (www.icmje.org) Uniform Format for Disclosure of Competing Interests are also available at www.haematologica.org.

References

1. Chelvanayagam G. A roadmap for HLA-A, HLA-B, and HLA-C peptide binding specificities. Immunogenetics. 1996;45(1):15-26.
2. Vyas JM, Van der Veen AG, Ploegh HL. The known unknowns of antigen processing and presentation. Nat Rev Immunol. 2008;8(8):607-18.
3. Kim Y, Sidney J, Pinilla C, Sette A, Peters B. Derivation of an amino acid similarity matrix for peptide: MHC binding and its application as a Bayesian prior. BMC Bioinformatics. 2009;10:394.
4. Zhao Y, Gran B, Pinilla C, Markovic-Plese S, Hemmer B, Tzou A, et al. Combinatorial peptide libraries and biometric score matrices permit the quantitative analysis of specific and degenerate interactions between clonotypic TCR and MHC peptide ligands. J Immunol. 2001;167(4):2130-41.
5. Zhang C, Bickis MG, Wu FX, Kusalik AJ. Optimally-connected hidden Markov models for predicting MHC-binding peptides. J Bioinform Comput Biol. 2006;4(5):959-80.
6. Mamitsuka H. Predicting peptides that bind to MHC molecules using supervised learning of hidden Markov models. Proteins. 1998;33(4):460-74.
7. Nielsen M, Lund O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinformatics. 2009;10:296.
8. Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee HG. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. Nature. 1991;351(6324):290-6.
9. Bade-Doeding C, Eiz-Vesper B, Figueiredo C, Seltsam A, Elsner HA, Blasczyk R. Peptide-binding motif of HLA-A*6603. Immunogenetics. 2005;56(10):769-72.
10. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics. 1999;50(3-4):213-9.
11. Lundegaard C, Lamberth K, Harndahl M,

Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. Nucleic Acids Res. 2008;36(Web Server issue):W509-12.
12. Reche PA, Glutting JP, Zhang H, Reinherz EL. Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. Immunogenetics. 2004;56(6):405-19.
13. DeLuca DS, Blasczyk R. Implementing the modular MHC model for predicting peptide binding. Methods Mol Biol. 2007;409: 261-71.
14. Parker KC, Bednarek MA, Coligan JE. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. J Immunol. 1994;152(1):163-75.
15. Burrows SR, Rossjohn J, McCluskey J. Have we cut ourselves too short in mapping CTL epitopes? Trends Immunol. 2006;27(1):11-6.
16. Tynan FE, Burrows SR, Buckle AM, Clements CS, Borg NA, Miles JJ, et al. T cell receptor recognition of a 'super-bulged' major histocompatibility complex class I-bound peptide. Nat Immunol. 2005;6(11): 1114-22.
17. Green KJ, Miles JJ, Tellam J, van Zuylen WJ, Connolly G, Burrows SR. Potent T cell response to a class I-binding 13-mer viral epitope and the influence of HLA micropolymorphism in controlling epitope length. Eur J Immunol. 2004;34(9):2510-9.
18. Bade-Doeding C, Elsner HA, Eiz-Vesper B, Seltsam A, Holtkamp U, Blasczyk R. A single amino-acid polymorphism in pocket A of HLA-A*6602 alters the auxiliary anchors compared with HLA-A*6601 ligands. Immunogenetics. 2004;56(2):83-8.
19. Barnstable CJ, Bodmer WF, Brown G, Galfre G, Milstein C, Williams AF, et al. Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens-new tools for genetic analysis. Cell. 1978;14(1):9-20.
20. Brodsky FM, Parham P, Barnstable CJ, Crumpton MJ, Bodmer WF. Monoclonal antibodies for analysis of the HLA system. Immunol Rev. 1979;47:3-61.
21. Hirosawa M, Hoshida M, Ishikawa M, Toya T. MASCOT: multiple alignment system for protein sequences based on three-way dynamic programming. Comput Appl Biosci. 1993;9(2):161-7.
22. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10.
23. Macdonald W, Williams DS, Clements CS, Gorman JJ, Kjer-Nielsen L, Brooks AG, et al. Identification of a dominant self-ligand bound to three HLA B44 alleles and the preliminary crystallographic analysis of recombinant forms of each complex. FEBS Lett. 2002;527(1-3):27-32.
24. Wynn KK, Fulton Z, Cooper L, Silins SL, Gras S, Archbold JK, et al. Impact of clonal competition for peptide-MHC complexes on the CD8+ T-cell repertoire selection in a persistent viral infection. Blood. 2008;111 (8):4283-92.
25. Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Cryst. 1993;26:795-800.
26. Storoni LC, McCoy AJ, Read RJ. Likelihood-enhanced fast rotation functions. Acta Crystallogr D Biol Crystallogr. 2004;60(Pt 3):432-8.
27. Zernich D, Purcell AW, Macdonald WA, Kjer-Nielsen L, Ely LK, Laham N, et al. Natural HLA class I polymorphism controls the pathway of antigen presentation and susceptibility to viral evasion. J Exp Med. 2004;200(1):13-24.
28. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr. 1997;53(Pt 3):240-55.
29. Afonine PV, Grosse-Kunstleve RW, Adams PD. A robust bulk-solvent correction and anisotropic scaling procedure. Acta Crystallogr D Biol Crystallogr. 2005;61(Pt 7):850-5.
30. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron-density maps and the location of errors in these models. Acta Crystallogr A. 1991;47 (Pt 2):110-9.
31. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004;60(Pt 12 Pt 1):2126-32.
32. Lamzin VS, Wilson KS. Automated refinement of protein models. Acta Crystallogr D Biol Crystallogr. 1993;49(Pt 1):129-47.
33. Merritt EA. Expanding the model: anisotropic displacement parameters in protein structure refinement. Acta Crystallogr D Biol Crystallogr. 1999;55(Pt 6):1109-17.
34. Vaguine AA, Richelle J, Wodak SJ. SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. Acta Crystallogr D Biol Crystallogr. 1999;55(Pt 1):191-205.
35. Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. Proteins. 1994;19(4):277-90.
36. Brunger AT. Version 1.2 of the Crystallography and NMR system. Nat Protoc. 2007;2(11):2728-33.
37. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 2007;35(Web Server issue): W375-83.
38. Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. Nucleic Acids Res. 2005;33(Web Server issue):W368-71.
39. Bade-Doeding C, DeLuca DS, Seltsam A, Blasczyk R, Eiz-Vesper B. Amino acid 95 causes strong alteration of peptide position Pomega in HLA-B*41 variants. Immunogenetics. 2007;59(4):253-9.
40. Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. Nat Struct Biol. 2003;10(12):980.
41. Probst-Kepper M, Hecht HJ, Herrmann H, Janke V, Ocklenburg F, Klempnauer J, et al. Conformational restraints and flexibility of 14-meric peptides in complex with HLA-B*3501. J Immunol. 2004;173(9): 5610-6.
42. Miles JJ, Borg NA, Brennan RM, Tynan FE, Kjer-Nielsen L, Silins SL, et al. TCR alpha genes direct MHC restriction in the potent human T cell response to a class I-bound viral epitope. J Immunol. 2006;177(10): 6804-14.
43. Tynan FE, Reid HH, Kjer-Nielsen L, Miles JJ, Wilce MC, Kostenko L, et al. A T cell receptor flattens a bulged antigenic peptide presented by a major histocompatibility complex class I molecule. Nat Immunol. 2007;8(3):268-76.
44. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22(12):2577-637.
45. Theodossis A, Guillonneau C, Welland A, Ely LK, Clements CS, Williamson NA, et al. Constraints within major histocompatibility complex class I restricted peptides: presentation and consequences for T-cell recognition. Proc Natl Acad Sci USA. 2010;107 (12):5534-9.
46. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res. 2006;34(Web Server issue): W116-8.
47. Kumar S, Nussinov R. Relationship between ion pair geometries and electrostatic strengths in proteins. Biophys J. 2002;83(3):1595-612.
48. Smith KD, Epperson DF, Lutz CT. Alloreactive cytotoxic T-lymphocyte-defined HLA-B7 subtypes differ in peptide antigen presentation. Immunogenetics. 1996;43(1-2):27-37.
49. Ghei M, Stroncek DF, Provenzano M. Analysis of memory T lymphocyte activity following stimulation with overlapping HLA-A*2402, A*0101 and Cw*0402 restricted CMV pp65 peptides. J Transl Med. 2005;3:23.
50. Pietersz GA, Li W, Apostolopoulos V. A 16-mer peptide (RQIKIWFQNRRMKWKK) from antennapedia preferentially targets the Class I pathway. Vaccine. 2001;19(11-12):1397-405.
51. Matsumura S, Kita H, He XS, Ansari AA, Lian ZX, Van De Water J, et al. Comprehensive mapping of HLA-A0201-restricted CD8 T-cell epitopes on PDC-E2 in primary biliary cirrhosis. Hepatology. 2002;36(5):1125-34.
52. Slezak SL, Bettinotti M, Selleri S, Adams S, Marincola FM, Stroncek DF. CMV pp65 and IE-1 T cell epitopes recognized by healthy subjects. J Transl Med. 2007;5:17.
53. Bell MJ, Burrows JM, Brennan R, Miles JJ, Tellam J, McCluskey J, et al. The peptide length specificity of some HLA class I alleles is very broad and includes peptides of up to 25 amino acids in length. Mol Immunol. 2009;46(8-9):1911-7.
54. Saper MA, Bjorkman PJ, Wiley DC. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 A resolution. J Mol Biol. 1991;219(2):277-319.
55. Bailey S. The CCP4 suite: programs for protein crystallography. Acta Cryst. 1994;50(Pt 5):760-3.
56. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol. 1971;55(3):379-400.