Targeting HLA class II in allogeneic stem cell transplantation

Balen, P. van

CHAPTER 5

Immunopeptidome analysis of HLA-DPB1 allelic variants reveals new functional hierarchies

P. van Balen, M.G.D. Kester, W. de Klerk, P. Crivello, E. Arrieta-Bolaños, A.H. de Ru, I. Jedema, Y. Mohammed, M.H.M. Heemskerk, K. Fleischhauer, P.A. van Veelen, J.H.F. Falkenburg

The Journal of Immunology 2020 (204): 3273-3282

ABSTRACT

HLA-DP alleles can be classified into functional T-cell epitope (TCE) groups. TCE-1 and TCE-2 are clearly defined, but TCE-3 represents still an heterogeneous group. Since polymorphisms in HLA-DP influence the presented peptidome, we investigated whether the composition of peptides binding in HLA-DP may be used to refine the HLA-DP group classification. Peptidomes of human HLA-DP typed B-cell lines were analyzed with mass spectrometry after immunoaffinity chromatography and peptide elution. Gibbs clustering was performed to identify motifs of binding peptides. HLA-DP peptide binding motifs showed a clear association with the HLA-DP allele specific sequences of the binding groove. Hierarchical clustering of HLA-DP immunopeptidomes was performed to investigate the similarities and differences in peptidomes of different HLA-DP molecules and this clustering resulted in the categorization of HLA-DP alleles into 3 DP peptidome clusters (DPC). The peptidomes of HLA-DPB1*09:01, 10:01 and 17:01 (TCE-1 alleles) and HLA-DPB1*04:01, 04:02 and 02:01 (TCE-3 alleles) were separated in two maximal distinct clusters, DPC-1 and DPC-3 respectively, reflecting their previous TCE classification. HLA-DP alleles categorized in DPC-2 shared certain similar peptide binding motifs with DPC-1 or DPC-3 alleles, but significant differences were observed for other positions. Within DPC-2, divergence between the alleles was observed based on the preference for different peptide residues at position 9. In summary, immunopeptidome analysis was used to unravel functional hierarchies among HLA-DP alleles, providing new molecular insights into HLA-DP classification.

INTRODUCTION

Unrelated donors for allogeneic stem cell transplantations who are HLA 10/10 matched with the recipient are mismatched for HLA-DP in 71-88% of cases.[1-4] Although frequently not taken into account in donor selection, mismatched HLA-DP can induce potent immune responses, consisting of graft-versus-leukemia reactivity and graft-versus-host disease (GVHD).[5-7] HLA-DPB1 alleles were previously categorized into T-cell epitope (TCE) groups based on in vitro experiments using recognition patterns of anti-HLA-DP directed T cells and amino acid (aa) sequences of the binding groove defining functional distance among the different HLA-DPB1 alleles.[8-10] Based on this classification, HLA-DP mismatches have been classified as permissive (mismatch within the same TCE group) or non-permissive (mismatch across different TCE groups) with predictive value for the outcome of transplantation. [8,11-14] The functional groups of TCE-1 alleles (HLA-DPB1*09:01, 10:01 and 17:01) and TCE-2 alleles (HLA-DPB1*03:01, 14:01 and 45:01) were clearly defined, but TCE-3 included any HLA-DPB1 allele not belonging to either group 1 or 2, and represents a relatively heterogeneous group. A first attempt towards the definition of functional relations within TCE-3 was made previously by subdividing TCE-3 into a group containing HLA-DPB1*02 and a group containing all other HLA-DPB1 alleles.[11,15]

It has been shown that polymorphisms within the peptide binding groove of HLA-DP molecules are more important for HLA-DP restricted alloreactivity than polymorphisms outside the peptide binding groove.[9,16] In addition, we have previously demonstrated that biologically relevant allo-HLA class I and II reactivity selected during a GVHD response were peptide specific.[17,18] Therefore, we hypothesize that differences in composition of peptides bound to the various HLA-DP molecules determine the potency to induce immune responses between mismatched HLA-DP molecules. By analyzing the peptidome of the 12 common HLA-DP molecules we investigated whether permissiveness and non-permissiveness with respect to alloreactivity in the context of allogeneic stem cell transplantation could be defined based on the similarity or differences in their respective immunopeptidomes. In this study, we analyzed the peptidome of HLA-DP molecules using immunoaffinity chromatography followed by mass spectrometry of the eluted peptides. Distinct peptide binding motifs were identified correlating with structural differences across HLA-DP molecules. We demonstrated new functional hierarchies based on similarities and differences in binding motifs of peptides binding to different HLA-DP molecules.

MATERIALS AND METHODS

EXPANSION AND LYSIS OF CELLS

To investigate peptides presented in HLA-DP molecules encoded by DPB1*09:01, 10:01, 17:01 (TCE1), DPB1*03:01, 14:01 (TCE2) and DPB1*01:01, 02:01, 04:01, 04:02, 05:01, 06:01 and 13:01 (TCE3), HLA-DP typed Epstein-Barr virus transformed B lymphoblastic cell lines (EBV-LCL) were expanded to 2-8x10^9 cells. HLA class II negative K562 cells were transduced with 7 frequent HLA-DP alleles using pLZRS vector HLA-DPA1-T2A-HLA-DPB1-IRES-ΔNGF-R combinations: DPA1*02:01/DPB1*01:01, DPA1*01:03/DPB1*02:01, DPA1*01:03/DPB1*03:01, DPA1*01:03/DPB1*04:01, DPA1*01:03/DPB1*04:02, DPA1*02:02/DPB1*05:01, DPA1*02:01/DPB1*14:01. After expansion cells were stored at −80 °C as dry cell pellet.

HLA-DP IMMUNOAFFINITY CHROMATOGRAPHY

To obtain anti-HLA-DP antibodies, hybridoma cell line B7/21 was expanded. Produced antibodies were purified using Prot-A Sepharose beads (GE Healthcare). Purified B7/21 antibodies were used to generate an immunoaffinity column (B7/21-Prot-A Sepharose 2.5 mg/ml).[19] Cell pellets were lysed in 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mMEDTA and 0.5% Zwittergent 3-12 (Merck) and supplemented with Complete® protease inhibitor (Merck). Concentration in the lysis buffer was 1×10^8 cells/ml. After 2h, the supernatant was pre-cleared with CL4B beads (GE Healthcare, Eindhoven, The Netherlands) and subjected to the immunoaffinity column with a flow rate of 2.5 ml/min. Bound peptide-HLA-DP complexes were eluted from the column and dissociated with 10% acetic acid (Merck). Peptides were separated from the HLA-DP molecules via passage through a 10 kDa membrane (Merck) and the filtrate was freeze-dried.

ANALYSIS OF PEPTIDE MIX USING MASS SPECTROMETRY

Q-Exactive mass spectrometer. Peptides were analysed using an Easy nLC1000 (Thermo, Bremen, Germany) coupled to a Q-Exactive mass spectrometer (MS, Thermo). Injection was done onto an homemade precolumn (100 µm × 15 mm; Reprosil-Pur C18-AQ 3 µm, Dr. Maisch, Ammerbuch, Germany) and elution via an homemade analytical column (15 cm × 75 µm; Reprosil-Pur C18-AQ 3um). The gradient was 0% to 30% solvent B (100 ACN/0.1 FA v/v) in 120 min. The analytical column was drawn to a tip of ~5 µm and acted as the electrospray needle of the MS source. The Q-Exactive MS was operated in top10-mode. Parameters were as follows: full scan, 70,000 resolution, 3,000,000 AGC target, max fill time 100 ms. Dynamic exclusion of 10s. For MS2 precursors were isolated with the quadrupole with an isolation width of 2 Da. HCD collision energy was set to 27V. MS2 scan resolution was 35,000 at a 1,000,000 AGC target, 128 ms max fill time. Apex trigger was set to 1–5 s, and allowed charges were 1–4.

Orbitrap Fusion LUMOS. Peptides were analyzed by on-line C18 nanoHPLC MS/MS with a system consisting of an Easy nLC 1200 gradient HPLC system (Thermo, Bremen, Germany),

and an Orbitrap Fusion LUMOS mass spectrometer (Thermo). Fractions were injected onto an homemade precolumn (100 µm × 15 mm; Reprosil-Pur C18-AQ 3 µm, Dr. Maisch, Ammerbuch, Germany) and eluted via a homemade analytical nano-HPLC column (15 cm × 75 µm; Reprosil-Pur C18-AQ 3 um). The gradient was run from 2% to 36% solvent B (20/80/0.1 water/acetonitrile/formic acid (FA) v/v/v) in 120 min. The nano-HPLC column was drawn to a tip of ~5 µm and acted as the electrospray needle of the MS source. The LUMOS MS was operated in data-dependent MS/MS (top-N mode with a cycle time of 3 s) with collision energy at 32 V and recording of the MS2 spectrum in the orbitrap. In the master scan (MS1) the resolution was 60,000, the scan range 300-1400, at an AGC target of 400,000 and maximum fill time of 50 ms. Dynamic exclusion after n=1 with exclusion duration of 20 s. Charge states 1 (range 800-1400), 2 (range 400-1200) and 3 (range 300-800) were included. For MS2 precursors were isolated with the quadrupole with an isolation width of 1.2 Da. HCD collision energy was set to 32 V. First mass was set to 110 Da. The MS2 scan resolution was 30,000 with an AGC target of 50,000 at maximum fill time of 100 ms.

DATA PROCESSING AND GIBBS CLUSTERING

Raw data were first converted to peak lists using Proteome Discoverer version 2.2 (Thermo Electron), and submitted to the Uniprot Homo sapiens database (20410 entries), using Mascot v. 2.2.07 (www.matrixscience.com) for protein identification. A false discovery rate of 1% (Target Decoy) was set together with Mascot Ion score above 10 for data collection and grouping. Peptide alignment and analysis of nested sets was done in Excel.[20] A nested set is defined as a collection of peptides originating from the same protein and sharing a sequence of 9 or more identical aa. The centers of the peptides belonging to the same nested set were not allowed to be more than 6 aa apart from each other. The Core Sequence was defined as the minimal peptide of a nested set. A Region Sequence was defined as tens of aa containing the identified peptide and therefore always have a fixed start and stop position and mainly a length of 10, 20, 30 or 40 aa. To investigate whether the eluted peptides frequently contained certain aa at certain positions, alignment and clustering with Gibbs sampling approach was performed using the GibbsCluster-1.1a Perl script (www.cbs.dtu.dk/services/).[21] Statistical analysis was performed using probability weighted Kullback Leibler distance (KLD). The height of the aa corresponds to their probability times their log-odds score, which is also the information contribution of the aa given and the Weight Matrix is used for analysis.

HIERARCHICAL CLUSTERING

To investigate the presence of identical peptides and the similarity between the peptidomes of different HLA-DP molecules, agglomerative hierarchical clustering was performed, in which observations are grouped into clusters. For this, the optimal grouping is found where similar observations are grouped together as clusters, but the different clusters are separate from one another. The similarity between the observation vectors was determined

by Euclidian and Manhattan distances.[22] The clustering solutions obtained by hierarchical clustering are visualized as a dendrogram, with the observation indices placed along the vertical axis, such that no two branches of the dendrogram cross. Their respective linkage value is denoted by the horizontal axis, such that the most similar observations are connected near the right end of the tree.[23]

ASSOCIATION BETWEEN PEPTIDE MOTIFS AND BINDING GROOVE OF HLA-DP

To quantify the association between aa in the groove of HLA-DP variants and specific aa in the bound peptide, we measured the distances between the binding motifs associated with each variant. In order to allow best compatibility and comparability we used weighted pairwise normalized distance (weighted Canberra) between each two variants and association with each position. The weighting corresponds to the normalized number of observations of the specific aa at a specific position. The higher the distance, the stronger the influence of having the HLA-DP variant on the binding motif at the corresponding position in the bound peptide.

RESULTS

BINDING MOTIFS OF PEPTIDES PRESENTED BY HLA-DP

To investigate the peptidome of peptides bound to HLA-DP, 19 HLA-DP typed EBV-LCL were expanded to 2-8x10^9 cells. Fifteen EBV-LCLs were homozygous for HLA-DPB1 and 4 were heterozygous (Table 1). After expansion, immunoaffinity chromatography using anti-HLA-DP B7/21 antibodies was performed followed by analysis of the peptidome using mass spectrometry, resulting in the detection of 2010-7260 unique peptides per HLA-DP allele derived from 587-1570 source proteins (Table 1). To investigate whether the eluted peptides frequently contained certain aa at specific positions, Gibbs sampling approach on all unique peptides per HLA-DP allele was performed.[21] Since the length of such an aa motif was not known, Gibbs clustering was used for analyzing potential motifs with lengths of 9 or 12 aa. For most EBV-LCL homozygous for HLA-DPB1, highest KLD were reached when clustering in one cluster was performed, meaning that one dominant motif could be detected in aa sequences of eluted peptides, both for a motif length of 9 and 12 aa. However, for peptides eluted from EBV-LCL homozygous for DPB1*05:01, two different motifs were identified. Figure 1 shows the motifs for all homozygous EBV-LCL, with the positions of aa depicted on the x-axis and the height of the depicted aa corresponding to their relative frequencies at these positions. In all analyzed samples a limited number of peptides were outliers which did not fit in the clustering using Gibbs sampling approach. Since clustering for motifs of lengths of 9 or 12 aa resulted in similar results (Supplementary Figure S1), it was concluded that the peptide binding motif of HLA-DP consisted primarily of 9 aa. As expected, clustering of peptides eluted from heterozygous EBV-LCL resulted in optimal clustering into two groups (Figure 2). Gibbs clustering of eluted peptides from EBV-LCL heterozygous for DPB1*04:01 and 10:01 and for 04:01 and 14:01 resulted in two motifs for each EBV-LCL, which were similar to the motifs of the peptidome of EBV-LCL homozygous for those alleles (Supplementary Figure S1). These results illustrate that heterozygous EBV-LCL can be used for the identification of binding motifs for HLA-DP alleles for which homozygous EBV-LCL were not available (e.g. DPB1*17:01, 06:01), provided that the other allele of the heterozygous EBV-LCL was of an HLA-DPB1 type with known binding motif and that both binding motifs are not similar to each other (Figure 2).

Table 1. HLA-DP typing of analyzed EBV-LCL and numbers of peptides eluted

EBV-LCL Sample ID	HLA-DPB1	HLA-DPA1	# unique eluted peptides	# source proteins for eluted peptides	median # peptides per protein (range)	average # peptides per protein
CCC 12493	DPB1*09:01	DPA1*02:01:01	4303	976	2 (1 - 89)	4.41
VNG 12286	DPB1*10:01	DPA1*02:01:01	5127	1122	2 (1 - 124)	4.57
ETV 11197	DPB1*14:01	DPA1*02:01:01	5868	1167	2 (1 - 113)	5.03
CLK 3987	DPB1*03:01	DPA1*01:03:01	4417	923	2 (1 - 129)	4.79
MPF 174	DPB1*03:01	DPA1*01:03:01	3733	849	2 (1 - 113)	4.40
TQP 8275	DPB1*05:01	DPA1*02:02:02	4211	822	2 (1 - 149)	5.12
WKD 3639	DPB1*05:01:01	DPA1*02:02:02	6349	1366	2 (1 - 140)	4.65
NAG 12292	DPB1*01:01:01	DPA1*02:01:02	2766	634	2 (1 - 71)	4.36
BZQ 12290	DPB1*13:01:01	DPA1*02:01:01	3896	838	2 (1 - 49)	4.65
FAQ 3528	DPB1*02:01:02	DPA1*01:03:01	5710	1126	2 (1 - 93)	5.07
REE 7929	DPB1*02:01:02	DPA1*01:03:01; DPA1*02:01	3577	836	2 (1 - 53)	4.28
HHC 5852	DPB1*04:01	DPA1*01:03:01	7260	1570	2 (1 - 145)	4.62
MHX 1699	DPB1*04:01	DPA1*01:03:01	4252	998	2 (1 - 70)	4.26
WEW 12494	DPB1*04:01	DPA1*01:03:01	4052	993	2 (1 - 58)	4.08
AST 5085	DPB1*04:02:01	DPA1*01:03:01	3186	760	2 (1 - 104)	4.19
UKM 7264	DPB1*04:01; DPB1*17:01	DPA1*01:03:01; DPA1*02:01	5067	1112	2 (1 - 87)	4.56
ASY 3436	DPB1*04:01; DPB1*10:01	DPA1*01:03:01; DPA1*02:01	2409	619	2 (1 - 51)	3.89
AZP 2877	DPB1*04:01; DPB1*14:01	DPA1*01:03:01; DPA1*02:01	4456	987	2 (1 - 67)	4.51
JTD 5000	DPB1*04:01:01; DPB1*06:01:01	DPA1*01:03:01	2010	587	2 (1 - 64)	3.42
Total			***82649***			

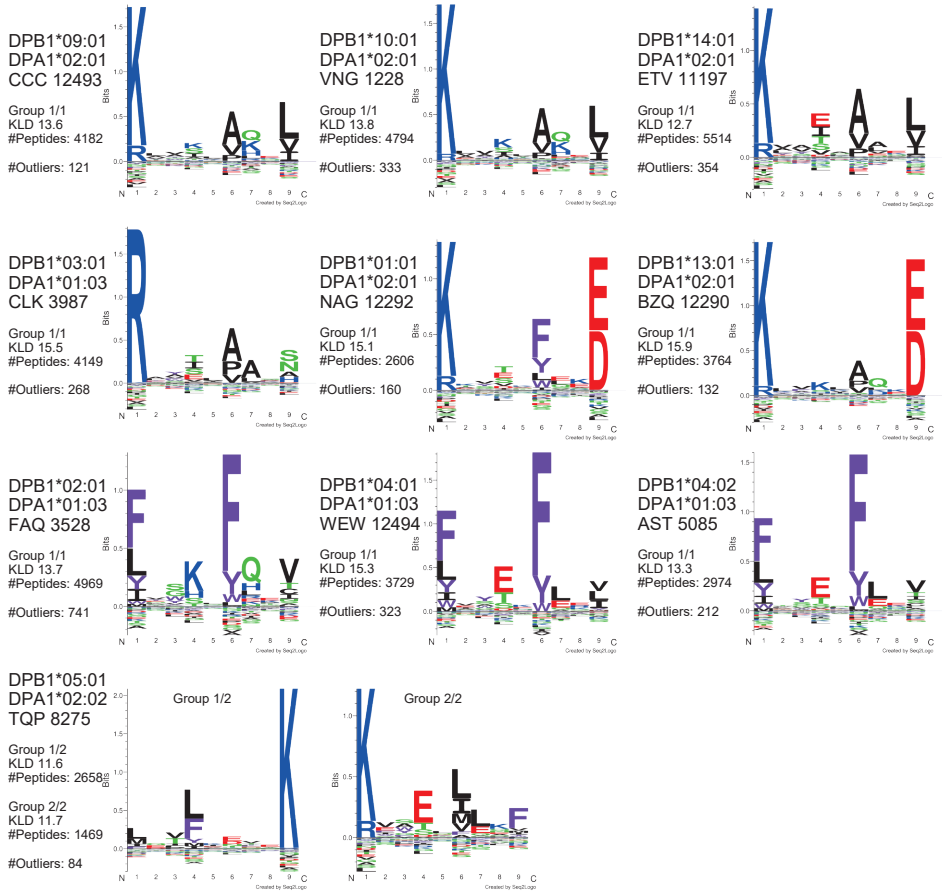


Figure 1. Motifs of peptides eluted from EBV-LCL homozygous for different HLA-DPB1 alleles. The position of aa in the peptide is depicted on the x-axis and the height of the depicted aa corresponds to their relative frequency at this position. Motifs were derived using Gibbs clustering of eluted peptides as described previously.[21] For the optimization of the fitness of the sequence alignment, Kullback–Leibler distance (KLD) was used. KLD was highest when clustering in one cluster was performed, meaning that one dominant motif could be detected in aa sequences of eluted peptides of homozygous EBV-LCL, except for DPB1*05:01 for which two dominant motifs were identified. Left from the y-axis Sample ID of EBV-LCL, number of peptides used in the clustering, number of outlier peptides not used in the clustering and KLD is shown. The sequence of depiction of motifs of peptides binding in different HLA-DP alleles follows the current classification into TCE groups.

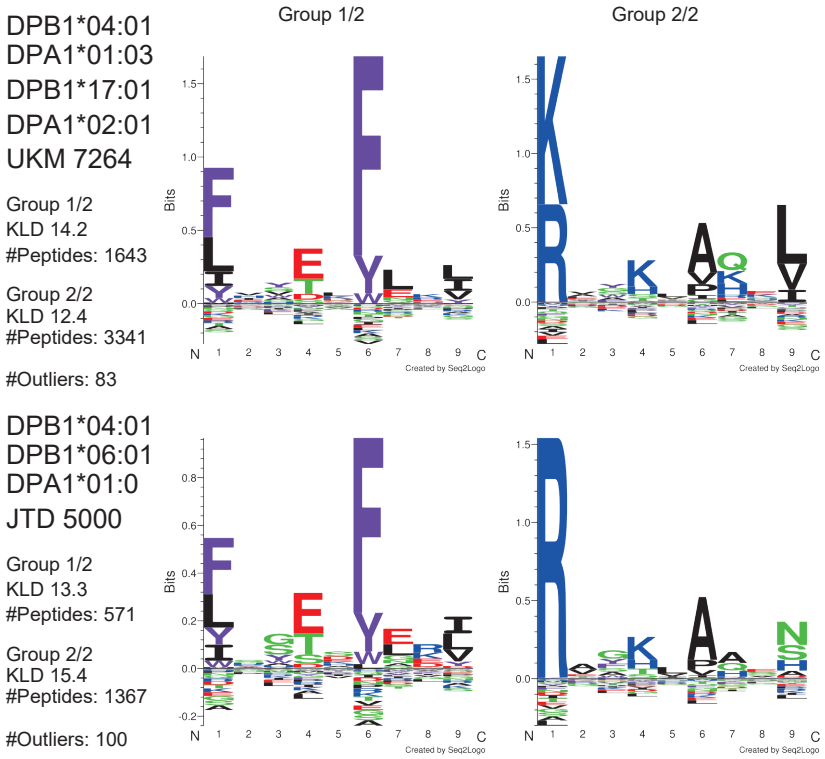


Figure 2. Motifs of peptides eluted from EBV-LCL heterozygous for different HLA-DPB1 alleles. The position of aa in the peptide is depicted on the x-axis and the height of the depicted aa corresponds to their relative frequency at this position. The motif was derived using Gibbs clustering of eluted peptides.[21] For the optimization of the fitness of the sequence alignment, Kullback–Leibler distance (KLD) was used. KLD was highest when clustering in two clusters was performed, meaning that two dominant motifs could be detected in aa sequences of eluted peptides of heterozygous EBV-LCL. Left from the y-axis Sample ID of EBV-LCL, number of peptides used in the clustering, number of outlier peptides not used in the clustering and KLD is shown.

To further validate the peptide binding motifs, the peptidome of HLA-DP transduced chronic myeloid leukemia cell line K562 was analyzed as alternative cell source of peptides for 7 different HLA-DP alleles.[24] The results demonstrated identical binding motifs as identified using EBV-LCL as cell source (Supplementary Figure S2), confirming the HLA-DP binding motifs, and demonstrating that cell source did not influence the binding motifs of different HLA-DP molecules.

All 3 alleles from TCE-1 (DPB1*09:01, 10:01, 17:01) had a similar motif with most likely aa K, K, A, Q and L at position 1 (p1), p4, p6, p7 and p9, respectively (Figure 1 and 2). This motif was shared by DPB1*14:01 from TCE-2 at p1, p6 and p9. However, its motif was different at p4

and p7. The motif of peptides binding in TCE-2 allele DPB1*03:01 was similar to the TCE-1 alleles at p1 and p6 and similar to DPB1*14:01 at p1, p6 and p7. Importantly, the peptide binding motifs of TCE-3 alleles were more heterogeneous. DPB1*04:01, 04:02 and 02:01 had largely superimposable motifs, with a dominant F at p1 and p6, whereas other TCE-3 alleles (DPB1*01:01, 05:01, 06:01 and 13:01) had significantly different motifs.

These results illustrate that, except for HLA-DPB1*05:01, clear binding motifs for different HLA-DP molecules could be generated using immunoaffinity chromatography followed by peptide elution, analysis by mass spectrometry and Gibbs clustering. While peptide binding motifs for TCE-1 alleles displayed high similarity, and motifs for TCE-2 alleles slightly differed from each other, DP alleles from group TCE-3 demonstrated great heterogeneity in their binding motifs.

NO INFLUENCE OF LENGTH VARIANTS ON BINDING MOTIFS

Since HLA class II ligands can have various different lengths, we investigated the potential influence of these length variants on the DP peptide binding motifs. For this, Gibbs clustering was performed on the dataset of all unique peptides, but also on a set that measured length variants of the same region as one peptide (i.e. core peptides). The HLA-DP ligandome consisted predominantly (49.4%) of peptides of 14-16 aa in length, the other half of the peptidome ranging from a minimum of 9 aa to a maximum of 47 aa (Figure 3). Two methods to investigate the influence of length variants were used and explained visually in Figure 3. In the first method, core sequences were determined, in which the overlapping sequence of length variants was only counted once in the clustering. To determine core sequences, nested sets of peptides consisting of a collection of peptides originating from the same protein region and sharing a sequence of 9 or more identical aa were first identified. By definition, the centers of the peptides belonging to the same nested set cannot be more than 6 aa apart from each other. The core sequence is defined as the minimal peptide of a nested set. Per HLA-DP allele 79-89% of unique peptides belonged to nested sets (Table 2). Peptides without length variants were counted as unique core sequences. 777-2258 core sequences could be detected per HLA-DP allele (Table 2). Gibbs clustering was performed on these core sequences and the same binding motifs were found as after clustering of all unique peptides (Figure 3). The second method to investigate the potential influence of length variants on the binding motif, was to assign the identified peptides to a ten-aa region in the protein and subsequently use each string of 10 aa just once for clustering (Table 2). This method resulted also in identical binding motifs as compared to clustering of all unique peptides (Figure 3). These results demonstrate that the identified binding motifs were not influenced by overrepresentation of peptides of which many length variants were present in the eluted peptide pool.

Table 2. Clustering of length variants of eluted peptides into nested sets and region sequences.

EBV-LCL Sample ID	HLA-DPB1	# unique peptides	# peptides in nested sets	# consensus core peptides	median # peptides per core (range)	mean # peptides per core	# region sequences	median # peptides in protein region (range)	mean # peptides in protein region
CCC 12493	DPB1*09:01	4303	3715	1358	2 (1 - 60)	3.17	1668	2 (1 - 28)	2.58
VNG 12286	DPB1*10:01	5127	4436	1634	2 (1 - 49)	3.14	2043	1 (1 - 35)	2.51
ETV 11197	DPB1*14:01	5868	5149	1726	2 (1 - 69)	3.4	2141	2 (1 - 53)	2.74
CLK 3987	DPB1*03:01	4417	3886	1185	2 (1 - 46)	3.73	1504	2 (1 - 32)	2.94
MPF 174	DPB1*03:01	3733	3233	1100	2 (1 - 57)	3.39	1308	2 (1 - 34)	2.85
TQP 8275	DPB1*05:01	4211	3661	1313	2 (1 - 43)	3.21	1647	2 (1 - 25)	2.56
WKD 3639	DPB1*05:01	6349	5329	2167	2 (1 - 34)	2.93	2572	1 (1 - 34)	2.47
NAG 12292	DPB1*01:01	2766	2377	888	2 (1 - 64)	3.11	1073	2 (1 - 47)	2.58
BZQ 12290	DPB1*13:01	3896	3476	1060	2 (1 - 34)	3.68	1312	2 (1 - 29)	2.97
FAQ 3528	DPB1*02:01	5710	5017	1634	2 (1 - 62)	3.49	1996	2 (1 - 43)	2.86
REE 7929	DPB1*02:01	3577	3044	1146	2 (1 - 37)	3.12	1359	2 (1 - 29)	2.63
HHC 5852	DPB1*04:01	7260	6167	2258	2 (1 - 49)	3.22	2665	1 (1 - 40)	2.72
MHX 1699	DPB1*04:01	4252	3566	1432	2 (1 - 69)	2.97	1736	1 (1 - 41)	2.45
WEW 12494	DPB1*04:01	4052	3472	1285	2 (1 - 33)	3.15	1559	1 (1 - 29)	2.6
AST 5085	DPB1*04:02	3186	2698	1031	2 (1 - 44)	3.09	1241	1 (1 - 31)	2.57
UKM 7264	DPB1*04:01;DPB1*17:01	5067	4325	1638	2 (1 - 45)	3.09	2014	1 (1 - 30)	2.52
ASY 3436	DPB1*04:01;DPB1*10:01	2409	1986	863	2 (1 - 38)	2.79	1014	1 (1 - 21)	2.38
AZP 2877	DPB1*04:01;DPB1*14:01	4456	3774	1452	2 (1 - 33)	3.07	1745	1 (1 - 30)	2.55
JTD 5000	DPB1*04:01;DPB1*06:01	2010	1592	777	1 (1 - 29)	2.59	865	1 (1 - 21)	2.32

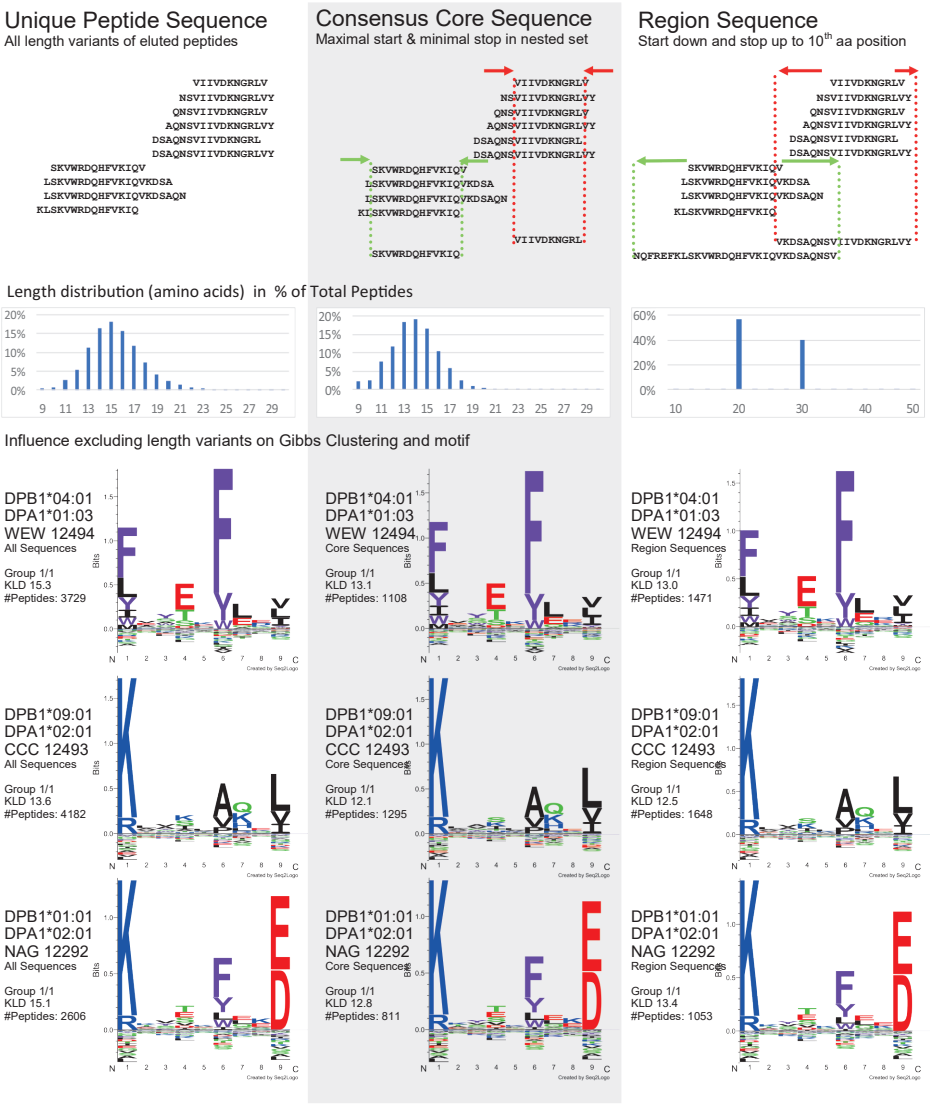


Figure 3. Influence of peptide length variants on the DP binding motif. Gibbs clustering was performed on all unique peptides, on core sequences and on region sequences. To determine core sequences, nested sets of peptides were identified first. A nested set is a collection of peptides originating from the same protein and sharing a sequence of 9 or more identical aa. The centers of the peptides belonging to the same nested set cannot be more than 6 aa apart from each other. The core sequence is defined as the minimal peptide of a nested set. To determine region sequences, peptides were assigned to a 10 aa region in the protein. Gibbs clustering performed on all unique peptides, core sequences and region sequences resulted in the same binding motifs, illustrating that the identified motifs for peptides binding in HLA-DP were not influenced by overrepresentation of peptides of which many length variants were present in the eluted peptide mix.

SIMILARITY IN PEPTIDOME OF HLA-DPB1 ALLELES

In addition to the analysis of the HLA-DP binding motifs, the composition of the HLA-DP peptidomes was analyzed by performing hierarchical clustering using Euclidian distance (shortest distance between two points) calculation.[22] Since Gibbs clustering of the HLA-DPB1*05:01 peptidome revealed two different binding motifs, this HLA allele was excluded from hierarchical clustering. The results of the hierarchical clustering are shown in Figure 4A. As expected, the peptidomes of different EBV-LCL expressing the same HLA-DP molecules displayed high similarity. The peptidomes of the TCE-1 alleles HLA-DPB1*10:01, 09:01 and 17:01 displayed more similarity to each other than the peptidomes of all other HLA-DP alleles. The peptidomes of TCE-3 alleles HLA-DPB1*02:01, 04:01 and 04:02 displayed similarity, but were most separate from the TCE-1 alleles. However, the peptidome of these TCE-3 alleles was very different from the peptidome of the other TCE-3 alleles. The peptidomes of TCE-3 alleles HLA-DPB1*13:01 and 01:01 cluster together more to HLA-DPB1*14:01, 03:01 and 06:01 than to the cluster containing HLA-DPB1*02:01, 04:01 and 04:02. Based on the results of this analysis, a new categorization in different DP peptidome cluster (DPC) groups could be made (Figure 4A). However, strict categorization into groups is not possible since the peptidomes can have similarities with several other HLA-DP peptidomes, also across the borders of a DPC group. DPB1*14:01 for instance, can be clustered into DPC group 1 or 2 depending on the method used. Indeed, when using hierarchical clustering based on Manhattan distance (sum of perpendicular distances on all axes between two points)[22], this allele clustered with the DPC-1 alleles instead of with DPB1*03:01 and 06:01 (Figure 4B), illustrating that the peptidome of DPB1*14:01 showed similarity to both DPC-1 and DPC-2 alleles. We therefore propose functional hierarchies instead of categorization into different groups. The peptidome of DPB1*09:01, 10:01 and 17:01 are similar and can be grouped together in DPC-1. These peptidomes are most different compared to the peptidomes of DPB1*04:01, 04:02 and 02:01, which can be clustered in DPC-3. All other HLA-DP peptidomes can be categorized in DPC-2, although this still represents a relatively heterogenous group. However, classification in functional hierarchies is preferred since strict categorization into groups is not possible.

ASSOCIATION BETWEEN PEPTIDE MOTIFS AND BINDING GROOVE OF HLA-DP

To investigate the relation between the binding motifs and the hypervariable regions in the HLA-DP molecules, we searched the literature for peptide residues known to interact with specific aa in the HLA-DP binding groove and aligned these. Based on a binding core of 9 aa, interactions between p1 and aa 31 of the HLA-DP alpha chain (31α) and aa 76 and 84-87 of the HLA-DP beta chain (76β, 84-87β), between p4 and 69β and 76β, between p6 and 11α and 11β, between p7 and 55β, 65β and 69β and between p9 and 9β, 35β, 36β and 55β are described.[8,25-31] Aa at these positions of the binding groove of different HLA-DP molecules are depicted in Figure 5A together with aa identified to be preferred at certain positions of the binding peptide based on the matrix score of the Gibbs clustering. The results show

clear association between aa in the binding groove with aa in the bound peptide. To quantify this association, we measured the distances between the binding motifs associated with each variant using weighted Canberra distance. The higher the distance, the stronger the influence of having the HLA-DP variant on the binding motif at the corresponding position. The relevant associations are depicted in Figure 5B.

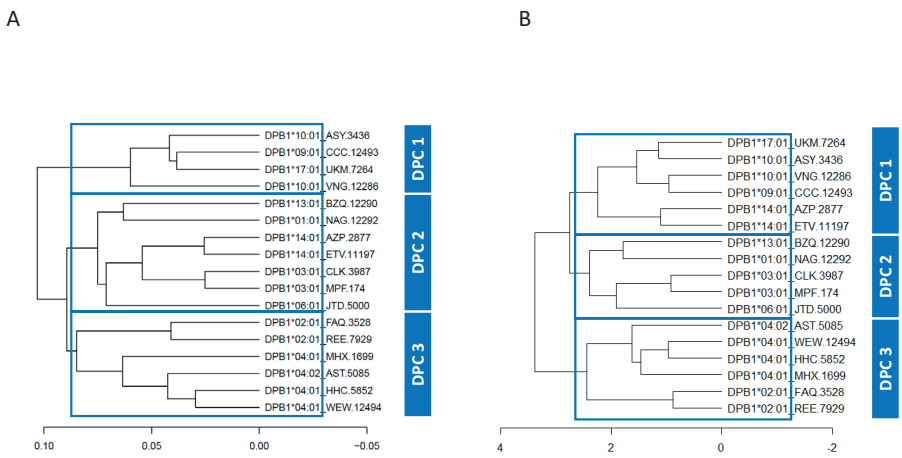


Figure 4. Hierarchical clustering dendrogram of similarity between peptidomes of different HLA-DP alleles. (A) Analysis using Euclidian distances.[22] The peptidome was most similar when the different EBV-LCL expressed the same HLA-DP molecule. The peptidomes of DPB1*10:01, 09:01 and 17:01 (DPC-1) were more similar to each other than to the peptidomes of all other HLA-DP alleles. The peptidomes of DPB1*02:01, 04:01 and 04:02 (DPC-3) cluster together. The peptidome of DPB1*06:01 was more similar to the peptidome of DPB1*03:01 and 14:01 within DPC-2, whereas within DPC-2 the peptidomes of DPB1*13:01 and 01:01 cluster together. **(B)** Analysis using Manhattan distances.[22] The results are nearly identical regarding clustering of HLA-DP alleles. However, with this approach, DPB1*14:01 was clustered together with DPB1*10:01, 09:01 and 17:01 instead of with DPB1*03:01 and 06:01.

Analyzing the data per hypervariable region interacting with an aa in the peptide, an association was observed between the aa in the binding groove and the preferred aa at that position in the peptide. For instance, a positively charged K or R at p1 interacts with the negatively charged DEAV motif in the binding groove. In addition the presence of M at position 31α coincided with preference of R at p1 instead of K. The diverging GGPM motif at 84-87β was associated with preferred F, W and Y at p1.

After correlating aa in the peptide with aa in the binding groove, we tried to correlate the results of the hierarchical clustering in Figure 4 with the results in Figure 5. All three HLA-DP alleles of DPC-1 are nearly identical in the hypervariable regions and preferred aa in the peptide, except for a difference in 76β, interacting with p1 and p4. This difference only slightly

altered the preferred aa in p1 and p4. The HLA-DP alleles of DPC-3, being DPB1*02:01, 04:01 and 04:02 are structurally similar for the hypervariable regions interacting with p1, p6 and p9. However, both their preferred aa in the bound peptide as well as their binding groove sequence differed significantly from DPC-1 alleles, most prominently at positions p1 and p6. The hierarchy within DPC-3 can be explained by differences in aa interacting with p4 and p7. Regarding DPC-2 alleles, the preferred aa in the binding peptide showed high similarity with DPC-1 alleles for p1 and p6 with the exception of DPB1*01:01 which differs from the DPC-1 alleles for p6. Within DPC-2 there are three dominant patterns at p9. HLA-DP alleles with the same preferred aa at p9 cluster together in the hierarchical clustering (Figure 4). Regarding DPB1*14:01, the preferred aa at p1, p6 and p9 were identical to the DPC-1 alleles. However, for p4 and p7 there was more resemblance between DPB1*14:01 and the peptides bound to DPC-2 allele DPB1*03:01, explaining the clustering of these peptidomes in Figure 4A.

These results illustrate that there is a clear association between peptide binding motifs of different HLA-DP alleles and the sequence of their binding grooves. Preferred aa at p1, p6 and p9 were associated with aa in the binding groove of HLA-DP and similarity at these positions seemed to have an important influence to the hierarchical clustering. Preferred aa at binding positions differ the most between DPC-1 and DPC-3. Within DPC-2, preferred aa at binding positions can be similar to DPC-1 alleles, especially for p1, but also to DPC-3 alleles, as is the case for p6 in DPB1*01:01. Preferred aa at p9 seemed to play an essential role in the hierarchical clustering within DPC-2.

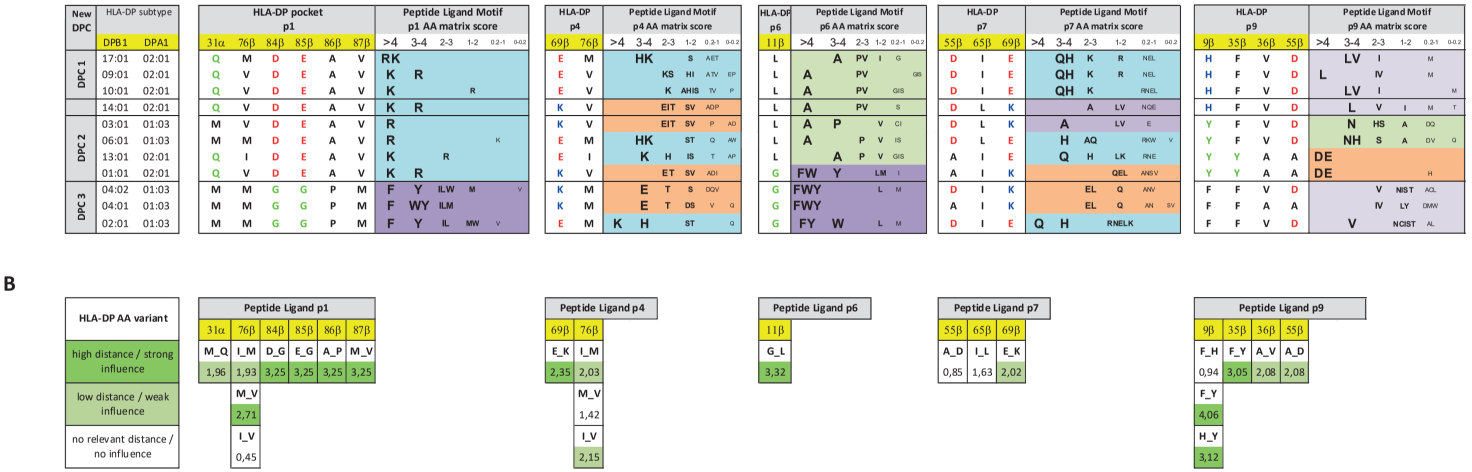


Figure 5. Association between aa polymorphism in the DP binding groove and preferred aa in the peptide. (A) Overview of aa in the peptide binding groove of different HLA-DP alleles and the preferred aa at specific positions of the bound peptide. Aa matrix score was a result of the Gibbs clustering and the higher this matrix score, the more frequent a certain aa was observed at this position. Colors are used to make differences and similarities more easily detectable. **(B)** Weighted pairwise normalized distance (weighted Canberra) between each two aa variants in the binding groove and aa in the bound peptide. The weighting corresponds to the normalized number of observations of the specific aa at a specific position. The higher the distance, the stronger the influence of having the HLA-DP variant on the binding motif at the corresponding position. A mixture modeling to allow hypothesis testing on the obtained distances resulted in separation between relevant and not relevant associations at 1.7 units on weighted Canberra distance.

DISCUSSION

The aim of our study was to investigate whether the composition of peptides bound to different HLA-DP alleles can be used to better understand functional clustering of HLA-DP alleles, previously categorized into TCE groups based on the relevance of HLA-DP variability on T-cell alloreactivity.[8,9] By analyzing HLA-DP peptide binding motifs and investigating similarities between the peptidomes of different HLA-DP molecules, we were able to produce an hierarchical clustering of HLA-DP alleles. Based on the results of this hierarchical clustering, a new categorization in 3 different DPC groups could be made. We were able to partly confirm, but also to refine the existing TCE group categorization. However, instead of strict categorization into closed groups, our data suggest a more gradual categorization. The peptidomes of DPB1*09:01, 10:01 and 17:01, previously classified as TCE-1 alleles, are most similar and can be grouped together in DPC-1. These peptidomes are most different from the peptidomes of DPB1*04:01, 04:02 and 02:01, previously assigned to TCE-3, which can be clustered in DPC-3. All other HLA-DP peptidomes, including some originally assigned to TCE-3, can be categorized in DPC-2, among which further stratification can be made, mainly based on similarities or differences at p9 of the bound peptides.

In unrelated stem cell transplantation, it has been shown that donor-recipient HLA-DP mismatching for alleles belonging to the same TCE group according to the previous definition (permissive mismatch), is associated with better clinical outcomes than mismatching for HLA-DP alleles from different TCE groups (non-permissive mismatches).[8,10,13,32] It could be suggested that in the setting of alloSCT patient-donor HLA-DP mismatches between alleles clustering together in hierarchical clustering are more permissive than mismatches between alleles not clustering together. It is thus tempting to speculate that previous associations might be further improved by our refined classification in DPC groups based on the peptide similarity between HLA-DP peptidomes. This speculation is based on the hypothesis that HLA-DP molecules which are more similar to each other and present more similar peptides in the binding groove, are less immunogenic to each other and therefore will be more permissive in case of a patient-donor mismatch than HLA-DP molecules that are clearly different from each other presenting more different peptides. Based on the new categorization proposed from our results, is it then expected that a mismatch between patient and donor is non-permissive if the mismatch is between DPC-1 and DPC-3 alleles. Whether other mismatches are permissive or non-permissive depends on the differences in the peptidome and whether these alleles cluster close to each other or not. We hypothesize that permissiveness is not a black and white phenomenon, but rather gradual based on similarities and differences in the peptidomes. Further studies are warranted to test this important point.

Our data allowed a new functional classification of HLA-DP alleles based on the relatedness of their peptide repertoires. Interestingly, the resulting clusters are very similar to the

evolutionary clades identified recently based on antigen recognition domain based variation in HLA-DPB1.[33] These authors found that HLA-DPB1 exhibits two deeply divergent conserved clades of alleles correlating with the rs9277534A/G polymorphism, a known expression marker.[34] The low-expression A clade included HLA-DPB1*02:01, 04:01 and 04:02 and showed reduced nucleotide diversity compared to the high-expression G clade.[33] Of note, low-expression allele HLA-DPB1*17:01 is not mentioned in their analysis, and the peptidome of this allele is clustered far away from HLA-DPB1*02:01, 04:01 and 04:02 in our study.

Some previous studies have reported DP peptide binding motifs. Using binding assays with synthetic biotinylated peptides on DPB1*09:01, a binding motif similar to the one characterized in the current study, could be identified.[35] Peptide binding motifs for DPB1*04:01, 04:02 and 02:01 determined in our study were similar to previously identified motifs at p1 and p6 and to lesser extend at p4 and p9.[36-42] While focused mainly on these alleles, we characterized in detail the binding motif of several previously not studied DP alleles. Our study resulted also in more detailed characterization of binding motifs than previous studies, not only for aa at p1 and p6, but also at p4, p7 and p9. In contrast to previous studies, our results were obtained after analysis of naturally processed and presented peptides in different HLA-DP molecules. In two recent publications, others found an identical binding motif for HLA-DPB1*06:01, 17:01 and 03:01 as we did.[42,43] For motifs of peptides bound to DPB1*01:01 and 05:01, there are conflicting data from previous studies. In contrast to our results, previous studies using synthetic peptides and single-substitution analogs revealed importance of F at p1 and p6 also for peptides binding in DPB1*01:01 and 05:01[37,38,44], while others assign important binding properties for K and R at p1 for peptides in DPB*05:01, as observed by us.[45] Our results showed clustering in two possible motifs of peptides binding in DPB1*05:01 (Figure 1). The motifs appeared to be mirror images. Based on these motifs, we hypothesize that a peptide can be bound in HLA-DPB1*05:01 both in the canonical orientation and in inversed alignment. This phenomenon has been described previously for class II-associated invariant chain peptides (CLIP) in HLA-DR1.[46] Further analyses using binding assays or crystallography are needed to confirm this hypothesis.

In this project, we focused on the role of the HLA-DP beta chain. However, also the alpha chain is known to have a role in T cell recognition[47] and in interaction with peptide.[25] Our results suggest that different aa in the alpha chain might also have an impact on the preferred aa at position p1 and p4 of the binding peptide. Although certain alpha and beta chain combinations occur together naturally, different alpha and beta chain combinations can be formed in heterozygous cells.[48-50] It is likely that pairing with another alpha chain can slightly alter the peptidome of HLA-DP alleles. The binding motifs identified in this study therefore only hold to the specific alpha-beta HLA-DP molecules investigated.

Taken together, the results of our study shed new light onto the influence of allelic variation in HLA-DP on the molecular nature of presented peptides. We show that functional clustering can be performed in DPC groups, based on peptidomics data, with potential impacts in clinical fields such as allogeneic stem cell transplantation. However, instead of strict categorization into groups, we argue that it is better to regard differences in HLA-DP peptidomes as gradual. This novel approach could thus be useful for improving our understanding of the functional heterogeneity of polymorphic HLA, and its consequences in different fields including transplantation, autoimmunity[51-55], infections[56], allergy[45] and cancer[57-59].

REFERENCES

1. Shaw BE, Marsh SG, Mayor NP, Russell NH, Madrigal JA. HLA-DPB1 matching status has significant implications for recipients of unrelated donor stem cell transplants. *Blood.* 2006;107(3):1220-1226.
2. Bettens F, Passweg J, Schanz U, et al. Impact of HLA-DPB1 haplotypes on outcome of 10/10 matched unrelated hematopoietic stem cell donor transplants depends on MHC-linked microsatellite polymorphisms. *Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation.* 2012;18(4):608-616.
3. Ludajic K, Balavarca Y, Bickeboller H, et al. Impact of HLA-DPB1 allelic and single amino acid mismatches on HSCT. *British journal of haematology.* 2008;142(3):436-443.
4. Shaw BE, Gooley TA, Malkki M, et al. The importance of HLA-DPB1 in unrelated donor hematopoietic cell transplantation. *Blood.* 2007;110(13):4560-4566.
5. Rutten CE, van Luxemburg-Heijs SA, van der Meijden ED, et al. Both permissive and nonpermissive HLA-DPB1 mismatches can induce polyclonal HLA-DPB1 specific immune responses in vivo and in vitro. *Blood.* 2010;115(1):151-153.
6. Rutten CE, van Luxemburg-Heijs SA, Halkes CJ, et al. Patient HLA-DP-specific CD4+ T cells from HLA-DPB1-mismatched donor lymphocyte infusion can induce graft-versus-leukemia reactivity in the presence or absence of graft-versus-host disease. *Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation.* 2013;19(1):40-48.
7. Stevanovic S, van Bergen CA, van Luxemburg-Heijs SA, et al. HLA class II upregulation during viral infection leads to HLA-DP-directed graft-versus-host disease after CD4+ donor lymphocyte infusion. *Blood.* 2013;122(11):1963-1973.
8. Zino E, Frumento G, Marktel S, et al. A T-cell epitope encoded by a subset of HLA-DPB1 alleles determines nonpermissive mismatches for hematologic stem cell transplantation. *Blood.* 2004;103(4):1417-1424.
9. Crivello P, Zito L, Sizzano F, et al. The impact of amino acid variability on alloreactivity defines a functional distance predictive of permissive HLA-DPB1 mismatches in hematopoietic stem cell transplantation. *Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation.* 2015;21(2):233-241.
10. Fleischhauer K, Shaw BE. HLA-DP in unrelated hematopoietic cell transplantation revisited: challenges and opportunities. *Blood.* 2017;130(9):1089-1096.
11. Crocchiolo R, Zino E, Vago L, et al. Nonpermissive HLA-DPB1 disparity is a significant independent risk factor for mortality after unrelated hematopoietic stem cell transplantation. *Blood.* 2009;114(7):1437-1444.
12. Fleischhauer K, Fernandez-Vina MA, Wang T, et al. Risk associations between HLA-DPB1 T-cell epitope matching and outcome of unrelated hematopoietic cell transplantation are independent of HLA-DPA1. *Bone marrow transplantation.* 2014;49(9):1176-1183.
13. Fleischhauer K, Shaw BE, Gooley T, et al. Effect of T-cell-epitope matching at HLA-DPB1 in recipients of unrelated-donor haemopoietic-cell transplantation: a retrospective study. *The Lancet Oncology.* 2012;13(4):366-374.
14. Arrieta-Bolanos E, Crivello P, Shaw BE, et al. In silico prediction of nonpermissive HLA-DPB1 mismatches in unrelated HCT by functional distance. *Blood Adv.* 2018;2(14):1773-1783.
15. Fleischhauer K, Zino E, Mazzi B, et al. Peripheral blood stem cell allograft rejection mediated by CD4(+) T

lymphocytes recognizing a single mismatch at HLA-DP beta 1*0901. *Blood.* 2001;98(4):1122-1126.

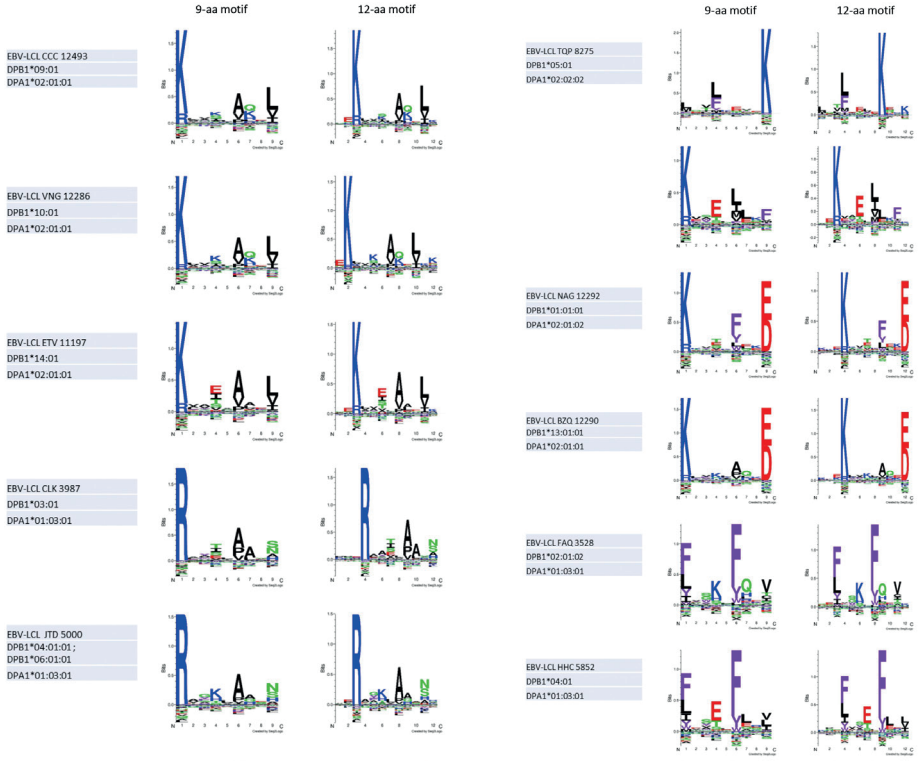
16. Crivello P, Heinold A, Rebmann V, et al. Functional distance between recipient and donor HLA-DPB1 determines nonpermissive mismatches in unrelated HCT. *Blood.* 2016;128(1):120-129.
17. Amir AL, van der Steen DM, Hagedoorn RS, et al. Allo-HLA-reactive T cells inducing graft-versus-host disease are single peptide specific. *Blood.* 2011;118(26):6733-6742.
18. Amir AL, Hagedoorn RS, van Luxemburg-Heijs SA, et al. Identification of a coordinated CD8 and CD4 T cell response directed against mismatched HLA Class I causing severe acute graft-versus-host disease. *Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation.* 2012;18(2):210-219.
19. Watson AJ, DeMars R, Trowbridge IS, Bach FH. Detection of a novel human class II HLA antigen. *Nature.* 1983;304(5924):358-361.
20. Mommen GP, Marino F, Meiring HD, et al. Sampling From the Proteome to the Human Leukocyte Antigen-DR (HLA-DR) Ligandome Proceeds Via High Specificity. *Molecular & cellular proteomics : MCP.* 2016;15(4):1412-1423.
21. Andreatta M, Lund O, Nielsen M. Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach. *Bioinformatics.* 2013;29(1):8-14.
22. Strauss T, von Maltitz MJ. Generalising Ward's Method for Use with Manhattan Distances. *PloS one.* 2017;12(1):e0168288.
23. Kimes PK, Liu Y, Neil Hayes D, Marron JS. Statistical significance for hierarchical clustering. *Biometrics.* 2017;73(3):811-821.
24. Lozzio CB, Lozzio BB. Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. *Blood.* 1975;45(3):321-334.
25. Yordanov V, Dimitrov I, Doytchinova I. Proteochemometrics-Based Prediction of Peptide Binding to HLA-DP Proteins. *Journal of chemical information and modeling.* 2017.
26. Urlacher A, Dormoy A, Tongio MM. DP epitope mapping by using T-cell clones. *Hum Immunol.* 1992;35(2):100-108.
27. Cesbron A, Moreau P, Milpied N, Harousseau JL, Muller JY, Bignon JD. Crucial role of the third and fourth hypervariable regions of HLA-DPB1 allelic sequences in the mixed lymphocyte reaction. *Hum Immunol.* 1992;33(3):202-207.
28. Naruse TK, Nose Y, Kagiya M, et al. Cloned primed lymphocyte test cells recognize the fourth, fifth, and sixth hypervariable regions at amino acid positions 65-87 of the DPB1 molecule. *Hum Immunol.* 1995;42(2):123-130.
29. Diaz G, Amicosante M, Jaraquemada D, et al. Functional analysis of HLA-DP polymorphism: a crucial role for DPbeta residues 9, 11, 35, 55, 56, 69 and 84-87 in T cell allorecognition and peptide binding. *International immunology.* 2003;15(5):565-576.
30. Berretta F, Butler RH, Diaz G, et al. Detailed analysis of the effects of Glu/Lys beta69 human leukocyte antigen-DP polymorphism on peptide-binding specificity. *Tissue Antigens.* 2003;62(6):459-471.
31. Rotzschke O, Falk K. Origin, structure and motifs of naturally processed MHC class II ligands. *Current opinion in immunology.* 1994;6(1):45-51.
32. Pidala J, Lee SJ, Ahn KW, et al. Nonpermissive HLA-DPB1 mismatch increases mortality after myeloablative unrelated allogeneic hematopoietic cell transplantation. *Blood.* 2014;124(16):2596-2606.
33. Klasberg S, Lang K, Gunther M, et al. Patterns of non-ARD variation in more than 300 full-length HLA-DPB1

alleles. *Hum Immunol.* 2019;80(1):44-52.

34. Petersdorf EW, Malkki M, O'HUigin C, et al. High HLA-DP Expression and Graft-versus-Host Disease. *The New England journal of medicine.* 2015;373(7):599-609.
35. Dong RP, Kamikawaji N, Toida N, Fujita Y, Kimura A, Sasazuki T. Characterization of T cell epitopes restricted by HLA-DP9 in streptococcal M12 protein. *Journal of immunology.* 1995;154(9):4536-4545.
36. Castelli FA, Buhot C, Sanson A, et al. HLA-DP4, the most frequent HLA II molecule, defines a new supertype of peptide-binding specificity. *Journal of immunology.* 2002;169(12):6928-6934.
37. Sidney J, Steen A, Moore C, et al. Five HLA-DP molecules frequently expressed in the worldwide human population share a common HLA supertypic binding specificity. *Journal of immunology.* 2010;184(5):2492-2503.
38. Andreatta M, Nielsen M. Characterizing the binding motifs of 11 common human HLA-DP and HLA-DQ molecules using NNAlign. *Immunology.* 2012;136(3):306-311.
39. Falk K, Rotzschke O, Stevanovic S, Jung G, Rammensee HG. Pool sequencing of natural HLA-DR, DQ, and DP ligands reveals detailed peptide motifs, constraints of processing, and general rules. *Immunogenetics.* 1994;39(4):230-242.
40. Diaz G, Canas B, Vazquez J, Nombela C, Arroyo J. Characterization of natural peptide ligands from HLA-DP2: new insights into HLA-DP peptide-binding motifs. *Immunogenetics.* 2005;56(10):754-759.
41. Chicz RM, Graziano DF, Trucco M, Strominger JL, Gorga JC. HLA-DP2: self peptide sequences and binding properties. *Journal of immunology.* 1997;159(10):4935-4942.
42. Racle J, Michaux J, Rockinger GA, et al. Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes. *Nature biotechnology.* 2019;37(11):1283-1286.
43. Abelin JG, Harjanto D, Malloy M, et al. Defining HLA-II Ligand Processing and Binding Rules with Mass Spectrometry Enhances Cancer Epitope Prediction. *Immunity.* 2019;51(4):766-779 e717.
44. Wang P, Sidney J, Kim Y, et al. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC bioinformatics.* 2010;11:568.
45. Kusano S, Kukimoto-Niino M, Satta Y, et al. Structural basis for the specific recognition of the major antigenic peptide from the Japanese cedar pollen allergen Cry j 1 by HLA-DP5. *Journal of molecular biology.* 2014;426(17):3016-3027.
46. Gunther S, Schlundt A, Sticht J, et al. Bidirectional binding of invariant chain peptides to an MHC class II molecule. *Proc Natl Acad Sci U S A.* 2010;107(51):22219-22224.
47. Lauterbach N, Crivello P, Wieten L, et al. Allorecognition of HLA-DP by CD4+ T cells is affected by polymorphism in its alpha chain. *Mol Immunol.* 2014;59(1):19-29.
48. Begovich AB, Moonsamy PV, Mack SJ, et al. Genetic variability and linkage disequilibrium within the HLA-DP region: analysis of 15 different populations. *Tissue Antigens.* 2001;57(5):424-439.
49. Hollenbach JA, Madbouly A, Gragert L, et al. A combined DPA1~DPB1 amino acid epitope is the primary unit of selection on the HLA-DP heterodimer. *Immunogenetics.* 2012;64(8):559-569.
50. al-Daccak R, Wang FQ, Theophille D, Lethielleux P, Colombani J, Loiseau P. Gene polymorphism of HLA-DPB1 and DPA1 loci in caucasoid population: frequencies and DPB1-DPA1 associations. *Hum Immunol.* 1991;31(4):277-285.
51. Shin DH, Baek IC, Kim HJ, et al. HLA alleles, especially amino-acid signatures of HLA-DPB1, might contribute to the molecular pathogenesis of early-onset autoimmune thyroid disease. *PloS one.* 2019;14(5):e0216941.

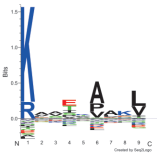
52. Gregersen JW, Erikstrup C, Ivarsen P, et al. PR3-ANCA-associated vasculitis is associated with a specific motif in the peptide-binding cleft of HLA-DP molecules. *Rheumatology (Oxford).* 2019.
53. Furukawa H, Oka S, Shimada K, et al. Independent association of HLA-DPB1*02:01 with rheumatoid arthritis in Japanese populations. *PloS one.* 2018;13(9):e0204459.
54. Lee JS, Park JK, Kim HJ, Lee HK, Song YW, Lee EB. Negatively-charged amino acids at the peptide-binding pocket of HLA-DPB1 alleles are associated with susceptibility to anti-topoisomerase I-positive systemic sclerosis. *Hum Immunol.* 2016;77(7):550-554.
55. Hilhorst M, Arndt F, Joseph Kemna M, et al. HLA-DPB1 as a Risk Factor for Relapse in Antineutrophil Cytoplasmic Antibody-Associated Vasculitis: A Cohort Study. *Arthritis Rheumatol.* 2016;68(7):1721-1730.
56. Koukoulioti E, Fischer J, Schott E, et al. Association of HLA-DPA1 and HLA-DPB1 polymorphisms with spontaneous HBsAg seroclearance in Caucasians. *Liver Int.* 2019;39(4):646-654.
57. Cheng L, Guo Y, Zhan S, Xia P. Association between HLA-DP Gene Polymorphisms and Cervical Cancer Risk: A Meta-Analysis. *Biomed Res Int.* 2018;2018:7301595.
58. Jia M, Han J, Hang D, et al. HLA-DP is the cervical cancer susceptibility loci among women infected by high-risk human papillomavirus: potential implication for triage of human papillomavirus-positive women. *Tumour Biol.* 2016;37(6):8019-8025.
59. Zhang X, Zheng C, Zhou ZH, et al. Relationship between HLA-DP gene polymorphisms and the risk of hepatocellular carcinoma: a meta-analysis. *Genet Mol Res.* 2015;14(4):15553-15563.

SUPPLEMENTARY FIGURES

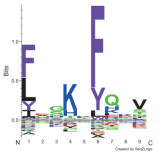


Supplementary Figure 1.
Binding motifs generated using Gibbs clustering for analyzing potential motifs with lengths of 9 or 12 animo acids (aa), with the positions of aa depicted on the x axis and the height of the depicted aa corresponding to their relative frequencies at these positions. For EBV LCL JTD 5000 only the clustering for HLA DPB1*06:01 is shown. Clustering for motifs of lengths of 9 or 12 aa resulted in similar results with core motif of 9 aa for all HLA DP alleles.

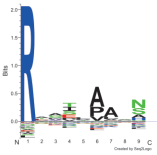
K562 - DPB1*14:01		
DPB1*14:01		
DPA1*02:01:01		
Motif Length: 9		
Group	Size	KLD
1	4513	11.8
Outliers	236	



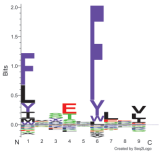
K562 - DPB1*02:01		
DPB1*02:01:01		
DPA1*01:03:01		
Motif Length: 9		
Group	Size	KLD
1	2915	13.0
Outliers	146	



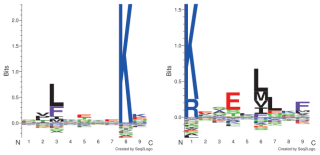
K562 - DPB1*03:01		
DPB1*03:01:01		
DPA1*02:01:01		
Motif Length: 9		
Group	Size	KLD
1	3451	14.8
Outliers	129	



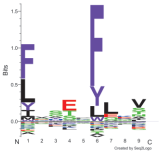
K562 - DPB1*04:01		
DPB1*04:01:01		
DPA1*01:03:01		
Motif Length: 9		
Group	Size	KLD
1	3451	13.7
Outliers	172	



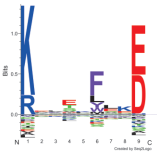
K562-DPB1*05:01		
DPB1*05:01:01		
DPA1*02:02:02		
Motif Length: 9		
Group	Size	KLD
1	3076	10.7
2	1872	10.8
Outliers	59	



K562 - DPB1*04:02		
DPB1*04:01:01		
DPA1*01:03:01		
Motif Length: 9		
Group	Size	KLD
1	4226	11.7
Outliers	143	



K562 - DPB1*01:01		
DPB1*01:01:01		
DPA1*02:01:01		
Motif Length: 9		
Group	Size	KLD
1	4063	11.8
Outliers	210	



Supplementary Figure 2.

Peptide binding motifs after analyzing the peptidome of HLA-DP transduced chronic myeloid leukemia cell line K562 as cell source of peptides for 7 different HLA-DP alleles. Positions of aa are depicted on the x-axis and the height of the depicted aa corresponds to their relative frequencies at these positions. Identical binding motifs were observed as identified using EBV-LCL as cell source.